

2014•2015
FACULTEIT BEDRIJFSECONOMISCHE WETENSCHAPPEN
*master in de toegepaste economische wetenschappen:
handelsingenieur in de beleidsinformatica*

Masterproef
Een data-analyse model voor management en project diensten

Promotor :
Prof. dr. Koenraad VANHOOF

Nickolas Janssens
*Scriptie ingediend tot het behalen van de graad van master in de toegepaste economische
wetenschappen: handelsingenieur in de beleidsinformatica*

2014•2015

FACULTEIT BEDRIJFSECONOMISCHE
WETENSCHAPPEN

*master in de toegepaste economische wetenschappen:
handelsingenieur in de beleidsinformatica*

Masterproef

Een data-analyse model voor management en project
diensten

Promotor :
Prof. dr. Koenraad VANHOOF

Nickolas Janssens

*Scriptie ingediend tot het behalen van de graad van master in de toegepaste economische
wetenschappen: handelsingenieur in de beleidsinformatica*



Een data-analyse model voor management en project diensten

*– Ontwikkelen van methodes om projecten/
projectteams te typeren (op basis van de
Dogglenet database)*

Masterproef

Nickolas Janssens

Master Handelsingenieur in de Beleidsinformatica

22 augustus 2015

Promotor: Prof. dr. Koen Vanhoof, Universiteit Hasselt

Begeleider: drs. Maarten Vanhoof, Universiteit Hasselt

Begeleider: dhr. Bart Plessers, Boondoggle

Voorwoord

Deze masterproef werd opgesteld in het kader van de masteropleiding Handelsingenieur in de Beleidsinformatica aan de Universiteit Hasselt. Hoewel dit onderzoek zich toespitst op de bedrijfssituatie van Boondoggle en de specifieke gegevens uit de Dogglenet database, kunnen de ontwikkelde methodes ook interessant zijn voor toepassing op andere data in andere (consultancy) contexten.

Ik wil graag mijn promotor, prof. dr. Koen Vanhoof, bedanken voor zijn begeleiding bij het schrijven van deze thesis. Tevens wens ik mijn begeleider, dr. Maarten Vanhoof, te bedanken voor zijn enthousiasme en de ondersteuning die hij bood tijdens dit onderzoeksproces. Ik dank ook Boondoggle voor het aanbieden van deze kans en in het bijzonder dank ik dhr. Bart Plessers voor zijn ("in-line") feedback op de vragen die ik op hem afvuurde.

Nickolas Janssens

Dessel, 22 augustus 2015

Samenvatting

Situering Bedrijven die op consultancybasis projecten afwerken, worden regelmatig geconfronteerd met de problematiek van het samenstellen en toewijzen van goede teams die tegelijk moeten kunnen voldoen aan de noden en de projecteisen van de klant, als aan de organisatie- en capaciteitsbeperkingen van het bedrijf. Die problematiek zal bovendien groter zijn wanneer er met variabele teams wordt gewerkt, dan wanneer enkel vaste teams kunnen worden ingezet. Het reclamebureau Boondoggle is zo'n bedrijf en zou graag meer inzicht verkrijgen in het samenstellen van hun teams voor de diverse projecten die ze uitvoeren. Met de jaren heeft het bedrijf een omvangrijke database opgebouwd die ze daartoe kunnen analyseren. Ze worden daarbij echter met één groot probleem geconfronteerd: de database bevat geen direct beschikbare indicatoren voor het succes van een project na afloop, noch voor het beschrijven van de projectsoort.

Het zal dus niet rechtstreeks mogelijk zijn te evalueren hoe een optimaal team voor een bepaald type project eruit ziet (of om zelfs maar enige zinvolle aanbevelingen naar teamgrootte of -samenstelling te doen). Deze thesis zal daarom trachten methodes te ontwikkelen om de projecten die Boondoggle in het verleden behandelde –en hun projectteams– te identificeren en hun kenmerken te beschrijven aan de hand van enkele indicatoren, zodoende een aantal team/project-types uit de data te kunnen afleiden. Het bedrijf kan deze informatie vervolgens aanwenden om hun teamkeuzes of teamprestaties te evalueren. Voor het typeren van de verschillende projecten en projectteams zal deze thesis zich baseren op het CRISP-DM-model om het onderzoeksproces in goede banen te leiden.

Data-preparatie De Dogglenet database is het product van een aantal data-migraties en upgrades. Bijgevolg zijn er heel wat projecten en projectteams waarvoor een deel van de informatie ontbreekt. Vooral met betrekking tot de timelogs, zijn er veel projecten die bestaan van voor de invoering van deze tabel. Om het informatiegehalte van de bevindingen zoveel mogelijk te verzekeren, is het belangrijk dat er uitsluitend volledige projecten beschouwd worden. De dataset wordt daarom gefilterd zodat enkel die projecten overblijven die aangevat én afgerond werden binnen het tijdsbestek van de timelog-observaties. Op basis van die overblijvende timelogs worden vervolgens de projectteams gereconstrueerd.

Ontwikkelde methodes Het typeren van een project wordt op drie vlakken beschouwd: de omvang van het project (scale), de diversiteit van het project (scope) en de financiële kant van het project. Per vlak worden een aantal attributen aangemaakt die verschillende

aspecten van het project met betrekking tot die categorie zullen belichten. Vervolgens worden twee globale indicatoren aangemaakt die een score (van 1 tot 5) toekennen aan een project om de *scale* en de *scope* van het project eenvoudig aan te kunnen duiden. Deze globale indicatoren worden berekend als een gewogen mix van de attributen die voor het belichten van de relevante deelaspecten werden opgesteld. De resulterende scores kunnen vervolgens gecombineerd worden om projecten een surrogaattypologie toe te kennen.

Met betrekking tot het bepalen van de projectomvang worden vijf indicatoren opgesteld: het aantal timelogs, die het verrichte werk met betrekking tot een project reflecteren; het aantal teamleden in het projectteam dat aan het project gewerkt heeft; het totale aantal aan het project gespendeerde uren; de looptijd van het project; en de omvang van het project zijn managementteam. Elke methode om de omvang van een project te beoordelen, belicht een ander aspect van die projectomvang en elke methode heeft ook zijn eigen beperkingen. Gecombineerd, echter, geven ze een goede globale indicatie van de projectomvang.

Voor het bepalen van de project(team)diversiteit werd vertrokken van uit drie assumpties. Ten eerste, hoe meer functies aan bod komen in een project, hoe diverser het project. Ten tweede, hoe meer functiegroepen vertegenwoordigd worden, hoe diverser het project. En ten derde, hoe meer de gepresteerde uren gespreid zijn over de verschillende vertegenwoordigde functiegroepen, hoe diverser het project. De indicatoren die daartoe werden opgesteld zijn (1) het aantal functies die werden uitgevoerd bij een project en (2) het aantal functiegroepen waarmee deze functies overeenkomen; (3) de user-functies mix en (4) de gepresteerde uren mix, gespreid over de verschillende functiegroepen.

Aangezien specifieke functies specifieke taken impliceren, kan op basis van deze user-functies/werkuren mix (tezamen met de andere typeringsindicatoren) ook een indicatie naar het soort project worden afgeleid. Uiteenlopende projecten, met andere doelomschrijvingen, zullen immers andere werkactiviteiten vereisen. Op het niveau van de activiteiten zelf kunnen, echter, geen analyses worden uitgevoerd wegens te weinig geformaliseerde data.

Tot slot werden ook enkele methodes voor het beoordelen van het financiële succes van een project opgesteld. Hierbij werden twee indicatoren gecreëerd die de omvang van een project in financiële termen uitdrukken, met name het totale budget en het totale gefactureerde bedrag. Deze indicatoren laten toe om, in combinatie met de eerder gecreëerde indicatoren, bepaalde aspecten van de projectefficiëntie te beoordelen en op die manier ook een eerste indicatie voor projectsucces te bepalen.

Inzetbaarheid Het nut van de globale en de deelaspect-indicatoren is vooral dat ze tijdens het onderzoeken van projectsucces of ontdekken van projectsoorten gebruikt kunnen worden om eenvoudig bepaalde projectkenmerken aan te duiden. Om het meeste potentieel uit de indicatoren te halen, kunnen ze worden gebruikt als input voor een data mining (clustering) proces om zo verschillende projecttypes te identificeren.

Inhoudsopgave

Voorwoord	i
Samenvatting	iii
Inhoudsopgave	v
Hoofdstuk 1: Inleiding	1
1. Probleemstelling	1
<i>Meer informatie over de situatie bij Boondoggle</i>	2
2. Onderzoeksopzet	3
Hoofdstuk 2: Preliminair Onderzoek	9
1. Definitie van een team	9
<i>Volgens de literatuur</i>	10
<i>Volgens Boondoggle</i>	11
<i>Implicaties voor het onderzoek</i>	12
<i>Accountdirector vs Accountmanager vs Projectmanager</i>	12
2. Voorstelling van de dataset.....	13
<i>Businessunits en types projecten</i>	17
Hoofdstuk 3: Reconstructie Projectteams	20
1. Data selectie.....	21
<i>Minimale datavereisten</i>	21
<i>Projecten</i>	23
<i>Attributen voor de reconstructie</i>	25
2. Reconstructie	25
Hoofdstuk 4: Projecten Typeren	29
1. Opbouw hoofdstuk.....	30
2. Voorbereiding tabellen	31
3. Omvang project.....	35
<i>Aantal timelogs</i>	32
<i>Totaal aantal gepresteerde uren op een project</i>	34
<i>Aantal teamleden</i>	36
<i>Looptijd</i>	38
<i>Aantal managers</i>	42
<i>Globale beoordeling van de omvang van een project</i>	46

4.	Diversiteit projectteam	51
	<i>Functies en Functiegroepen</i>	51
	<i>Mix van functie-gerelateerde werkl</i> ast.....	56
	<i>Activiteiten</i>	58
	<i>Globale projectdiversiteitsindicator</i>	58
	<i>Conclusie</i>	60
5.	Financiële omvang-indicatoren.....	62
	<i>Projectefficiëntie analyseren</i>	62
	Hoofdstuk 5: Projecten karakteriseren: Projectsoort en projectsucces	65
1.	Data Mining.....	65
2.	Process Mining	67
	Besluit	69
	Literatuurbronnen	71
	Bijlage A: Mailvragen	73
	Bijlage B: Queries	77

Hoofdstuk 1

Inleiding

Bedrijven die op consultancybasis projecten afwerken, worden regelmatig geconfronteerd met de problematiek van het samenstellen en toewijzen van goede teams die tegelijk moeten kunnen voldoen aan de noden en de projecteisen van de klant, als aan de organisatie- en capaciteitsbeperkingen van het bedrijf. Die problematiek zal bovendien groter zijn wanneer er met variabele teams wordt gewerkt, dan wanneer enkel vaste teams kunnen worden ingezet. Deze thesis onderzoekt bovenvermelde problematiek in het kader van de specifieke bedrijfsdata van het reclamebureau Boondoggle.

1. Probleemstelling

Boondoggle is een bedrijf dat marketing-, reclame- en PR-diensten verleent en dit doet bij wijze van consultancy. Sinds haar ontstaan in 2007 is het bedrijf flink gegroeid. De tijd dat de bedrijfsleiding vanuit het buikgevoel teams en projecten kon organiseren is voorbij en zonder een formele aanpak voor het samenstellen van teams in gebruik, blijkt het niet evident te zijn om goede teams samen te stellen, rekening houdend met de werklast voor de individuele teamleden en hun bestaande taken en plannings.

Een deel van die problematiek ligt hem in het feit dat projecten sterk kunnen uiteenlopen qua eisen naar de oplevering toe, en daaraan gekoppeld ook qua noden vanuit het bedrijf om aan die eisen te voldoen. Bijgevolg dient voor elk project een team op maat te worden samengesteld en dit zowel naar omvang als naar samenstelling. Eén project kan bijvoorbeeld met een klein, voornamelijk technisch team worden afgehandeld, terwijl voor een ander project verschillende werknemers uit diverse domeinen zullen nodig zijn. De opleveringstermijn kan bij die teamkeuze ook een rol spelen. Het samenstellen van een team is echter ook onderworpen aan de capaciteitsbeperkingen van het bedrijf. Er zijn immers slechts een gegeven aantal werknemers beschikbaar met een zekere skill set of achtergrond en het afstemmen van de werkplanningen voor deze werknemers kan niet verenigbaar zijn met andere taken of projecten waar zij reeds aan werken.

Meer informatie over de situatie bij Boondoggle

Het bedrijf Boondoggle vormt een interessant observatieonderwerp. Enkele jaren geleden hebben ze namelijk de overstap gemaakt van een systeem met vaste teams (en dus vaste partners) naar een systeem met variabele teams, waarbij telkens een team op maat van het project wordt samengesteld om een flexibele dienstverlening te kunnen bieden. Hoewel de flexibiliteit inderdaad toenam, bracht de nieuwe situatie ook extra moeilijkheden met zich mee.

Vroeger, wanneer er met vaste teams gewerkt werd, oordeelden de managers van Boondoggle telkens men een nieuw project binnenhaalde, welke van hun (beschikbare) teams het meest geschikt was om aan dit project te werken. Dit deden ze veelal vanuit hun buikgevoel. Ze konden immers intuïtief vrij goed inschatten hoeveel tijd en budget er moest worden toegemeten aan de verschillende projectfasen. Dit systeem van vaste teams maakte dat het toewijzen van personeel aan een project relatief eenvoudig was.

Ook voor de werknemers waren er voordelen verbonden aan dit systeem. Aangezien men steeds met dezelfde partners werkte, kon men niet alleen een sterke persoonlijke band opbouwen, men leerde ook omgaan met elkaars manier van werken en men leerde de sterktes en zwaktes kennen van de anderen en van zichzelf. Zo konden partners goed op elkaar inspelen en anticiperen, en dit, samen met de sterkere vertrouwensband, resulteerde vaak in een goede samenwerking om zo een project af te leveren binnen de gestelde deadlines en binnen de lijnen van het budget.

De oude manier van werken bleef echter niet haalbaar naarmate het bedrijf groeide. Doordat het aantal projecten toenam, specifieke projecteisen meer en meer verschilden, en personeelsleden steeds meer gespecialiseerde kennis en vaardigheden bezaten, werd de nood voor variërende teams steeds groter. Vaker waren meer de kwaliteiten van enkele specifieke personen nodig, zonder dat ook per se de partners van de oude teams gewenst waren. Men stapte dan ook af van het systeem van de vaste teams en beoordeelt voortaan voor elk nieuw project hoe een optimaal team er voor dat project uit zou zien. Teams kunnen onder het nieuwe systeem dus variëren zowel qua samenstelling als qua omvang.

Deze nieuwe werkwijze bracht echter aanzienlijk meer organisatiemoeilijkheden met zich mee. Personeelsleden zouden ook klagen van een hogere werkdruk en een grotere planlast. Het feit dat werknemers nu in meerdere teams tegelijk kunnen zitten en gelijklopende projecten elkaar in de weg kunnen staan, maakt dat het voor een persoon moeilijk kan zijn om al zijn taken goed te coördineren en de verschillende projectplanningen en deadlines te respecteren. Meer zelfs,

door het afwijken van de planning door één persoon kan het zijn dat latere projectfasen ook met vertraging pas kunnen worden aangevat. Vanwege dit cascade-effect, samen met de hogere plannings- en organisatiemoeilijkheden, vormt het systeem van variabele teams een groter risico om deadlines of budgetten te overschrijden. De werknemers ervaren deze oefening in het combineren van deadlines en het stellen van prioriteiten als een verhoogde werkdruk, hetgeen uiteindelijk resulteert in een groter verloop van personeelsleden bij het bedrijf.

2. Onderzoeksopzet

Aangezien het werken met variabele teams geen sinecure is, zou het bedrijf graag meer inzicht verkrijgen in het samenstellen van hun teams. Echter, aangezien de database waarop het onderzoek zal steunen geen onmiddellijke indicatoren bevat over het succes van een project na afloop, noch over de projectsoort, zal het niet mogelijk zijn te evalueren hoe een optimaal team voor een bepaald type project eruit ziet of welke teamtypes algemeen de beste prestaties opleveren. In plaats daarvan zal deze thesis trachten methodes te ontwikkelen om de projectteams te identificeren van de projecten die Boondoggle in het verleden behandelde en hun kenmerken te beschrijven, zodoende teams zoveel mogelijk te kunnen typeren, zodat het bedrijf deze informatie kan aanwenden in een aanpak om hun teamkeuzes of teamprestaties te evalueren of om ongewone afwijkingen te detecteren. De centrale onderzoeksvraag luidt bijgevolg:

CENTRALE ONDERZOEKSVRAAG

Hoe is het mogelijk om op basis van de beschikbare Dogglenet data en literatuur methodes te ontwikkelen om projecten/projectteams te typeren?

Om het beantwoorden van deze centrale onderzoeksvraag te vereenvoudigen, kan ze worden opgesplitst in een aantal deelvragen waarop eerst een antwoord gezocht zal moeten worden alvorens een globaal antwoord kan worden geformuleerd. De eerste stap zal dan zijn te onderzoeken of het überhaupt mogelijk is om teams te reconstrueren op basis van de dataset die beschikbaar werd gesteld. Als dit niet mogelijk is, zal verder onderzoek ook zinloos zijn. Deelvraag 1 wordt bijgevolg geformuleerd als:

DEELVRAAG 1

Kunnen we projectteams reconstrueren op basis van de beschikbare Boondoggle data?

Alvorens echter teams gereconstrueerd kunnen worden, dient eerst te worden gedefinieerd wat er precies onder een team verstaan wordt. Dit zal vanuit twee perspectieven worden benaderd: Enerzijds zal een korte literatuurstudie worden uitgevoerd, en anderzijds zal de

visie van het bedrijf ook in rekening worden genomen, zodoende een werkbare definitie op te stellen. Vandaar de eerste sub-deelvraag bij deelvraag 1:

SUB-DEELVRAAG 1.1

Wat is een team?

Eenmaal deze definitie vastgesteld, moet er data worden geselecteerd om de projectteamreconstructie op uit te voeren. Hierbij is het belangrijk na te gaan welke data er allemaal in de database zitten en te evalueren welke van deze gegevens nuttige info kunnen bijdragen in de verdere stappen van de analyse. Sub-deelvraag 2 luidt bijgevolg:

SUB-DEELVRAAG 1.2

Omvat de Dogglenet database de juiste data voor het reconstrueren van projectteams?

Indien het antwoord op deze vraag positief is, kan worden overgegaan tot het reconstrueren van de projectteams. Hierbij moet er worden nagedacht over hoe de gegevens verbonden aan deze reconstructie het best kunnen worden opgeslagen. Daarbij moet er meteen ook al aandacht worden besteed aan de volgende fase van het onderzoek, waar er nieuwe informatie met betrekking tot de projectteams zal worden gezocht. Er moet dus geanticipeerd worden hoe deze informatie zal kunnen worden bijgevoegd aan de basisgegevens van de projectteams. Daarom wordt de derde sub-deelvraag geformuleerd als volgt:

SUB-DEELVRAAG 1.3

Hoe kunnen de gegevens met betrekking tot de gereconstrueerde projectteams het best worden opgeslagen?

De volgende stap in het onderzoek betreft het zoeken naar mogelijkheden om de gereconstrueerde projectteams (of meer algemeen: projecten) te beschrijven. Aangezien de dataset geen indicatoren bevat over het succes van een project na afloop, noch over de projectsoort, gaat deze thesis op zoek naar alternatieve mogelijkheden om de projecten/projectteams te typeren. Het gaat dan om een exploratieve analyse van mogelijke methodes en interessante attributen die kunnen helpen in het karakteriseren van projecten en projectteams. Vandaar de tweede deelvraag:

DEELVRAAG 2

Hoe kunnen we projectteams beschrijven?

Het zoeken naar deze mogelijkheden zal gebeuren aan de hand van een iteratief proces, volgens het stramien van de CRISP-DM methode (Chapman et al., 2000). Dit acroniem staat voor 'Cross-Industry Standard Process for Data Mining' en slaat op een werkwijze om een data mining-onderzoek succesvol aan te pakken. De methode werd ontwikkeld als

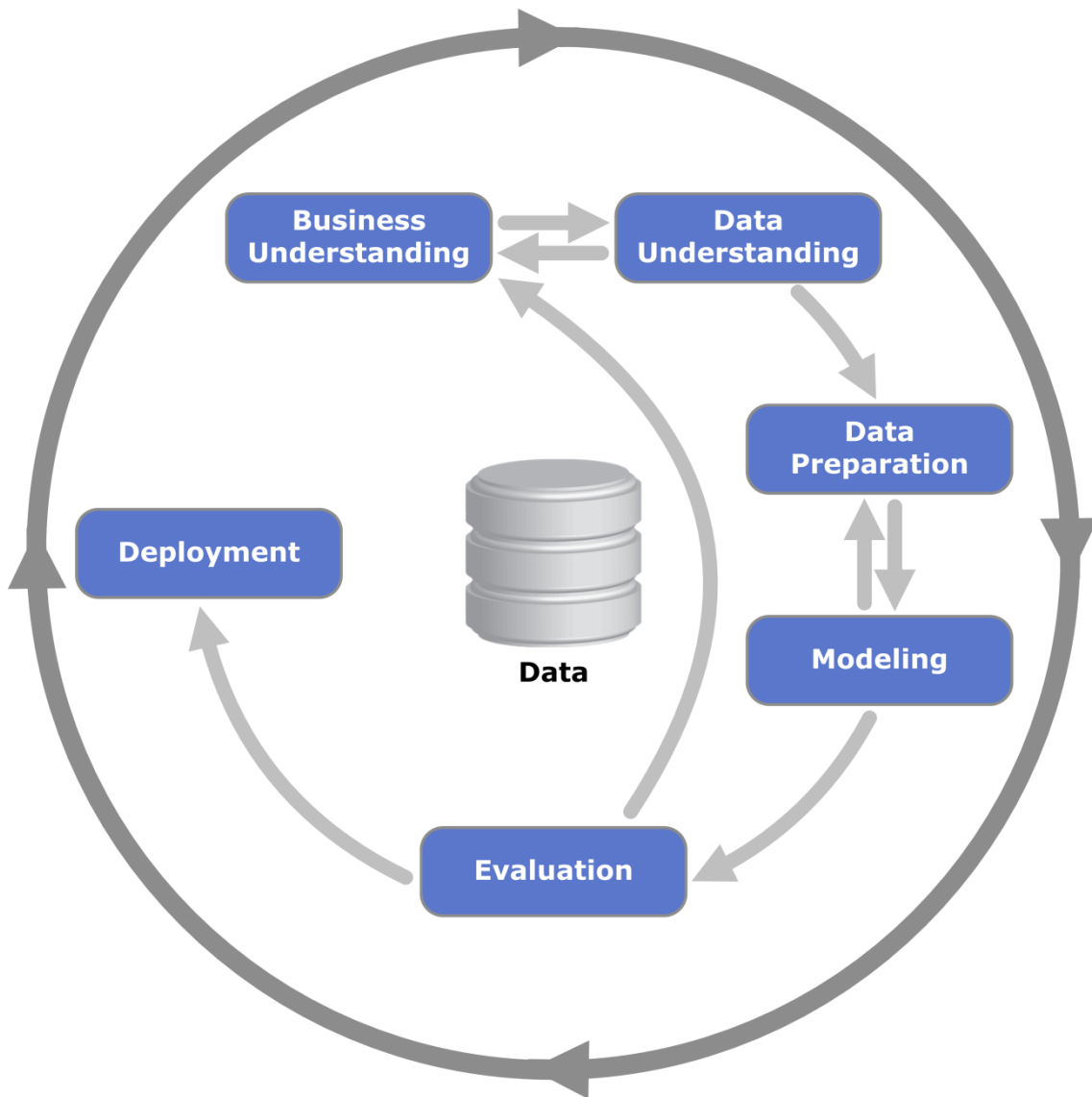
een aanpak tot data mining specifiek en de problemen die zich daarbij kunnen voordoen, maar hoeft niet uitsluitend tot dat onderzoeksdomein gelimiteerd te zijn. Zo kan de CRISP-DM-methode ook breder gezien worden als een strategie voor Knowledge Discovery in het algemeen. We zullen de methode dan ook in zijn bredere zin gebruiken om inzichten te ontdekken uit de Dogglenet dataset met betrekking tot projectteams.

Zoals men in *Figuur 1* kan zien, bestaat de CRISP-DM methode uit 6 verschillende fases. Deze fases maken deel uit van een onderzoekscyclus. Het cyclische karakter is typisch voor een knowledge discovery-onderzoek. Ook tussen de verschillende fases onderling kan het nodig zijn om meerdere malen heen en weer te gaan. Echter, in logische volgorde zijn de fases (1) inzicht verwerven in de bedrijfswerking, (2) inzicht verwerven met betrekking tot de data(base), (3) de data prepareren teneinde ze geschikt te maken voor de gekozen knowledge discovery-methode, (4) de gegevens modelleren aan de hand van de gekozen knowledge discovery-methode, (5) de resultaten evalueren, en (6) de resulterende informatie aanwenden in de bedrijfswerking of in vervolgonderzoek.

De zes fases in meer detail besproken (Chapman et al., 2000):

1. *Business Understanding*: De eerste fase beoogt een verdieping in de doelstellingen en de vereisten voor het onderzoek vanuit een zakelijk perspectief. Deze inzichten moeten dan worden vertaald in een probleemstelling en een onderzoeksplan dat gevolgd zal moeten worden om de onderzoeksdoelstellingen te realiseren.
2. *Data Understanding*: Deze fase behelst het vertrouwd raken met de data. Dit omvat dan het vaststellen van problemen met de kwaliteit van de data (en de eventueel opvallende afwezigheid van gegevens), het inzicht verwerven in de structuur van de data, en het ontdekken van interessante subsets van gegevens.
3. *Data Preparation*: Deze fase omvat alle activiteiten die nodig zijn om de initiële, ruwe dataset om te vormen tot een voor het voorziene gebruik geschikte dataset.
4. *Modeling*: Onder deze fase wordt een modelleringstechniek gekozen en toegepast. Er zal ook aandacht besteed worden aan het afstellen van de verschillende parameters vereist voor het toepassen van die techniek.
5. *Evaluation*: Deze fase beoogt een evaluatie van de bekomen uitkomsten van de modellering en ook de doorlopen stappen worden gereviseerd om na te gaan of ze de vooropgestelde (bedrijfs)doelstellingen wel behalen.
6. *Deployment*: Het ontdekken van kennis uit de data is typisch niet het einde van het onderzoek. De resultaten zullen ook nog geordend en gepresenteerd moeten worden op zo'n manier dat ze geïnterpreteerd kunnen worden en bruikbaar zijn voor de klant.

In deze thesis zullen fases 1 en 2 doorlopen worden bij wijze van een algemeen preliminair onderzoek en niet als onderdeel van de cyclus. Ze komen overeen met hoofdstukken 1 en 2 van deze paper. Daarna zullen de volgende fasen wel cyclisch doorlopen worden, met een grote aandacht voor fase 4, daar deze thesis onderzoeken wil welke methoden er zoal bestaan en toepasbaar zijn op de data om projectteams te typeren.



Figuur 1: CRISP-(DM) methode

Met het CRISP-DM-model in het achterhoofd, kunnen nu de nodige sub-deelvragen beschreven worden die betrekking zullen hebben op het tweede deel van dit onderzoek. Vooral de knowledge discovery cyclus kan worden aangevat echter, moeten we eerst notie nemen van welke methodes er mogelijk gebruikt kunnen worden voor het typeren van projectteams. Eenmaal dit bekend is, kan voor elke methode apart de onderzoeks cyclus, startende vanaf fase 3, doorlopen worden. Sub-deelvraag 2.1 luidt bijgevolg:

SUB-DEELVRAAG 2.1

Welke methodes zijn er (mogelijk) om de projectteams te typeren en wat laten ze ons toe te onderzoeken?

Enmaal een onderzoeksmethode gekozen werd, zal er moeten worden nagegaan welke attributen er kunnen dienen als input voor de methode. Er zal met andere woorden bekeken worden welke data, wanneer ze gecombineerd of gemanipuleerd worden, nuttige informatie kunnen bijdragen aan het onderzoek. De tweede sub-deelvraag formuleren we daarom als:

SUB-DEELVRAAG 2.2

Welke extra attributen moeten (nog) worden meegenomen voor de analyse?

De volgende stap omvat het processen van de data teneinde ze bruikbaar te maken voor de analysemethode. Dit houdt in dat problemen met de kwaliteit van de gegevens (aangaande volledigheid en uniformiteit) worden aangepakt en dat er wordt nagegaan of de data in een beter geschikt formaat moeten worden getransformeerd. Daarna kan de data nog gesorteerd of geaggregeerd worden al naargelang de verwachte input voor de gekozen methode. Sub-deelvraag 3 kan daarom als volgt geformuleerd worden:

SUB-DEELVRAAG 2.3

Hoe moet de data worden gemanipuleerd teneinde ze geschikt te maken voor de gekozen methode voor het typeren van de projectteams?

Eens de vorige vragen zijn geadresseerd, kan de analyse worden uitgevoerd en kunnen vervolgens de verschillende mogelijkheden en moeilijkheden —analysemogelijkheden, interpretatiemogelijkheden, nuttigheid en inzetbaarheid van de informatie, aandachtspunten en beperkingen— worden onderzocht zodat een antwoord op de tweede deelvraag kan worden geformuleerd, met name: Hoe kunnen aan de hand van de geselecteerde methode projectteams getypeerd worden? Dit proces herhalen we dan voor de verschillende gekozen methodes.

