

ANALY TEAKA

IS480 Project Proposal Scanteak Analytics System

analyteaka

Version 1.4
19 June 2018

Team Members:

- Solomon **TEO** Kok How (Solomon.teo.2016) - Project Manager & System Architect
- **ONG** Chin Hao (ong.chinhao.2016) - Backend Developer
- **HONG** Yang (hongyang.2016) - Data Analyst
- Jeryl **SOH** Yen Cheng (Jeryl.soh.2015) - UI/UX & Front-End Developer
- Larry **THING** Bo Xuan (larry.thng.2016) - QA & Co-Backend Developer
- David **CHUA** Wee Song (david.chua.2016) - QA & Co-Data Analyst

Faculty Supervisor:

- Name¹

¹ To be assigned by course coordinator
IS480 - IS Application Project

Sponsor and/or Clients :

Scanteak

Department

- Jamie (jamie@scanteak.com.sg) - Regional Marketing Director
- Eric (erichan@scanteak.com.sg) - Data Analyst

Project Overview

1.1 Project Description:

Our client, Scanteak, is a leading furniture chain store retailer in Singapore with more than 100 retail outlets and flagship stores in Germany, Taiwan, Brunei, Japan, Canada, and USA. Despite having achieved turnovers of more than \$100 million annually, Scanteak, which actively seeks to remain relevant in the highly competitive market, is constantly looking for ways to drive transformation in the face of digital disruption.

Hence, we intend to digitally transform Scanteak's traditional way of doing business by deploying various forms of data analytics. By providing Scanteak with valuable insights drawn from data analytics, we hope to benefit both the senior management as well as sales executives by empowering them to make better and more informed business decisions.

Ultimately, by leveraging on Scanteak's wealth of offline retail data, our project aims to provide Scanteak with **descriptive and predictive data analytics** in three primary areas – **customers, staffs, and stores**. In doing so, we believe Scanteak will be capable of allocating the right **staffs** to the right **stores** to provide right **customers** with the right **products**. This will, in turn, improve the overall customer experience and capture maximal business value for our client.

1.2 Stakeholders:

Sponsor	Jamie, Regional Marketing Director, Scanteak Jamie will provide us with the high-level business requirements which she hopes to achieve through our system.
Users	<p>Senior management team The senior management team is responsible for supervising the overall business process and making core business decisions.</p> <p>Store managers The store managers are responsible for taking charge of a store outlet and ensuring that sales executives are meeting their sales targets.</p> <p>Sales executives The sales executives are responsible for attending to walk-in customers and providing them with appropriate recommendations so that they are more likely to make a purchase.</p>
Advisors/Practitioners/Mentors	Eric, Scanteak Eric advises Jamie on technology-related matters. He will assist in providing us with historical data.

1.3 Deliverables:

Outcomes: Our team will deliver a Live Dashboard System which will be integrated with our client's existing fully-functional system and be released for live usage.

Value Statement: The Live Dashboard System will firstly, present our client with an overview of various business performance indicators (e.g. product sales figure, staff performance measure, store performance index) and their correlation to different contributing factors (e.g. day of the week, age group, expected salary), and secondly, it will provide the sales executives a predictive sales model (i.e. which product to recommend to which type of customer based on identifiable traits and observations). As such, the Live Dashboard System will provide our client with useful insights which can be used to optimize their internal business process, reduce the time spent on making business decision and potentially increase their revenue.

1.4 Scope:

Core Function	Secondary Function
<u>Bootstrap Module</u> <ul style="list-style-type: none">• Uploading of data• Data cleaning/filtering	<u>Analytics & Reporting Module</u> <ul style="list-style-type: none">• Weekly/Monthly report• Exporting of report to pdf
<u>Customer Profiling Module</u> <ul style="list-style-type: none">• Customer profiling (age, race, district etc)• Customer categorisation	<u>Staff Profiling Module</u> <ul style="list-style-type: none">• Staff profiling/performance report
<u>Store Profiling Module part 1</u> <ul style="list-style-type: none">• sales data by product and item category group	<u>Machine Learning Module part 1</u> <ul style="list-style-type: none">• Machine learning for customer clustering
<u>Store Profiling Module part 2</u> <ul style="list-style-type: none">• sales data by product, peak sales period and outlets providing store insights	<u>Machine Learning Module part 2</u> <ul style="list-style-type: none">• Predictive analytics for customer and store
<u>Data Visualisation Module</u> <ul style="list-style-type: none">• Dashboard layout• Display insights	

Customer profiling module

This module is responsible for providing descriptive analytics for different customers and their identifiable traits. It is also responsible for categorizing and profiling customers based on their identifiable traits. It will provide the foundation for predictive analytics (e.g. recommended product for a customer of a certain age and race).

