SINGAPORE MANAGEMENT UNIVERSITY

ANLY482 - ANALYTICS PRACTICUM

# MID TERM REPORT

TEAM MST

HENG Yi Teng Mabel
Siti HAMIDAH Binte Abdul Hamid
Tarika GUPTA

# Contents

# 1. Recap & Objective

Our sponsor, TrustSphere is a software company that provides relationship analytics solutions. Their products deliver insights that help clients across the globe improve key business issues including sales force effectiveness, enterprise-wide collaboration and corporate governance. The company engaged our team to utilize our technical and analytical capabilities to help them understand and tackle **their business problem of little growth in sales and a longer than ideal sales cycle.**

While the field of Sales Analytics has received plenty attention in the past, recent studies reveal that few companies have also delved into the area of Sales People Analytics. Salespeople communications to potential clients, especially in the B2B sphere, are wholly relied upon for marketing the company's product. Furthermore, Steward et al. (2010) found that higher-performing salespeople also regularly activated their internal company networks, to coordinate a team of experts tailored to serve a particular customer. Just sales figures to evaluate salespeople performance covers a very narrow perspective as it disregards cycle time and in-progress pitches, therefore our team has defined our scope as **to analyse the sales team's internal and external communications to gain insight into their relationships with internal and external parties and to identify the sales stages that act as bottlenecks in the sales process.**

# 2. Data Provided by the Client

For this project, our team is working with two sets of data provided to us by TrustSphere:

## 2.1 Daily email communication data (MAIN DATASET)

This dataset contains year-to-date (up till 31 August 2016) records of daily email communication data of all 19 TrustSphere sales people across the globe. This data includes the following variables. (109,771 Records)

***Date:*** *Includes the date and time of a particular email being sent*

***Originator address:*** *Sender email address*

***Recipient address:*** *Receiver email address*

***Direction:*** *Nature of communication (internal, inbound or outbound)*

***MsgID:*** *Unique message ID of emails sent*

| Date | Originator Address | Recipient Addre | Direction | Subject | Msgid | Size |
|---|---|---|---|---|---|---|
| 31/8/16 23:56 | tom.butler@trustsphere.cor | mark.padginton@tr | internal | RE: TquilaANZ CRM Diagnosti | <AM4PR0201MB1923867721618F2B03C7D15D98E30@AM4PR0201MB1923;eurprd02;prod;outlook;com> | |
| 31/8/16 23:54 | mark.padginton@trustspher | tom.butler@trustsp | internal | TquilaANZ CRM Diagnostic da | <8772BB27-2927-4EA4-BBEF-AEC18C545A38@trustsphere;com> | |
| 31/8/16 23:54 | mark.padginton@trustspher | ricardo.gonzalez@tr | internal | TquilaANZ CRM Diagnostic da | <8772BB27-2927-4EA4-BBEF-AEC18C545A38@trustsphere;com> | |
| 31/8/16 23:52 | mark.padginton@trustspher | cameron.hook@tqu | outbound | Re: SalesForce Integration | <D6F364EA-A083-4D22-9C07-A8FC56F7325F@trustsphere;com> | |
| 31/8/16 23:52 | mark.padginton@trustspher | tom.butler@trustsp | internal | Re: SalesForce Integration | <D6F364EA-A083-4D22-9C07-A8FC56F7325F@trustsphere;com> | |
| 31/8/16 23:51 | mark.padginton@trustspher | warren.tait@trustsp | internal | ESRI | <1164AE5A-313C-4518-99F1-9C59AB322AFB@trustsphere;com> | |
| 31/8/16 23:51 | cameron.hook@tquilaanz.cc | mark.padginton@tr | inbound | Fwd: SalesForce Integration | <CAB+Ok+4 mzVF1t9B YLp444PxJ41zuS- pOyNP11OjdtZ+c4Jg@mail;gmail;com> | |
| 31/8/16 23:38 | warren.tait@trustsphere.co | kathleen.tan@trust | internal | Accepted: A&NZ Strategy Call | <061a2df6db744109bb550965c4fac5f3@HE1PR0201MB1465;eurprd02;prod;outlook;com> | |
| 31/8/16 23:38 | warren.tait@trustsphere.co | tabitha.paton-durra | internal | Accepted: A&NZ Strategy Call | <061a2df6db744109bb550965c4fac5f3@HE1PR0201MB1465;eurprd02;prod;outlook;com> | |
| 31/8/16 23:38 | warren.tait@trustsphere.co | grace.siew@trustsp | internal | Accepted: A&NZ Strategy Call | <061a2df6db744109bb550965c4fac5f3@HE1PR0201MB1465;eurprd02;prod;outlook;com> | |
| 31/8/16 23:38 | warren.tait@trustsphere.co | manish.goel@trusts | internal | Accepted: A&NZ Strategy Call | <DB5PR0201MB1605FF38833725266E593FE9E3E30@DB5PR0201MB1605;eurprd02;prod;outlook;com> | |

***Email Subject:*** *Email subject header*

*Figure 1 - Original Dataset*

## 2.2 Staff List

The dataset lists all of TrustSphere staff (57) with the following variables.