Dit hoofdstuk behandelde fase 1 van de CRISP-DM-methode reeds gedeeltelijk. In het volgende hoofdstuk zal verder worden gegaan op die *business understanding*-fase en zal ook de *data understanding*-fase worden besproken. Hoofdstuk 3 behandelt dan de reconstructie en voorstellingwijze van de projectteams, waarna in het volgende hoofdstuk verschillende methodes voor het typeren van projectteams aan bod zullen komen.

Hoofdstuk 2

Preliminair Onderzoek

In dit hoofdstuk zal de basis van het onderzoek gelegd worden. Allereerst wordt nagegaan hoe een team gedefinieerd moet worden en wat de kenmerken van zo'n team moeten zijn. Vervolgens zal de database beschreven worden en verkennen we de data. Op die manier trachten we meer inzicht te verkrijgen in de projectwerking en het gebruik van de database. Bovendien zullen ook enkele vragen worden voorgelegd aan Boondoggle om extra klaarheid te scheppen over de bedrijfswerking en sommige keuzes uit de database.

1. Definitie van een team

Het onderzoeksproces in de constructie van informatie met betrekking tot de projectteams kan niet succesvol worden doorlopen zonder een duidelijk afgebakend idee van wat een team precies uitmaakt. Er zijn immers verschillende definities mogelijk, waarbij enkele elementen kunnen verschillen of een andere invulling kunnen hebben. Zo kan een team bijvoorbeeld een tijdsdimensie omvatten. Dit kan dan worden begrepen als hetzij een groep mensen die op hetzelfde moment samenwerken, ofwel gedurende een bepaalde periode samenwerken, dan wel minstens een bepaalde duur hebben samengewerkt. Een tweede mogelijkheid is dat de definitie van een team enkel steunt op een gemeenschappelijk einddoel waar iedereen naartoe werkt en geen tijdsdimensie omvat.

Eerst dient daarom te worden gedefinieerd wat de kenmerken zijn van een team of welke de (minimum)vereisten zijn om van een team te kunnen spreken in de context van consultancywerk op basis van variabele projectteams. Tot deze definitie zullen we komen door het onderwerp vanuit twee perspectieven te benaderen. Enerzijds zal een korte literatuurstudie worden uitgevoerd, en anderzijds zal de visie van het bedrijf ook in rekening worden genomen. Deze zullen dan met elkaar worden gecontrasteerd om tot een algemene definitie te komen. Toch volstaat dit niet. Om een werkbare definitie op te stellen zal ook rekening gehouden moeten worden met de beperkingen uit de dataset. De realiteit van de dataset zal immers bepalen wat onderzocht kan worden en wat niet. Dit laatste zullen we in het volgende deel onderzoeken.

Volgens de literatuur

Er is in de literatuur al heel wat onderzoek gedaan naar het functioneren van teams. Vaak is men daarbij geïnteresseerd in de groepsdynamische aspecten verbonden aan teams (bijvoorbeeld betreffende leiderschap, verticale of horizontale gezags- en verantwoordelijkheidsstructuren, e.d.). Vele van deze werken starten hun onderzoek met het definiëren van een team. Zo wordt al eens verwezen naar de definitie van Sundstrom, McIntyre, Halfhill, en Richards (2000) die stellen dat een team een verzameling is van individuen die onderling afhankelijk zijn en die bovendien gezamenlijk verantwoordelijk zijn voor specifieke resultaten van hun organisatie. Wanneer het dan over projectteams specifiek gaat, verwijzen zij ook nog naar enkele extra eigenschappen die Munster (2007) kernachtig opsomt als 'gedefinieerd', 'gespecialiseerd', 'in tijd beperkt' en 'worden aan het eind ontbonden'.

Een andere definitie van een team wordt gegeven door Salas, Dickinson, Converse, en Tannenbaum (1992) en werd naar het Nederlands vertaald door Rijnbergen (2007) als een te onderscheiden reeks van twee of meer mensen die dynamisch, van elkaar afhankelijk en adaptief naar een gemeenschappelijk en gewaardeerd doel/objectief/missie toe werken, die specifieke rollen of functies hebben, en die een beperkte levensduur van lidmaatschap hebben. Salas et al. leggen in hun definitie dus meer de nadruk op het gemeenschappelijke doel en de gezamenlijke afhankelijkheid, waarbij ieder teamlid een eigen unieke en kritieke rol vervult en interactie aangaat met andere teamleden om zo de gezamenlijke doelen van het team te bereiken (Rijnbergen, 2007), terwijl Sundstrom et al. (2000) ook aandacht hebben voor de gedeelde verantwoordelijkheid die daar bij hoort. Verder zijn beide definities ten aanzien van projectteams vrij gelijklopend.

Dat de ultieme definitie van een projectteam waarschijnlijk een unie van de voorgaande twee zal zijn, blijkt uit een derde en laatste formulering die stelt dat de definitie van een team die van een groep overstijgt door de gerichtheid waarmee op effectieve wijze en met gezamenlijke verantwoordelijkheid wordt gewerkt aan het behalen van vooraf gestelde doelen en taken (Katzenbach & Smith, 1993) (via Rijnbergen, 2007). Met het oog op het reconstrueren van projectteams stileren we uit de bovengenoemde definities nu de kenmerken waaraan zo'n projectteam moet voldoen.

Kenmerken van een projectteam volgens de literatuur:

- Twee of meer individuen
- Leden hebben gespecialiseerde kennis, vervullen specifieke rollen
- Interactie
- Onderlinge afhankelijkheid
- Gezamenlijke verantwoordelijkheid
- Gericht werken naar gemeenschappelijke, gestelde doelen
- Beperkte (bestaans-)duur en ontbinding na afloop

Volgens Boondoggle

Nu het uit de literatuur gestileerde antwoord gekend is, kan het ook interessant zijn na te gaan wat de visie van Boondoggle is ten aanzien van een projectteam. Hebben zij op beleidsniveau een interne definitie opgesteld? Zijn er formele of informele guidelines? Of zijn er misschien minimumvereisten waaraan een projectteam moet voldoen? Om dit uit te vinden, werden deze (en andere) vragen voorgelegd aan Bart Plessers, multimedia director bij Boondoggle en tevens het aanspreekpunt voor deze thesis binnen het bedrijf. (Zie bijlage A voor de betreffende mails.)

De projecten die Boondoggle aanneemt kunnen sterk uiteenlopen. Bijgevolg zijn er maar weinig "standaard" projecten. Komt daarbij dat het bedrijf ook nauwelijks guidelines heeft met betrekking tot het vormen van projectteams, noch qua minimum bezetting als qua rolverdeling. Minimaal zal er altijd wel een accountmanager in een projectteam zitten, maar verder ligt er niets vast. Gezien de projecten nogal uit elkaar kunnen liggen, wordt ad hoc dus het meest geschikte team samengesteld, waarbij die samenstelling bepaald wordt door de inhoud van het project en de beschikbaarheid van personeel.

Boondoggle heeft echter ook enkele speciale businessunits die een aparte manier van werken kennen. Bij projecten van de 'Agency'-businessunit worden enkel accounts (= accountdirectors of accountmanagers) vast toegewezen aan een klant, maar verder wordt niet strikt een bepaald soort team samengesteld. Bij de 'Bright'-projecten, daarentegen, wordt zoveel mogelijk met klantenteams gewerkt. Dit houdt in dat voor de verschillende projecten voor een bepaalde klant overwegend met dezelfde teamcombinaties gewerkt zal worden en dat projecten zoveel mogelijk bij dezelfde mensen terecht zouden moeten komen, hierbij wel rekening houdend met de beschikbaarheid van die mensen. Voor nieuwe projecten wordt er ad hoc een team samengesteld en voor alle projecten geldt dat teams worden ontbonden na realisatie van de project-einddoelstellingen.

Hoe verhoudt bovenstaande toelichting van Boondoggle zich nu tot wat er uit de literatuur geleerd werd? Een team bestaat volgens het bedrijf uit twee of meer mensen, waaronder ten minste één accountmanager. Teams worden op maat van de inhoud van het project samengesteld en mensen worden daarbij gekozen op basis van hun expertise (en beschikbaarheid). Bovendien worden teams na afloop van een project ontbonden (maar wordt er wel voor gezorgd dat opvolgprojecten voor een bepaalde klant min of meer bij dezelfde mensen terecht komen). De idee van Boondoggle ligt dus zeker in lijn met die van uit de literatuur. Toch, wat betreft de gedeelde verantwoordelijkheid van de verschillende teamleden, kunnen we in het geval van Boondoggle niet besluiten dat die verantwoordelijkheid voor het eindresultaat effectief door alle leden gedragen wordt; voornamelijk dan vanwege de verplichte aanwezigheid in elk team van een accountmanager.

Implicaties voor het onderzoek

De implicaties voor dit onderzoek zijn dat de gereconstrueerde teams uit minimaal twee personen zullen moeten bestaan. Als er 'teams' bestaande uit slechts één persoon worden gevonden, dan zullen zij als opmerkelijk worden aangeduid en kan er eventueel verder onderzocht worden naar de precieze oorzaak of omstandigheden van deze opmerkelijke beslissing. Misschien zou het in dergelijke gevallen kunnen gaan om projecten die wel geïntialiseerd werden in de database, maar uiteindelijk niet zijn uitgevoerd.

Ten tweede, gegeven dat een groep specifieke personen geselecteerd worden om een rol op te nemen binnen het team, en gegeven dat teams worden ontbonden aan het eind van hun opdracht, moet er aan een gegeven project precies één afgebakende groep van mensen kunnen worden gelinkt. Deze strategie zou echter kunnen falen voor projecten van het Bright-type, aangezien deze projecten een doorlopend karakter hebben.

Gegeven de (mogelijks) specifieke aard van de projecten behorende tot de verschillende businessunits, dringt er zich verder onderzoek op naar de functies van de verschillende businessunits en de specifieke aard van hun projecten. Er zal moeten worden vastgesteld of de diverse projecten van de diverse businessunits in hun geheel wel interessant zijn voor het onderzoek. Dit verder onderzoek zal aan het eind van dit hoofdstuk beschouwd worden.

Tot slot kan het zijn dat er teams zullen aangetroffen worden zonder een accountmanager. Dit zou dan, althans volgens de hierboven geformuleerde beleidsvisie, een afwijking betekenen ten opzichte van de guidelines inzake het samenstellen van een projectteam. Dergelijke anomalieën kunnen interessant zijn voor nader onderzoek.

Accountdirector vs Accountmanager vs Projectmanager

Bovenstaande tekst roept vragen op over wat het onderscheid tussen een accountdirector en een accountmanager precies is. Dit is algeheel niet duidelijk. Projecten uit de dataset hebben soms beide toegewezen gekregen, maar lang niet altijd. Komt daarbij kijken dat sommige projecten ook nog een derde manager hebben, met name een projectmanager. Uit de database leiden we af dat alle projecten een AccountManagerID hebben, echter, twee derde van de projecten heeft geen AccountDirectorID, en vier vijfde van de projecten heeft geen ProjectManagerID. Er is bovendien geen verband tussen het ontbreken van een accountdirector en het ontbreken van een projectmanager bij een project. De vraag wat dan het precieze verschil tussen deze drie functies is, werd voorgelegd aan Bart Plessers. (Zie bijlage A)

De accountdirector is de eindverantwoordelijke voor een bepaalde *klant*. Eén klant kan meerdere projecten (gehad) hebben bij Boondoggle. De accountdirector staat dan in voor de klanttevredenheid en het stimuleren van nieuwe businessopportunities met die klant. Een accountdirector beheert typisch ook meerdere klanten tegelijk.

De accountmanager neemt over van de accountdirector van zodra er een nieuw *project* gestart wordt met een klant. De accountmanager helpt met het opstellen van budgetten en zorgt voor administratieve ondersteuning van de accountdirector. De accountmanager vormt dus een team met de accountdirector, maar waar die laatste op klantbasis wordt toegewezen, wordt de accountmanager op projectbasis toegewezen. Gezien zijn taakomschrijving is de naamkeuze voor de accountmanager misschien wat verwarrend, daar men zijn functie intuïtief misschien eerder met de term projectmanager zou verbinden.

Projectmanagers volgens Boondoggle, echter, worden pas bij een project in het leven geroepen wanneer dat project te groot in omvang of te technisch wordt. De projectmanager neemt dan de interne organisatie van het project over. Bij niet-technische projecten en kleinschaligere projecten (zoals affiches, radio- of tv-spotten) blijft deze taak aan de accountmanager behouden. Wel voegt Bart Plessers er nog aan toe dat het onderscheid tussen de drie functies in feite meer een grijze zone is. Zo neemt in uitzonderlijke gevallen de accountdirector alle drie de rollen tezamen op en durven bij niet-structurele klanten de hoedjes wel eens te wisselen.

2. Voorstelling van de dataset

De opzet van deze thesis is een exploratief onderzoek te voeren naar mogelijke methodes om projectteams te typeren. Dit onderzoek zal zich daarbij baseren op de “Dogglenet”-database. Dit is een database die van Boondoggle verkregen werd en die het bedrijf haar data aangaande hun projecten bevat. Meer specifiek omvat de dataset zowel de werkingsdata die nodig is voor de bedrijfsactiviteit en het projectwerk, als een logboek waarin de personeelsleden registeren waaraan ze wanneer gewerkt hebben. In dit tweede deel van het hoofdstuk volgt een korte voorstelling van deze dataset en onderzoeken we de structuur, de omvang, de mogelijkheden en de beperkingen van de dataset. Daarbij zullen ook meteen enkele opvallende afwijkingen en problemen geadresseerd worden.

De structuur van de Dogglenet-database wordt gevisualiseerd in *Figuur 2*. Het betreft een schema dat door het bedrijf zelf werd opgesteld. Opvallend aan dit schema zijn de kleurgroepen die de diverse inhoud van deze dataset duidelijk onderscheiden. Ze tonen namelijk de verschillende tabellen die relateren aan een bepaald aspect van de projectwerking (waarvan de naam in de meeste gevallen voorgesteld wordt door de grootste hub-tabel binnen de kleurgroep waaraan de kleinere tabellen vervolgens gelinkt zijn). Zo leert het schema ons dat de database informatie bevat over de verschillende projecten van Boondoggle, de budgetten en de kosten, over de personeelsleden, hun functies, en de timelogs die ze invullen (om bij te houden waaraan wie wanneer werkte), en als laatste ook over de klanten van het bedrijf.

De database bevat 50 tabellen hoewel het schema in *Figuur 2* er slechts 46 toont. Echter, niet alle tabellen zullen interessant zijn voor dit onderzoek. De vier tabellen die niet vermeld worden, zijn 'BudgetInsertTemplates', 'Departments', 'OfficialHolidays' en 'TimelogEntry-Templates'. Waarom werden deze tabellen uit het schema weggelaten? Het lijkt erop dat de twee "Templates"-tabellen het overblijfsel zijn van een experiment – zij zijn bovendien voor dit onderzoek niet meteen nuttig. Ook de OfficialHolidays-tabel is voor dit onderzoek niet interessant en de reden voor het weglaten zal bijgevolg niet nagegaan worden. De Departments-tabel, echter, bevat mogelijk wel interessante gegevens voor dit onderzoek, met name voor het analyseren van de samenstelling van projectteams, wanneer men personen wil groeperen op basis van hun activiteit binnen het bedrijf bijvoorbeeld. De data uit de Departments-tabel wordt getoond in *Tabel 1*.

ID	Code	Name	SeqNr
1	TECHIES	Technical Deve- lopment	3
2	ACCOUNTS	Accounts	2
4	GRAPHS	Multimedia Deve- lopment	4
6	STAFF	Staff	6
8	CONTENT	Creation	5
9	EXTERN	Extern	7
10	MARSALES	Marketing & Sales	8
11	CRM	CRM	9
15	STRAT	Strategy	10
16	DESIGN	Design	11

Tabel 1: Data uit Departments-tabel

ID	Code	Name	SeqNr
1	MANAGEMENT	Account and Project Mgmt.	1
2	STRATEGY	Strategy	2
3	CREATION	Creation	3
4	DESIGN	Design	4
5	ANALYSIS	Analysis	5
6	DEVPROD	Development & Production	6
7	OTHER	Other	7

Tabel 2: Data uit FunctionGroups-tabel

In de database bestaat er een gelijkaardige tabel als de Departments-tabel, met name de FunctionGroups-tabel, dewelke wordt afgebeeld in *Tabel 2*. Ze heeft dezelfde structuur en een gelijkaardige inhoud. Het verschil tussen de twee tabellen zit hem in het feit dat de eerste tabel betrekking heeft op de indeling van de verschillende personeelsfuncties binnen de structuur van het bedrijf Boondoggle zelf (dus ook functies die niet rechtstreeks aan projecten verbonden kunnen worden) en de tweede tabel op de indeling van de functies binnen projectteams. Voor dit onderzoek zullen we in eerste instantie dus met de tweede tabel werken, hoewel ook de informatie uit de eerste tabel een interessante insteek zou kunnen opleveren. Bijgevolg mogen de tabellen die niet in het schema staan genegeerd worden, met eventuele uitzondering van de Departments-tabel.

De gegevens uit de database hebben betrekking op 9584⁽¹⁾ projecten, gespreid over een periode van zo'n zeven jaar, van 30-10-2006 tot 17-10-2013. Voor de meeste van die projecten betreft het afgeronde projecten, maar sommige en dan voornamelijk de meest recente

Aanmerkingen bij de cijfergegevens:

¹ *Projecten*

De Projects-tabel bevat 9584 projecten. Er zijn echter slechts 9582 distincte projecten. De betekenis en de implicaties hiervan worden verder besproken in het volgende hoofdstuk onder het kopje Projecten.

projecten zijn nog work-in-progress op het moment dat deze versie van de Dogglenet-database werd opgeslagen. Opvallend is ook dat er heel wat projecten in de dataset zitten die schijnbaar gestart werden op 30-10-2006 – de vroegste datum uit de set. Het betreft hier uiteraard geen overvloed aan nieuwe projecten, maar wel een migratie van de gegevens uit een oude database naar deze “nieuwe” database. Verder omvat de database 504⁽²⁾ gebruikers (users, personeelsleden), al dan niet nog in dienst van het bedrijf, die tezamen gewerkt hebben voor reeds 2861⁽³⁾ klanten. Sinds 28-01-2010 vullen de personeelsleden ook een logboek in, hetgeen resulteert in een tabel die 238744⁽⁴⁾ timelogs beslaat. Deze tabel zal de basis vormen voor onze projectteam-reconstructie.

De Dogglenet-database is dus behoorlijk uitgebreid, maar ondanks dat de database teruggaat tot het jaar 2006, zal slechts een deel van die data maar bruikbaar zijn in dit onderzoek, omdat het uitvalspunt voor de constructie van de project-teaminformatie de Time-logs-tabel zal zijn. Deze tabel gaat namelijk slechts terug tot 2010. Voor de projecten van voor die periode zullen geen teams gereconstrueerd kunnen worden. Zo wordt er een belangrijk deel van de database en potentieel interessante informatie genegeerd. Immers, van de 9584 projecten in de database vallen er maar 3191 binnen de periode van de timelogs. Toch, gegeven de vele records in de Timelog-tabel en gegeven het (nog steeds grote) aantal observeerbare projecten, zal er genoeg broninformatie zijn om voldoende teams te kunnen reconstrueren.

Naast de beperking van het tijdsbestek van de observeerbare data (en de reeds gekende problemen van ontbrekende direct beschikbare projectsucces-maatstaven, en een indicatie van de projectsoort), heeft de dataset nog een enkele andere beperking. Vele daarvan kunnen worden teruggebracht op de timelogs. Ze zijn inconsistent, aangezien ze vrij uiteenlopende proporties kunnen aannemen, van zeer nauwgezet gerapporteerde taakuitvoeringen tot blokken van tientallen uren. Ook kunnen sommige van hun velden vrij worden ingevuld (of zo lijkt het althans). Bovendien ontbreekt er vaak ook planning-informatie.

² *Gebruikers*

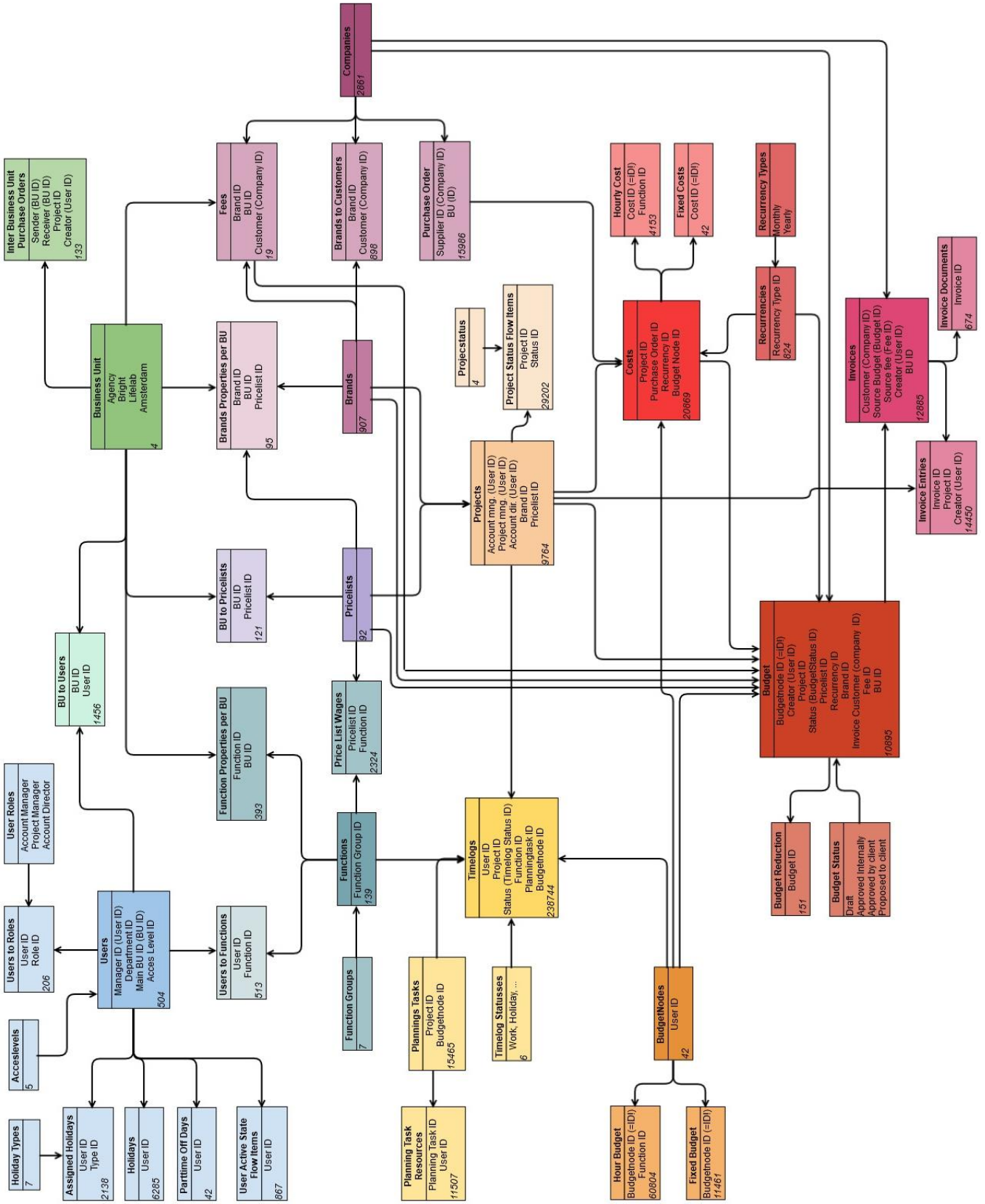
De Users-tabel bevat 504 rijen, maar de laatste userID is 505. Er werd dus één rij uit de tabel verwijderd, namelijk de rij met userID 467.

³ *Klanten*

De Companies-tabel bevat 2861 rijen. Het laatste ID is ook 2861 en er zijn bovendien 2861 verschillende Codenamen, maar de tabel bevat slechts 2756 verschillende bedrijfsnamen. Er zijn dus namen die dubbel voorkomen. Bijvoorbeeld de “123 Comedy Club – Vanpoucke St” staat tweemaal in de tabel. Daarnaast zijn er ook bedrijven die onder een andere schrijfwijze meerdere malen vermeld staan. Bijvoorbeeld “10 milligram” en “10mg”. Het precieze aantal unieke klanten van het bedrijf zal dus lager liggen dan 2765.

⁴ *Timelogs*

De Timelogs-tabel bevat 238744 logs. Echter, de ID-nummering loopt op tot 240455. Er werden dus 1711 rijen uit de tabel verwijderd.



Figuur 2: Schema van de "Doggenet"-database

Businessunits en types projecten

In het eerste deel van dit hoofdstuk hebben we onderzocht wat de beleidsvisie van Boondoggle ten aanzien van (de definitie van) projectteams was. We hebben nagegaan of het bedrijf minimumvereisten of guidelines inzake het vormen van projectteams had. Daarbij werd een paar keer verwezen naar twee aparte categorieën van projecten en zijn de namen van 'Agency' en 'Bright' gevallen. Deze twee namen verwijzen naar twee van Boondoggle's businessunits, maar er zijn nog drie zulke units waarvan nog geen melding werd gemaakt. De bedoeling van deze sectie is om de betekenis van de verschillende businessunits en de eventuele specifieke aard van hun projecten verder uit te diepen.

Men kan vaststellen dat projecten in de database worden ingedeeld bij een bepaalde businessunit. In *Tabel 3* kan je de betreffende indeling van de projecten aflezen. Een vraag die zich daarbij opdringt is of deze indeling een soort typering van de projecten reflecteert. Staat elke businessunit voor een bepaalde soort projecten of slaat het eerder op de geografische locatie waar een project werd binnengehaald of uitgevoerd? Uit onderstaande tabel kan men afleiden dat Boondoggle vijf zulke businessunits heeft. Twee daarvan, Boondoggle Amsterdam en Boondoggle Ireland lijken inderdaad buitenlandse filialen te betreffen. Wat ook meteen opvalt is dat Boondoggle Agency veruit de meeste projecten heeft lopen en dat Boondoggle Ireland er geen heeft.

Business UnitID	Business UnitCode	Business Unit	#Projecten
1	Y	Boondoggle Agency	8042
2	B	Boondoggle Bright	837
3	F	Boondoggle LifeLabs	16
4	A	Boondoggle Amsterdam	689
5	I	Boondoggle Ireland	0
			9584

Tabel 3: Indeling projecten volgens BusinessUnit

Er werd aan Boondoggle gevraagd wat precies de betekenis van of het onderscheid tussen de verschillende businessunits is. In een reactie op deze en bovenstaande vragen zei Bart Plessers dat het verhaal met de businessunits een beetje dubbel is. (Voor de betreffende mail: zie bijlage A, mail van 12/08/2015.)

Alle businessunits zijn ontstaan uit het moederbedrijf Boondoggle. Ze hebben elk hun eigen rapportering wat betreft omzet en kosten en hebben bovendien een eigen management. Boondoggle Ireland en Boondoggle Amsterdam zijn inderdaad twee buitenlandse filialen, maar het filiaal in Ireland is sinds enkele jaren reeds gesloten. Vandaar dat er (in deze versie van de dataset) geen projecten aan deze businessunit gelinkt zijn. Boondoggle Amsterdam, daarentegen, is tegenwoordig ondergebracht onder Havas (een fusie vond plaats tussen Havas Worldwide en Boondoggle Amsterdam tot het nieuwe "Havas-Boondoggle Amsterdam").

Boondoggle Amsterdam heeft het dogglenet nooit goed gebruikt, weet Bart Plessers ook te vertellen. Om die reden zijn er waarschijnlijk aan de businessunit Amsterdam eveneens weinig projecten gelinkt.

'Agency' en 'Bright' zijn beide in Leuven gebaseerde filialen. Zij onderscheiden zich vooral door het type projecten die ze uitvoeren. De Agency-businessunit werkt aan kortlopende campagnes, terwijl de Bright-businessunit zich bezig houdt met permanente "touchpoints" zoals CRM en het ontwikkelen en onderhouden van corporate websites en apps. 'Lifelabs' was dan weer een initiatief waar mensen uit Leuven in konden instappen. Zij zouden dan 4/5^{de} voor Boondoggle werken en 1/5^{de} voor Lifelabs. Daarbij trachtten ze op eigen houtje innoverende producten op te zetten om zo winst voor henzelf te genereren. Dit was vooral een PR-stunt voor Boondoggle.

Afgezien van misschien de Agency- en Bright-projecten reflecteert de BusinessUnitID dus niet meteen het soort project. Een tweede manier waarop de businessunits en mogelijks in meer detail ook het type project gekenmerkt kunnen worden, is via de eerste letters van de codenamen van de verschillende projecten. *Tabel 4* toont een kleine greep uit de projecten van de Projects-tabel. (Voor het doeleinde van dit voorbeeld worden enkel de ProjectID, Code en BusinessUnitID kolommen uit de Projects-tabel getoond; de andere kolommen werden geweggelaten.)