Customer categorization

Based on the historical sales data provided by Scanteak, we are going to work out the customers' race (from name), gender (from name), age (from NRIC), income level (based on their housing district), and if they are return customers (based on the past transaction records).

Customer profiling

Moving forward from customer categorization, which isolates various identifiable traits (age, race, gender etc.), we are going to generate several profiles/personas based on a combination of identifiable traits.

Examples of descriptive analytics would include:

- Age/race/gender/housing district composition of customers
- Customer return rate
- Sales and quantity sold for each identifiable trait or profile
- Main payment method

Examples of business questions that will be answered:

- What is the average amount spent by customers between the age 40-45?
- Are they mostly new customers or return customers?
- What is the main target profile (40-year-old Chinese male)?

Store profiling module 1

This module is responsible for providing descriptive analytics for different products and their respective categories. It will provide the foundation for predictive analytics (e.g. recommended product and quantity allocation for each store). -> do we want to mention this in our proposal?

Examples of descriptive analytics would include:

- Sales figure and quantity sold for each product and category
- Sales composition of each product and category

- Grouping of products which are commonly purchased together

Examples of business questions that will be answered:

- What is the best performing product/category for this outlet?
- How many products do customers usually buy in a single purchase?
- What other products do customers usually buy when they purchase a coffee table?

Store profiling module 2

This module is responsible for providing descriptive analytics for different stores and their respective locations. It will provide the foundation for predictive analytics (e.g. products to be recommended to customers of a specific store).-> do we want to mention this in our proposal?

Examples of descriptive analytics would include:

- Number of customers for each store
- Number of sales and sales figure for different months/days/periods
- Customer (age, race, gender, etc.) composition of each store

Examples of business questions that will be answered:

- What is the best performing store/location?
- What is the best performing month/day/period?
- Which is the race/age/gender demographic for each store?

Staff profiling module

This module is responsible for providing descriptive analytics for individual sales executives at respective outlets.

Examples of descriptive analytics would include:

- Revenue generated by each sales executive
- Number of items from each product type sold by each sales executive
- Number of deals closed from each type of customer (based on traits/personas) by each sales executive

Examples of business questions that will be answered:

- Who is the best performing sales executive?
- What is the best-selling product type for each sales executive?
- Is the sales executive's performance consistent? What's the possible reason?
- What is the customer composition (based on traits/personas) for each sales executive?

Bootstrap module

A new sales system, meant to replace Scanteak's legacy system, is currently being developed by Scanteak's in-house developers. As Scanteak's new sales system is still in the midst of completion, our bootstrap module is responsible for allowing users to upload customer data exported from the sales system. Once Scanteak's new sales system has been completed, the manual bootstrapping of CSV files will be phased out and the bootstrap module will be modified to interact with the new sales system directly through API calls.

Sample data:

Image removed for privacy issues
Fig 1 : Sample CSV output from sales system

Image removed for privacy issues
Fig 12 : Sales system interface

Steps for bootstrapping

1. Stream uploaded csv into data frame
2. Change all column types into its respective value. e.g str -> datetime
3. Infer all columns required
 1. Gender and ethics based on first name and last name
 2. Residential district based on postal code
 3. Age based on NRIC.

4. Data cleaning by removing duplicate row (double entry, invalid rows)
5. Convert into datastore request objects
6. Uploading data to datastore

Data visualization module - This module will make use of Bootstrap, flask and Dash by Plotly to generate charts based on data output generated by customer and store profiling module.

Analytics & reporting module - This module provide weekly and monthly report based on data provided by the store and customer profiling module and provide recommendations based on the data. This module will make use of data visualization module to generate reports for end users based on their roles (sales executives, senior management, etc.).

Machine learning module

This module will contain the machine learning system. As we are using Python as our main programming language, we will be utilizing libraries such as – SciPy, NumPy, matplotlib, pandas, Scikit-learn to help us complete this module. Using the training dataset (6 months' worth of offline retail data) we have prepared, we will train the system to provide predictive analytics for both customers and stores.

The entire process can be automated, whereby the system will retrieve raw sales data from the in-house sales system, process the raw sales data in the machine learning module before handing it over to the analytics & reporting module.