***Name***

***Hierarchy***

***Department***

***Position***

***Location***

# 3. Scope of Work

After repeated interaction with the Sales team our team has decided to divide the scope into the following sections

## 3.1 Relationship Report

a. Analyses the number and strength of Internal and External Relationships Salespeople have developed over the analysis period.
b. Takes into account the frequency and recency of emails exchanged by and with the sales person to highlight their communication and collaboration effort.

## 3.2 Clients and Sales Stages

a. Reports the sales progress for each account up-till 31st August. That is, how many account are active and what stage of the sales cycle they're in.
b. Evaluates the performance of each salesperson depending on how many accounts they have in each stage, their response trends in each stage, progress from historical communication indications etc.
**c.** Provides a postmortem on inactive accounts, reporting on how long communications with a client lasted, what stage did communications end and which salespeople were responsible.

## 3.3 Social Network Analysis

a. Analyses collaboration trends, that is, what department do salespeople interact with more, for example does interacting with C-Suite employees correlate with better performance?
b. Examines overlap trends, to see if multiple salespeople interact with the same client, and whether abnormalities exist within the overlap.
c. Examines how networks differ with relation to location, whether their lies a communication gap between teams based in different regions.

## 4. Methodology

| Area of Interest | Analyses | Purpose |
|---|---|---|
| Relationship Report | Data Transformation, Summary Statistics and Visualisations | - To condense email data into unique Salesperson-Person X relationships<br>- Classify if the relationships are Internal/External<br>- Establish the the frequency and recency of emails exchanged in each relationship<br>- Establish the timeline of emails exchanged within the relationship |
| Clients & Stages | Data Transformation, Text Mining, Summary Statistics | - To condense email data into unique records of email threads (1 record -> 1 email thread)<br>- Use text mining to determine which thread the email belongs to.<br>- Establish the timeline of emails exchanged within an email thread.<br>- Calculating average response time within the thread. |
| Salespeople Network | Social Network Analysis | - To observe what departments salespeople communicate with<br>- To observe how salespeople from different locations interact with each other<br>- To observe overlap between salespeople |

# 5. Classifications and Metrics

Following our exploratory analysis of the data and the scope and repeated meetings with the client we have established certain classifications and metrics that ease reporting of evaluations of the sales performance.

## 5.1 Stages and clients

1. <u>Sales Stages</u>

    a. Prospecting Stage: When a prospecting email is sent out to a prospective client. They may or may not respond.

    b. Meeting Stage: When the client has responded favourably to the prospecting stage and TrustSphere has a scheduled pitch meeting with the client.

    c. POC Stage: When the client has agreed to commission a product trial.

    d. After POC Stage: Follow ups, quotations, contracts etc.

The classification of these stages was provided to us by TrustSphere. Their commission structure rewards sales people after they get a client into the Meeting and POC Stage. Therefore all emails after the POC stage are classified into the the ‘After POC’ stage.

2. <u>Active and Inactive Clients</u>

    a. Active Client: Contact has taken place within the past 30 days

    b. Inactive Client: No contact has taken place in the past 30 days.

3. <u>Failed Prospects</u>

Failed Prospect is a classification that indicates what percentage of prospecting emails sent out did not make it to the meeting stage.

## 5.2 Relationship Report

1. Hot & Cold Relationships

    a. Hot Relationship: Last contact was made less than 3 days ago.

    b. Cold Relationship: Last Contact was made more than 3 days ago.

2. Strong & Weak Relationships

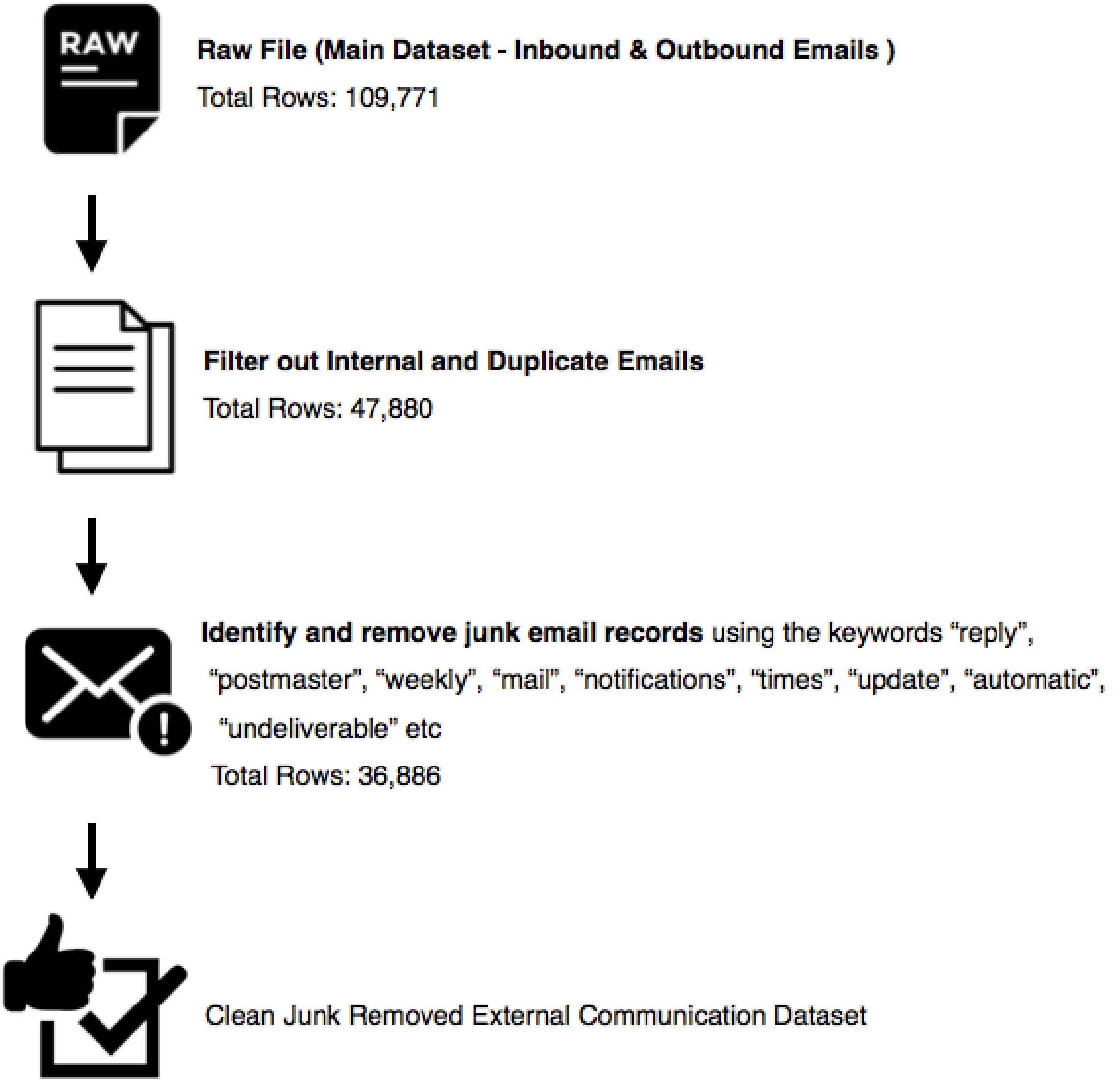
    a. Strong Relationship: Above average number of emails exchanged AND is a hot relationship