ProjectID	Code	BusinessUnitID
6269	P-BDBE-7023	1 ("Agency")
6589	P-COLA-7370	1
8184	P-BELF-9163	2 ("Bright")
9198	PA-HEIN0-9198	4 ("Amsterdam")
9199	PA-HEIN0-9199	4
8955	PB-BELF-8955	2
9393	PB-BDBR-9393	2
8938	PF-LABS-8938	3 ("LifeLabs")
6982	P-LABS-7836	3
9264	PY-COLA-9264	1
9173	PY-COCA-9173	1

Tabel 4: Selectie van projecten en hun codenamen

Zoals men uit bovenstaande tabel kan afleiden, zijn de codenamen van projecten opgebouwd volgens een bepaald stramien. De eerste letters of lettergroepen schijnen het soort project te indiceren. Indien men de projecten zou indelen op basis van de eerste letters van de projectcodenaam dan bekomt men een overzicht zoals in *Tabel 5*. Wanneer men deze tabel vergelijkt met *Tabel 3*, blijkt duidelijk de overeenstemming tussen (de tweede letter van) de projectcode en de businessunitcode. Het enige verschil zijn projecten waarvan de codenaam enkel met "P-" start. Zou het kunnen dat dit op een soort standaardprojecten slaat, terwijl de andere eerste-letter-codes op speciale projecten van de businessunits duiden? Immers, *Tabel 5* mapt op *Tabel 3* door de "standaardprojecten" over de andere 5

categorieën te verspreiden. Bovendien zou het kunnen dat het type project nog verder gespecificeerd wordt aan de hand van verdere letter- of cijfercombinaties verder in de projectcodes. De vraag die dient gesteld te worden is dus: Gaat het voor die simpelweg 'P-' projecten inderdaad om standaardprojecten, of betreft het eerder een niet-consistente naamgeving? Zo ja, kan op deze manier misschien al een eerste stap gezet worden richting het bepalen van de projectsoort.

Projecttype ?	Projectcode begint met	#Projecten
"Standaard" ?	P-	8863
Amsterdam	PA-	83
Bright	PB-	260
LifeLabs	PF- of P-LABS-	14
Ierland	PI-	0
Agency	PY-	364
		9584

Tabel 5: Indeling projecten volgens begin van projectcodenaam

De reactie van Boondoggle: De eerste letters van de projectcode slaan enkel op de businessunits die verantwoordelijk zijn voor het project. Er zijn bovendien geen standaardprojecten (althans niet volgens de logica van de codenamen). Waar die 'P-projecten' dan wel vandaan komen is het volgende: Het concept van de businessunits bestaat nog maar sedert enkele jaren in de Dogglenet database. Voordien werden alle projecten gewoon als "P-" aangeduid. Bovendien was er vroeger per nationaal filiaal een apart Dogglenet. In 2011 werd dan al die data samengebracht onder één gedeelde database – aldus Bart Plessers. (Deze migratie van data en creatie van nieuwe attributen, samen met de notie dat Boondoggle Amsterdam dogglenet nooit echt goed gebruikt heeft, kunnen ook verklaren waarom de database over het algemeen inconsistente gegevens bevat.)

Op de vraag of er dan op een andere manier wel een onderscheid gemaakt wordt naar projecttypes binnen de dataset antwoordde Bart Plessers negatief. Er is geen eenduidige manier om types projecten te onderscheiden. Dat is een groot gebrek van Dogglenet. Als er al een manier moest zijn, dan is dat bij wijze van een patroon of soort fingerprint die type projecten achterlaten. Hij voegde er nog aan toe dat, moest dit inderdaad het geval zijn, hij dat dan erg boeiend zou vinden om weten. – Laat dat nu exact zijn waar deze thesis naar op zoek gaat: methodes om projecten en projectteams te typeren.

We kunnen dus besluiten dat wanneer men de verschillende projectsoorten wil kunnen typeren, het nodig zal zijn om alle projecten, ongeacht hun overeenkomstige businessunit, mee te nemen in de verdere analyse. Bovendien is het interessanter om zich te baseren op resultaten uit het dataonderzoek en de karakterisering, dan op noties van de beleidspersonen om te besluiten of bepaalde projecten al dan niet interessant zouden zijn voor verder onderzoek. Dit moet zo maar blijken uit de data. Om die reden zullen alle projecten meegenomen worden in de verdere analyse.

Hoofdstuk 3

Reconstructie Projectteams

In dit hoofdstuk zal de reconstructie van de projectteams behandeld worden. Alvorens echter die reconstructie kan plaatsvinden dient eerst te worden geverifieerd of de dataset wel de juiste data daarvoor bevat en daarna kan er geselecteerd worden welke data interessant zijn voor de reconstructie. Eens dit gebeurd is, zal bekeken worden hoe het reconstructieproces kan worden aangepakt. Eerst voor één project, daarna hoe dit kan worden uitgebreid naar alle relevante projecten. Daartoe moet er een optimale manier gezocht worden om de gegevens met betrekking tot de verschillende projectteams voor te stellen. Bij het bepalen van deze voorstellingswijze moet er meteen ook al rekening gehouden worden met de hoe de informatie voor en uit de verschillende teamtyperingswijzen opgeslagen kan worden.

1. Data selectie

Het reconstrueren van projectteams begint met een goede selectie van de gegevens die gebruikt zullen worden in dit reconstructieproces. Alvorens personen aan projecten gelinkt zullen worden, moet er immers onderzocht worden welke projecten interessant zijn voor dit onderzoek en welke niet. Eerst zullen we echter bekijken welke gegevens er minimaal nodig zijn voor de reconstructie van een projectteam en of die ook in de dataset aanwezig zijn. Bovendien kan er met het oog op de volgende stap in dit onderzoek besloten worden welke attributen er nog interessant kunnen zijn en bijgevolg ook geselecteerd moeten worden.

Minimale datavereisten

De eerste stap van de data-selectie bestaat eruit na te gaan welke gegevens er minimaal voor nodig zijn om projectteams te reconstrueren en of deze ook in de database aanwezig zijn, zodoende er zeker van te zijn dat het überhaupt mogelijk is om een projectteam te reconstrueren. Het spreekt voor zich dat er minstens een oplijsting van de uitgevoerde projecten en een oplijsting van de werknemers moet zijn. Waar we echter echt naar op zoek zijn is een manier om die twee met elkaar in verband te brengen. Deze kan gevonden worden in de Timelogs-tabel, waarin de personeelsleden bijhouden aan welke taak ze gewerkt hebben voor welk project, gedurende een bepaalde periode (van de dag). Als we bijgevolg

voor een gegeven ProjectID alle uitgevoerde taken opvragen via deze tabel, dan krijgen we een lijst zoals die in *Tabel 6*.^[Q1] ⁽⁵⁾

ID	ProjectID	UserID	StatusID	FunctionID	DurationHrs	BookDay	(1)	(2)	(3)	(4)	(5)
209037	9021	381	1	17	7.50000	2013-04-18					...
209065	9021	263	1	11	2.50000	2013-04-18					...
209203	9021	263	1	11	0.50000	2013-04-19					...
209481	9021	171	1	17	8.00000	2013-04-17					...
209482	9021	171	1	18	8.00000	2013-04-18					...
...											

Tabel 6: Voorbeeld timelogs voor ProjectID '9021'

(1) Comment, (2) EnterTime, (3) LastChangeTime, (4) PlanningTaskID, (5) BudgetNodeID

Bovenstaande tabel leert ons dat aan project 9021 onder andere de gebruikers (of beter personeelsleden) met de identificatienummers 381, 263 en 171 meegewerkt hebben. Let wel, bovenstaande tabel toont slechts een deel van de informatie. Om eenvoudig alle personeelsleden (of beter projectteamleden) te vinden die aan het project hebben meegewerkt, voeren we een nieuwe query^[Q2] uit. Deze geeft als resultaat:

$$UserID's = \{101, 159, 171, 199, 222, 258, 263, 298, 299, 304, 313, 337, 381, 401, 466\}$$

Wanneer men echter bovenstaande tabel grondig bekijkt, valt het op dat sommige personeelsleden meer dan één functie uitoefenen binnen het project. Zo bekleed user 171 zowel de functies 17 (technisch ontwikkelaar) als 18 (grafisch ontwikkelaar). Bij de reconstructie van de projectteams is het essentieel dat ook deze informatie wordt meegenomen. Het is immers nutteloos te weten wie er deel uitmaakte van een projectteam, zonder ook zijn rol binnen dat team te kennen. Een aangepaste query^[Q3] levert dan het volgende op:

$$UserID + FunctionID's = \{(101, 1), (159, 19), (171, 17), (171, 18), (199, 6), (222, 86), (258, 89), (263, 11), (298, 19), (298, 97), \dots\}$$

In dit voorbeeld zie je dat zowel gebruiker 171 als gebruiker 298 twee verschillende functies bekleeden binnen het team. De meeste gebruikers zullen echter maar één rol vertolken binnen een projectteam. Wat betreft de overige attributen uit de TimeLogs-tabel, deze zijn niet essentieel, hoewel StatusID wel een interessante parameter is. De StatusID geeft namelijk aan wat voor log een nieuwe toevoeging aan de tabel betreft. Als deze 1 is, gaat het om een werk-gerelateerde log, terwijl logs met andere StatusID's gerelateerd zijn aan afwezigheden en vakanties. Het is dus zinvol om de TimeLogs-tabel te filteren zodat enkel logs met StatusID = 1 zullen overblijven. In het volgende deel gaan we na of we die tabel nog verder dienen te filteren, maar dan met betrekking tot de projecten.

⁵ Relevante gebruikte queries worden toegevoegd in Bijlage B en zullen in de tekst worden aangeduid met een volgnummer [Qx]

Projecten

De tweede stap van de data-selectie omvat het identificeren van de projecten die interessant kunnen zijn voor ons onderzoek. Het uiteindelijke doel van dit onderzoek is om projectteams te karakteriseren (en om opvallende of algemene kenmerken te identificeren om zo eventuele teamtypes te kunnen ontdekken). Om dit te doen slagen, moet de informatie van waaruit vertrokken wordt wel accuraat en volledig zijn. Dit impliceert dat de projecten waarop de teams betrekking hebben aan enkele voorwaarden moeten voldoen.

Projecten moeten zijn:

- uniek
- voltooid
- volledig

Projecten moeten uniek zijn in die zin dat aan één gegeven project geen twee of meer verschillende teams gekoppeld kunnen worden. Dit was één van de voorwaarden uit de projectteamdefinitie. Hiertoe moeten we kunnen verzekeren dat elk project uniek identificeerbaar is. Aangezien dat elk nieuw project dat het bedrijf aanneemt, toegevoegd wordt aan de Projects-tabel, krijgt het ook een uniek ID-nummer aan de hand waarvan in andere tabellen naar dit project gerefereerd kan worden (zo ook in de TimeLogs-tabel). Bijgevolg is aan de voorwaarde voldaan, zolang elk nieuw project maar aan de Projects-tabel wordt toegevoegd. Problemen ontstaan er pas wanneer er projecten zijn die doorlopend blijven bestaan of die een permanente activiteit met kleine deelprojectjes hebben. Dergelijke projecten zullen echter in de volgende stap van het projecten filteren worden opgevangen.

Volledigheidshalve moeten we wel vermelden dat er twee keer twee projecten in de database zitten die dezelfde codenaam delen. Normalerwijs zou ook deze projectcode uniek moeten zijn. De betreffende projecten worden getoond in *Tabel 7*. Opnieuw, de IDs voor deze projecten zijn wel uniek en ook de gegevens uit de andere kolommen getuigen dat het werkelijk verschillende projecten betreft. Opvallend is wel dat het viermaal over een project voor Boondoggle Nederland (BrandID = 2) gaat. Desalniettemin mogen alle vier de projecten ongemoeid meegenomen worden in de verdere analyse.

ID	Code	Name	AccMgrID	EnterTime	LastChangeTm	CreatorID	BrandID
8381	P-BDNL-4488	Wedstrijd Zomer	415	2008-08-12	2008-12-08	415	2
4059	P-BDNL-4488	Digibaro-meter	10	2007-07-03	2008-03-11	10	2
8443	P-BDNL-4554	Blauwe Maandagen	417	2009-03-30	2010-06-22	197	2
4115	P-BDNL-4554	Microsoft EMEA	197	2007-08-01	2008-01-21	187	2

Tabel 7: Tweemaal twee projecten met een gedeelde projectcodenaam

(Enkele kolommen werden weggelaten uit dit overzicht en sommige kolomnamen en velden werden ingekort!)

Het tweede zorgpunt van deze projectfiltering betreft de volledigheid van de gegevens aan de hand waarvan de projectteams gereconstrueerd zullen worden. Immers, om te voorkomen dat halve waarheden met elkaar zullen vergeleken worden, mogen enkel volledige projecten beschouwd worden. Dit houdt in dat projecten zowel gestart als afgerond moeten zijn binnen de observeerbare periode uit de dataset.

De periode die we kunnen observeren voor het reconstrueren van de projectteams wordt bepaald door de omvang van de TimeLogs-tabel. Deze tabel bevat logs beginnende vanaf boekingsdag 28-01-2010 tot en met de laatste dag uit de dataset (de dag waarop de versie van de database werd opgeslagen). De eerste records (of logs) uit die tabel betreffen voornamelijk projecten die al in zekere mate zijn afgewerkt. Met betrekking tot deze projecten zullen we dus maar een deel de relevante informatie kunnen ophalen (namelijk, van het werk dat nog gedaan zal worden binnen het tijdsbestek van de database). Er moet dus gezocht worden naar een manier om enkel die projecten te identificeren die aangevat werden op 28-01-2010 of daarna.

Die mogelijkheid wordt gevonden in de Projects-tabel. Zoals je in *Tabel 7* kan zien, bevat deze tabel een kolom "EnterTime" die bijhoudt wanneer het project werd ingegeven. De Projects-tabel filteren op invoegdatums die groter of gelijk zijn aan 28-01-2010^[Q4] levert bijgevolg een selectie op van de projecten die binnen het bereik van de TimeLogs-tabel gestart werden.

Eenzelfde werkwijze als die om te vinden wanneer een project aangevat werd, kan echter niet gebruikt worden voor te besluiten wanneer een project werd afgerond. De LastChangeTime signaleert immers enkel wanneer de laatste wijziging aan een project gebeurde, maar het gaat daarom nog niet om de finale wijziging. In plaats daarvan zal in de ProjectStatusFlowItems-tabel moeten worden nagegaan wanneer de status van een project gewijzigd werd in 'Finished project'. Deze tabel vormt namelijk de link tussen de Projects- en de ProjectStatusses-tabel en registreert de verschillende werkfasen die een project doorloopt. Enkel de projecten die in de vorige stap geselecteerd werden (de "aangevatte" projecten), komen in aanmerking. Het resultaat is een lijst van uitsluitend volledige, voltooide projecten, uitgevoerd tussen 28-01-2010 en het moment van opslaan van de Dogglenet-versie.^[Q5]

Zoals eerder geredeneerd in hoofdstuk 2, herhalen we hier nogmaals de beslissing om geen projecten vroegtijdig uit de werkdataset te filteren omwille van een andere reden dan het niet uniek, volledig of voltooid zijn. Indien bepaalde projecten sterk afwijkend zouden zijn of niet interessant voor het karakteriseren, dan moet dit zo blijken uit de data en niet louter uit een opvatting over de projecten. Om die reden zullen alle van de in de vorige paragraaf gefilterde projecten meegenomen worden in de verdere analyse.

Attributen voor de reconstructie

Nu de lijst van alle bruikbare ProjectIDs beschikbaar is, kan er onderzocht worden welke attributen er interessant kunnen zijn om mee te nemen in de projectteams-reconstructie. Waar het in de vorige sectie ging om het filteren van specifieke rijen uit bepaalde tabellen, zal het in het volgende deel de bedoeling zijn kolommen te selecteren.

De keuze van deze kolommen (of “attributen”) voor het reconstrueren van de projectteams is vrij voor de hand liggend. Enerzijds moeten alle teamleden aan hun projecten gekoppeld kunnen worden. Daarom zal de combinatie van ProjectID en UserID minimaal nodig zijn. Daarnaast hebben we ook al gesteld dat het zinloos is om personeelsleden aan een project te linken zonder hun functie binnen dat project te kennen. Ook de FunctionID is bijgevolg onontbeerlijk.

2. Reconstructie

In het eerste deel van dit hoofdstuk hebben we vastgesteld dat de dataset inderdaad de nodige attributen bevat voor het reconstrueren van de projectteams. Bovendien hebben we de dataset gefilterd zodat enkel nog volledige, afgewerkte projecten overblijven. Deze hebben we aan de hand van enkele queries opgeslagen in een nieuwe (tijdelijke) hulptabel: de ProjectsCOMPLETED-tabel. We hebben ook vastgesteld dat de manier waarop de projectteams gereconstrueerd kunnen worden, is om uit de TimeLogs-tabel de verschillende project/user-combinaties te stileren. Momenteel, echter, bevat die TimeLogs-tabel nog alle records die er over de hele bestaansduur van de tabel in werden opgeslagen. Ook deze tabel zullen we dus moeten filteren zodat enkel de relevante records overblijven. Daartoe maken we een tweede hulptabel aan, met als naam TimelogsFILTERED. Voor deze tweede hulptabel dienen enkel die timelogs geselecteerd te worden waarvan de ProjectID overeenkomt met een ID uit de ProjectsCOMPLETED-tabel. De query voor deze bewerking kan worden teruggevonden in bijlage B. ^[Q6]

Vanuit deze hulptabel kunnen dan eenvoudig de verschillende projectteams gereconstrueerd worden van al de volledige, afgewerkte projecten. In het eerste deel van dit hoofdstuk hebben we geargumenteed dat een projectteam minimaal gekenmerkt wordt door het ProjectID, de gerelateerde UserIDs, maar ook de FunctionIDs van de verschillende teamleden, aangezien één teamlid verschillende rollen kan opnemen binnen een team. De data uit de gefilterde timelogstabel dient bijgevolg gegroepeerd te worden op deze drie attributen. De timelogstabel bevat echter ook nog heel wat andere attributen, vele daarvan die niet geaggregeerd kunnen worden. In *Tabel 6* kan men de verschillende attributen uit de TimeLogs-tabel aflezen.

De meeste van deze andere attributen hebben betrekking op de kenmerken van één instantie uit de tabel. Zo duidt ID op het volgnummer van een gegeven instantie in de time-logstabel. De StatusID beschrijft wat voor soort log de instantie betreft, echter alle logs die aan projectwerk gerelateerd zijn, zullen een StatusID van 1 hebben. Ook dit attribuut is bijgevolg niet interessant om mee te nemen. De attributen EnterTime en LastChangeTime beschrijven wanneer de log werd ingevoerd in de tabel of wanneer er laatst een wijziging aan werd aangebracht. De BookDay beschrijft dan weer op welke dag de ingegeven log betrekking heeft. (Er kan immers wat spatie zitten tussen de dag dat een personeelslid een bepaalde taak verrichtte en het moment waarop hij deze activiteit in het logboek neerschrijft.) Al deze data zijn irrelevant in het verdere onderzoek, daar we vooral op zoek zijn naar informatie van een hogere granulariteit. Het zou daarnaast ook gewoon zinloos zijn deze datums te aggregeren. Het spreekt bovendien voor zich dat het evenzeer zinloos zou zijn om het Comment-attribuut te willen aggregeren.

Blijven nog de Duration_Hours, PlanningTaskID en BudgetNodeID attributen over. Ook deze laatste twee zijn niet interessant om mee te nemen, niet in het minst omdat ze zo zelden werden ingevuld. Wel interessant is het om bij Boondoggle om eens na te gaan waarom deze (schijnbaar belangrijke velden) nauwelijks worden ingevuld. Bart Plessers zegt hierover het volgende: “Alleen als de accountmanagers en projectmanagers hun planning juist maken krijgen de mensen timelogsuggesties met een budgetreferentie. In het andere geval geven de mensen manueel hun timelog in en is er geen link met planning en budget.” De meeste records uit de tabel werden dus manueel ingevuld, waarbij de PlanningTaskID en BudgetNodeID geen in te vullen velden zijn. Integendeel, het ziet ernaar uit deze velden voorgedefinieerd kunnen worden zodat gebruikers timelogsuggesties zullen krijgen die ze enkel maar hoeven te accepteren. De gepresteerde uren, zoals die worden bijgehouden onder het Duration_Hours attribuut, ten slotte, vormen wel een zeer interessante bron van informatie waarmee later in het onderzoek zeker nog aan de slag gegaan zal worden. De waarden van dit attribuut zullen mee geaggregeerd worden, wat effenaf een zinvolle actie is.

Wanneer we bovenstaande selectie uitvoeren^[Q7] bekommen we een resultaat zoals in *Tabel 8*. Deze tabel toont een fragment uit de nieuw gecreëerde ProjectTeams-tabel. Dergelijke tabel geeft meteen een overzicht van de teamleden per project en wanneer een teamlid bovendien meerdere functies (tegelijk of achtereenvolgens) bekleed heeft binnen het project dan wordt dit ook duidelijk aangegeven doordat voor die user de verschillende door hem uitgeoefende functies meerdere lijnen zullen beslaan. De vierde kolom uit deze tabel geeft het totaal aantal gepresteerde uren aan per teamlid per functie die hij binnen het project heeft opgenomen. Om met het voorbeeld uit het eerste deel van dit hoofdstuk verder te werken, kan men zien dat voor project met ID '9021' personeelslid met ID '171' twee verschillende functies heeft gedragen binnen project. Enerzijds was hij technisch developer (FunctionID = 17) in welke hoedanigheid hij 16 uren heeft gewerkt voor het project, en anderzijds heeft hij ook als grafisch developer (FunctionID = 18) 26 uren aan het project gepresteerd.

ProjectID	UserID	FunctionID	TotalUserFnctHrsOnProject
...
9020	381	17	34.00
9021	101	1	8.00
9021	159	19	5.00
9021	171	17	16.00
9021	171	18	26.00
9021	199	6	2.00
9021	222	86	9.50
9021	258	89	2.00
9021	263	11	6.50
9021	298	19	16.00
9021	298	97	32.50
9021	299	3	5.00
9021	304	29	3.50
9021	313	4	144.00
9021	337	10	1.00
9021	381	17	46.00
9021	401	87	19.00
9021	466	6	1.50
9022	101	1	8.00
...

Tabel 8: Een fragment uit de nieuw gecreëerde ProjectTeams-tabel

Men kan zich afvragen waarom ervoor gekozen werd om met de weinigzeggende IDs te werken voor de projecten, de gebruikers en hun functies. Vanuit het eerste gedacht zou men er misschien voor opteren om de IDs te vervangen door hun overeenkomstige en meer betekenisvolle (code)namen. Echter, in de wereld van computertaal en van databases blijkt het werken met numerieke data vaak veel eenvoudiger dan het werken met tekstuele gegevens. Zo zouden voor de gebruikersnaam, wil men de numerieke UserID vervangen door de naam van die persoon, er al minstens twee velden gecreëerd moeten worden om die persoon voor te stellen, namelijk diens voornaam en achternaam. Echter, zelfs dan nog kunnen personen niet uniek identificeerbaar zijn, bijvoorbeeld wanneer er twee personen met dezelfde naam werken in het bedrijf. Om die reden is het aangegeven te werken met IDs en bovendien vereenvoudigt dit het rekenwerk voor de computer of programmeur vaak. Wil men dus de naam of codenaam van een specifiek attribuut uit de tabel weten, dan moet men dit apart opzoeken in de diverse tabellen.

In het volgende deel van dit onderzoek zullen verschillende methodes en invalshoeken bekeken worden om de projectteams uit de ProjectTeams-tabel te karakteriseren. Hiertoe zullen reeds aanwezige en nieuwe attributen worden aangesproken en met elkaar in verband worden gebracht zodoende nieuwe informatie (van een hogere granulariteit) met betrekking tot de projectteams te vergaren. Deze informatie kan dan worden bijgevoegd in de bovenstaande ProjectTeams-tabel, of in een tweede nieuwe tabel die nog gecreëerd zal worden.

Hoofdstuk 4

Projecten typeren

De voorgaande hoofdstukken van deze thesis handelden over het verkennen van het bedrijf en meer nog over het verkennen de dataset. De omvang (en voor een stuk reeds ook de mogelijkheden en tekortkomingen) van de Dogglenet database werden geanalyseerd en vervolgens werd een afgebakend, betrouwbaar deel van de data geselecteerd op basis waarvan een heel aantal projectteams gereconstrueerd werden. We beschikken nu over een set van projecten en de teams die aan die projecten verbonden zijn, bestaande enkel uit de elementaire basisinfo daarover. De volgende stap in het onderzoek, en tevens de opzet van dit hoofdstuk, zal zijn om die projecten en projectteams te beschrijven. Daartoe zal de dataset van de projectteams aangevuld worden met extra informatie, dewelke vervolgens zal worden samengevat in enkele attributen die de projecten en projectteams kunnen beschrijven.

Het analyseren van de projecten of projectteams kan gebeuren aan de hand van verschillende tools. Elke tool, echter, zal toelaten om andere aspecten van de projectteamwerking te onderzoeken. In dit hoofdstuk is de opzet het beschrijven van de projecten en de projectteams, teneinde deze te kunnen typeren. Veel van de informatie die daarvoor nodig is, kan gecreëerd worden onder de vorm van nieuwe (samenvattende) attributen, dewelke met behulp van SQL gegenereerd kunnen worden. Deze tool is een verassend krachtig instrument om databases te manipuleren en zal ons toelaten via diverse invalshoeken (en op verschillende niveaus) informatie met betrekking tot de projectteams te vergaren.

Andere tools die er bestaan om de projectteamwerking te analyseren –zoals Data Mining en Proces Mining– zullen in het volgende hoofdstuk besproken worden. Vaak zullen deze tools een reeds voorbereide dataset vereisen om op verder te bouwen. Ook die kan in zekere mate met behulp van SQL bekomen worden. De query-taal laat immers toe om data op verschillende manieren te filteren of te aggregeren en sommige van de tools uit het volgende hoofdstuk zullen meer betekenisvolle resultaten kunnen opleveren wanneer ze op dergelijke data kunnen ingrijpen. In dit hoofdstuk zullen daarom verschillende manieren onderzocht worden hoe met behulp van SQL interessante informatie over de projecten en projectteams kan bekomen worden met het uiteindelijke doel om een aantal projecttypes uit de data te kunnen stileren.

1. Opbouw hoofdstuk

Het onderzoek in dit hoofdstuk zal opgebouwd worden rond drie topics. Eerst zal gezocht worden naar methodes om de omvang van een project te meten. Aangezien er geen directe aanduiding van de projectgrootte in de database staat, zal er naar alternatieve indicatoren voor die grootte onderzocht moeten worden. Hierbij zal dan, bijvoorbeeld, onder andere de teamgrootte in rekening worden gebracht. De tweede topic waarop zal worden toegespitst betreft de samenstelling van de projectteams en de mate waarin die de diversiteit van een project reflecteert. Met diversiteit wordt hier niet naar een mix van persoonskenmerken verwezen, maar wel naar de 'scope' van het project. Ook de teamsamenstelling zal vanuit verschillende invalshoeken geanalyseerd worden. Tot slot wordt er onderzocht hoe de efficiëntie van de projectteams beter beoordeeld kan worden door middel van enkele nieuwe financiële omvangsindicatoren te ontwikkelen. Een vraag die hier bijvoorbeeld onderzocht kan worden is: Werd het project binnen het budget afgerond?

Bij de zoektocht naar informatie relaterend aan deze drie topics zullen we dan gaandeweg attributen selecteren, bewerken en berekenen, en net zoals het woord 'projectteams' evenzeer slaat op projecten als op teams, dienen we hierbij opzoek te gaan naar zowel attributen die projecten beschrijven als attributen die teams beschrijven. Het is immers niet zinvol om teams te karakteriseren, zonder ze tegen het licht te houden van de projecten waarvoor ze werden samengesteld.

Bij het analyseren van de data moeten we aandacht hebben voor het feit dat informatie op verschillende niveaus verzameld kan en zal moeten worden. Zo kan soms het simpelweg bekijken van de verschillende waarden voor een gegeven attribuut al een interessante insteek bieden. In dit geval betreft men een kolom uit database bij het onderzoek en relateert men deze kolom aan (één van de focuskolommen uit) het onderzoeksobject. De waarden van deze kolom worden hier dus ongemoeid gelaten. Dit is informatie op het laagste granulaire niveau. Vaker echter, zal het interessanter zijn informatie op een hoger niveau te bekijken. In dit geval gaat het dan over samenvattende info of vergelijkende info. In het eerste geval wordt er een aggregatiefunctie uitgevoerd op een attribuut. In het tweede geval gaat het om een vergelijking tussen twee attributen aan de hand van een bewerking. Alle drie de informatietypes zullen in deze analyse aan bod komen. Dat is ook logisch. Wil men immers complexere informatie bekomen, dan dient deze te worden opgebouwd uit bouwstenen van lagere orde informatie.

2. Voorbereiding tabellen

Vooraleer het onderzoek naar goede indicatoren voor de projectgrootte kan worden aangevat, moeten er eerst twee nieuwe tabellen worden aangemaakt om de nieuwe informatie in op te slaan. We starten vanaf de ProjectTeams-tabel die in het vorige hoofdstuk werd aangemaakt om de basisgegevens van de gereconstrueerde projectteams in op te slaan. Een fragment uit die tabel werd toen getoond in *Tabel 9* – We herhalen voor de eenvoudigheid hieronder nogmaals een deel van die tabel. Bij het aanmaken van de ProjectTeams-tabel hebben we bepaald dat de FunctionID een onontbeerlijk attribuut is voor het identificeren van de projectteams. Het is immers niet nuttig te weten welke personen er in een team zitten zonder ook hun functie binnen dat team te kennen. Bovendien kunnen personen meerdere functies uitoefenen en dient dus geweten te zijn welke rollen een persoon binnen het project heeft aangenomen.