Examples of predictive analytics:

- Recommended products for different customer profiles
- Recommended price range for different customer profiles
- Recommended products to be displayed for different stores
- Best selling type of item and item category for different stores

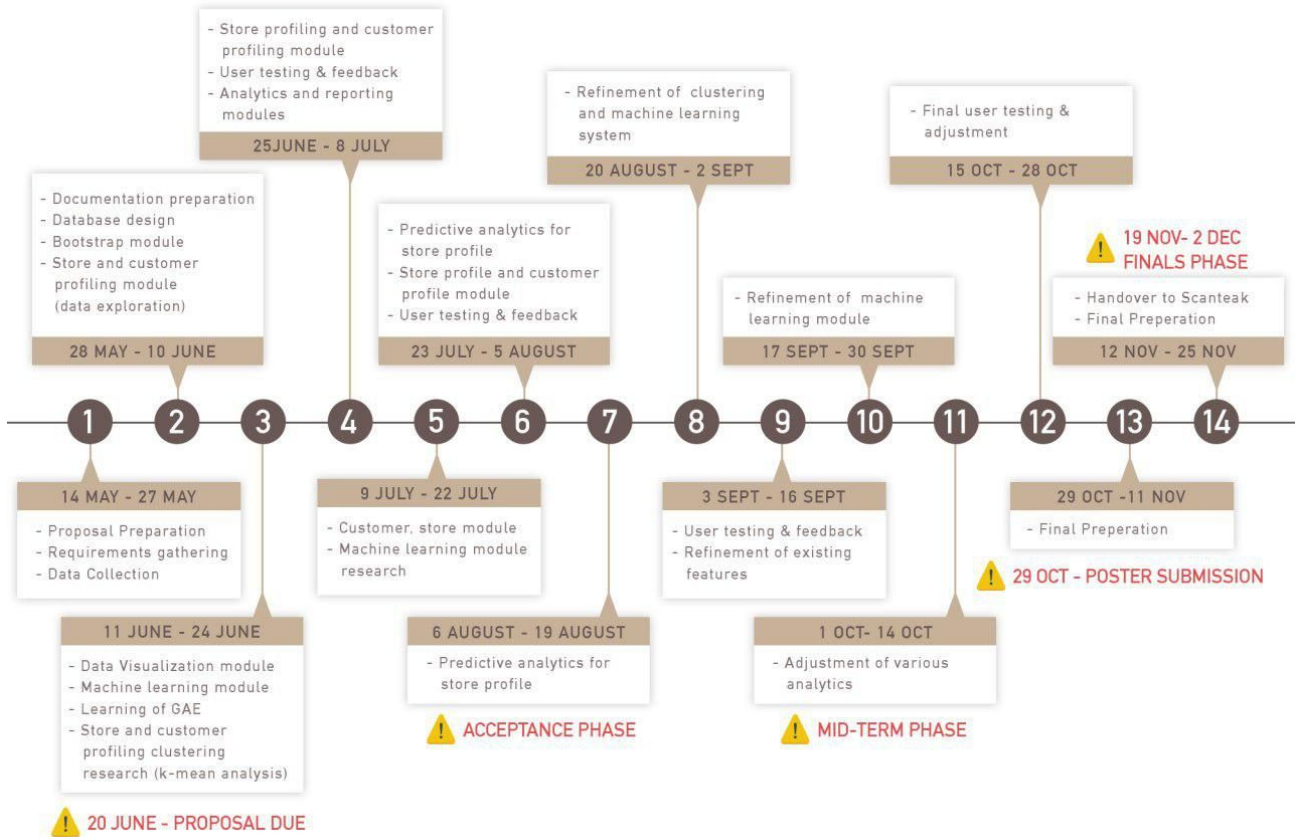
Examples of business questions that will be answered:

- What kind of furniture (based on price range) should you recommend to a certain customer profile (e.g. 30-year-old Chinese male at the Suntec Branch)
- What kind of customers (e.g. Chinese) are you expecting at a certain branch (e.g. Suntec) and what kind of furniture (e.g. Oriental-style) should you display?

Project Plan

This plan describes the resource, schedule and major risks. The plan provides a basic execution description of the project based on current knowledge of the project.

o Project milestone:



o Risks:

By taking into consideration the estimated probability of occurrence as well as the potential impact from the risk impact assessment, a quantitative impact measure is obtained. Based on the team and clients risk threshold and appetite, the measures will be grouped and the appropriate action to be taken will be identified.