    b. Weak Relationship: Below average number of emails exchanged AND/OR Cold Relationship

# 6. Data Cleaning & Transforming

## 6.1 Stages and Clients

### 6.1.1. REMOVING JUNK AND INTERNAL EMAILS



### 6.1.2. CONDENSING EMAIL RECORDS INTO UNIQUE THREAD RECORDS

Following The Data Cleaning in Step 1, we wished to reduce the emails to unique threads, i.e. 1 record for each thread, instead of each email

**Identifying unique email threads using**

- Salesperson
- Contact person
- Subject Thread

Total Rows: 20,090

### 6.1.3. COMPUTED FIELDS

**Direction of Emails**: Number Inbound, Number Outbound emails in each thread

**Average Response Time:** Average response time for an email reply

Out of these 20090 unique salesperson – contact – subject thread email communications, there were some with and without replies. Further steps needed to be taken to remove irrelevant emails.

### 6.1.4. FURTHER REMOVAL OF IRRELEVANT EMAILS

Two key aspects were used to determine if each unique salesperson – contact communication is relevant

- Is there at least 1 reply within each unique salesperson – contact relationship?

- Do the subject threads between each unique salesperson – contact relationship contain specific keywords (“Relationship Analytics”, “POC”?)

**There needs to be 1 reply between a salesperson-client for the communication to be relevant.**

The Sales cycle consists of 4 stages. Email subject headers may change during these stages. There are no specific templates or subject headers used except for the following 2 keywords (“Relationship analytics”, “POC”). There would also be some subject headers that required a reply and those that don’t. **Our team has concluded that within each sales process, there should be a minimum of 1 reply between a salesperson and his contact**. We have categorized such unique salesperson – contact relationship with at least 1 reply for any thread to be **relevant**.

**The protocol keywords of either “Relationship Analytics” or “POC” need to be present in communication threads for a salesperson-client communication to be relevant.**

Although a unique salesperson – contact relationship may have at least 1 reply throughout their email exchange, their communication exchange may not be relevant to sales or the process of acquiring a potential client. TrustSphere has a number of partners whom are important to them. As such, this communication may be email exchanges between TrustSphere and their partners, which are irrelevant. For each unique salesperson – contact – thread record, our team has identified if the thread has one of the following keywords (“Relationship analytics”, “POC”). **A unique salesperson – contact communication is only relevant if there is at least 1 reply and at least 1 thread with stated keywords throughout their communication exchange.**

### 6.1.5. DETERMINE STAGES

At the end of this steps, a total of **1814 / 20090** unique salesperson – contact – thread records are identified as relevant. These unique records contain a total of 5864 emails sent and received.

Our next step is to look at each unique salesperson – contact communication and determine their sales stages. TrustSphere has provided the team with a list of their partners. For each of these 1814 unique salesperson – contact – thread records, out team has categorised them into the following groups:

1. Prospecting
2. Meeting
3. POC
4. After POC

The team will consult with TrustSphere to ensure these categorisations are accurate. The team will also delve into the use of text analytics for further insights and alternative categorisations methodologies.