ProjectID	UserID	FunctionID	TotalUserFnctHrsOnProject
...
9020	381	17	34.00
9021	171	17	16.00
9021	171	18	26.00
...

Tabel 9: Fragment uit ProjectTeams-tabel (HERHAALD)

In totaal maakt de database melding van 139 verschillende functies. 52 daarvan staan aangeduid als gearchiveerd. Dit wil zeggen dat ze momenteel niet meer gebruikt worden. Gegeven deze grote verscheidenheid aan functies, echter, zal het minder interessant zijn om de teamsamenstelling op het niveau van de functies te bekijken. Nuttiger zou zijn om deze functies op een hoger level te analyseren. Dit kan gedaan worden via de functiegroepen. De functiegroepen overkoepelen namelijk sets van functies die min of meer samen horen. Via de functiegroepen kan dan de mix van mensen in een projectteam worden geanalyseerd.

Een andere zienswijze om de mix van mensen in een projectteam te analyseren, is om de teamleden te groeperen op hun plaats binnen het bedrijf, ongeacht hun rol bij een project. Dit is mogelijk via de DepartmentID. Terwijl de FunctionID's specifiek op rollen bij een project slaan, hoeven de DepartmentID's niet uitsluitend aan projecten te relateren. Ze kunnen op elke positie binnen de bedrijfswerking van Boondoggle (front-end of back-end) duiden. Deze typering zal wellicht minder interessant zijn als projectenkaracterisering, maar kan toch voor het beleid en de structuur van de bedrijfsindeling bij Boondoggle eventueel onverwachte inzichten opleveren.

Een derde wijze waarop de teamleden van een project kunnen worden ingedeeld, is op basis van hun hoofd-businessunit (hun voornaamste “werkplek”, zeg maar) binnen het bedrijf. Merk op dat ook projecten een BusinessUnitID hebben, dat aanduidt binnen welke bedrijfs-

unit het project wordt uitgevoerd. Het hoeft echter nog niet zo te zijn omdat een project binnen een bepaalde businessunit wordt afgehandeld, dat dan ook alle projectteamleden dezelfde businessunit als hun hoofd-businessunit (of voornaamste “werkplek”) zullen hebben. Daarom kan het interessant zijn om na te gaan of deze twee businessunitID’s doorgaans overeenstemmen, dan wel cross-businessunit collaboraties reflecteren.

De drie hierboven aangehaalde wijzen om de mix van personeelsleden binnen een projectteam op een meer geaggregeerd level te analyseren, worden gecombineerd in één query die de ProjectTeams-tabel zal uitbreiden tot de ProjectTeamsEXTENDED-tabel.^[Q8] Ter wille van de leesbaarheid zal de TotalUserFnctHrsOnProject-kolom uit de oorspronkelijke ProjectTeams-tabel (zie *Tabel 9*) in de illustratie in deze tekst even worden weggelaten, en zal bovendien de BusinessUnitID van het project als extra kolom mee in die illustratie worden opgenomen (hoewel dit attribuut eigenlijk niet thuis hoort in de ProjectTeamsEXTENDED-tabel). Het resultaat wordt getoond in *Tabel 10*. Wat betreft het resultaat van de query kunnen we bovendien stellen dat het aantal rijen in de ProjectTeamsEXTENDED-tabel nog steeds overeenkomt met het aantal rijen in de ProjectTeams-tabel (namelijk 16215 rijen) en dat er daarnaast geen enkele NULL-waarden in de nieuw toegevoegde kolommen staan. De query kon bijgevolg de verschillende attributen perfect één op één mappen en de database vertoont bovendien geen lacunes of inconsistenties met betrekking tot deze attributen.

ProjectID	UserID	FunctionID	FunctionGroupID	DepartmentID	UserMain_ BusinessUnitID	Project_ BusinessUnitID
...
9021	263	11	6	16	1	1
9021	298	19	2	15	1	1
9021	298	97	7	15	1	1
9021	299	3	1	2	1	1
9021	304	29	4	16	1	1
9021	313	4	1	2	1	1
9021	337	10	4	16	1	1
9021	381	17	6	1	1	1
9021	401	87	4	16	1	1
9021	466	6	3	8	1	1
9022	101	1	1	2	1	2
9022	337	10	4	16	1	2
9022	339	2	1	2	2	2
9022	380	3	1	2	2	2
9022	385	17	6	1	2	2
...

Tabel 10: Een fragment uit de ProjectTeamsEXTENDED-tabel (met de TotalUserFnctHrsOnProject-kolom weggelaten en met ProjectBusinessUnitID als extra toegevoegde kolom –louter ter illustratie)

Informatie die men rechtstreeks uit bovenstaande tabel kan afleiden is bijvoorbeeld dat persoon 298 bij het project 9021 twee verschillende functies uitoefende. Deze functies zijn bovendien dermate verschillend dat ze onder een verschillende functiegroep vallen, namelijk functiegroepen 2 (= strategy) en 7 (= other). DepartmentID en UserMainBusinessUnitID

zijn echter niet aan de functie verbonden, maar wel aan de persoon zelf. Voor elke lijn van persoon 298 binnen deze tabel zal het departement dus 15 (= strategy) en de hoofd-bedrijfseenheid 1 (= Agency) zijn.

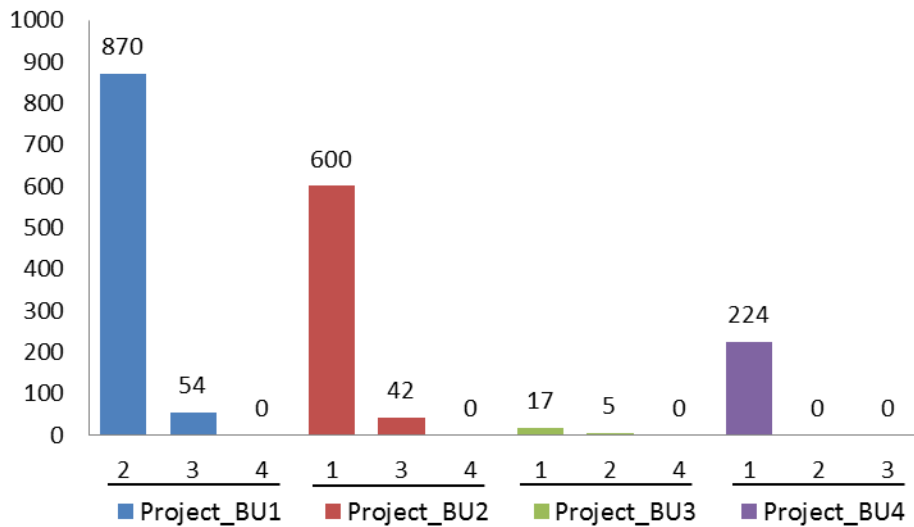
Een tweede gegeven wat men rechtstreeks kan afleiden uit de tabel is dat de hoofd-businessunit van de teamleden inderdaad kan verschillen van de businessunit van het project. Project 9022 werd bijvoorbeeld uitgevoerd binnen businessunit 2 (= Bright). Alle instanties uit de tabel gerelateerd aan dit projectnummer zullen bijgevolg een ProjectBusinessUnitID van 2 hebben. In de kolom daarnaast, echter, kan men aflezen dat niet alle users als hoofd-businessunit 2 hebben. Integendeel, personen 101 en 337 blijken voornamelijk voor businessunit 1 (= Boondoggle Agency) te werken. Cross-businessunit collaboraties zijn dus mogelijk. Dit is echter geen grote verrassing. Men kan zich afvragen waarom dit attribuut in de eerste plaats anders in het leven werd geroepen.

Wat wel interessanter kan zijn om te onderzoeken is in welke mate dergelijke cross-businessunit collaboraties zich voordoen. Wat is bijvoorbeeld de frequentie waarmee beide kolommen van elkaar verschillen? Uit de database kunnen we afleiden dat voor 1905 instanties de User_MainBusinessUnitID niet overeenkomt met de Project_BusinessUnitID. De personen die meerdere functies uitoefenen binnen een projectteam worden hierbij echter meerdere malen geteld. Het aantal personen per project waarvoor die personen hun hoofd-businessunit niet overeenstemt met de businessunit waar het project voor wordt uitgevoerd, komt overeen met 1812 instanties. Voor een totaal van 16215 instanties in de dataset, betekent dit dat gemiddeld 11,2% van de teamleden bij projecten uit een andere businessunit komen. Het gemiddelde is hier echter een bedrieglijke indicator, aangezien een dergelijke mismatch tussen de beide businessunitID's zich bij slechts 867 uit een totaal van 2475 projecten voordeed, of nog: voor 35% van de projecten.

Men kan zich ook afvragen, wanneer de UserMainBusinessUnitID afwijkt van de ProjectBusinessUnitID, of het dan steeds gaat om een persoon uit de Agency-businessunit die bijspringt in een project van een andere unit, of dat dergelijke collaboratie in alle richtingen kan werken. Uit *Tabel 11* kan men afleiden dat het businessunit-overschrijdend werken in alle richtingen kan, met uitzondering van de personeelsleden uit businessunit 4 ("Amsterdam") die dit niet deden. De bevindingen worden grafisch ook nog eens voorgesteld in *Figuur 3*.

Project_BusinessUnitID	User_MainBusinessUnitID	AantalTeamleden
1	2	870
1	3	54
2	1	600
2	3	42
3	1	17
3	2	5
4	1	224

Tabel 11: Afwijkende hoofd-businessunit's van teamleden vs. businessunit project



Figuur 3: cross-businessunit collaboraties voorgesteld per project-businessunit

De volgende stap van de analyse verlegt de focus van instanties op het team-niveau naar het project-niveau, waarbij getracht zal worden om uit de eerder low-level-info uit de ProjectTeamsEXTENDED-tabel, verschillende vormen van hogere-orde-informatie te genereren. Er zal op zoek gegaan worden naar attributen die de omvang van een project en de diversiteit binnen een projectteam kunnen beschrijven. Het gaat hier dus om een aantal kengetallen die per project zullen samenvatten hoe de mix van personeelsleden binnen het project kan worden uitgedrukt en hoe de projectgrootte kan worden gemeten. Dergelijke hogere-orde-informatie kan niet langer toegewezen worden aan de instanties uit de ProjectTeams-tabel, maar zal op niveau van het totale project moeten worden gepresenteerd. Er zal bijgevolg een nieuwe tabel moeten worden aangemaakt om de project(team)info in op te slaan, waarbij één (en slechts één) lijn uit die tabel overeen zal komen met één project.

Voor de creatie van deze nieuwe tabel –die we ProjectINFO zullen noemen– zal vertrokken worden vanuit de ProjectsCOMPLETED-tabel. Dit is namelijk de tabel die alle volledige en afgewerkte projecten bevat. Omdat bij het creëren van de nieuwe tabel vertrokken zal worden vanuit de ProjectsCOMPLETED-tabel, zal meteen ook al een selectie gemaakt kunnen worden van kolommen die rechtstreeks interessant kunnen zijn voor de teamsamenstelling en projectgrootte te helpen beoordelen. Uiteraard zal de ProjectID opnieuw minimaal geselecteerd moeten worden om de verschillende projecten te kunnen identificeren. Bovendien zal het ook interessant zijn om de (Project-)BusinessUnitID mee te nemen in de nieuwe tabel om zo een onderscheid te kunnen maken naar het “soort” project. Dit attribuut zal ons toelaten om achteraf te kunnen filteren of groeperen op de eenheid waar het project werd uitgevoerd. Vervolgens zullen ook de AccountManagerID-, ProjectMangerID- en AccountDirectorID-kolommen overgenomen worden in de nieuwe tabel. Geen van de andere kolommen uit de ProjectsCOMPLETED-tabel kunnen voor deze eerste topics uit de analyse op zich meteen interessante informatie bijdragen, hoewel men ervoor kan opteren om hier

de BrandID of IsInternal attributen mee te nemen naar de nieuwe tabel teneinde projecten een zekere context te kunnen geven. Indien men bovenstaande in een query giet,^[Q9] bekomt men een resultaat zoals getoond in *Tabel 12*.

ProjectID	BrandID	IsInternal	PrBUnitID	AccountMngrID	ProjectMngrID	AccountDrtrorID
...
5989	633	0	1	135	117	NULL
5990	612	0	1	233	193	204
5991	1	1	1	134	NULL	NULL
5992	576	0	1	134	NULL	264
...

Tabel 12: Een fragment uit de basis ProjectsINFO-tabel (waarop verder gebouwd zal worden)

Deze ProjectsINFO-tabel zullen we vanaf nu systematisch uitbreiden met kengetallen over de omvang van de projecten uit die tabel. Ook in de volgende delen van dit hoofdstuk zal deze tabel verder worden aangevuld met info over de andere topics die nog aan bod zullen komen.

3. Omvang project

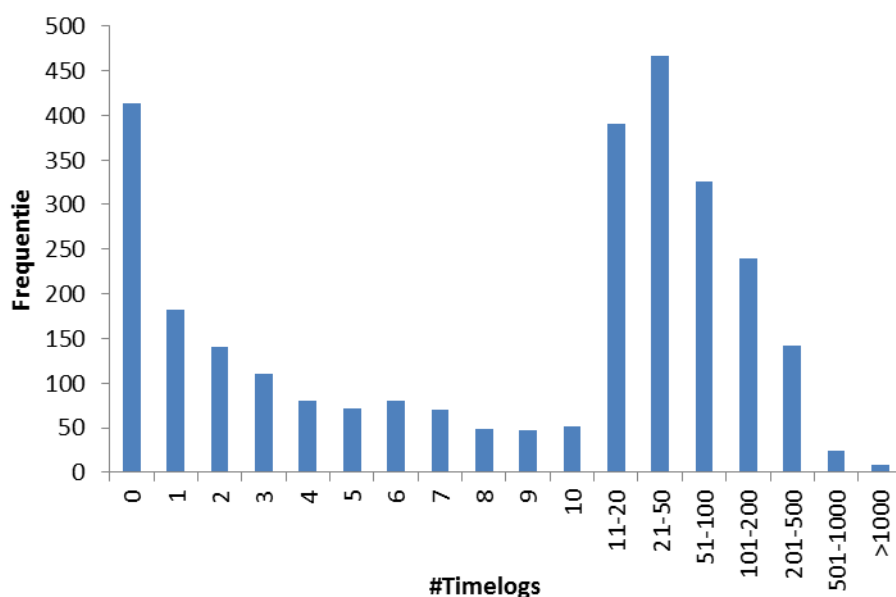
De Dogglenet database bevat geen rechtstreekse indicator die de omvang van een project beschrijft. Nochtans, indien men projecten wil typeren, dan is de omvang zeker één van de pijlers die hierin een rol zal spelen. In deze sectie zullen we daarom bekijken hoe de omvang van een project kan worden afgeleid. Een eerste manier waarop dit kan, is via het tellen van het aantal timelogs die er voor een bepaald ProjectID in de TimeLogsFILTERED-tabel werden gemaakt. De idee hierachter is dat het aantal timelogs enigszins de hoeveelheid verricht werk of het aantal uitgevoerde taken reflecteert. Toch is dit geen erg precieze indicator omdat personen hun uitgevoerde activiteiten zowel nauwgezet in fijne stukjes als in één grote brok kunnen rapporteren. Het resultaat voor dit nieuwe attribuut wordt voorgesteld in *Tabel 13*.^[Q10]

ProjectID	BrandID	IsInternal	PrBUnitID	#TimeLogs	(...)
...
5988	634	0	1	1	(...)
5989	633	0	1	NULL	(...)
5990	612	0	1	15	(...)
5991	1	1	1	100	(...)
5992	576	0	1	14	(...)
...

Tabel 13: Fragment uit de met het aantal timelogs uitgebreide ProjectsINFO-tabel

Zoals men in die tabel kan zien, kan het aantal timelogs per project aanzienlijk verschillen. Langs de ene kant zijn er projecten met veel (bv. 100, of meer) timelogs. Aan het andere uiteinde zijn er ook projecten met zeer weinig of zelfs geen logs. Aan project 5988 kan bijvoorbeeld maar 1 log gelinkt worden. Die betreffende log vermeldt als activiteit “meeting @ Dexia”. Het project waaraan die log gelinkt is, werd uitgevoerd voor Dexia SA (= Brand 634) en heeft als projectnaam de omschrijving “Guidelines: internet-intranet”. Het lijkt hier inderdaad om een zeer klein project te gaan, maar men kan zich afvragen of er niet meer activiteiten voor het project werden uitgevoerd. Misschien ontbreekt er een deel van de informatie? Vooraleer hier meer over uit te weiden, zullen we eerst het volgende project uit de tabel bespreken. Voor project 5989, werden er namelijk geen activiteiten gelogd in de TimeLogs-tabel. Gaat het hier om een spookproject? Het lijkt erop van niet. De omschrijving voor het project luidt namelijk “Brouwerij Haacht briefing”. Daar deze naam van het project meer weg heeft van de omschrijving van een activiteit lijkt het hier wederom om een zeer klein project te gaan, maar werd er ditmaal niet de moeite genomen om de gepaarde activiteit ook te loggen in de database.

Om een beter idee te krijgen van het aantal timelogs die voor de verschillende projecten geregistreerd werden, werd een frequentieverdeling voor het #TimeLogs-attribuut opgesteld, dewelke getoond wordt in *Figuur 4*. Van de 2889 projecten die er in totaal in de ProjectsINFO-tabel staan, zijn er 414 projecten waarvoor er geen activiteiten in de TimeLogs-tabel gelogd werden. Nog eens 879 projecten hebben in totaal slechts 10 of minder gelogde activiteiten. Slechts één vierde (740 uit 2889) van de projecten heeft meer dan 50 logs. Grofweg kunnen we dus stellen dat ongeveer de helft van de projecten in de dataset vrij tot zeer kleine projecten betreffen en dat er over het algemeen eerder weinig timelogs aan projecten verbonden kunnen worden.



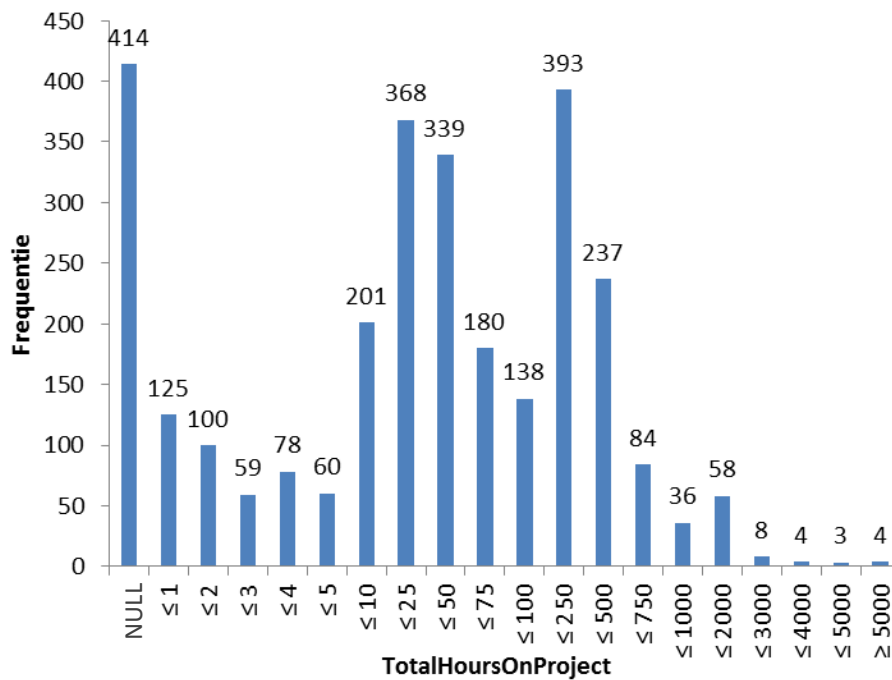
Figuur 4: Frequentieverdeling van het aantal timelogs per project

Gegeven de algemene trend van het eerder beperkte aantal logs per project en gegeven dat er bovendien heel wat projecten zijn zonder zelfs maar één gelogde activiteit, lijken er twee zaken aan de hand te zijn hier. Enerzijds lijken ze bij Boondoggle de policy te hebben om grote projecten op te splitsen in kleinere deelprojecten en deze als een apart “project” te vermelden in de database. Anderzijds kan het zijn dat personeelsleden een zekere laksheid vertonen met betrekking tot het invullen van de timelogs en dat bijgevolg een deel van de activiteiten die voor een project werden uitgevoerd niet in de TimeLogs-tabel geregistreerd staan. Beide scenario's zijn niet meteen het meest welkome nieuws voor het onderzoek in deze thesis. Het ontbreken van uitgevoerde activiteiten in de timelogs, laat ons immers met een onvolledig beeld van de werkelijkheid achter. Het opsplitsen van projecten in deelprojecten zorgt er dan weer voor dat we met een gefragmenteerde werkelijkheid aan de slag moeten en dat een globale evaluatie van de projecten niet mogelijk zal zijn. Er is in de database immers geen constructie die aantoont welke deelprojecten al dan niet samen horen (mocht het inderdaad zo zijn dat in de database enkel deelprojecten getoond worden). Deels uit noodzaak, deels omwille van de niettemin informatieve waarde van de dataset zal toch worden verder gewerkt met de dataset zoals ze is en zullen geen projecten gefilterd worden. Ook de projecten zonder timelogs zullen behouden blijven als referentie voor welke projecten geen logs hebben.

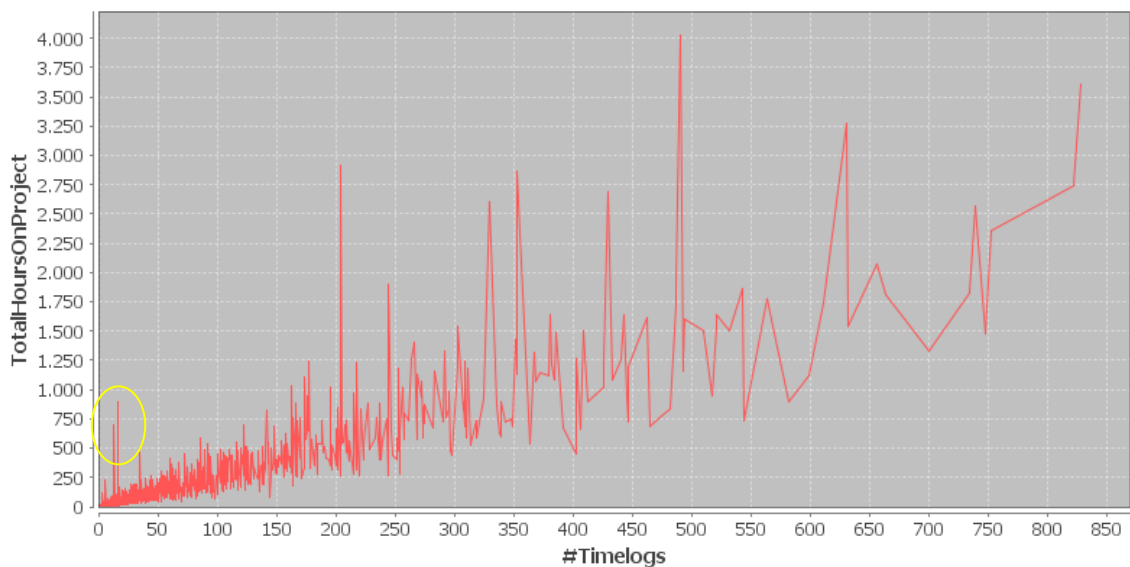
Het aantal timelogs hoeft dus geen goede indicator voor de projectgrootte te zijn. Een groter aantal timelogs kan immers betekenen dat het inderdaad een groter project betreft, of dat de teamleden hun activiteiten zorgvuldiger (“vollediger”) gelogd hebben, ofwel dat ze hun taken in een meer gefragmenteerde vorm (als kleinere sub-taken) hebben gelogd. Een bijkomstige indicator voor de projectgrootte moet dus worden gezocht, en deze kan opnieuw in de TimeLogs-tabel worden gevonden bovendien. Elke timelog registreert namelijk hoelang een bepaalde persoon aan een bepaald project gewerkt heeft. Indien men bijgevolg het totaal aantal gepresteerde uren op een project berekent, bekomt men een tweede indicator van de projectgrootte, die men bovendien kan contrasteren met de eerste. Ook deze nieuwe indicator is echter niet zonder gebreken. Gezien het (vermoedelijke) probleem van de activiteiten die niet gelogd werden in de timelogs-tabel (zoals gedemonstreerd door de vele “lege” projecten) zal opnieuw geen volledig beeld van de verschillende projectgroottes gegenereerd kunnen worden. Niettegenstaande biedt deze nieuwe indicator een interessante blik op de projecten. Een fragment uit de met het TotalHoursOnProject uitgebreide ProjectsINFO-tabel wordt getoond in *Tabel 14*. Op het eerste zicht lijken het aantal timelogs en het totaal aantal uren gewerkt op het project zich min of meer evenredig te verhouden. Om echter een beter beeld te krijgen op de precieze relatie tussen de twee indicatoren werd eerst een frequentieverdeling opgesteld voor het nieuwe attribuut om een idee te krijgen over de spreiding en het bereik van de waarden voor dit attribuut. Deze wordt getoond in *Figuur 5*. Opnieuw valt het grote aantal kleine waarden op. De projecten zonder info (de NULL-projecten) genegeerd, werden voor 623 van de 2889 projecten, grofweg 1/5^e

ProjectID	#TimeLogs	TotalHoursOnProject	(...)
...
5988	1	3.00000	(...)
5989	NULL	NULL	(...)
5990	15	34.00000	(...)
5991	100	133.00000	(...)
5992	14	35.50000	(...)
...

Tabel 14: Fragment uit de ProjectsINFO-tabel die werd uitgebreid met het totaal aantal gepresteerde uren per project.



Figuur 5: Frequentieverdeling van het totaal aantal gepresteerde uren per project



Figuur 6: Gesorteerde (#TimeLogs, TotalHours)-paren tegen elkaar geplott

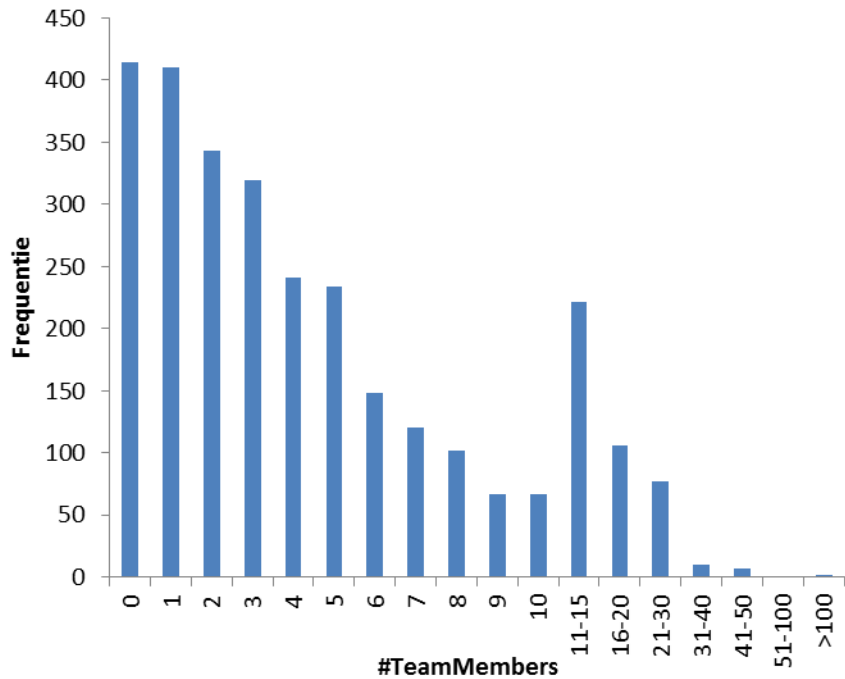
van de projecten, maar 10 uren of minder, in totaal, aan het project gewerkt. Aan het andere uiteinde zijn er projecten met duizenden gepresteerde uren. De vier absolute uitlopers, echter, (de projecten die meer dan 5000 gepresteerde uren hebben) zijn projecten 7127 (34.849 uren), 6513 (21.425 uren), 6635 (6296 uren) en 9414 (5062 uren). Vervolgens kunnen beide indicatoren tegen elkaar worden uitgezet om na te gaan of hun relatie inderdaad een zekere evenredigheid vertoont, zoals getoond in *Figuur 6*. Om deze grafiek te bekomen werden de (#TimeLogs, TotalHoursOnProject)-paren tweevoudig gesorteerd. Eerst op het aantal timelogs en vervolgens op het totale aantal aan het project gepresteerde uren. Daarna werden de negen grootste uitschieters en de NULL-projecten uit de lijst verwijderd en werden de twee variabelen tegen elkaar uitgezet. Uit de grafiek blijkt dat de twee variabelen inderdaad een goede evenredigheid vertonen, met de grotere uitschieters bij de grotere projecten. Hoe verder naar rechts in de grafiek, echter, hoe “ijler” het aantal observaties. Er zijn immers aanzienlijk meer kleine projecten. Dit impliceert dat uitschieters links in de grafiek meer zijn significant (zoals die rond 15à20). Ondanks de goede evenredigheid kunnen beide projectgrootte-indicatoren dus toch een andere voorspelling geven van de werkelijke projectomvang.