Impact / probability	Very low	Low	Moderate	High	Very high
	0.05	0.1	0.2	0.4	0.8
0.9	0.05	0.09	0.18	0.36	0.72
0.7	0.04	0.07	0.14	0.28	0.56
0.5	0.03	0.05	0.10	0.2	0.4
0.3	0.02	0.03	0.06	0.12	0.24
0.1	0.01	0.01	0.02	0.04	0.08

Risk responses involve a combination of the following activities: Avoiding, Mitigating and Accepting. Prior to the start of each phase of the project, all risks will be considered and preventive measures will be put in place as part of mitigation. Additional action will be taken in accordance to the descriptive bands that are detailed below.

Green (Conditional Management) – Green regions will involve continuing with the project as per normal. Acceptance will be the main response in this band, active steps will only be taken when indicators of the risk arising begin to present themselves.

Orange (Passive Management, Mitigating) – Yellow regions will involve continuing with the project with passive risk management. Mitigation will be the main response in this band. During each end of iteration review, we will take note of whether the risk is still a possible threat and given the circumstances at each checkpoint, we will decide whether to invoke contingency measures and subsequently, risk management measures.

Red (Active Management, Avoiding and Mitigating) – Red regions will involve continuing with the project with active risk management. Avoidance will be the main response in this band. Contingency measures will be set in place and during each end of iteration review, we will take note of whether the risk is still a possible threat and given the circumstances at each checkpoint, we will decide whether to directly invoke risk management measures

Type	Risk	Likelihood	Impact	Mitigation	Strategy
Technical risk	Majority of our team is not proficient in python, google app engine, flask and machine learning.	High - 0.9	Very high - Potential delay in the schedule due to production delay.	Spent time going through bootcamp training for python and machine learning to build on our foundation. We will also take various modules such as visual analytics, AA, EWS and networking to build on our competency.	Active management
Resource Management Risk	Most of the team are not well verse in analytics, which is the core feature of our system. As well as limited knowledge on consumer behavior and digital marketing.	High - 0.9	High - Could affect our production timeline and analytics result.	Spent time going through bootcamp training for analytics and consult professors on analytics. Also take courses related to digital marketing and consumer behavior.	Active management
Human Risk	Potential schedule clash due to the different in class timetable.	Medium - 0.5	Moderate - Schedule need to be re-adjusted if	Project manager to better plan the schedule. Ensuring that	Passive management

	Additionally, FYP with internship concurrently means we could only work on our project at night and during the weekends.		there's a delay in the execution of task.	everyone will have a common free slot for work (e.g 1 day block slot every week)	
Client management risk	Client unable to provide csv containing sensitive data (NRIC, name) due to privacy compliances	Medium - 0.3	Moderate - Would require a work around to analysis the data	Discuss with clients and capture what we need during bootstrap stage (e.g determine age, gender and race) and hash the values such that we can still analyze it without having access or storing the raw csv file.	Passive management
Client management risk	Employees may not comfortable with using technology and figuring out how the application works.	Medium - 0.3	Low - Potential lack of adoption.	We will conduct adequate research into scenarios and personas, multiple rounds of prototyping with sufficient client feedback. Additionally, run through a tutorial during the first launch to highlight the importance of the features and use of the application. As well as getting the support of the upper management.	Conditional management
Human Risk	Team members fall sick to the point(ie hospitalization).	low - 0.1	Very low - contribution to project is affected	Always ensure that the wellbeing of the team members is preserved by encouraging healthy diet, exercise and work life balance. Additionally, conduct peer sharing sessions	Conditional management

				so that everyone is aware of what is being done by each other such that taking over will be easy in such situations.	
--	--	--	--	--	--

o **Resource and reference:**

We will be doing research on analytics, machine learning, Google App Engine system and consumers behavior analysis. We will get our reference from these online courses to level up our knowledge:

1. **Introduction to Python** - <https://courses.edx.org/courses/course-v1:MITx+6.00.1x+2T2017/course/>
2. **Machine Learning** - <https://www.udemy.com/machinelearning/>
3. **D3 Visualization** - <http://adilmoujahid.com/posts/2015/01/interactive-data-visualization-d3-dc-python-mongodb/>
4. **Clustering of user** - <https://datascience.stackexchange.com/questions/5838/i-am-trying-to-classify-cluster-users-profile-but-dont-know-how-with-my-attribute>
5. **Predictive analytics module** - <https://www.udemy.com/become-a-citizen-data-scientist-marketing-perspective/>
6. **Flask Tutorial** - <https://www.youtube.com/user/schafer5/playlists>

o **Technology**