### 6.1.6. FINAL DATASET

| A | B | C | D | E | F | G | H | I | J | |
|---|---|---|---|---|---|---|---|---|---|---|
| Salesperson | Contact | Thread | Incoming Count | Outgoing Count | Date First In/Out | Date Last In | Date Last Out | Average Resp | Total Response Ler | Phase |
| michael.knight@trustsphere.com | ajuliano@landmarkventures.com | TrustSphere - Relationship Analytics and Knowledge Management | 1 | 1 | 42453.89479 | 42453.89479 | 42453.90439 | 14 | 14 | Prospecting |
| michael.knight@trustsphere.com | ajuliano@landmarkventures.com | Connecting | 0 | 1 | 42453.904 | | 42453.904 | | | Prospecting |
| michael.knight@trustsphere.com | jeff.marshall@havas.com | Follow up from you inquiry | 3 | 0 | 42459.95078 | 42461.11356 | | | | Meeting |
| michael.knight@trustsphere.com | jeff.marshall@havas.com | POC Guide for Havas | 1 | 0 | 42468.90168 | 42468.90168 | | | | POC |
| michael.knight@trustsphere.com | jeff.marshall@havas.com | POC | 2 | 3 | 42473.62376 | 42474.5963 | 42474.5978 | 701 | 1402 | POC |
| michael.knight@trustsphere.com | cindy.hodnett@brookfield.com | Follow up from VM regarding Linkswithin | 3 | 0 | 42565.94313 | 42566.68287 | | | | Meeting |
| michael.knight@trustsphere.com | cindy.hodnett@brookfield.com | Brookfield-Trustsphere Meeting | 1 | 1 | 42566.68269 | 42566.68269 | 42599.50637 | 47266 | 47266 | Meeting |
| michael.knight@trustsphere.com | cindy.hodnett@brookfield.com | TrustSphere Presentation 8\\\/17\\\/16 | 4 | 1 | 42591.90271 | 42598.75714 | 42598.8423 | 123 | 123 | Meeting |
| michael.knight@trustsphere.com | cindy.hodnett@brookfield.com | Meeting with TrustSphere - Confirmed | 1 | 1 | 42599.74586 | 42599.74586 | 42599.75377 | 11 | 11 | Meeting |

*Figure 2 - Cleaned & Transformed Sales Stage Dataset*

***Salesperson:*** TrustSphere Sales Staff

***Contact:*** Email address of person a salesperson is in communication with

***Thread:*** ThreadSubject

***Incoming count:*** Total count of inbound emails in the thread

***Outgoing count:*** Total count of outbound emails in the thread

***Date First In / Out:*** Date email is first sent or received for a particular thread

***Date Last In:*** Date email is last received for a particular thread

***Date Last Out:*** Date email is last sent for a particular thread

***Average Response Time:*** Average time taken for an email reply

***Total Response Time:*** Total time taken for all email replies

***Relevant:*** True if a Salesperson – Contact communication has at least 1 reply in their entire email communication. False if otherwise.

***Keyword:*** True if a Salesperson – Contact communication with at least 1 thread with the keywords "relationship analytics" or "poc" in their entire email communication. False if otherwise.

***Partner:*** True if Contact is a TrustSphere partner. False if otherwise.

***Phase:*** Stage of the sales process (Prospecting, Meeting, POC, After POC)

## 6.2 Relationship Report

### 6.2.1. CLASSIFYING EMAILS INTO MONTHS

The relationship report was created using emails of the whole data set of 7 months, as well as separately for each month from January to August to observe time series changes.

**Raw File (Main Dataset, Inbound & Outbound Emails)**

Total Rows: 109,771

The Dataset was then distributed across 8 worksheets

Combined: All email communications from Jan 1 to Aug 31 - **109,771 Rows**

January: All email communications from Jan 1 to Jan 31 - **10,831 Rows**

February: All email communications from Feb 1 to Feb 29 - **12,038 Rows**

March: All email communications from Mar 1 to Mar 31 - **12,755 Rows**

April: All email communications from Apr 1 to Apr 30 - **12,611 Rows**

May: All email communications from May 1 to May 31 - **17,097 Rows**

June: All email communications from June 1 to June 30 - **14,176 Rows**

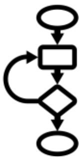
July: All email communications from Jul 1 to Jul 30 - **14,576 Rows**

August: All email communications from Aug 1 to Aug 31 - **15,694 Rows**

### 6.2.2. DATA TRANSFORMATION AND COMPUTED COLUMNS FOR EACH RELATIONSHIP (UNIQUE ORIGINATOR - RECEIVER COMBINATION)

The data was then transformed from 1 record being 1 email communication between an originator and a receiver to 1 record being a salesperson-contact person relationship. The computed columns included number of inbound and outbound emails sent and received between the Salesperson and Contact person. The number of days since the last email was sent and received by the salesperson from the contact person etc.