Een derde evidente manier om de projectgrootte te beoordelen is door het aantal teamleden in het projectteam van het project na te gaan. Hiertoe vertrekken we vanuit de eerder gecreëerde ProjectTeams-tabel. Een fragment uit het resultaat van de uitgevoerde query^[Q11] kan men bekijken in *Tabel 15*.

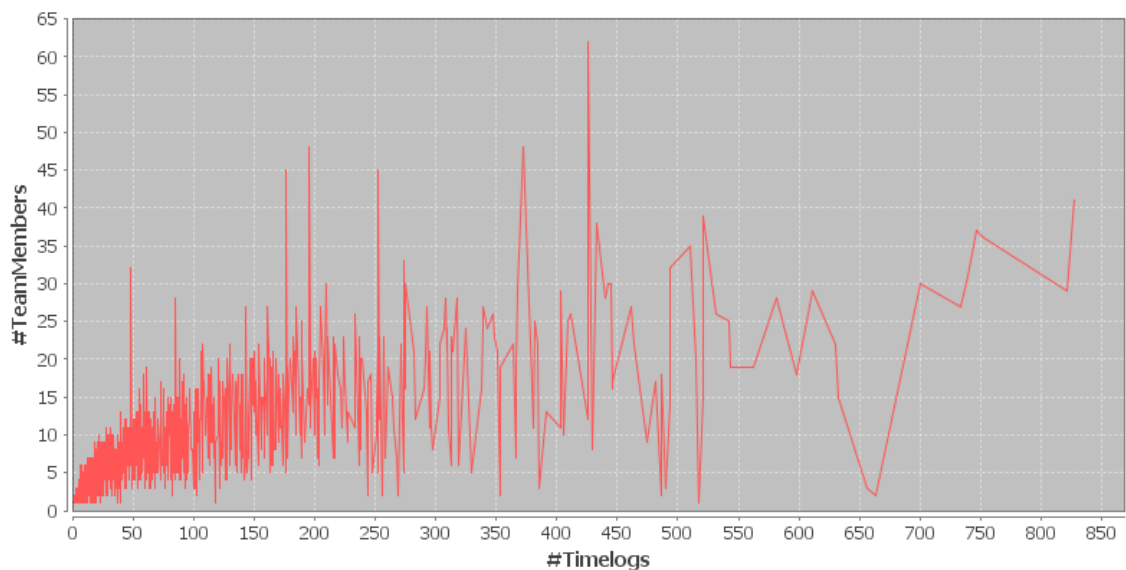
ProjectID	BrandID	IsInternal	PrBUnitID	#TimeLogs	#TeamMembers	(...)
...
5988	634	0	1	1	1	(...)
5989	633	0	1	NULL	NULL	(...)
5990	612	0	1	15	3	(...)
5991	1	1	1	100	5	(...)
5992	576	0	1	14	5	(...)
...

Tabel 15: Fragment uit de met het aantal teamleden uitgebreide ProjectsINFO-tabel

Uit deze query komen geen echt onverwachte dingen naar boven. De instanties waar het aantal TimeLogs NULL was, zullen ook voor het aantal teamleden NULL hebben. De projecten met slechts één timelog, hebben bovendien ook slechts één persoon in het projectteam. Er zijn echter nog meer teams bestaande uit slechts één persoon, maar op 16 uitschieters na komen deze allemaal overeen met projecten die 10 of minder timelogs hebben. Verder neemt het aantal teamleden per project min of meer evenredig toe met het aantal timelogs per project, zoals gedemonstreerd wordt in *Figuur 7* en *Figuur 8*, maar de grotere spreiding van de observaties rond de evenredigheids-regressielijn bij de grafiek in *Figuur 8* ten opzichte van de grafiek in *Figuur 6* valt wel op. Het aantal teamleden is dus minder sterk gecorreleerd met het aantal timelogs (en via de grote correlatie tussen het aantal timelogs en het totaal aantal uren per project kan men voor die tweede hetzelfde besluiten).



Figuur 7: Frequentieverdeling van het aantal teamleden per project



Figuur 8: Gesorteerde ($\#TimeLogs, \#TeamMembers$)-paren tegen elkaar uitgezet

Uit de bovenstaande grafiek werden opnieuw de negen grootste uitschieters volgens het aantal timelogs verwijderd. Men kan zich echter afvragen hoezeer de projecten met de meeste timelogs overeenkomen met de projecten met de grootste teams. Om dit te vergelijken werd *Tabel 16* opgesteld waarbij de top 20 grootste projecten gerangschikt werden enerzijds volgens het aantal timelogs en anderzijds volgens het aantal teamleden. De blauwgeaccenteerde ProjectIDs (in beide zuilen) duiden de overeenstemmende projecten uit beide rangschikkingen aan. Slechts 8 van de 20 projecten (uit de verschillende rangschikkingen) komen overeen. Dit geeft aan dat de twee manieren om de grootte van een project te be-

oordelen niet per se tot dezelfde conclusie zullen leiden. Deze uitspraak wordt nog bijge- staan door de aanwezigheid van enkele opmerkelijke uitschieters in deze top-20-vergelijking. Zoals men in de eerste zuil van de tabel kan zien, zijn er enkele projecten die veel meer timelogs hebben in verhouding tot het aantal teamleden dan andere projecten. (De betref- fende cellen zijn omkaderd.) Deze opmerkelijkheid leidde tot een nieuwe analyse, namelijk de top 20 van de grootste dergelijke uitschieters, dewelke getoond wordt in de derde zuil van onderstaande tabel. Deze uitschieters zijn de minst evenredige paren volgens de #TimeLogs- en #TeamMembers-verdelingen, in die zin dat ze grootste verhouding geven wanneer men het aantal timelogs per project deelt door het aantal teamleden per project. De instanties van deze top 20 staan in de tabel gesorteerd op hun verhouding, met de grootste ratio bovenaan.

Top 20 grootste projecten o.b.v. het aantal timelogs			Top 20 grootste projecten o.b.v. het aantal teammembers			Top 20 uitschieters gesor- teerd op ratio logs/leden		
PrID	#Logs	#Members	PrID	#Logs	#Members	PrID	#Logs	#Members
7127	8947	113	7127	8947	113	6553	517	1
6513	7138	113	6513	7138	113	7125	663	2
6614	1519	48	6527	426	62	6536	486	2
6751	1429	16	6614	1519	48	7157	656	3
9414	1346	27	7151	372	48	9394	353	2
8522	1242	16	6539	195	48	6510	490	3
6467	1121	30	7160	253	45	7152	269	2
6635	1036	41	6516	176	45	6528	386	3
9415	827	41	6635	1036	41	9395	256	2
6615	821	29	9415	827	41	9417	244	2
7948	752	36	6949	521	39	6335	118	1
6585	747	37	6616	433	38	6751	1429	16
6618	739	31	6585	747	37	7127	8947	113
9396	734	27	7948	752	36	8522	1242	16
6224	700	30	7081	510	35	9393	330	5
7125	663	2	6950	274	33	6513	7138	113
7157	656	3	8049	494	32	7166	274	5
6530	632	15	6512	48	32	5948	429	8
7900	630	22	6618	739	31	6500	319	6
7043	611	29	6430	521	31	7255	475	9

Tabel 16: Vergelijking top 20 grootste projecten volgens #TimeLogs en volgens #TeamMembers, + de 20 meest afwijkende projecten ("minst evenredige" uitschieters)

Een vierde factor die eventueel verbonden kan worden aan de projectomvang is de looptijd van een project. Deze kan berekend worden door het aantal dagen te tellen tussen het mo- ment waarop het project werd aangevat en het moment waarop het project werd afgerond. Voor het moment van aanvatten zal de EnterTime van het project uit de Projects-tabel gekozen worden. Deze geeft misschien niet helemaal precies aan wanneer het project werke- lijk werd begonnen –een project kan immers later pas zijn ingegeven in de database– maar algemeen kan men er toch van uitgaan dat een project eerst moet worden aangemaakt in de database vooraleer er op gewerkt zal worden. Bovendien bevat de ProjectStatusses-tabel,

zoals weergegeven in *Tabel 17*, geen beter alternatief voor de precieze aanvangsdatum te bepalen. Als indicator voor wanneer een project werd afgerond hebben we in hoofdstuk 3 (query 5) reeds bepaald dat de *StartTime* uit de *ProjectStatusFlowItems*-tabel van projectinstanties met status 5 een goede keuze is. Die *StartTime* registreert immers wanneer een project een nieuwe status toegewezen kreeg, en een status van '5' staat daarbij voor "afgewerkt project". Dus, de *StartTime* van een gegeven project met *StatusID* '5' geeft aan wanneer het project werd afgerond. (De betreffende *StartTime*'s hebben we in hoofdstuk 3 ook reeds in de *ProjectsCOMPLETED*-tabel gekopieerd.) Bijgevolg kan de *ProjectDuration*, of looptijd van het project, berekend worden door de *EnterTime* van een project af te trekken van de *StartTime* uit de *ProjectsCOMPLETED*-tabel: $ProjectDuration = EnterTime - StartTime$. Een probleem hierbij is wel dat beide kolommen uit het rechterlid van de vergelijking als datatype 'datetime2' hebben. De output in het linkerlid zou echter de looptijd in dagen moeten zijn, voorgesteld door een integer. Met de functie *DATEDIFF()* kan dit probleem worden aangepakt.^[Q13]

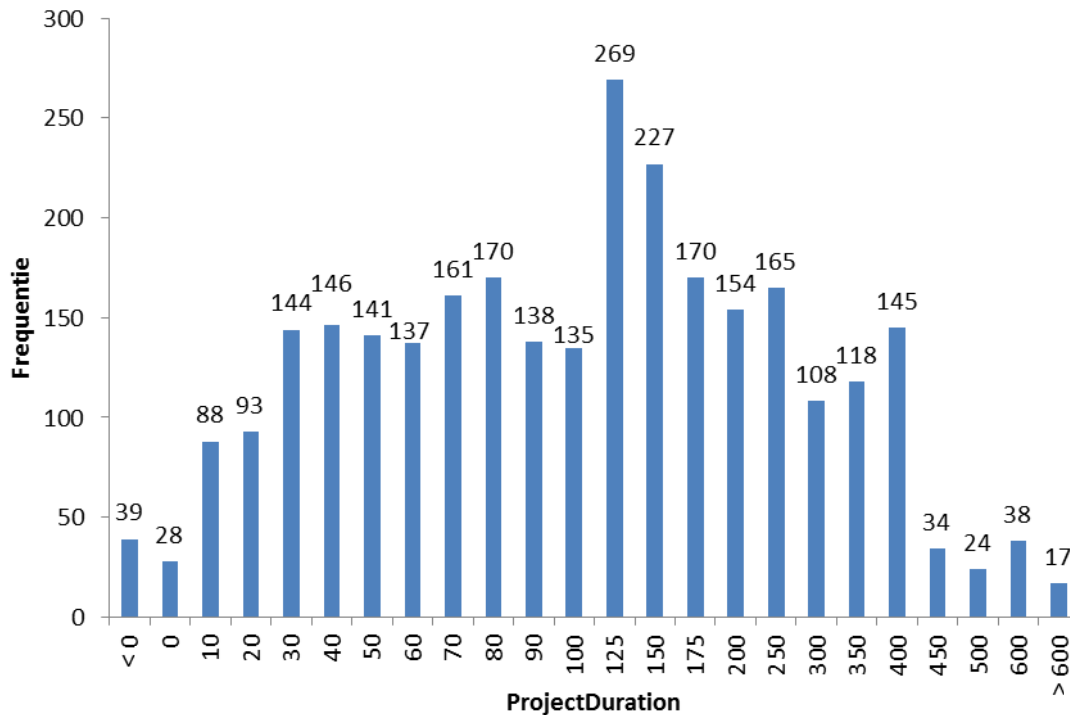
ID	Code	Name	SeqNbr
2	PRS	Prospect	1
4	PRD	Production	2
5	FNP	Finished project	3
6	ARC	Archived	4

Tabel 17: ProjectStatusses-tabel

In *Tabel 18* kan men een fragment uit de *ProjectsINFO*-tabel zien nadat deze werd uitgebreid met de *ProjectDuration*. Enkele dingen vallen op in deze tabel. Ten eerste kunnen projecten zonder overeenkomstige *timelog*-registraties toch op hun omvang getypeerd worden door middel van de looptijd, zoals gedemonstreerd voor project 5989. Ten tweede lijken de looptijd en het totaal aantal aan het project gependeerde uren in zekere mate gelijk op te lopen, maar zoals projecten 5990 en 5992 aantonen, kunnen voor projecten met een gelijkaardig aantal gepresterde uren, de looptijden toch flink uit elkaar liggen. Een eerste vraag die zich hierna opwerpt is wat de range aan waarden is die het *ProjectDuration*-attribuut zoal aanneemt en wat de frequentie is waarmee die waarden geobserveerd werden. Daartoe werd de frequentietabel in *Figuur 9* opgesteld.

ProjectID	#TimeLogs	#TeamMembers	TotalHoursOnProject	ProjectDuration	(...)
...
5988	1	1	3.00000	15	(...)
5989	NULL	NULL	NULL	128	(...)
5990	15	3	34.00000	135	(...)
5991	100	5	133.00000	328	(...)
5992	14	5	35.50000	99	(...)
...

Tabel 18: Fragment uit de met de looptijd uitgebreide ProjectsINFO-tabel



Figuur 9: Frequentieverdeling van de looptijd per project

Wat meteen opvalt uit *Figuur 9* is de grotere spreiding van de verschillende observaties over de mogelijke waarden, dan bij de eerdere omvang-indicatoren die we berekend hebben. Bovendien zijn niet langer 414 instanties nul, zoals bij de andere indicatoren het geval was (en waar dit getal overeenkwam met het aantal projecten zonder logs). In dit geval lijkt het echt om zeer korte projecten te gaan, met name, projecten die binnen één dag werden afgerond. Een andere mogelijke verklaring voor de 0-projecten, kan zijn dat het projecten betreffen die wel werden geïnitieerd, maar niet uitgevoerd en toch de status van afgewerkt werden gegeven. Het eerste scenario is echter waarschijnlijker, gegeven de grote hoeveelheid projecten met bijvoorbeeld 1 timelog en een zeer klein aantal gepresteerde uren.

Een derde opmerkelijkheid uit de frequentieverdeling is dat er 39 projecten bestaan met een negatieve looptijd. Om de reden hiervoor te achterhalen, bekijken we enkele van de projecten met een negatieve looptijd, dewelke getoond worden in *Tabel 19*. Uit de Bright-projecten in de eerste regels van de tabel, kunnen we afleiden dat hoewel ze op 2011 en 2012 slaan, ze pas in 2013 werden ingegeven in de database. Waarschijnlijk werd er tijdens dat ingeven de afwerkingsdatum manueel ingesteld om overeen te komen met de werkelijke datum waarop dit gebeurde. Mogelijks kan het ook zijn dat de EnterTime toch niet het meest geschikte attribuut is voor het bepalen van de aanvangsdatum in de berekening van de looptijd. Echter, uit *Tabel 20* kan men aflezen dat project 9395 (= het eerste project uit *Tabel 19*) bijvoorbeeld geen alternatieve indicator heeft voor het bepalen van de aanvang van een project. We behouden bijgevolg het ProjectDuration-attribuut zoals we dat gedefinieerd hebben en negeren de instanties met een negatieve looptijd (voor dit attribuut).

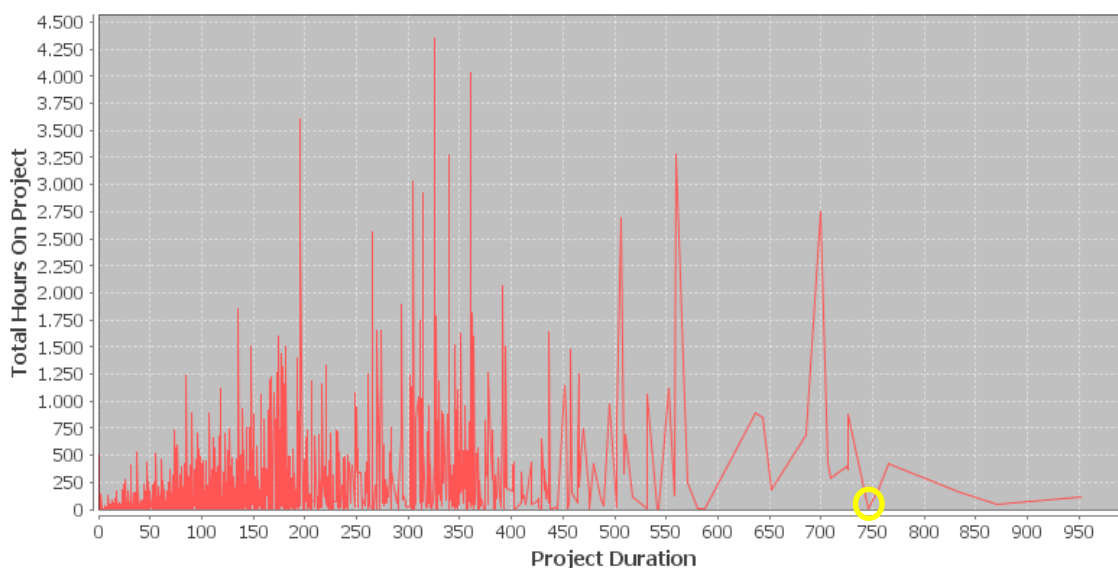
Pr.ID	Code	Name	Ingevoegd	LastChanged	Afgerond	Looptijd
9395	PB-BDBR-9395	Bright management 2011	2013-06-18	2013-07-01	2011-12-31	-535
9396	PB-BDBR-9396	Boondoggle Bright 2011	2013-06-18	2013-07-01	2011-12-31	-535
6370	P-LABS-7134	Drum Trainer 2009	2010-10-21	2010-11-03	2009-12-30	-295
9414	PB-BDBR-9414	Boondoggle Bright 2012	2013-06-19	2013-07-01	2012-12-31	-170
9417	PB-DELH-9417	Overall geboekte uren Michel en Bart 2012	2013-06-19	2013-09-25	2012-12-31	-170
9394	PB-BDBR-9394	Bright management 2012	2013-06-18	2013-07-01	2012-12-30	-170
9095	PB-DELH-9095	Fee Delhaize 2013 DWW - WB	2013-03-21	2013-03-21	2013-01-02	-78
9096	PB-DELH-9096	FEE DELHAIZE 2013 DLL - WB	2013-03-21	2013-03-21	2013-01-02	-78
7399	P-HAAC-8299	Primus strategy & concept dev. wielrennen	2012-03-12	2012-06-07	2011-12-31	-72
9642	PB-IGLD-9642	SLA Iglo DE 2013 (as from May 5)	2013-08-21	2013-10-14	2013-06-30	-52

Tabel 19: Tien (uit de 39) projecten met een negatieve looptijd

ProjectID	ProjectStatusID	Status	StartTime
9395	4	Production	2011-01-01 00:00:00.0000000
9395	5	Finished project	2011-12-31 00:00:00.0000000
9395	6	Archived	2012-01-01 00:00:00.0000000

Tabel 20: Alle StatusID's (statusovergangen) voor project 9395

Een tweede vraag die zich opwerpt is wat precies het verschil is tussen de ProjectDuration en de TotalHoursOnProject-attributen, en hoe deze twee zich verhouden tot elkaar. De ProjectDuration meet de looptijd van een project in het aantal *dagen* en verschilt van het totaal aantal aan het project gependeerde uren (zoals voorgesteld door het TotalHoursOnProject-attribuut) in dat het de spreiding van die gepresteerde uren (of van de uitgevoerde taken) over de tijd heen reflecteert. Bovendien kan worden nagegaan hoezeer deze twee gecorreleerd zijn. Daartoe werden de (ProjectDuration, TotalHoursOnProject)-paren uit de dataset tweevoudig gesorteerd, eerst op looptijd, vervolgens op gepresteerde uren, en werden de allergrootste uitlopers voor de TotalHoursOnProject tijdelijk uit de dataset gefilterd. Vervolgens werden de observaties voor deze twee attributen tegen elkaar uitgezet, zoals wordt getoond in *Figuur 10*.



Figuur 10: Gesorteerde (ProjectDuration, TotalHoursOnProject)-paren tegen elkaar uitgezet

Uit *Figuur 10* komt niet meteen een sterke correlatie naar voor tussen de looptijd van een project en het totaal aantal aan dat project gespendeerde uren. Vooral wanneer de vergelijking tussen deze twee attributen in de figuur gecontrasteerd wordt met die tussen de andere indicatoren voor de omvang van een project, zoals getoond in *Figuur 6* en *Figuur 8*, valt de afwezigheid van zo'n significante correlatie goed op. Voor de projecten met een korte looptijd observeren we veelal wel minder gepresteerde uren, maar dat is ook logisch aangezien aan een project van 2 dagen simpelweg geen 500 uren gewerkt kunnen worden. Naarmate de projectduur toeneemt, echter, kunnen de observaties voor het totaal aantal gepresteerde uren min of meer alle richtingen uitgaan. Vooral in het meest rechtse deel van de grafiek valt het aantal lage TotalHoursOnProject-waarden, in tegenstelling tot wat men zou verwachten, duidelijk op.

Het ProjectDuration-attribuut toont dus inderdaad veeleer de spreiding van het gepresteerde werk dan de hoeveelheid werk. Toch is ook dit een interessante indicator om te beschouwen in een verregaandere analyse, enerzijds omdat het toelaat de omvang van een project zonder timelog-referenties te schatten. Anderzijds omdat het gebruikt zou kunnen worden als een soort van indicator voor de dringendheid waarmee een project werd afgewerkt wanneer men deze indicator contrasteert met de bevindingen van andere indicatoren voor de projectomvang.

Tot slot zullen we, vooraleer we overgaan tot het opstellen van een globale beoordeling voor de projectomvang, nog een vijfde, meer "theoretische" manier bespreken aan de hand waarvan de projectomvang beoordeeld zou kunnen worden. In hoofdstuk 2 werd het onderscheid tussen de accountmanager, de accountdirector en de projectmanager toegelicht. Er werd toen besloten dat elk team een accountmanager heeft (en inderdaad zou moeten hebben) en dat naarmate projecten groter of complexer worden, er ook een projectmanager zal toegewezen worden aan het team. De aanwezigheid van een projectmanager zou dus een indicator kunnen zijn voor de grootte van een team. Het ProjectManagerID-attribuut staat echter al in ProjectsINFO-tabel en bijgevolg kunnen we meteen nagaan of deze theorie ook klopt. Voor de grootste projecten uit de tabel, zowel beoordeeld op basis van de teamgrootte als op basis van het aantal timelogs, echter, blijkt er nagenoeg nooit een ProjectManagerID ingevuld te zijn. Deze theorie klopt dus alvast niet. In hoofdstuk 2 werd tevens ook gezegd dat de accountdirector soms ook de rol van de andere twee managers opneemt en dat de hoedjes ook wel eens durven wisselen. Een tweede theorie zou dan kunnen zijn dat hoe meer van de drie soorten managers een project heeft, hoe groter het project zal zijn. Om het aantal managers eenvoudiger te kunnen afleiden, wordt een nieuwe query^[Q12] opgesteld. De geldigheid van deze hypothese zullen we nagaan op de blauw-gearceerde projecten uit *Tabel 16*, aangezien we deze projecten het meest algemeen kunnen beschouwen als de grootste projecten uit de dataset. Het resultaat van deze query wordt getoond in *Tabel 21*.

ProjectID	AccountManagerID	AccountDirectorID	ProjectManagerID	#Managers
6513	10	NULL	NULL	1
7127	10	NULL	NULL	1
7948	192	134	NULL	2
9414	138	138	NULL	2
9415	138	138	NULL	2
6585	300	286	NULL	2
6614	300	286	NULL	2
6635	135	204	134	3
6467	309	138	138	3

Tabel 21: Het aantal managers voor de 9 algemeen grootste projecten (dat zijn de blauw-gearceerde projecten uit Tabel 16)

Afgezien van de twee allergrootste projecten (6513 en 7127) hebben alle projecten, naast de (verplicht aanwezige) accountmanager ook een accountdirector, maar hebben slechts twee projecten een projectmanager. We kunnen dus slechts voorzichtig positief zijn over de mate waarin het aantal managers de grootte van een project reflecteert. Bijgevolg is het aantal managers geen betrouwbare indicator voor de projectomvang. Afgezien daarvan kan dit attribuut op zich eventueel wel een interessante insteek zijn voor het karakteriseren van projecten. Bijgevolg zal dit attribuut toch opgenomen worden in de ProjectsINFO-tabel.^[Q13]

Globale beoordeling van de omvang van een project

In dit deel van het hoofdstuk werd er op zoek gegaan naar mogelijkheden om de omvang van een project te bepalen. Diverse indicatoren werden daarbij beschouwd: het aantal time-logs die het verrichte werk met betrekking tot een project reflecteren; het aantal teamleden in het projectteam dat aan het project gewerkt heeft; het totale aantal aan het project gespendeerde uren; de looptijd van het project; en de omvang van het project zijn managementteam. Geen enkele indicator op zich bood echter voldoende volledige informatie om een gedegen uitspraak te kunnen doen over de werkelijke omvang van een project, maar belichtte telkens een ander aspect van die projectomvang. Een globale beoordeling van de omvang van een uitgevoerd project zal bijgevolg niet door één attribuut kunnen gegeven worden. Willen we echter projecten toch aan de hand van één attribuut kunnen typeren op hun omvang, en niet op slechts een deelaspect van die omvang, dan zal dus een nieuw attribuut gecreëerd moeten worden dat de informatie uit de verschillende indicatoren van de omvang samenbrengt in één waarde.

Voor de creatie van de globale projectomvang-indicator, zullen alle vijf de indicatoren die we eerder berekend hebben in rekening worden genomen, ongeacht het feit dat sommige van deze indicatoren minder doorslaggevend of relevant bleken te zijn voor het bepalen van de projectomvang. Toch zal er wel een onderscheid gemaakt worden in de mate waarin elke indicator doorweegt op het globale resultaat. Dit zullen we doen door de waarden van elke indicator te verdelen over een aantal scores. Door middel van de hoogte van die scores zal dan de doorslagkracht van elk attribuut afgesteld kunnen worden. Voor het uiteindelijke

resultaat herleiden we de som van die scores tot een geheel getal tussen 1 en 5 (beide grenzen inbegrepen). We creëren ditmaal dus geen onbegrensd attribuut, maar wel een soort selectie-attribuut, op basis waarvan dan eenvoudig op teamgrootte gefilterd kan worden.

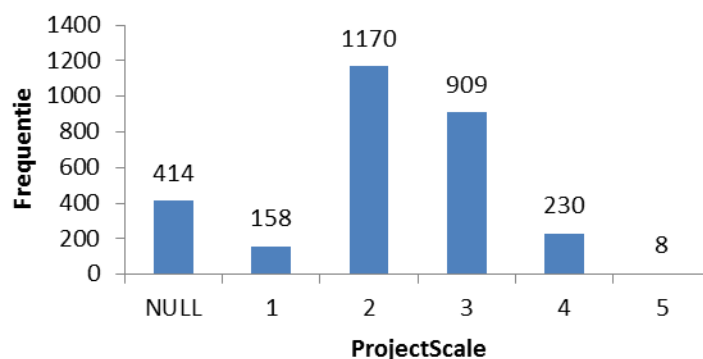
De allocatieverdeling op basis waarvan de verschillende attribuut-observaties naar een score zullen worden gemapped, wordt gepresenteerd in *Tabel 22*. De allocatiesleutels die hierbij voor de verschillende attributen gebruikt werden, zijn deels gebaseerd op de frequentieverdelingen die eerder in dit hoofdstuk getoond werden, en deels op de eigen interpretatie van de informatiewaarde voor het analyseren van de projecttypes. Het is echter niet zinvol deze sleutels te zeer aan te passen aan de observatiedata, teneinde deze mooi evenredig over het aantal categorieën te spreiden, bijvoorbeeld. Er zitten in de dataset immers aanzienlijk meer kleine projecten dan grote projecten. Wil men een zekere informatiewaarde meegeven aan de categorieën en de (uiteindelijke, globale) omvangindicator, dan moeten deze in de eerste plaats een realistisch beeld van verschillende projectgroottes geven, ook al zullen bepaalde grootte-categorieën dan nauwelijks vertegenwoordigd zijn in de database. Bovendien zijn de allocatiesleutels gekozen vanuit de eigen inschatting van de beste informatiewaarde. Echter, beter geplaatste personen die meer inzicht hebben in de aard van de verschillende projecten die bij Boondoggle de revue passeren, zullen misschien andere keuzes maken met betrekking tot die verdeelsleutels. Deze kunnen bij een eventuele implementatie van het werk uit deze thesis, dan uiteraard altijd nog afgestemd worden.

INDICATOR	Range van waarden die voorkomen in de dataset	Verdeelsleutel/Scorebepaling
#TeamMembers	NULL, 1 — 113	1 → 1
		2-5 → 2
		6-15 → 3
		16-40 → 4
		> 40 → 5
#TimeLogs	NULL, 1 — 8947	1-10 → 1
		11-50 → 2
		51-200 → 3
		201-500 → 4
		>500 → 5
TotalHoursOnPr	NULL, 0 — 34.849	0-10 → 1
		>10-50 → 2
		>50-200 → 3
		>200-1000 → 4
		>1000 → 5
ProjectDuration	-535 – 0 — 951	≤ 0 → 0
		1-30 → 1
		30-365 → 2
		>365 → 3
#Managers	1 — 3	0,1 → 0
		2,3 → 1

Tabel 22: Bepalen van de scores voor de verschillende indicatoren

De eerste drie indicatoren uit de tabel zullen de voornaamste componenten vormen voor het bepalen van de globale projectomvang-score. Het `#TeamMembers` meet immers de omvang van de projectteams die nodig waren om de projecten tot een goed einde te brengen. Het `#TimeLogs` vormt dan weer een indicator voor het aantal uitgevoerde taken, en de `TotalHoursOnProject` geeft aan hoeveel uren het uitvoeren van die taken in beslag heeft genomen. Alle drie de indicatoren kunnen een score krijgen tussen 1 en 5. `ProjectDuration` is in mindere mate van belang, omdat het vooral de spreiding van die taken in rekening brengt. (Toch betreft het hier geen onbelangrijk attribuut.) Daarom zal de score slechts kunnen oplopen tot een maximum van 3. Bovendien voorzien we voor deze indicator ook een 0-waarde, dewelke gebruikt zal worden om het kleine aantal negatieve observaties op te vangen. Tot slot zal ook het `#Managers`-attribuut betrokken worden bij de globale omvang-indicator, hoewel slechts met een beperkte doorweging. Omdat alle projecten een accountmanager hebben (en zouden moeten hebben) wordt het `#Managers = 1` gelijkgeschaald met 0, aangezien dit niets aan de vergelijking bijdraagt.

Eens de scores voor de verschillende (sub-)indicatoren berekend zijn, zullen deze worden opgeteld, waarna ze herschaald zullen worden zodat de uiteindelijke score 1, 2, 3, 4 of 5 zal bedragen. (Wanneer een project de maximumscore behaalt volgens elke indicator, dan zal dit een totaalscore van 19 opleveren. Bijgevolg zal de som van de behaalde sub-scores gedeeld worden door 19 en vervolgens vermenigvuldigd met 5, waarna ze naar een geheel getal zullen worden afgerond.) Het probleem van de 414 instanties met NULL-waarden voor de meeste van de indicatoren, zal worden aangepakt door voor de globale indicator opnieuw NULL-waarden toe te kennen aan de betreffende instanties. De betreffende queries kan men terugvinden in bijlage.^[Q15] Een selectie van de resultaten wordt getoond in *Tabel 23*. Uit de tabel blijkt dat de verschillende gekozen parameters voor de score-toewijzingen vrij goede resultaten opleveren in termen van grootte-scores, nochtans zijn er enkele projecten (zoals project 6510) die misschien niet helemaal de juiste globale score hebben gekregen. Sommige parameters kunnen dus eventueel nog bijgesteld worden. Tot slot willen we weten hoe de projecten uit de dataset ingedeeld werden volgens de nieuwe indicator. *Figuur 11* toont deze indeling. We kunnen opnieuw besluiten dat de dataset voornamelijk kleine projecten bevat.



Figuur 11: Frequentieverdeling van de projecten volgens de nieuw-gecreëerde projectomvang-indicator

Proj. ID	#TeamM	#TeamM SCORE	#TimeL	#TimeL SCORE	TotalHours	TotalHours SCORE	Proj.Dur.	ProjDur SCORE	#Mngrs	#Mngrs SCORE	GLOBAL SCALE SCORE	(...)
8040	NULL	NULL	NULL	NULL	NULL	NULL	0	0	2	1	NULL	(...)
7254	NULL	NULL	NULL	NULL	NULL	NULL	531	3	3	1	NULL	(...)
7889	1	1	5	1	10.00000	1	29	1	3	1	1	(...)
8721	1	1	1	1	0.50000	1	3	1	3	1	1	(...)
9364	1	1	1	1	0.25000	1	23	1	2	1	1	(...)
9118	1	1	8	1	12.50000	2	25	1	1	0	1	(...)
7613	3	2	10	1	8.50000	1	28	1	1	0	1	(...)
6335	1	1	118	3	147.00000	3	109	2	1	0	2	(...)
7029	5	2	9	1	56.50000	3	39	2	3	1	2	(...)
6945	6	3	18	2	28.00000	2	22	1	3	1	2	(...)
6285	6	3	28	2	36.00000	2	140	2	1	0	2	(...)
6510	3	2	490	4	4030.97996	5	360	2	1	0	3	(...)
7399	10	3	53	3	76.50000	3	-72	0	2	1	3	(...)
7506	10	3	64	3	268.00000	4	26	1	1	0	3	(...)
6512	32	4	48	2	74.25000	3	360	2	1	0	3	(...)
7207	17	4	164	3	183.00000	3	352	2	2	1	3	(...)
9393	5	2	330	4	2601.50000	5	196	2	2	1	4	(...)
5983	16	4	120	3	185.50000	3	432	3	2	1	4	(...)
8113	26	4	234	4	761.41666	4	283	2	2	1	4	(...)
7900	22	4	630	5	3276.01666	5	339	2	2	1	4	(...)
6516	45	5	176	3	662.75000	4	360	2	1	0	4	(...)
7021	18	4	598	5	1116.86665	5	553	3	2	1	5	(...)
9415	41	5	827	5	3605.48323	5	195	2	2	1	5	(...)
7127	113	5	8947	5	34848.81634	5	406	3	1	0	5	(...)
...

Tabel 23: Overzicht van de verschillende score-indicatoren voor enkele geselecteerde rijen met betrekking tot de project scale

In dit deel van het hoofdstuk werden verschillende methodes ontwikkeld om de omvang van een project te typeren. Het resultaat is dat men nu zowel kan terugvallen op een globale score om met behulp van één indicator naar de projectomvang te filteren, als op de verschillende berekende observatietotalen en de bijhorende sub-scores om meer de deelaspecten van de projectomvang te kunnen analyseren. Bij het bepalen van de verschillende score-allocaties zijn we echter manueel op zoek gegaan naar goede parameters. We zouden hier graag ook nog even aanstippen dat het afstellen van de parameters voor het toekennen van die scores echter ook met behulp van data mining technieken zou kunnen gebeuren. Hierover zal in het volgende hoofdstuk nog verder worden uitgeweid. Eerst, echter, zullen we onderzoeken hoe de projecten of projectteams, naast hun omvang ook naar hun diversiteit getypeerd kunnen worden. Dit doen we in het volgende deel van dit hoofdstuk.

4. Diversiteit projectteam

In het vorige deel van dit hoofdstuk werden een aantal indicatoren opgesteld met betrekking tot de omvang van een project, met de bedoeling om projecten naar hun grootte te kunnen typeren. Naast die projectomvang kan het echter ook interessant zijn om projecten te typeren op basis van hun diversiteit. De diversiteit van een project kan verstaan worden als de mix van functies binnen een projectteam (d.i. de samenstelling van het projectteam) en de mix van de door hen uitgevoerde activiteiten. Aangezien bepaalde functies met bepaalde taken verbonden kunnen worden, zal dergelijke mix dus de diversiteit van het project (en deels ook de soort van de voor het project uitgevoerde taken) kunnen beschrijven. Dit moet ons alweer een stap dichterblijngen in de richting van het afleiden van enkele projecttypes uit de dataset. Waar in het vorige deel van dit hoofdstuk dus gezocht werd naar indicatoren om de “scale” van de projecten te beoordelen, zal in dit deel op zoek gegaan worden naar methodes om de “scope” van die projecten te beschrijven.

Functies en Functiegroepen

De eerste wijze waarop de diversiteit geanalyseerd kan worden, is via de verschillende functies die de teamleden binnen het project uitoefenden. We hebben eerder in dit hoofdstuk al geoordeeld dat vanwege de grote verscheidenheid aan functies het focussen op dit attribuut te detaillistisch zou zijn. Om die reden werd toen de functiegroep aan de ProjectTeams-tabel toegevoegd. *Tabel 24* toont de verschillende functiegroepen zoals die gedefinieerd staan in de Dogglenet-database. Aangezien er slechts 7 verschillende zulke groepen zijn, laat dit attribuut dus toe om de functiemix op een hoger level te analyseren en te benoemen door middel van een beperkt aantal categorieën.

ID	Code	Name
1	MANAGEMENT	Account and Project Management
2	STRATEGY	Strategy
3	CREATION	Creation
4	DESIGN	Design
5	ANALYSIS	Analysis
6	DEVPROD	Development & Production
7	OTHER	Other

Tabel 24: De verschillende functiegroepen uit Dogglenet.

Aan de hand van de functies en de functiegroepen kan de teamsamenstelling binnen een project op twee verschillende manieren beoordeeld worden. Enerzijds kan een attribuut gecreëerd worden dat de diversiteit van het project reflecteert door het aantal verschillende functies of het aantal functiegroepen die aan bod komen binnen dat project te tellen. Anderzijds kan ook de mix van functies binnen de teams worden voorgesteld door voor elk van

de zeven functiegroepen apart te tellen hoeveel teamleden een functie (of rol) uitoefenen binnen het project die onder die functiegroep vallen. Daarbij moet men aandacht hebben voor het feit dat één teamlid meerdere functies kan uitoefenen binnen een projectteam.

Om te beginnen zullen twee attributen gecreëerd worden die, op basis van de uitgeoefende functies en de functiegroepen, de diversiteit aan uitgevoerde taken en uitgeoefende rollen binnen het project zullen weergeven. De achterliggende assumptie van waaruit hier vertrokken wordt, is dat hoe meer functies er in een gegeven project aan bod komen, hoe diverser het project en dus hoe complexer het project zal zijn. De functie van een werknemer reflecteert immers het specifieke takenpakket dat hij gewoonlijk uitvoert. Verschillende functies betekenen verschillende takenpakketten en bijgevolg zou de diversiteit aan functies binnen een project, de scope of complexiteit van dat project kunnen reflecteren. (Mogelijk zelfs, kunnen deze attributen in zekere mate eveneens als een soort voorspeller gebruikt worden voor de grootte van een project.) De twee attributen die hiertoe zullen worden aangemaakt zijn enerzijds het aantal functies en anderzijds het aantal functiegroepen die voorkomen in een gegeven project.^[Q16]

Vanwege het grote aantal functies in de database zal het eerste attribuut –het aantal functies– niet veel meer zijn dan een simpele indicatie van hoeveel rollen er uitgeoefend zijn bij een project. Door het tweede attribuut –het aantal functiegroepen– naast het eerste te plaatsen, echter, kan geanalyseerd worden of die functies allen min of meer in dezelfde lijn liggen, dan wel diverse domeinen beslaan. Zo kan een groot aantal functies gespreid over één of twee functiegroepen, bijvoorbeeld, duiden op een project met een nauwere doelomschrijving of focus, terwijl een groot aantal functies gespreid over de meeste of zelfs alle functiegroepen kan indiceren dat het een complex en breed project betreft. In *Tabel 25* worden enkele geselecteerde rijen uit de ProjectINFO-tabel getoond, nadat deze werd aangevuld met de twee nieuwe attributen: #Functies en #Functiegroepen per project.

ProjectID	#Managers	#TimeLogs	#TeamMembers	#Functies	#FGroups	#Depart	(...)
...
5989	2	NULL	NULL	NULL	NULL	NULL	(...)
7325	3	17	1	1	1	6	(...)
7844	2	100	13	11	4	3	(...)
7666	1	100	13	10	6	4	(...)
6836	3	403	29	18	4	8	(...)
8116	3	493	29	18	7	1	(...)
7127	1	8947	113	34	7	2	(...)
...

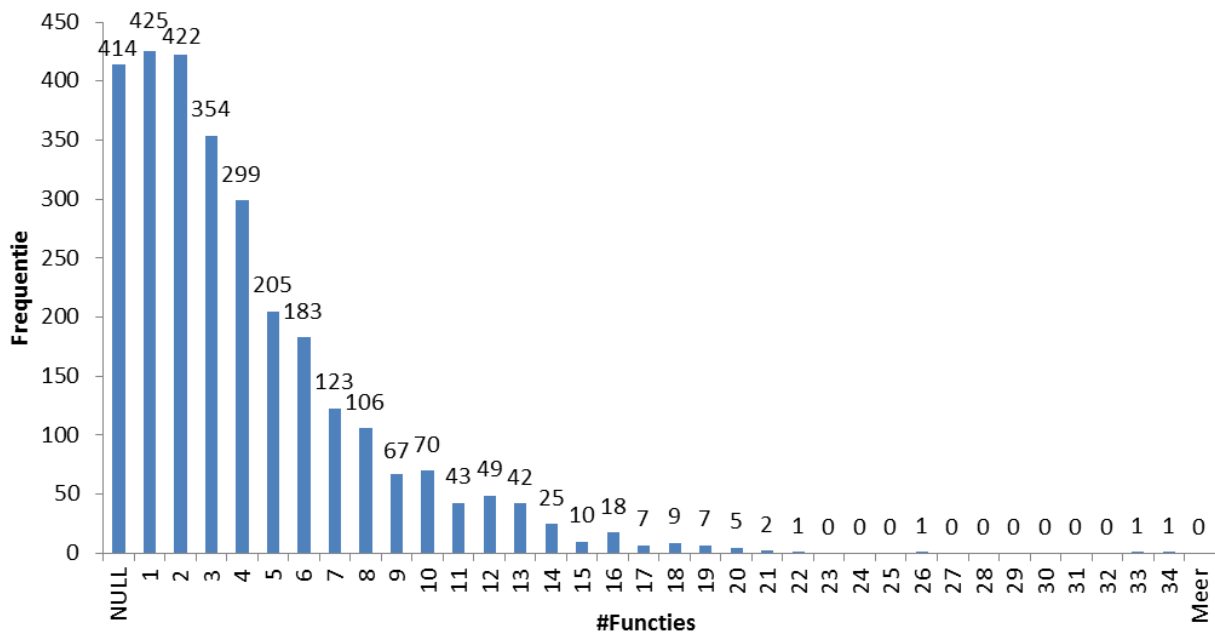
Tabel 25: Enkele geselecteerde instanties uit de ProjectsINFO-tabel na uitbreiding met het aantal functies, het aantal functiegroepen en het aantal departementen per project

De tabel toont enkele opvalligheden. Ten eerste tonen projecten 5989 en 7325 aan dat het aantal managers dat een project heeft, zoals voorgesteld door het #Managers-attribuut, niet overeen hoeft te komen met het aantal teamleden die een management(-achtige) func-

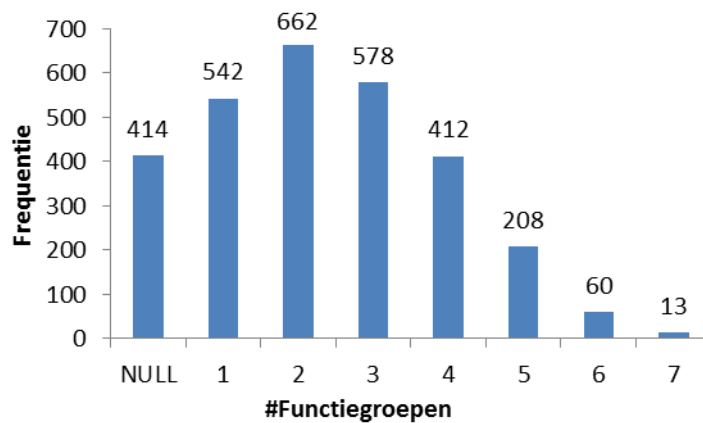
tie uitoefenen, zoals voorgesteld door de functiegroep 1 (zie *Tabel 24*). Project 5989 heeft immers geen overeenkomstige timelogs en dus geen functies (uit functiegroep 1), maar toch twee managers. Project 7325 is dan weer een voorbeeld van één van de vele eenmansprojecten, maar heeft bovendien drie managers. Deze mismatch doet toch wat vreemd aan, vooral omdat functiegroep 1 als omschrijving “Account and Project Management” heeft (zie *Tabel 24*). Mogelijks kan die mismatch wel worden toegeschreven aan het ontbreken van relevante timelogs (die niet in Dogglenet werden gelogd, zoals eerder in dit hoofdstuk reeds besproken werd). Van de vele eenmansprojecten (410 in totaal) waren er bovendien slechts 12 projecten waarbij de projectwerker meer dan 1 functie uitoefende. In die gevallen ging het om een totaal van 2 uitgeoefende functies en in 8 van de 12 instanties behoorden die ook tot 2 verschillende functiegroepen. Kleine projecten zijn dus veelal niet divers.

De volgende twee instanties uit *Tabel 25* tonen twee projecten die even groot zijn, zowel qua aantal timelogs als qua aantal teamleden. Het eerste van die twee projecten vertoont een groter aantal uitgeoefende functies. Het tweede project, echter, vertoont een groter aantal functiegroepen. Als men moest bepalen welke van deze twee projecten het meest divers was, zou de keuze uitgaan naar het tweede project, omdat de functiegroepen diversiteit indiceren op een hoger level, toch kan er niet met volledige zekerheid een uitspraak gedaan worden over de complexiteit van beide projecten. Ook voor de volgende twee instanties uit de voorbeeldtabel, namelijk projecten 6836 en 8116, geldt dat ze ongeveer even groot zijn qua aantal timelogs. Bovendien zijn de projecten ook even groot in termen van het aantal teamleden, en in termen van het aantal verschillende functies die die teamleden gezamenlijk uitoefenen. Van deze twee projecten kan project 8116, met zijn vertegenwoordiging van alle 7 functiegroepen, wel duidelijk als het meer complexe project bestempeld worden. De laatste instantie uit de voorbeeldtabel, tot slot, betreft het grootste project uit de dataset. Het zal niet verbazen dat dit project alle 7 de functiegroepen vertegenwoordigt ziet. Meer zelfs, voor de meeste grote projecten geldt dat ze 7, 6 of 5 verschillende functiegroepen tellen, terwijl tussen de projecten met een klein aantal functiegroepen (1, 2 of 3) nauwelijks grote projecten staan. De assumptie dat het aantal functiegroepen een voorspeller zou kunnen zijn voor de grootte van een project zou dus enigszins kunnen kloppen.

Om een beter idee te krijgen van de relatie tussen het aantal functies die er zoal worden uitgeoefend bij een project en het aantal functiegroepen waartoe die functies behoren, werden drie figuren opgemaakt. *Figuur 12* toont de frequentieverdeling van het aantal functies per project. In lijn met de observaties van bij de vorige project-beschrijvende indicatoren, zijn het groot aantal lage waarden die geobserveerd werden. Inderdaad, de database bevat relatief meer kleine projecten. Wat er hier dus afgeleid lijkt te kunnen worden is dat die kleine projecten veelal geen complexe projecten zullen betreffen, vanwege het beperkte aantal functies die eraan te pas komen. *Figuur 13*, die de frequentieverdeling van het aantal vertegenwoordigde functiegroepen per project verbeeldt, toont dan weer dat die functies bij de meeste projecten onder een beperkt aantal functiegroepen kunnen worden ingedeeld.



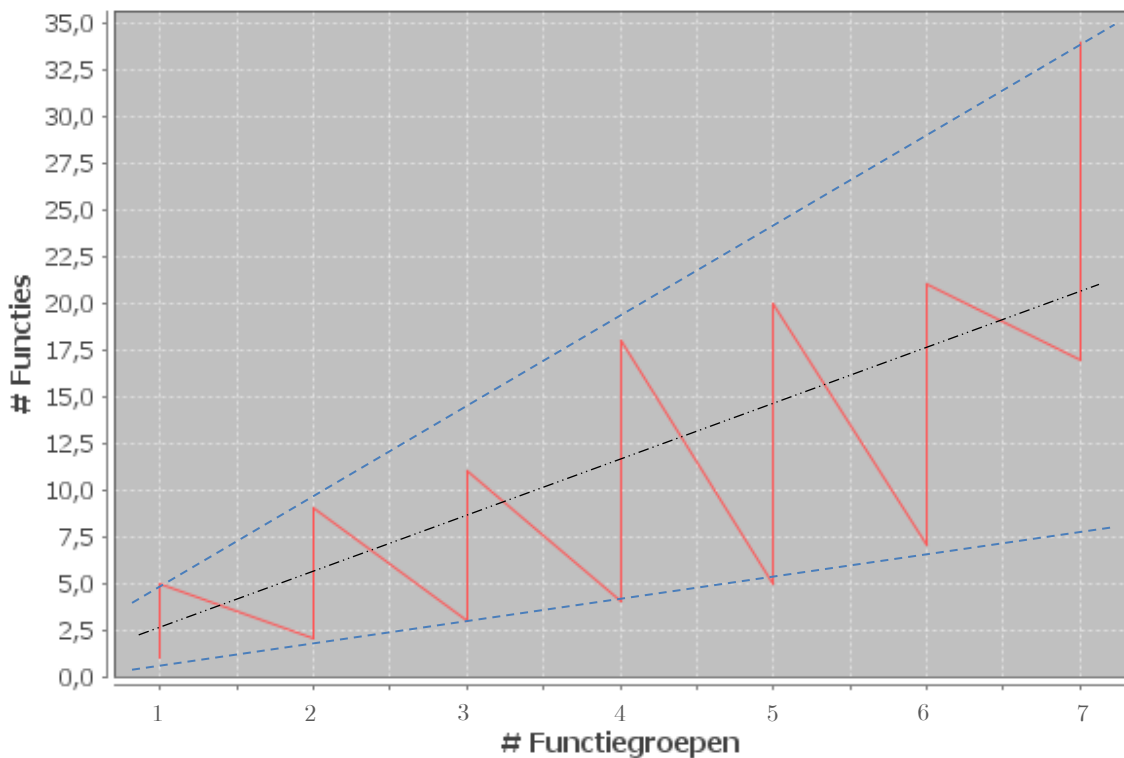
Figuur 12: Frequentieverdeling van het aantal uitgeoefende functies per project



Figuur 13: Frequentieverdeling van het aantal vertegenwoordigde functiegroepen per project

Aangezien de grote massa van de observaties in zowel *Figuur 12* als *Figuur 13* —de NULL-waarden buiten beschouwing gelaten— aan de linkerkant van de grafiek, bij de lage waarden, geconcentreerd zit, suggereert het combineren van de informatie uit beide grafieken dat hoe kleiner het aantal functies per project, des te minder functiegroepen ze zullen vertegenwoordigen. Waar deze uitspraak altijd zal kloppen voor (heel) kleine projecten omwille van louter algebraïsche redenen, hoeft dit echter niet zo te zijn voor projecten vanaf bijvoorbeeld 4 vertolkte functies. Om de uitspraak beter op haar waarheidsgehalte te onderzoeken werd de grafiek in *Figuur 14* opgesteld. In deze grafiek wordt, per mogelijk aantal functiegroepen (1-7), de geobserveerde range aan aantal overeenkomstige functies per project voorgesteld (door de verticale delen van de rode curve). Uit de figuur blijkt dat er in-

derdaad een correlatie is tussen de twee attributen in die zin dat met het aantal functiegroepen, algemeen genomen, ook het aantal functies zal toenemen. Echter, ook de range van geobserveerde waarden voor het aantal uitgeoefende functies neemt toe met het aantal functiegroepen, hetgeen suggereert dat er slechts een milde correlatie tussen de twee attributen bestaat. Dit impliceert dat beide attributen beschouwd zullen moeten worden, wil men de diversiteit van een project gedegen beoordelen.



Figuur 14: Het aantal functies uitgezet tegen het aantal functiegroepen die voorkomen per project

Tijdens het analyseren van de projectteams kan men eventueel ook het aantal departementen willen weten dat door de verschillende teamleden in een project vertegenwoordigd worden. Dit attribuut werd dan ook mee toegevoegd aan de ProjectsINFO-tabel, toen de attributen voor het aantal functies en het aantal functiegroepen werden aangemaakt. Er zal in dit onderzoek echter niet verder worden ingegaan op dit attribuut.

Tevens kan hier opnieuw het #Managers-attribuut aangehaald worden als een indicator voor de diversiteit van de projecten. Er werd immers gesteld dat een projectmanager bij een project betrokken zal worden van zodra het project te omvangrijk of te technisch werd. De aanwezigheid van een projectmanager of een volledig managementteam (van 3) zou dus kunnen duiden op een meer divers project. Echter, om gelijkaardige redenen als in het vorige deel van dit hoofdstuk zal opnieuw het #Managers-attribuut slechts miniem significant zijn als een voorspeller voor de projectdiversiteit.

Mix van functie-gerelateerde werklast

De drie attributen die we net gecreëerd hebben, zijn eerder ruwe, high-level indicatoren voor de projectdiversiteit. Een tweede, meer gedetailleerde manier waarop die diversiteit (of teamsamenstelling) geanalyseerd kan worden is door de *mix* van functies binnen de teams te onderzoeken en deze te contrasteren met hun aandeel aan gepresterde uren per functie(groep). Deze mix kan worden voorgesteld door vooreerst voor elk van de zeven functiegroepen apart te tellen hoeveel teamleden binnen het project een functie uitoefenen die onder die functiegroep valt. Om dit te bestuderen worden zeven nieuwe kolommen toegevoegd aan de ProjectsINFO-tabel^[Q17], waarbij elke kolom dan het aantal Users (teamleden) per FunctieGroep voorstelt, per project. Het resultaat wordt gepresenteerd in *Tabel 26*.

PrID	#U	#M	#F	#FG	#UFG1	#UFG2	#UFG3	#UFG4	#UFG5	#UFG6	#UFG7	(...)
...		
9395	2	2	2	1	2	NULL	NULL	NULL	NULL	NULL	NULL	(...)
9396	27	2	13	5	10	1	NULL	5	1	13	NULL	(...)
9413	10	3	7	4	3	NULL	1	2	NULL	4	NULL	(...)
9414	27	2	13	5	10	3	NULL	4	3	11	NULL	(...)
9415	41	2	13	6	10	4	NULL	5	5	16	2	(...)
9416	NULL	3	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	(...)
9417	2	2	2	2	1	NULL	NULL	NULL	NULL	1	NULL	(...)
...		

Tabel 26: Fragment uit ProjectsINFO-tabel met aantal Users in Functiegroepen 1-7 per project

Het verschil met het vorige wat we onderzocht hebben, is dat hier, per project, per functiegroep, het aantal verschillende personeelsleden (“users”) geteld wordt en niet het aantal verschillende functies. Nochtans zal via deze methode de mix van functies geanalyseerd kunnen worden, zij het op een hoger level via de functiegroepen. Daarbij wordt het aantal gebruikers weliswaar uitgesmeerd over de zeven verschillende functiegroepen. Toch hoeft de som van de verschillende #UFG-kolommen (per project) niet gelijk te zijn aan het aantal teamleden in kolom #U, aangezien een teamlid meerdere functies kan uitoefenen. In dergelijk geval zal dat teamlid meerdere malen geteld worden. Bijgevolg kan de som van de kolommen in het rechterlid van de tabel groter (maar nooit kleiner) zijn dan het aantal teamleden (#U, in de linkerkant van de tabel) die aan een project hebben meegewerkt.

De tweede stap in het ontwikkelen van deze methode omvat het berekenen van het aandeel aan gepresterde uren per functiegroep. Het totaal aantal aan het project gependeerde uren, zoals we dat in het vorige deel van dit hoofdstuk berekend hebben, zal hier dus uitgespreid worden over de verschillende functiegroepen volgens het aantal werkuren dat de verschillende teamleden gepresteerd hebben in de hoedanigheid van hun specifieke functiegroep-rollen. Om dit te bestuderen worden opnieuw zeven nieuwe kolommen toegevoegd aan de ProjectsINFO-tabel^[Q18] waarbij elke kolom ditmaal het aantal Hours (uren) per FunctieGroep voorstelt, per project. Het resultaat wordt gepresenteerd in *Tabel 27*.

Pr.ID	TotalHrs	#FG	#UFG1	#HFG1	#UFG2	#HFG2	#UFG3	#HFG3	#UFG4	#HFG4	#UFG5	#HFG5	#UFG6	#HFG6	#UFG7	#HFG7	(...)
8040	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	(...)
8721	0.50	1	1	0.50	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	(...)
6335	147.00	1	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	1	147.00	NULL	NULL	(...)
7029	56.50	2	4	52.50	1	4.00	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	(...)
9393	2601.50	2	3	2593.00	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2	8.50	NULL	NULL	(...)
7889	10.00	2	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	1	1.00	1	9.00	(...)
6510	4030.98	2	1	768.00	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2	3262.98	(...)
5983	185.50	3	5	37.50	NULL	NULL	NULL	NULL	2	43.75	NULL	NULL	9	104.25	NULL	NULL	(...)
7207	183.00	4	7	85.00	NULL	NULL	2	13.50	2	27.50	NULL	NULL	6	57.00	NULL	NULL	(...)
7399	76.50	4	3	20.00	2	11.50	3	34.50	NULL	NULL	NULL	NULL	2	10.50	NULL	NULL	(...)
7506	268.00	5	1	19.00	4	185.50	3	42.50	1	18.00	NULL	NULL	1	3.00	NULL	NULL	(...)
6512	74.25	5	10	24.00	1	0.50	5	10.00	4	8.50	NULL	NULL	12	31.25	NULL	NULL	(...)
9415	3605.48	6	10	2150.60	4	210.67	NULL	NULL	5	256.15	5	35.50	16	916.73	2	35.83	(...)
7900	3276.02	6	5	132.25	4	49.50	4	71.50	2	6.50	NULL	NULL	4	32.50	3	2983.77	(...)
8113	761.42	6	8	339.17	4	83.50	2	13.00	4	125.50	NULL	NULL	8	196.25	1	4.00	(...)
7127	34848.82	7	36	18924.00	9	1529.50	20	5367.25	13	1935.25	3	130.50	34	5394.87	6	1567.45	(...)
...