The following process was repeated for all 8 worksheets:



**Condense email communications to unique originator email and receiver email combination (Using Summary Tables - JMP 12)**

| Originator Address | Recipient Address | N Rows |
|---|---|---|
| 123js sitf.org.sg@mail135.wdc02.mcdlv.net | annabel.koh@trustsphere.com | 1 |
| 123js sitf.org.sg@mail226.atl61.mcsv.net | dev.menon@trustsphere.com | 1 |
| 123js sitf.org.sg@mail27.suw15.mcsv.net | annabel.koh@trustsphere.com | 1 |
| 123js sitf.org.sg@mail32.atl71.mcdlv.net | dev.menon@trustsphere.com | 1 |
| 123js sitf.org.sg@mail38.suw17.mcsv.net | annabel.koh@trustsphere.com | 1 |
| 123js sitf.org.sg@mail76.atl11.rsgsv.net | annabel.koh@trustsphere.com | 1 |
| 123js sitf.org.sg@mail80.atl51.rsgsv.net | dev.menon@trustsphere.com | 1 |
| 123js sitf.org.sg@mail95.atl71.mcdlv.net | dev.menon@trustsphere.com | 1 |
| 1salome@my.ibm.com | shaun.keating@trustsphere.com | 1 |
| 25603@ontramail.com | ajay.rana@trustsphere.com | 6 |
| 2ksaurabh@gmail.com | manmeet.nayyar@trustsphere.com | 1 |

*Figure 3 - Unique relationships with frequency of emails sent from originator to recipient*

Now that we have a single record of each originator and recipient id and the number of emails sent from Originator to Recipient (N Rows in Figure 3). We compute -

**Compute how many emails were sent to Originator from Recipient. (Excel VLOOKUP)**

**Compute how many days ago the last email was sent and received by the originator from the recipient. (Excel VLOOKUP)**

*Combined - How many days since 31st August contact was made.

Individual Month - How many days since the last date of that month contact was made.

**Classify into internal and External Relationships**

**Select the records that have the concerned salespeople in the Originator Address and filter out the rest.**

The filtering process has a distinct advantage:

1. The relationships where Salespersons are not the originator are either Spam or Internal personnel that have never been replied to (TrustSphere does not get inquiry emails that go un-replied.) Both Spam and unreciprocated internal relationships automatically get cleared out.

### 6.3.3. FINAL DATASET

***Originator Address:*** Sales Person

***Receiver Address:*** Contact Person

***Type:*** Whether it is an external/internal relationship

***Outbound Count:*** Number of Emails sent from Salesperson to Contact Person

***Inbound Count:*** Number of Emails received by Salesperson from Contact Person

***Total Frequency:*** Sum of Inbound and Outbound Count. Total emails exchanged between Salesperson and Contact Person

***Last Out***: Number of days ago the last email from sent from Salesperson to Contact Person*

***Last In***: Number of days ago the last email was received by Salesperson from Contact Person*

***Last Contact***: The lesser of Last in and Last Out. The number of days ago the last communication was sent or received.*

***Table:*** Which month does the record belong to

***Hot/Cold:*** Whether the relationship is hot or cold (Refer to Classifications and Metrics)

***Strong/Weak:*** Whether the relationship is strong or weak (Refer to Classifications and Metrics)

| A | B | C | D | E | F | G | H | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Originator / | Recipient A | Outbound | Inbound | Total Frequ | Last Out | Last In | Last Contac | Type | HOT/COLD | STRONG/W | Table |
| annabel.ko | ac@kerrycc | 1 | 1 | 2 | 23 | 23 | 23 | External | COLD | WEAK | January |
| annabel.ko | adhiraj@ta | 2 | 1 | 3 | 3 | 3 | 3 | External | HOT | WEAK | January |
| annabel.ko | allen@sitf. | 5 | 2 | 7 | 13 | 17 | 13 | External | COLD | WEAK | January |
| annabel.ko | amanda@c | 3 | 1 | 4 | 11 | 11 | 11 | External | COLD | WEAK | January |
| annabel.ko | andywoo@ | 1 | 1 | 2 | 5 | 5 | 5 | External | COLD | WEAK | January |
| annabel.ko | antoine@s | 2 | 1 | 3 | 19 | 20 | 19 | External | COLD | WEAK | January |
| annabel.ko | bng@axwa | 1 | 1 | 2 | 4 | 4 | 4 | External | COLD | WEAK | January |
| annabel.ko | casey_oww | 2 | 1 | 3 | 19 | 20 | 19 | External | COLD | WEAK | January |

*Figure 4 - Final Relationship Report*

# 7. Future Tasks and Deliverables

Now that our Datasets are ready our next steps are to -

**A. Visualise Data and Derive Insights**

We will be Visualising the data as deemed fit by our scope of work. After observing trends through visualisations we intend to report derived insights and recommendations through a final project report and presentation.

**B. Create a Dashboard to Report Insights to Sales Manager**

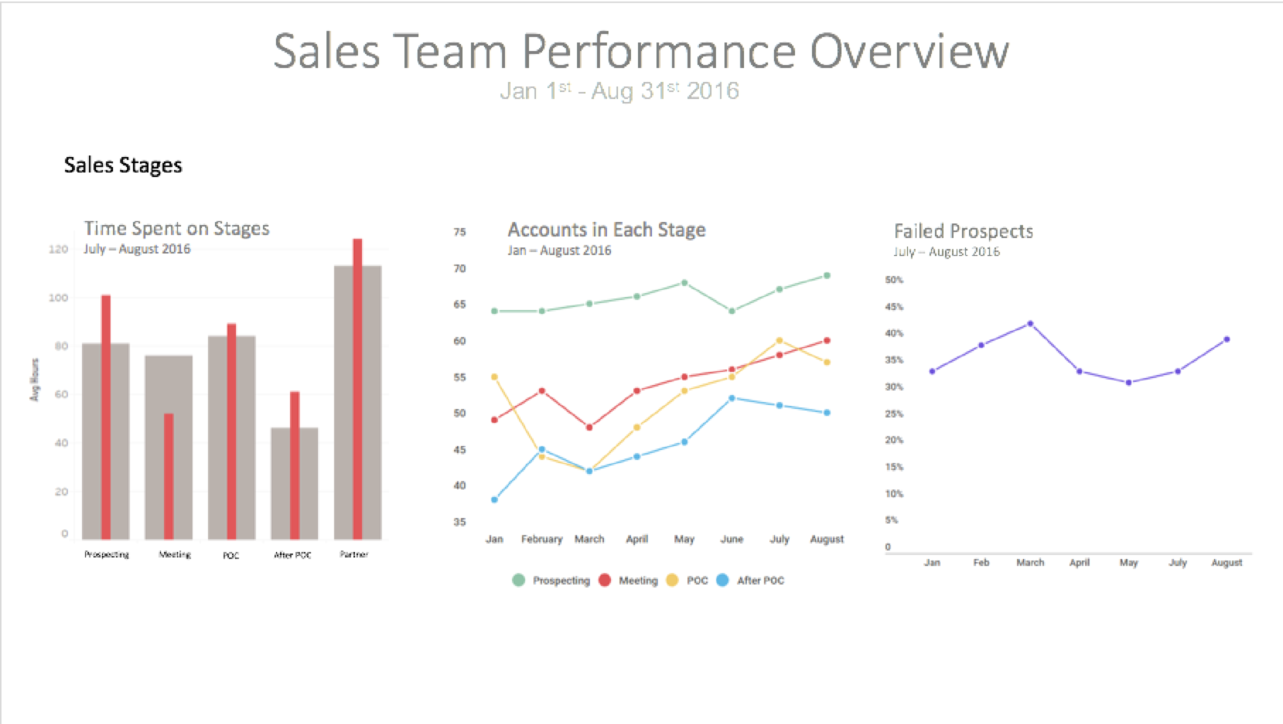
- Using Javascript and D3 we will create a web based dashboard to report insights to the sales insights to the Sales Director. (Mock Up created on Powerpoint, Appendix A).