Tabel 27: Fragment uit ProjectsINFO-tabel met aandeel gepresteerde werkuren (Hours) per Functiegroepen 1-7, per project.

Activiteiten

Het derde en meest detaillistische niveau, waarop de projectdiversiteit onderzocht kan worden is op niveau van de voor het project uitgevoerde taken. Niet alleen zou een analyse van de activiteiten voor een project een precies beeld kunnen geven over de omvang en de diversiteit van dat project; de projectsoort op zich zou hieruit zelfs kunnen worden afgeleid. Jammer genoeg, echter, is de manier waarop personeelsleden hun activiteiten loggen niet geformaliseerd. Twee problemen doen zich hierbij voor. Enerzijds kunnen activiteiten zowel als kleine sub-taken gelogd worden, als in één blok (bijvoorbeeld aan het einde van reeks van samen hordende taken). Anderzijds zijn de activiteitsomschrijvingen niet gestandaardiseerd, maar kunnen personeelsleden vrij intypen wat ze willen als comment. Dit maakt het onmogelijk om stereotype taken te onderscheiden. Het zal dus niet mogelijk zijn om de diversiteit of omvang van een project via deze weg te bepalen. Bijgevolg zal ook de projectsoort niet achterhaald kunnen worden. Tevens andere interessante onderzoeksmogelijkheden blijven hierdoor onbenut. Met het oog op het beoordelen van de project- of projectteamperformance, bijvoorbeeld, zal men niet kunnen bepalen hoe lang bepaalde taken gemiddeld duren om vervolgens een vergelijkende prestatieanalyse te kunnen uitvoeren.

Globale projectdiversiteitsindicator

Vanuit het ontbreken van projecttyperende indicatoren in de database werden in dit deel van het hoofdstuk diverse mogelijkheden opgesteld om de diversiteit, of de *scope* van een project te bepalen. De diverse attributen die daarbij werden beschouwd, analyseren die scope op verschillende informatieniveaus. Zo werden er eerst enkele high-level indicatoren opgesteld zoals het aantal functies die werden uitgevoerd bij een project en het aantal functiegroepen waarmee deze functies overeenkomen. Daarnaast werden ook het aantal managers en het aantal vertegenwoordigde departementen in overweging genomen, maar de relevantie van deze laatste twee attributen werd vervolgens afgedaan als te beperkt. Vervolgens hebben we onze focus verlegd naar meer mid-level indicatoren zoals de user-functions mix en de gepresteerde uren mix, gespreid over de verschillende functiegroepen, per project. Een derde, detaillistischer informatieniveau dat we wilden aanboren was dat van de uitgevoerde activiteiten. Helaas zijn we hier op een dood spoor gestuikt. Hoewel elk van de gevonden indicatoren iets meer vertelt over de scope van een project, zal elk op zich niet zoveel inzicht bijbrengen. Daarom zal de informatie uit de meest diverse indicatoren samengebracht worden in één globale indicator. Op basis van dit attribuut kan de scope van een project dan eenvoudig worden afgeleid.

Voor de creatie van de globale projectomvang-indicator, zullen enkel de meest significante diversiteitsindicatoren geselecteerd worden. Deze zijn het #Functions, het #Function-Groups en het aantal uren per functiegroep (#HFG's). Hierbij zal worden uitgegaan van

drie assumpties: Ten eerste, hoe meer functies aan bod komen in een project, hoe diverser het project. Ten tweede, hoe meer functiegroepen vertegenwoordigd worden, hoe diverser het project. En ten derde, hoe meer de gepresteerde uren gespreid zijn over de verschillende vertegenwoordigde functiegroepen, hoe diverser het project. Enkel wanneer deze drie attributen (of drie assumpties) samengenomen worden, zullen een goede indicatie geven van de scope van een project. Toch hoeft niet elk attribuut even zwaar door te wegen op het globale resultaat. We zullen opnieuw de waarden van elke indicator verdelen over een aantal scores volgens een allocatieverdeling die wordt voorgesteld in *Tabel 28*. Door middel van de hoogte van die scores zal dan de doorslagkracht van elk attribuut afgesteld kunnen worden. Voor het uiteindelijke resultaat herleiden we de som van die scores opnieuw tot een getal tussen 1 en 5. Op die manier kan dus een soort selectie-attribuut gecreëerd worden op basis waarvan vervolgens eenvoudig op de projectscope gefilterd kan worden.

INDICATOR	Range van waarden die voorkomen in de dataset	Verdeelsleutel/Scorebepaling
#Functions	NULL, 1 — 34	1-2 → 1
		3-5 → 2
		6-10 → 3
		11-15 → 4
		> 15 → 5
#Function-Groups	NULL, 1 — 7	1, 2 → 1
		3 → 2
		4 → 3
		5 → 4
		6, 7 → 5
#HFG	NULL, 0,25 — 18.924	
TotalHours	NULL, 0,25 — 34.849	
HFG-Spread	NULL, 0 — 1	<0,25 → 4
		<0,50 → 3
		<0,75 → 2
		≤ 1 → 1

Tabel 28: Allocatie van scores voor de drie diversiteitsindicatoren

Zoals men in de bovenstaande tabel kan zien, wordt voor het bepalen van de spreiding van de gepresteerde uren over de verschillende functiegroepen nog een nieuw attribuut opgesteld, met name de HFG-Spread. Voor het berekenen van dit attribuut wordt teruggevallen op een eenvoudige spreidingsmaatstaf uit de statistiek, namelijk de ‘variatiebreedte’, die berekend kan worden door de kleinste observatie uit een reeks in mindering te brengen van de grootste observatie uit die reeks. De “reeks” in dit geval zijn de zeven waarden (NULL meegerekend) van de verschillende #HFG-attributen per project. Vervolgens wordt dat verschil genormaliseerd tot een waarde tussen 0 en 1 door dit te delen door het totaal aantal aan het project gepresteerde uren:

$$\frac{\text{grootste \#HFG} - \text{kleinste \#HFG}}{\text{TotalHours}}$$

Resulterende waarden die dicht bij nul liggen reflecteren een grotere spreiding van de werkuren over de verschillende functiegroepen. Wanneer een project de waarde 1 zou aannemen, dan betekent dit dat alle werk op slechts één functiegroep gebeurde. Opnieuw zijn de allocatiesleutels die hier gebruikt werden deels gebaseerd op de frequentieverdelingen die eerder in dit hoofdstuk getoond werden, en deels op de eigen interpretatie van de informatiewaarde voor het analyseren van de projecttypes. De globale score voor het classificeren van projecten op hun diversiteit (= de scope van een project) kan men aflezen in *Tabel 29*. De betreffende queries vind men in bijlage B.^[Q19,Q20] De projecten uit de onderstaande tabel zijn dezelfde als deze in *Tabel 27*. Uit die tabel kan men desgewenst de verschillende #HFG-waarden aflezen.

Pr.ID	TotalHours	#F	#F_score	#FG	#FG_score	#HFG-Spread_score	GLOBAL SCOPE SCORE (...)
8040	NULL	NULL	NULL	NULL	NULL	NULL	NULL (...)
8721	0.50000	1	1	1	1	1	1 (...)
6335	147.00000	1	1	1	1	1	1 (...)
7029	56.50000	4	2	2	1	1	1 (...)
9393	2601.50000	4	2	2	1	1	1 (...)
7889	10.00000	2	1	2	1	1	1 (...)
6510	4030.97996	2	2	2	1	2	2 (...)
5983	185.50000	8	5	3	2	3	4 (...)
7207	183.00000	10	5	4	3	3	4 (...)
7399	76.50000	6	3	4	3	3	3 (...)
7506	268.00000	8	3	5	4	2	3 (...)
6512	74.25000	16	5	5	4	3	4 (...)
9415	3605.48323	13	5	6	5	2	4 (...)
7900	3276.01666	12	5	6	5	1	4 (...)
8113	761.41666	14	5	6	5	3	5 (...)
7127	34848.81634	34	5	7	5	2	4 (...)
...

Tabel 29: Overzicht van de verschillende score-indicatoren voor enkele geselecteerde rijen met betrekking tot de project scope

Conclusie

Het nut van deze nieuwe attributen uit dit hoofdstuk is dat ze toelaten om de mix aan uitgevoerde functies voor een project te analyseren. Door het aantal personen die een functie van een bepaalde functiegroep uitoefenen binnen het project te contrasteren met het aantal gepresteerde uren voor de functiegroep kan de relatieve impact van het werk uit die functiegroep op het resultaat voor het project beter worden ingeschat. Veel personen die tezamen slechts enkele uren werk geleverd hebben binnen één functiegroep, zullen waarschijnlijk geen project-definiërende activiteiten hebben uitgevoerd aan het project. De combinatie van deze attributen laat dus toe om de hogere-level-indicatoren zoals #FunctionGroups, #Functions en #TeamMembers met meer perspectief te bekijken. Meer zelfs, aangezien

specifieke functies specifieke taken(pakketten) impliceren, kan op basis van deze attributenmix ook voorzichtig het soort project worden afgeleid. Uiteenlopende projecten met andere doelomschrijvingen zullen andere werkactiviteiten vereisen. Echter op niveau van de activiteit zelf zal geen analyse kunnen worden uitgevoerd, wegens te weinig geformaliseerde data. Een schatting naar de projectsoort zal dus enkel op basis van de gewogen functiewerkuren-mix (tezamen met de andere typeringsindicatoren) gemaakt kunnen worden.

Manueel het aantal personen en het aantal gepresteerde uren voor de verschillende functiegroepen gaan vergelijken, echter, is een moeizame taak en zal wellicht niet het grootste potentieel uit de data halen. Waar de kracht van deze attributen zit, is om ze als input te gebruiken in een dataminingproces. Via “clustering”, een descriptieve mining techniek, bijvoorbeeld, kan getracht worden de data te vatten in enkele onderscheiden gehelen, zodat elk geheel bepaalde kenmerken gemeenschappelijk heeft (Agrawal et al., 1998). Met behulp van deze techniek zouden zo de meest gangbare projectsoorten blootgelegd kunnen worden. De uitvoering van deze en dergelijke analyses valt echter buiten dit onderzoek. In het volgende hoofdstuk zal echter nog worden teruggekomen op mogelijkheden voor vervolgonderzoek (aan de hand van andere onderzoeksmethodes, waaronder Data Mining) die zich zullen baseren op de dataset die in dit hoofdstuk voorbereid werd.

Aan het einde van dit hoofdstuk beschikken we nu over een hele reeks indicatoren voor de projectomvang en de projectdiversiteit te beoordelen. Twee globale indicatoren werden gestileerd, de ene voor de *scale*, de andere voor de *scope* van een project te identificeren op basis van een score van 1 tot 5. Aan de hand van deze indicatoren en scores wordt het zo mogelijk gemaakt om projecten een surrogaat type toe te kennen in een verdere analyse, en kan een onderzoeksobject nu ook eenvoudiger afgebakend worden op verschillende niveaus aan de hand van diverse karakteristieken – bijvoorbeeld: projecttype, omvang, teamgrootte, complexiteit of functiemix. De combinatie van de twee globale indicatoren, plus de bijbehorende deelaspect-indicatoren, laten tevens toe vragen te onderzoeken als: Hebben de “grootste” projecten ook de langste looptijd? Of: Heeft de projectcomplexiteit een impact op de looptijd?

5. Financiële omvang-indicatoren

In het derde deel van dit hoofdstuk werd op zoek gegaan naar mogelijke manieren om de omvang van projecten te beoordelen op basis van team- en tijd-gerelateerde grootte-indicatoren. De omvang van een project zou echter ook in financiële termen kunnen worden uitgedrukt. In deze sectie zullen enkele van die indicatoren beschouwd worden.

Het eerste dergelijke classificatie-attribuut is een indicator die aangeeft wat het totale budget is dat voor een project beschikbaar was. Dit klinkt als een evidentie observatie, maar in tegenstelling tot wat men zou verwachten kan het totale budget niet rechtstreeks uit de database worden afgelezen. De budget-informatie is namelijk opgeslagen in meerdere tabellen. Zo staan de financiële bedragen vermeld in de BudgetNodes-tabel, terwijl de context-informatie en ook de link naar het betreffende project in een andere tabel staan. De BudgetNodes-tabel is bovendien recursief. Zo bestaat een budget uit meerdere nodes (voorgesteld door rijen in de tabel) dewelke hiërarchisch gesorteerd staan, zodat elke lagere lijn zijn voorganger als parent-node heeft. De top-node van een budget heeft bijgevolg geen parent en vat de financiële gegevens van de onderliggende nodes samen. We zijn dus geïnteresseerd in het verbinden van die top-nodes aan de relevante projecten. Bij het reconstrueren van het TotalBudget per project, moet men er bovendien aandacht voor hebben om de bedragen exclusief BTW te selecteren.^[Q21] De grootte van een budget kan een indicator zijn voor de grootte van een project, maar zonder een verdere analyse van de gemaakte kosten per project, hoeft deze indicatie niet per se te kloppen.

Een tweede financieel attribuut dat interessant kan zijn om projecten te classificeren is de uiteindelijke totale prijs voor de klant van het project. Overheen de looptijd van het project worden afzonderlijke onderdelen van dat project namelijk apart gefactureerd. Er is waarschijnlijk dus niet slechts één factuur die aan het einde van de rit naar de klant wordt gezonden maar doorlopend tussentijdse facturen voor het reeds gepresteerde werk. Om dus een beeld te krijgen van de globale opbrengst van een project moeten deze verschillende facturen behorende tot één project gegroepeerd worden. Dit kan gedaan worden met een relatief eenvoudige query^[Q22] die een nieuw attribuut TotalBillingAmount zal creëren. Ook bij deze bewerking geldt echter het aandachtspunt van de BTW, wil men geen appels met peren gaan vergelijken. Voor de factuurbedragen in de TotalBillingAmount werd echter nog geen BTW verrekend. De TotalBillingAmount kan opnieuw een indicator zijn voor de omvang van een project.

Projectefficiëntie analyseren

De twee financiële omvang-indicatoren die net gecreëerd werden, in combinatie met de verschillende omvangs- en diversiteitsindicatoren uit de voorgaande delen van dit hoofdstuk laten toe heel wat aspecten gerelateerd aan de efficiëntie van projecten te analyseren. Op

die manier kan bovendien een eerste indicatie bekomen worden over het succes van een project (wat een ontbrekende –direct afleesbare– informatie uit de Dogglenet database is). Volgens Shrnhur et al. (1997) dient het succes van een project immers op vier dimensies beoordeeld te worden. Eén daarvan is de efficiëntie waarmee het project werd uitgevoerd. Anderzijds zal voor een volledige beoordeling van het succes van een project ook de impact op de klant en de impact op het bedrijf (in termen van zowel directe als lange termijn voordelen en opportuniteiten) bepaald moeten worden, zoals getoond wordt in *Figuur 15*.



Figuur 15: De vier dimensies van project succes (Shrnhur et al., 1997)

De performance van een project kan dan beoordeeld worden door de verschillende indicatoren die gecreëerd werden met elkaar te combineren of te contrasteren. Om de sterkte van de indicatoren in een performance-analyse te illustreren zullen kort enkele voorbeelden worden aangehaald. Ten eerste kunnen de twee indicatoren die hierboven besproken werden met elkaar gecontrasteerd worden om de mate te bepalen waarin de financiële werkelijkheid voor een project afwijkt van de geplande uitkomst: $\text{Budget-reality gap} = \text{TotalBudget} - \text{TotalBillingAmount}$. Het totale budget staat hier immers voor de voorziene kosten en is tevens het bedrag dat in de offerte aan de klant werd voorgelegd. Het totale factuur bedrag slaat dan weer op het effectieve bedrag dat Boondoggle zal ontvangen voor de gepresteerde diensten. Dit is slechts een indicatie voor de werkelijke ‘budget-reality gap’ omdat sommige kosten misschien niet gefactureerd kunnen worden.

Een tweede voorbeeld van hoe de indicatoren gebruikt kunnen worden in het beoordelen van de performance is bij het bepalen van de gemiddelde loonkost per project. Wanneer de totale gefactureerde loonkosten geplaatst worden tegenover het totaal aantal aan het project gepresteerde uren dan kan de gemiddelde loonkost per project berekend worden. Indien deze hoger is dan een standaard (die eveneens uit de data kan worden afgeleid als het gemiddelde der gemiddelden) dan kan men onderzoeken of bepaalde groepen werknemers (functions, functiongroups) meer vertegenwoordigd zijn dan anderen. Wanneer dit zo is kan worden nagegaan welke functies (of functie-mix) grotere kostendrijvers zullen zijn voor een project. Interessant om hier te onderzoeken zijn bijvoorbeeld de vele kleine projecten (met projectteams van 1 of 2 personen) die toch een managementteam van 2 of 3 managers hebben.

Hoofdstuk 5

Projecten karakteriseren: Projectsoort en Projectsucces

In het vorige hoofdstuk werden een heel aantal indicatoren opgesteld voor het beschrijven van verschillende kenmerken (zoals de omvang en de diversiteit) van de projecten en hun projectteams. Wanneer die indicatoren gecombineerd worden, wordt het mogelijk om de complexiteit en de mix van uitgeoefende functies binnen een project(team) te beoordelen. Uit de combinatie van deze twee kan dan weer een beschrijving van een projecttype worden afgeleid. Men zou dit manueel kunnen doen, door opnieuw enkele nieuwe attributen te creëren die de samengevatte informatie uit de eerdere indicatoren combineert en vervolgens nogmaals samenvat in enkele categorieën. Deze werkwijze zal echter niet alleen veel werk vragen, het resultaat van die bewerkingen zal ook in grote mate beïnvloed worden door de keuzes die gemaakt worden bij het bepalen van de parameters. Beter kan men in deze volgende stap van het onderzoek terugvallen op een onderzoekstool zoals K-Nime of Weka en diverse clustering of classificatiemethodes uit het data mining onderzoeksdomein toepassen. In dit hoofdstuk onderzoeken we hoe met behulp van data mining de dataset aan de hand van de gecreëerde indicatoren en de eventuele andere context-attributen verder beschreven kan worden.

1. Data Mining

Data mining is een toegepast onderzoeksdomein dat zich bezig houdt met de analyse van veelal grote, uit de praktijk stemmende datasets om onverwachte relaties te ontdekken en om de gegevens op nieuwe manieren samen te vatten die zowel betekenisvol als interessant zijn voor de data eigenaar (Hand et al., 2001). Het betreft echter een erg ruim onderzoeksdomein met tal van applicaties. Eén daarvan is 'descriptive modeling'. Dergelijke modellering laat toe de belangrijkste kenmerken van de data te beschrijven, door de gegevens samen te vatten op zo'n manier dat de belangrijkste aspecten van de data bestudeerd kunnen worden zonder daarbij gehinderd te worden door de ruis en de omvang van de dataset. Hierbij wordt er een model gemaakt, waarin verschillende gegevenspatronen aan de dag

kunnen worden gelegd (Hand et al., 2001). De descriptieve modelleringsmethodes waar we hier echter in geïnteresseerd zijn, dienen de data te kunnen groeperen op basis van gemeenschappelijke kenmerken. De methode die dat toelaat is 'clustering'. Dit is een (niet-gesuperviseerde) classificatie van patronen of observaties uit de dataset in groepen op basis van gedeelde kenmerken (Jain et al., 1999).

Men zou clustering rechtstreeks kunnen toepassen op de geselecteerde kolommen uit de ProjectsINFO-tabel die werd aangemaakt in het vorige hoofdstuk, aangezien deze tabel reeds heel wat projectkenmerken beschrijft. Vooral op basis van de verschillende (ook financiële) omvangsindicatoren en de *aantal functies*- en *aantal functiegroepen*-diversiteitsindicatoren zou men clusters van verwante projecten kunnen onderscheiden. Echter, om ook de userfuncties/werkuren-mix goed te kunnen analyseren kan men de data eerst beter onderzoeken op frequente patronen van dergelijke mixen via de techniek van 'frequent itemset mining'. Hiertoe kunnen de #UFG- en #HFG-attributen wel best eerst worden voorbereid door ze ofwel te binariseren (waarbij 0: geen uitgeoefende functies binnen de betreffende functiegroep voor een gegeven project; en 1: wel één of meerdere dergelijke functies), ofwel te categoriseren (door het onderbrengen van de waarde in een aantal scores (of 'bins')). Na het toepassen van frequent itemset mining, kunnen de gevonden patronen gebruikt worden in het clustering-proces om zo alle van de in het vorige hoofdstuk opgestelde indicatoren te kunnen betrekken in de typering van de data.

De beschreven methodes laten toe om de meest gangbare projectsoorten uit de data bloot te leggen, bovendien zullen ook projectteam-patronen kunnen worden onderscheiden. Data mining is dus beslist een belangrijke onderzoekstool om projecten of projectteams te typeren en het loont om verdere analyses aan de hand van deze tool te ondernemen.

Vanuit de via frequent pattern mining gevonden patronen kunnen ook associatieregels worden opgesteld. Hieruit kan dan vervolgens een predictiemodel worden afgeleid om voor toekomstige projecten, aan de hand van de aangegeven projectkenmerken, te voorspellen hoe een goed team er voor dat project zal uitzien. De kwaliteit van die voorspelling is echter wel geheel afhankelijk van de associatieregels die eerder gevonden zullen moeten worden. Echter, gegeven de twijfels over de volledigheid van de informatie bij het reconstrueren van de projectteams (vanwege het mogelijk ontbreken van activiteiten die niet in de timelogs-tabel gelogd werden) en het feit dat de parameters die gehanteerd werden bij het opstellen van de verschillende score-indicatoren in het vorige hoofdstuk voor een deel gebaseerd zijn op een eigen interpretatie van de projectwerking, kan de kwaliteit van de ontdekte associatieregels niet optimaal blijken te zijn. Bijgevolg zullen de voorspellingen gebaseerd zijn op gekleurde informatie en zullen ze misschien niet helemaal de gewenste resultaten opleveren. De kunst hier betreft dus het finetunen van de parameters in de verschillende modellen.

2. Process Mining

Process mining is een onderzoeksgebied dat als doel heeft om procesgerelateerde informatie af te leiden uit logboekgegevens (van der Aalst, 2011). Via technieken uit dit onderzoeksgebied kan men onder andere de performantie van projectteams onderzoeken. Project succes wordt, zoals in hoofdstuk 4 reeds werd verteld, voor een deel bepaald door de efficiëntie waarmee het project werd afgehandeld. Efficiëntie kan dan weer begrepen worden als het binnen het budget en binnen de planning afronden van een project. Het is met betrekking tot die tweede efficiëntiemaatstaf dat proces mining van nut kan zijn. Immers, via proces mining kan het afwijken van de werkelijke uitvoering van een project ten opzichte van de geplande uitvoering geanalyseerd worden.

Jammer genoeg bevat de Dogglenet-database niet voldoende planningsinfo om dergelijk onderzoek hier relevant te maken. Wil Boondoggle in de toekomst dus haar projectteams evalueren op de mate waarin ze kunnen vasthouden aan de planning, dan raden we aan om vaker een planning te koppelen aan de uit te voeren projecten, of althans om deze vaker in de database te registeren.

Besluit

Gegeven het probleem van het ontbreken van eenvoudig beschikbare indicatoren voor het bepalen van de projectsoort of het succes van een project in de Dogglenet database, is het niet rechtstreeks mogelijk om op basis van de data uit de database een evaluatie op te maken van de eerder uitgevoerde projecten, teneinde een beter inzicht te verkrijgen in het samenstellen en selecteren van projectteams om projecten succesvol of efficiënt af te werken.

Deze thesis heeft vanuit die problematiek onderzocht hoe het mogelijk is om op basis van de beschikbare data alternatieve methodes te ontwikkelen om projecten en projectteams te typeren. In een eerste stap daartoe hebben we succesvol een groot aantal projectteams gereconstrueerd. Vervolgens werden diverse indicatoren voor een projecttypering opgesteld.

De indicatoren laten toe om de omvang en de diversiteit van projecten te analyseren. Door de verschillende attributen met elkaar te contrasteren, wordt het mogelijk om de complexiteit en mix van uitgeoefende functies binnen een project(team) te beoordelen. Met behulp van clustering, een data mining-techniek, kunnen vervolgens project(team)soorten onderscheiden worden.

Tijdens het onderzoeksproces werden op enkele problemen of tekortkomingen uit de Dogglenet database gestuikt, naast het ontbreken van een aanduiding voor het projecttype of een beoordeling van projecten na hun afloop. De belangrijkste problemen houden verband met de timelogs. Enerzijds hebben we geconstateerd dat deze niet altijd worden ingevuld, als werd aangetoond door de vele projecten zonder overeenkomstige logs (–voor de reeds gefilterde projecten die werden uitgevoerd sinds het invoegen van de timelogs). Daardoor kan het zijn dat niet alle activiteiten die voor een project werden uitgevoerd, beschouwd worden, zodat een onvolledig beeld van de projecten gecreëerd zal worden.

Daarnaast geldt voor de timelogs die wel werden ingevuld dat de wijze van rapporten weinig consistentie vertoont. Logs kunnen immers zowel een fractie van een uur beslaan, als (in een beperkter aantal gevallen) tientallen uren beslaan. Bovendien zijn de activiteiten/taakomschrijvingen niet gestandaardiseerd. Dit zorgt ervoor dat we geen onderzoek kunnen voeren op niveau van de activiteiten. De beste manier waarop de uitgevoerde taken voor een project kunnen worden benaderd, wil men niet handmatig alle comments gaan ontcijferen (en ook deze kunnen soms heel weinigzeggend zijn), is door het analyseren van de uitgeoefende functiemix die in een project aan bod kwam.

De database bevat ook een groot aantal zeer kleine projecten. Mogelijk gaat het hier om projecten waarvan de meeste activiteiten niet gelogd werden. Daarnaast lijkt het er echter ook op dat Boondoggle de policy heeft om grotere projecten op te splitsen in kleinere deelprojecten (die dan als een afzonderlijk project in de database werden ingevoerd). Een eenduidige mogelijkheid om in dat geval de samen horende projecten te identificeren werd niet gevonden.

Tot slot is er voor de meeste projecten geen uitgewerkte planning beschikbaar in de database. Ook in de timelogs is het planningtaskID en tevens ook het budgetnodeID zelden ingevuld. Om toekomstige mogelijkheden voor analyse en de eventuele interessante inzichten die uit dat onderzoek zouden kunnen volgen voor het bedrijf, te vergroten, kan een aanbeveling gemaakt worden naar Boondoggle om het gebruik van de timelogs of het invullen daarvan nog meer te standaardiseren.

Literatuurbronnen

Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. *SIGMOD '98 Proceedings of the 1998 ACM SIGMOD international conference on Management of data, Vol.27, Nr.2*, pp. 94-105.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT press.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), pp. 264-323.

Katzenbach, J.R. & Smith, D.K. (1993). *The wisdom of teams: Creating the high-performance organization*. Boston: Harvard Business Press.

Monster, T.J. (2007). *Persoonlijkheidskenmerken en creatieve onrust* [afstudeerscriptie doctoraalopleiding bedrijfskunde, strategie en organisatie]. Open Universiteit Nederland.

Rijnbergen, T. (2007). *Samenwerking in teams: De impact van verticaal en gedeeld taak-en relatiegericht leiderschap en groepsontwikkeling op teamprestatie* [masterthesis organisatiepsychologie]. Universiteit Utrecht.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. *Teams: Their training and performance*. (pp. 3-29). Westport, CT, US: Ablex Publishing.

Shrnhur, A. J., Levy, O., & Dvir, D. (1997). Mapping the dimensions of project success. *Project management journal*, 28 (2), pp. 5-13.

Sundstrom, E., McIntyre, M., Halfhill, T., & Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics: Theory, Research, and Practice, Vol.4, Nr.1*, pp. 44-67.

van der Aalst, W. (2011). *Discovery, conformance and enhancement of business processes*. Springer.

Bijlage A: Mailvragen

Mail van 01/07/2015:

Hebben jullie op beleidsniveau een bepaalde definitie of standaard van wat een projectteam is of hoe een team er (minstens) uit moet zien? Ik denk dan bijvoorbeeld aan guidelines inzake teamgrootte of samenstelling, of dat er per project naar een team op maat wordt gezocht of dat jullie met enkele vaste combinaties werken, en dergelijke. Dit zou interessant zijn voor mij om de ideeën op beleidsniveau af te toetsen tegen de werkelijkheid die uit de data zal blijken.

“By default mag je er van uit gaan dat adhoc het meest geschikte team wordt samengesteld. Wij hebben maar weinig “standaard” projecten. Team samenstelling wordt dus bepaald door inhoud van het project en beschikbaarheid van mensen.

Bij de bright projecten (code start met PB-) daar wordt meer met klanten teams gewerkt. Projecten van een bepaalde klanten zouden overwegend bij dezelfde mensen terecht moeten komen. Beschikbaarheid kan daar natuurlijk root in het eten gooien.”

Mail van 17/07/2015:

Zijn er specifieke rollen die altijd in een projectteam vervuld moeten zijn? Ik denk bijvoorbeeld aan: Er moet altijd minstens 1 projectmanager zijn, 2 designers, ... (Ik noem zo maar iets). Of hebben jullie guidelines betreffende een minimum bezetting/rolvertolking binnen een team?