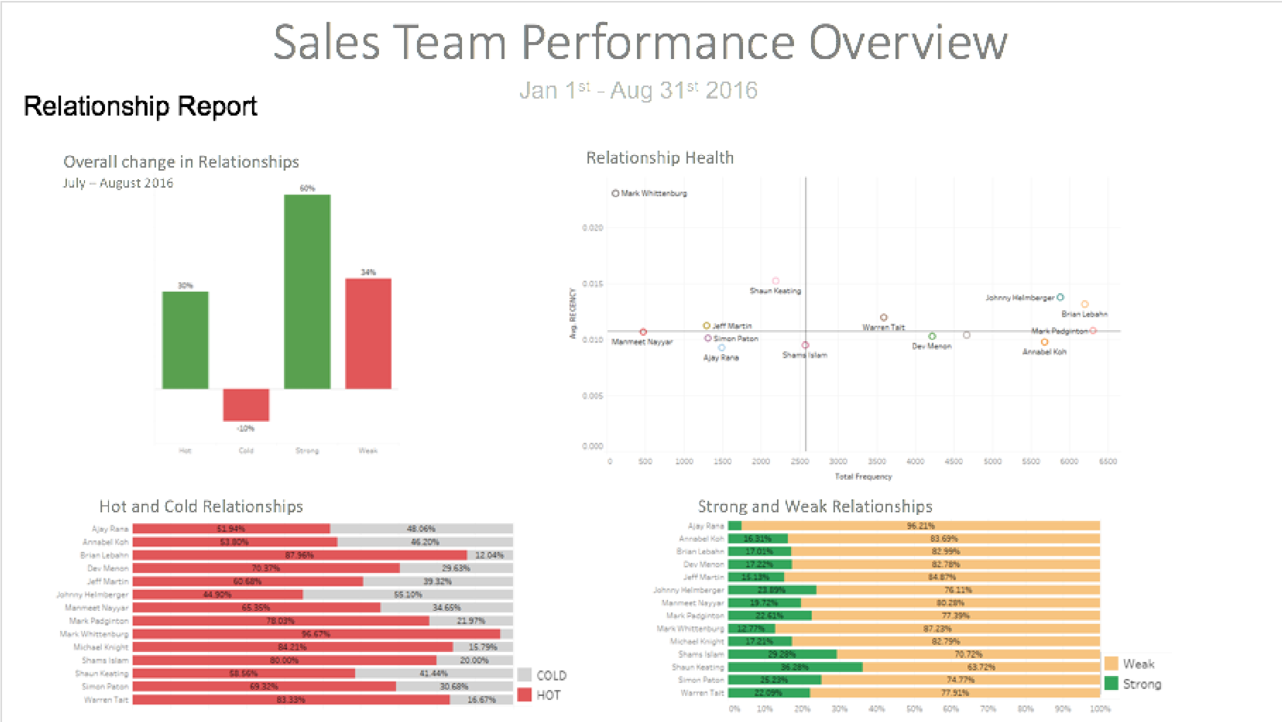
———————————————————————————————————

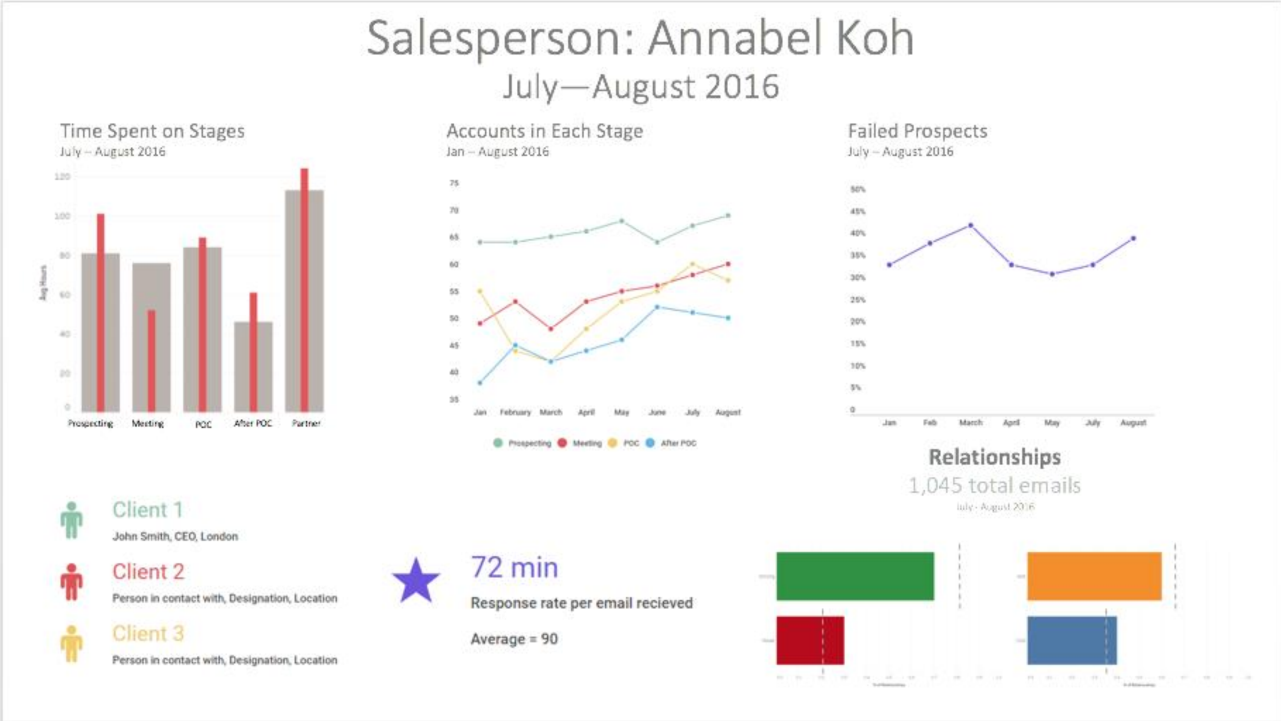
# 8. Appendix

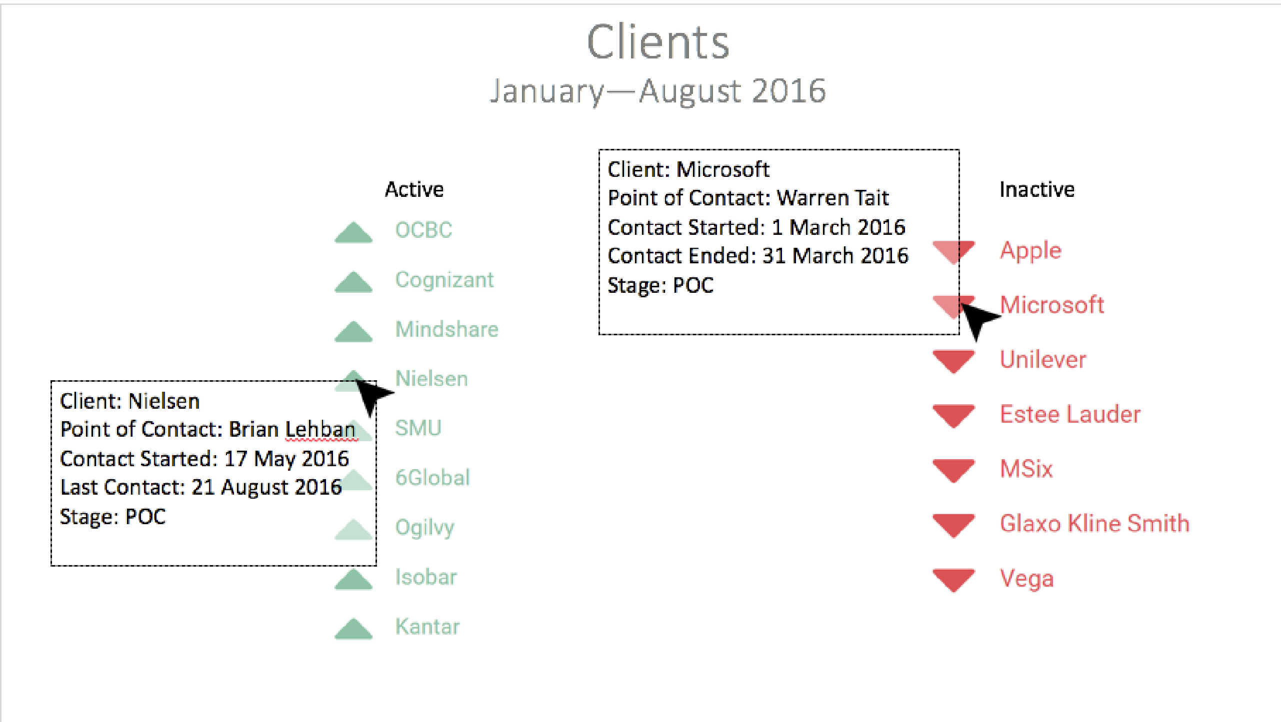
## Appendix A

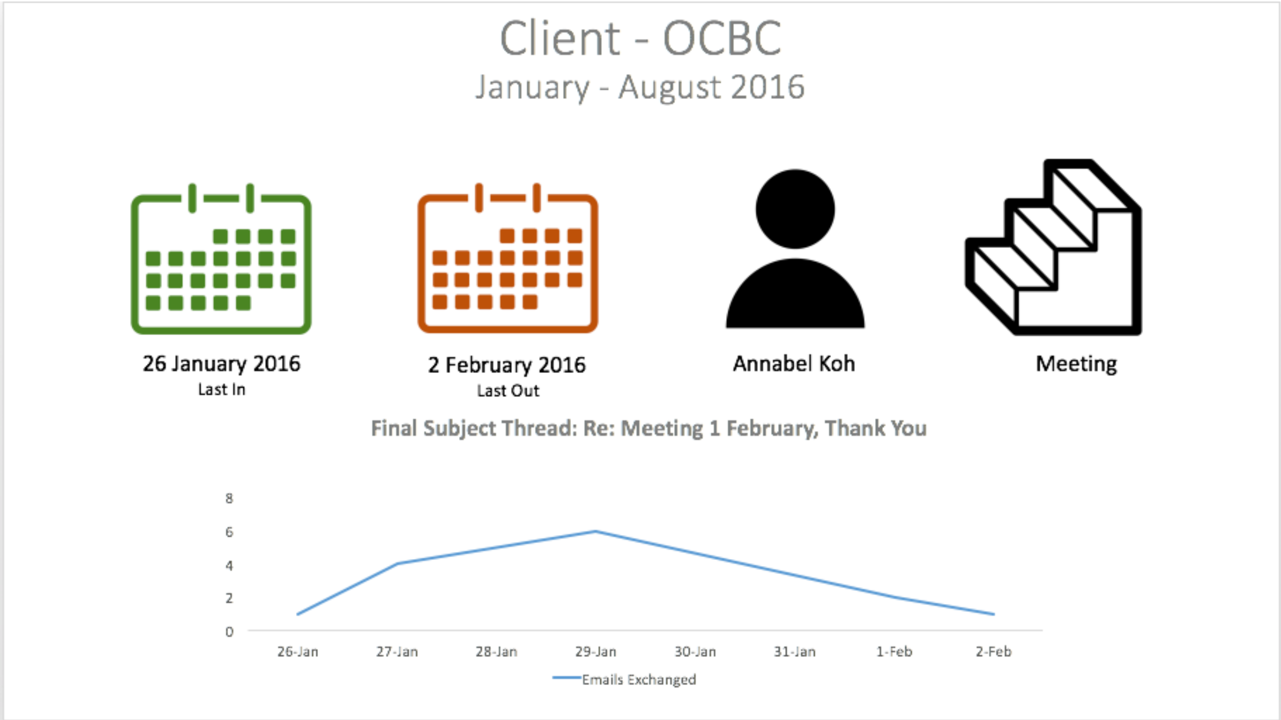
## Dashboard Initial Drawings




Sales Team Performance Overview
Relationship Report
Jan 1st – Aug 31st 2016
Overall change in Relationships
July – August 2016
Relationship Health
Hot and Cold Relationships
COLD
HOT
Strong and Weak Relationships
Weak
Strong


Salesperson: Annabel Koh
July—August 2016
Time Spent on Stages
July – August 2016
Prospecting
Meeting
POC
After POC
Partner
Accounts in Each Stage
Jan – August 2016
Prospecting
Meeting
POC
After POC
Failed Prospects
July – August 2016
Relationships
1,045 total emails
Client 1
John Smith, CEO, London
Client 2
Person in contact with, Designation, Location
Client 3
Person in contact with, Designation, Location
72 min
Response rate per email recieved
Average = 90


Clients
January—August 2016
Active
OCBC
Cognizant
Mindshare
Nielsen
SMU
6Global
Ogilvy
Isobar
Kantar
Client: Nielsen
Point of Contact: Brian Lehban
Contact Started: 17 May 2016
Last Contact: 21 August 2016
Stage: POC
Client: Microsoft
Point of Contact: Warren Tait
Contact Started: 1 March 2016
Contact Ended: 31 March 2016
Stage: POC
Inactive
Apple
Microsoft
Unilever
Estee Lauder
MSix
Glaxo Kline Smith
Vega


Client - OCBC
January - August 2016
26 January 2016
Last In
2 February 2016
Last Out
Annabel Koh
Meeting
Final Subject Thread: Re: Meeting 1 February, Thank You
8
6
4
2
0
26-Jan
27-Jan
28-Jan
29-Jan
30-Jan
31-Jan
1-Feb
2-Feb
Emails Exchanged