“Er zijn bitter weinig guidelines, projecten liggen nogal uit elkaar. Minimum heb je een account manager of een account director, daar ga je nooit buiten kunnen. Verder wordt alles bepaald door het type project. Everything goes.”

Bestaan alle projectteams slechts voor een beperkte tijd of zijn er teams die doorlopend blijven bestaan omdat een klant/bedrijf een min of meer permanente activiteit van jullie vraagt? (Hiermee wil ik informeren of alle teams slechts een beperkte bestaansduur hebben en aan het eind, dus na afloop van het project, ontbonden worden.)

“We hebben 2 business units (gehad) Bright en Agency: bij agency waren er geen vaste teams. Daar werden enkel accounts vast toegewezen aan een klant, maar verder niemand strict. Bij bright bestaan wel klanten teams en zal je per project steeds dezelfde combinaties zien verschijnen. Dat is ook geen vaste waarheid, maar 90% van de gevallen voldoen daaraan.”

Wordt er eerst een concreet einddoel bepaald en dan een team op maat samengesteld, of vertrekken jullie van een team die dan de einddoelen zelf zullen opstellen in samenspraak met de klant? (doel > team VS. team > einddoel?)

“Beide situaties bestaan. Wij verkiezen team > einddoel, maar er zijn nu eenmaal ook die andere projecten.”

Mail van 02/08/2015

[...]

Een tweede topic waar ik nog met vragen over zit is wat het onderscheid is tussen een AccountManager, een ProjectManager en een AccountDirector. Wie doet wat precies? Waarop slaat “account”? En is de Director de uiteindelijke verantwoordelijke?

“Haha, wie zal het zeggen ☺ Dat is hier eigenlijk een grijze massa.

Account director: is eindverantwoordelijk voor een bepaalde klant bvb. Belfius. Hij zorgt voor klanten tevredenheid en new bizz op die klant. Een account director heeft meerder klanten.

Account manager: gaat het overnemen van de account director van zodra een project verkocht is. Een account director helpt ook met opstellen van budgetten en de administratieve ondersteuning van een Account Director. Account managers zitten in een team met de account director en worden op project basis toegewezen.

Project managers: wanneer een project te technisch wordt of te groot in omvang wordt er een project manager aan toegewezen. Die neemt de interne organisatie van het project dan voor zich. Bij niet technische projecten zoals affiches of radio of TV zal de account manager deze rol opnemen.

In uitzonderlijke gevallen, doet de account director ook alle drie de rollen samen. Bij niet structurele klanten durven de hoedjes wel eens te wisselen.”

Mail van 12/08/2015

In de database worden projecten ingedeeld bij een bepaalde businessunit. Reflecteert dit het type project of slaat het eerder op een geografische locatie? In onderstaande tabel kan je die betreffende indeling van de projecten aflezen.

BusinessUnit	#projecten
Boondoggle Agency	8042
Boondoggle Amsterdam	689
Boondoggle Ireland	0
Boondoggle Bright	837
Boondoggle LifeLabs	16
	9584

“Dat is een beetje dubbel. Dat zijn business units die hun eigen rapportering hebben wat betreft omzet en kosten en dus ook hun eigen management. Dat is ook geografisch, behalve Bright en Agency zitten beiden in Leuven en hebben verschillende projecten.”

Wat is precies het onderscheid tussen de verschillende businessunits? Zou je me kort iets meer kunnen vertellen voornamelijk over het onderscheid tussen de types Agency, Bright, LifeLabs? Ik neem aan dat Amsterdam en Ireland gewoon projecten uit buitenlandse filialen betreffen.

“Alles is ontstaan uit Boondoggle. BD Ireland en BD Amsterdam zijn buitenlandse filialen, waarvan Ireland enkele jaren geleden gesloten is. Amsterdam is nu ondergebracht in Havas en heeft het dogglenet nooit goed gebruikt.

Agency en Bright zijn beiden Leuven-based filialen. Zij onderscheiden zich vooral door type projecten. Agency doet kort lopende campagnes, Bright doet permanent touchpoints zoals CRM, corporate websites en apps.

Lifelabs is een initiatief waar mensen uit Leuven konden instappen. Zij werkten dan 4/5^{de} voor Boondoggle en 1/5^{de} voor Lifelabs. Zij trachtten op eigen houtje innoverende producten op te zetten en zo winst voor hun eigen te genereren. Is vooral een PR project geweest.”

Als de BusinessUnitID niet het soort project reflecteert, maken jullie op een andere manier dan wel een onderscheid naar projecttype in de dataset?

“Nee, dat is een groot gebrek in dogglenet, er is geen eenduidige manier om type projecten te onderscheiden. Als die er al is dan is dat door een patroon of soort finger-

print dat type projecten achterlaten. Moest dat zo zijn zou ik dat erg boeiend vinden om weten."

Zou deze andere manier dan misschien via de eerste letters van de projectcodenaam kunnen zijn? Als je namelijk de codenaam van projecten bekijkt, dan kun je projecten indelen op basis van de eerste letters van die codenaam. In onderstaande tabel zie je welke indeling ik bedoel.

Type Project	Code begint met	#projecten
<i>Standaard ?</i>	P-	8863
<i>Amsterdam</i>	PA-	83
<i>Bright</i>	PB-	260
<i>LifeLabs</i>	PF- of P-LABS-	14
<i>Ierland</i>	PI-	0
<i>Agency</i>	PY-	364
		9584

"Nee, dat slaat enkel op de business units die verantwoordelijk zijn voor het project."

Volgens deze indeling zijn er ook "standaardprojecten" (code start gewoon met P-). Deze tabel mapt op de eerste tabel door de "standaardprojecten" over de andere 5 categorieën te verspreiden. (Wel ja, 4 categorieën eigenlijk, want Ierland blijft 0.) Gaat het voor die 'P-' projecten inderdaad om niet-speciale (of standaard-)projecten, of eerder om een niet-consistente naamgeving?

"Er zijn geen standaardprojecten. Het concept business unit bestaat nog maar enkele jaren in dogglenet, voordien waren alle projecten gewoon P-. Vroeger waren er per nationaal filiaal een eigen dogglenet, in 2011 is al die data samen gegaan in 1 db met 1 dogglenet."

Bijlage B: Queries

Q1: Alle gegevens uit Timelog-tabel relaterend aan één project (met ProjectID = 9021)

```
SELECT *
FROM [Dogglenet2011].[dbo].[TimeLogs]
WHERE ProjectID = 9021
```

Q2: Wie werkte in het team van één gegeven project (met ProjectID = 9021)?

```
SELECT distinct UserID
FROM [Dogglenet2011].[dbo].[TimeLogs]
WHERE ProjectID = 9021
```

Output:

UserID's = {101, 159, 171, 199, 222, 258, 263, 298, 299, 304, 313, 337, 381, 401, 466}

Q3: Welke functies beoefenen de teamleden van één gegeven project binnen dat project (met ProjectID = 9021)?

```
SELECT UserID, FunctionID
FROM [Dogglenet2011].[dbo].[TimeLogs]
WHERE ProjectID = 9021
GROUP BY UserID, FunctionID
```

Output:

UserID + FunctionID's = {(101,1), (159,19), (171,17), (171,18), (199,6), (222,86), (258,89), (263,11), (298,19), (298,97), ... }

Q4: Welke projecten werden gestart/ingevoerd na 28-01-2010 (start timelogs)?

```
SELECT *
INTO ProjectsSTARTED
FROM Dogglenet2011.dbo.Projects
WHERE EnterTime >= '2010-01-28 00:00:00.000000'
```

Output:

Het resultaat van deze query wordt opgeslagen in een nieuwe hulptabel

Q5: Welke van de eerder geselecteerde "gestarte" projecten werden afgerond binnen de observeerbare periode van de dataset?

```
SELECT B.*, A.StartTime
INTO ProjectsCOMPLETED
FROM [Dogglenet2011].[dbo].[ProjectStatusFlowItems] as A
JOIN [Dogglenet2011].[dbo].[ProjectsSTARTED] as B
    on A.ProjectID = B.ID
WHERE ProjectID IN (
    SELECT ID
    FROM [Dogglenet2011].[dbo].[ProjectsSTARTED] )
    AND StatusID = 5
```

Q6: Filteren TimeLogs-tabel, zodat deze enkel de met de ProjectsCOMPLETED overeenstemmende logs zal bevatten.

```
SELECT *
INTO TimeLogsFILTERED
FROM [Dogglenet2011].[dbo].[TimeLogs]
WHERE ProjectID IN (
    SELECT ID
    FROM [Dogglenet2011].[dbo].[ProjectsCOMPLETED] )
    /* AND StatusID = 1 */
ORDER BY ProjectID
```

Notie:

De clause StatusID = 1 zou overbodig moeten zijn, daar alle logs met betrekking tot het projectwerk StatusID 1 zouden moeten hebben, echter om eventuele afwijkingen op te vangen vermelden we deze filter toch in de query.

Q7: Query ter creatie van de (basis) ProjectTeams-tabel.

```
SELECT ProjectID, UserID, FunctionID,
    sum(Duration_Hours) as TotalUserFnctHrsOnProject
INTO ProjectTeams
FROM [Dogglenet2011].[dbo].[TimeLogsFILTERED]
GROUP BY ProjectID, UserID, FunctionID
ORDER BY ProjectID, UserID, FunctionID
```

Q8: Query ter creatie van de extended ProjectTeams-tabel.

```
SELECT A.*,
    B.FunctionGroupID,
    C.DepartmentID, C.MainBusinessUnitID as UserMain_BusinessUnitID,
-- D.BusinessUnitID as Project_BusinessUnitID
INTO ProjectTeamsEXTENDED
FROM Dogglenet2011.dbo.ProjectTeams as A
JOIN Dogglenet2011.dbo.Functions as B
    on A.FunctionID = B.ID
JOIN Dogglenet2011.dbo.Users as C
    on A.UserID = C.ID
```

(query continued on next page...)

```
-- JOIN Dogglenet2011.dbo.ProjectsCOMPLETED as D
-- on A.ProjectID = D.ID
```

Notie:

De Project_BusinessUnitID werd louter toegevoegd om een mogelijke afwijking van de (User_)MainBusinessUnitID's ten opzichte van de BusinessUnitID van het project te illustreren in deze thesis, maar hoort in principe niet thuis in de ProjectTeamsEXTENDED-tabel.

Q9: Query voor het creëren van de basis ProjectsINFO-tabel.

```
SELECT ID as ProjectID,
       BrandID, IsInternal, BusinessUnitID as PrBUnitID,
       AccountManagerID, ProjectManagerID, AccountDirectorID
INTO Dogglenet2011.dbo.ProjectsINFO_tempBasic
FROM Dogglenet2011.dbo.ProjectsCOMPLETED
```

Q10: Query die het aantal timelogs per project (#TimeLogs) toevoegt aan de Projects-INFO-tabel (zij het wel via een copy out in een nieuwe tijdelijke tabel)

```
SELECT ProjectID, count(ID) as #TimeLogs
INTO Dogglenet2011.dbo.tempTable
FROM Dogglenet2011.dbo.TimeLogsFILTERED
GROUP BY ProjectID;

SELECT A.*, B.#TimeLogs
INTO Dogglenet2011.dbo.ProjectsINFO_temp1
FROM Dogglenet2011.dbo.ProjectsINFO_tempBasic as A
LEFT JOIN Dogglenet2011.dbo.tempTable as B
on A.ProjectID = B.ProjectID;

DROP TABLE Dogglenet2011.dbo.tempTable;
--DROP TABLE Dogglenet2011.dbo.ProjectsINFO_tempBasic;
```

Q11: Query die het totaal aantal gepresteerde uren per project berekent en toevoegt aan de Projects-INFO-tabel (zij het wel via een copy out in een tweede nieuwe tijdelijke tabel)

```
SELECT ProjectID,
       sum(TotalUserFnctHrsOnProject) as TotalHoursOnProject
INTO Dogglenet2011.dbo.tempTable
FROM Dogglenet2011.dbo.ProjectTeamsEXTENDED
GROUP BY ProjectID
ORDER BY ProjectID;

SELECT A.*, B.TotalHoursOnProject
INTO Dogglenet2011.dbo.ProjectsINFO_temp2
FROM Dogglenet2011.dbo.ProjectsINFO_temp1 as A
LEFT JOIN Dogglenet2011.dbo.tempTable as B
on A.ProjectID = B.ProjectID;

DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp1;
```

Q12: Query die het aantal teamleden per project (#TeamMembers) toevoegt aan de Projects-INFO-tabel (zij het wel via een copy out in een derde nieuwe tijdelijke tabel)

```
SELECT ProjectID, count(distinct UserID) as #TeamMembers
INTO Dogglenet2011.dbo.tempTable
FROM [Dogglenet2011].[dbo].[ProjectTeams]
GROUP BY ProjectID;
```

```
SELECT A.*, B.#TeamMembers
INTO Dogglenet2011.dbo.ProjectsINFO_temp3
FROM Dogglenet2011.dbo.ProjectsINFO_temp2 as A
LEFT JOIN Dogglenet2011.dbo.tempTable as B
on A.ProjectID = B.ProjectID;
```

```
DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp2;
```

Q13: Query die de looptijd (ProjectDuration) van een project berekent en toevoegt aan de ProjectsINFO-tabel.

```
SELECT A.*, DATEDIFF(day,B.EnterTime,B.StartTime) as ProjectDuration
INTO Dogglenet2011.dbo.ProjectsINFO_temp4
FROM Dogglenet2011.dbo.ProjectsINFO_temp3 as A
JOIN Dogglenet2011.dbo.ProjectsCOMPLETED as B
on A.ProjectID = B.ID
```

```
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp3;
```

Q14: Query die het aantal managers (uit de set van {accountmanager, accountdirector, projectmanager}) per project berekent.

```
SELECT ID as ProjectID,
AccountManagerID, AccountDirectorID, ProjectManagerID,
sum(CASE WHEN AccountManagerID is not null THEN 1 ELSE 0 END +
CASE WHEN AccountDirectorID is not null THEN 1 ELSE 0 END +
CASE WHEN ProjectManagerID is not null THEN 1 ELSE 0 END
) as #Managers
FROM Dogglenet2011.dbo.Projects
WHERE ID IN (7127,6513,6614,9414,6467,6635,9415,7948,6585)
GROUP BY ID, AccountManagerID, AccountDirectorID, ProjectManagerID
ORDER BY #Managers
```

Q15: Query die het aantal managers (uit de set van {accountmanager, accountdirector, projectmanager}) per project aan de ProjectsINFO-tabel toevoegt (zij het wel via een copy out in een vijfde nieuwe tijdelijke tabel).

```
SELECT ID as ProjectID,
sum(CASE WHEN AccountManagerID is not null THEN 1 ELSE 0 END +
CASE WHEN AccountDirectorID is not null THEN 1 ELSE 0 END +
CASE WHEN ProjectManagerID is not null THEN 1 ELSE 0 END
) as #Managers
INTO Dogglenet2011.dbo.tempTable
```

```

FROM [Dogglenet2011].[dbo].[Projects]
GROUP BY ID, AccountManagerID, AccountDirectorID, ProjectManagerID;

SELECT A.*, B.#Managers
INTO Dogglenet2011.dbo.ProjectsINFO_temp5
FROM Dogglenet2011.dbo.ProjectsINFO_temp4 as A
LEFT JOIN Dogglenet2011.dbo.tempTable as B
      on A.ProjectID = B.ProjectID;

DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp4;

```

Q15: Query die de sub-score attributen en het globale score-attribuut voor het bepalen van de projectomvang berekent.

```

SELECT ProjectID,
       BrandID, IsInternal, PrBUnitID,
       AccountManagerID, ProjectManagerID, AccountDirectorID,

       #Managers,
       CASE WHEN #Managers > 1 THEN 1
            ELSE 0 END
       as #Managers_score,

       #TeamMembers,
       CASE WHEN #TeamMembers > 40 THEN 5
            WHEN #TeamMembers > 15 THEN 4
            WHEN #TeamMembers > 5 THEN 3
            WHEN #TeamMembers > 1 THEN 2
            WHEN #TeamMembers = 1 THEN 1
            ELSE NULL END
       as #TeamMembers_score,

       #TimeLogs,
       CASE WHEN #TimeLogs > 500 THEN 5
            WHEN #TimeLogs > 200 THEN 4
            WHEN #TimeLogs > 50 THEN 3
            WHEN #TimeLogs > 10 THEN 2
            WHEN #TimeLogs > 0 THEN 1
            ELSE NULL END
       as #TimeLogs_score,

       TotalHoursOnProject,
       CASE WHEN TotalHoursOnProject > 1000 THEN 5
            WHEN TotalHoursOnProject > 200 THEN 4
            WHEN TotalHoursOnProject > 50 THEN 3
            WHEN TotalHoursOnProject > 10 THEN 2
            WHEN TotalHoursOnProject > 0 THEN 1
            ELSE NULL END
       as TotalHours_score,

       ProjectDuration,
       CASE WHEN ProjectDuration > 365 THEN 3
            WHEN ProjectDuration > 30 THEN 2

```

```

        WHEN ProjectDuration > 0 THEN 1
        ELSE 0 END
    as ProjectDuration_score,

INTO  Dogglenet2011.dbo.ProjectsINFO_temp6
FROM  Dogglenet2011.dbo.ProjectsINFO_temp5;

-----

SELECT A.*,

       cast( round( ( #TeamMembers_score
                    + #TimeLogs_score
                    + TotalHours_score
                    + ProjectDuration_score
                    + #Managers_score ) *5.0/19, 0) as int)
       as GlobalProjectSize

INTO  Dogglenet2011.dbo.ProjectsINFO_temp7
FROM  Dogglenet2011.dbo.ProjectsINFO_temp6 as A;

-----

DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp5;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp6;

```

Q16: Query voor toevoegen aantal verschillende functies per project en aantal verschillende functiegroepen per projecten – aan ProjectsINFO-tabel.

```

SELECT ProjectID,
       count(distinct FunctionID) as #Functions,
       count(distinct FunctionGroupID) as #FunctionGroups,
       count(distinct DepartmentID) as #Departments
INTO  Dogglenet2011.dbo.tempTable
FROM  Dogglenet2011.dbo.ProjectTeamsEXTENDED
GROUP BY ProjectID
ORDER BY ProjectID;

SELECT A.*, B.#Functions, B.#FunctionGroups
INTO  Dogglenet2011.dbo.ProjectsINFO_temp8
FROM  Dogglenet2011.dbo.ProjectsINFO_temp7 as A
LEFT JOIN  Dogglenet2011.dbo.tempTable as B
         on  A.ProjectID = B.ProjectID;

DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp7;

```

Q17: Query die het aantal teamleden per functiegroep (voor FG van 1 tot 7) per project berekent.

```

SELECT ProjectID, count(distinct UserID) as #UinFG_1
INTO  Dogglenet2011.dbo.tempTable
FROM  [Dogglenet2011].[dbo].[ProjectTeamsEXTENDED]
WHERE FunctionGroupID = 1
GROUP BY ProjectID, FunctionGroupID;

```

```

SELECT A.*, B.#UinFG_1
INTO  Dogglenet2011.dbo.ProjectsINFO_temp9a
FROM  Dogglenet2011.dbo.ProjectsINFO_temp8 as A
LEFT JOIN  Dogglenet2011.dbo.tempTable as B
          on  A.ProjectID = B.ProjectID;

```

```

DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp8;

```

```

-----
SELECT ProjectID, count(distinct UserID) as #UinFG_2
INTO  Dogglenet2011.dbo.tempTable
FROM  [Dogglenet2011].[dbo].[ProjectTeamsEXTENDED]
WHERE FunctionGroupID = 2
GROUP BY ProjectID, FunctionGroupID;

```

```

SELECT A.*, B.#UinFG_2
INTO  Dogglenet2011.dbo.ProjectsINFO_temp9b
FROM  Dogglenet2011.dbo.ProjectsINFO_temp9a as A
LEFT JOIN  Dogglenet2011.dbo.tempTable as B
          on  A.ProjectID = B.ProjectID;

```

```

DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp9a;

```

```

-----
SELECT ProjectID, count(distinct UserID) as #UinFG_3
INTO  Dogglenet2011.dbo.tempTable
FROM  [Dogglenet2011].[dbo].[ProjectTeamsEXTENDED]
WHERE FunctionGroupID = 3
GROUP BY ProjectID, FunctionGroupID;

```

```

SELECT A.*, B.#UinFG_3
INTO  Dogglenet2011.dbo.ProjectsINFO_temp9c
FROM  Dogglenet2011.dbo.ProjectsINFO_temp9b as A
LEFT JOIN  Dogglenet2011.dbo.tempTable as B
          on  A.ProjectID = B.ProjectID;

```

```

DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp9b;

```

```

-----
(gelijkaardige code voor functiegroepen 4, 5 en 6)
-----

```

```

SELECT ProjectID, count(distinct UserID) as #UinFG_7
INTO  Dogglenet2011.dbo.tempTable
FROM  [Dogglenet2011].[dbo].[ProjectTeamsEXTENDED]
WHERE FunctionGroupID = 7
GROUP BY ProjectID, FunctionGroupID;

```

```

SELECT A.*, B.#UinFG_7
INTO  Dogglenet2011.dbo.ProjectsINFO_temp9
FROM  Dogglenet2011.dbo.ProjectsINFO_temp9f as A
LEFT JOIN  Dogglenet2011.dbo.tempTable as B
          on  A.ProjectID = B.ProjectID;

```

```

DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp9f;

```

Q18: Query die het gepresteerde uren per functiegroep (voor FG van 1 tot 7) per project berekent.

```
SELECT ProjectID, sum(TotalUserFnctHrsOnProject) as #HinFG_1
INTO  Dogglenet2011.dbo.tempTable
FROM  Dogglenet2011.dbo.ProjectTeamsEXTENDED
WHERE FunctionGroupID = 1
GROUP BY ProjectID, FunctionGroupID;
```

```
SELECT A.*, B.#HinFG_1
INTO  Dogglenet2011.dbo.ProjectsINFO_temp11a
FROM  Dogglenet2011.dbo.ProjectsINFO_temp10 as A
LEFT JOIN  Dogglenet2011.dbo.tempTable as B
         on  A.ProjectID = B.ProjectID;
```

```
DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp10;
```

```
-----
SELECT ProjectID, sum(TotalUserFnctHrsOnProject) as #HinFG_2
INTO  Dogglenet2011.dbo.tempTable
FROM  Dogglenet2011.dbo.ProjectTeamsEXTENDED
WHERE FunctionGroupID = 2
GROUP BY ProjectID, FunctionGroupID;
```

```
SELECT A.*, B.#HinFG_2
INTO  Dogglenet2011.dbo.ProjectsINFO_temp11b
FROM  Dogglenet2011.dbo.ProjectsINFO_temp11a as A
LEFT JOIN  Dogglenet2011.dbo.tempTable as B
         on  A.ProjectID = B.ProjectID;
```

```
DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp11a;
```

(gelijkaardige code voor functiegroepen 3, 4, 5 en 6)

```
-----
SELECT ProjectID, sum(TotalUserFnctHrsOnProject) as #HinFG_7
INTO  Dogglenet2011.dbo.tempTable
FROM  Dogglenet2011.dbo.ProjectTeamsEXTENDED
WHERE FunctionGroupID = 7
GROUP BY ProjectID, FunctionGroupID;
```

```
SELECT A.[ProjectID]
      ,A.[BrandID]
      ,A.[IsInternal]
      ,A.[PrBUnitID]
      ,A.[AccountManagerID]
      ,A.[ProjectManagerID]
      ,A.[AccountDirectorID]
      ,A.[#Managers]
      ,A.[#Managers_score]
      ,A.[#TeamMembers]
      ,A.[#TeamMembers_score]
      ,A.[#TimeLogs]
      ,A.[#TimeLogs_score]
```

```

,A.[TotalHoursOnProject]
,A.[#TotalHours_score] as TotalHours_score
,A.[ProjectDuration]
,A.[ProjectDuration_score]
,A.[GlobalProjectSize]
,A.[#Functions]
,A.[#FunctionGroups]
,A.[#Departments]
,A.[#UinFG_1]
,A.[#HinFG_1]
,A.[#UinFG_2]
,A.[#HinFG_2]
,A.[#UinFG_3]
,A.[#HinFG_3]
,A.[#UinFG_4]
,A.[#HinFG_4]
,A.[#UinFG_5]
,A.[#HinFG_5]
,A.[#UinFG_6]
,A.[#HinFG_6]
,A.[#UinFG_7]
,B.[#HinFG_7]
INTO Dogglenet2011.dbo.ProjectsINFO_temp11
FROM Dogglenet2011.dbo.ProjectsINFO_temp11f as A
LEFT JOIN Dogglenet2011.dbo.tempTable as B
      on A.ProjectID = B.ProjectID;

DROP TABLE Dogglenet2011.dbo.tempTable;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp11f;

```

Q19: Query die het verschil tussen de grootste #HFG- en de kleinste #HFG-observatie uit een rij berekent, deze normaliseert tot een waarde tussen 0 en 1, en deze vervolgens opslaat in het nieuwe #HFG-Spread-attribuut.

```

CREATE FUNCTION RowMinimum (
      @value1 decimal(38, 5),
      @value2 decimal(38, 5)
) RETURNS int
AS
BEGIN
    DECLARE @Result decimal(38, 5)
    SET @value1 = ISNULL(@value1,0)
    SET @value2 = ISNULL(@value2,0)
    IF @value1 <= @value2
        SET @Result = @value1
    ELSE
        SET @Result = @value2
    RETURN @Result
END
GO

CREATE FUNCTION RowMaximum (
      @value1 decimal(38, 5),
      @value2 decimal(38, 5)
) RETURNS int
AS
BEGIN
    DECLARE @Result decimal(38, 5)
    SET @value1 = ISNULL(@value1,0)

```



```

        SET @value2 = ISNULL(@value2, 0)
        IF @value1 >= @value2
            SET @Result = @value1
        ELSE
            SET @Result = @value2
    RETURN @Result
END
GO

```

Q20: Query die de sub-score attributen en het globale score-attribuut voor het bepalen van de projectdiversiteit berekent.

```

SELECT ProjectID,
    BrandID, IsInternal, PrBUnitID,
    AccountManagerID, ProjectManagerID, AccountDirectorID,
    #Managers, #Managers_score, #TeamMembers, #TeamMembers_score,
    #TimeLogs, #TimeLogs_score, TotalHoursOnProject, TotalHours_score,
    ProjectDuration, ProjectDuration_score,

    #Functions,
    CASE WHEN #TeamMembers > 15 THEN 4
         WHEN #TeamMembers > 10 THEN 3
         WHEN #TeamMembers > 5 THEN 2
         WHEN #TeamMembers > 0 THEN 1
         ELSE NULL END
    as #Functions_score,

    #FunctionGroups,
    CASE WHEN #FunctionGroups IN (6,7) THEN 5
         WHEN #FunctionGroups = 5 THEN 4
         WHEN #FunctionGroups = 4 THEN 3
         WHEN #FunctionGroups = 3 THEN 2
         WHEN #FunctionGroups IN (1,2) THEN 1
         ELSE NULL END
    as #FunctionGroups_score,

    CASE WHEN #HFG-Spread < 0.20 THEN 5
         WHEN #HFG-Spread < 0.40 THEN 4
         WHEN #HFG-Spread < 0.60 THEN 3
         WHEN #HFG-Spread < 0.80 THEN 2
         WHEN #HFG-Spread <= 1.0 THEN 1
         ELSE NULL END
    as #HFG-Spread_score,

    #UinFG_1, #HinFG_1,
    #UinFG_2, #HinFG_2,
    #UinFG_3, #HinFG_3,
    #UinFG_4, #HinFG_4,
    #UinFG_5, #HinFG_5,
    #UinFG_6, #HinFG_6,
    #UinFG_7, #HinFG_7

INTO Dogglenet2011.dbo.ProjectsINFO_temp13
FROM Dogglenet2011.dbo.ProjectsINFO_temp12;
-----
SELECT A.*,
    cast( round(
        ( #Functions_score
        + #FunctionGroups_score
        + HFG_score ) *5.0/14, 0) as int)

```

```

        as GlobalProjectDiversity
INTO  Dogglenet2011.dbo.ProjectsINFO_temp14
FROM  Dogglenet2011.dbo.ProjectsINFO_temp13 as A;
-----
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp11;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp12;
DROP TABLE Dogglenet2011.dbo.ProjectsINFO_temp13;

```

Q21: TotalBillingAmount

```

SELECT [ProjectID], sum(Amount_EURO) as TotalBillingAmount
FROM  [Dogglenet2011].[dbo].[InvoiceEntries]
Group by ProjectID
Order by ProjectID

```

Q22: TotalBudget

```

SELECT A.ProjectID, sum(B.TotalAmount_EURO) as TotalBudget
FROM  Dogglenet2011.dbo.Budgets as A
JOIN  Dogglenet2011.dbo.BudgetNodes as B
on A.ID = B.ID
WHERE A.StatusID = 3
GROUP BY A.ProjectID
ORDER BY A.ProjectID

```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:

Een data-analyse model voor management en project diensten

Richting: **master in de toegepaste economische wetenschappen:
handelsingenieur in de beleidsinformatica**

Jaar: **2015**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Janssens, Nickolas

Datum: **24/08/2015**