

Final Year Project Proposal

**SINGAPORE
POLYTECHNIC**



Class: DFI/FT/3A/03

Leader: Jasper Guay (P1622014)

Members: Donovan Tan Yi Khye (P1622212)

Tan Mei Xuan (P1622407)

Mabel Lim Pei Xuan (P1622465)

Nurasilah Bte Ramli (P1644614)

Manoj Kumar (P1622238)

Contents

Introduction	3
Task List.....	4
Work Allocation.....	4
Project Timeline	5
Analysis of Power BI against Tableau.....	7
Power BI	7
Tableau.....	7
Conclusion	8
Web scraping.....	9
Data Analysis	9
Database Prototype	11
References	13

Introduction

In recent years, consumer's preferences have shifted towards healthier food choices through public education efforts made by the Health Promotion Board. There are several reasons why salad is so popular: Its healthy, offers a variety of choices and is considered a relatively quickly prepared food for takeaway. As such, large amounts of salad shop patrons are the corporate employees during lunch time who are time packed yet in need of a hearty meal.

Caffeine, being the world's most popular drug, draws in a large number of consumers who constantly want to get their daily dosage of caffeine. In addition, cafes are also widely considered as a social gathering point where co-workers, friends, couples and even families would gather to have coffee, be it dining in or taking out. Thus, our group believes that the salad and coffee industry is thriving and there are many business opportunities for The Caffeine Experience. However, with increased consumer demand comes with increased competition.

As a cafe, The Caffeine Experience primarily serves coffee and tea, plus, they have a section for customisable salads. At the same time, they are a social enterprise that seeks to employ and help ex-convicts get their feet back and adjust to society again, while doing their best to provide quality customer service and beverages, as well as food. While doing charity work, they are in the end, still a business that seeks to serve customers and fulfil their needs. Just like any other enterprise, they have been facing the same problems as any normal business would encounter.

Our team's purpose in this project is to help the Caffeine Experience tackle these business problems such as meeting the needs of the customers as well as identifying possible flaws that might have slowed down the business. One solution we believe might help the Caffeine Experience would be dashboards that can help them point out areas of improvement.

With dynamic dashboards for The Caffeine Experience, information on these dashboards will display better visualization and more relevant context, thus allowing intelligent decision-making processes. We believe that through careful and wise decision making, The Caffeine Experience will be able to thrive in business.

Task List

- Project planning
- Proposal preparation
- Acquiring skills on Python fundamentals
- Developing a web scraper
- Clean and process data
- Analyze the processed data
- Dashboard designing
- Gallery walk preparation
- Final report writing
- Final presentation preparation

Work Allocation

Project Manager - Jasper

Role: Planning, directing the team and drafting deliverables. Oversees operations and provide technical advice to fellow consultants. Primary handler of paperwork tasks and key communicator between supervisor (Mr Joshua Tan) and the team. All team members are currently picking up Python programming language.

Lead Consultant - Mei Xuan

Role: Leader in the technical field of programming and analytics. Provides IT related solutions for the team, heads the technical operations of the project and closely coach other fellow consultants in the project. Mei Xuan is extremely proficient and efficient in the aspect of coding.

Consultant - Mabel

Role: Consultant in the technical field of programming and analytics. Works closely with Mei Xuan to come up with technical infrastructures that meet requirements of the project. Mabel is also innovative in the design aspect and will be main contributor to design related needs of the project.

Consultant - Donovan

Role: Consultant in both technical and business aspect of the project. Donovan works closely with Jasper to analyze The Caffeine Experience requirements and continually improve on deliverables drafting, discusses the project flow with Jasper and also provide technical operations assistance to other fellow consultants.

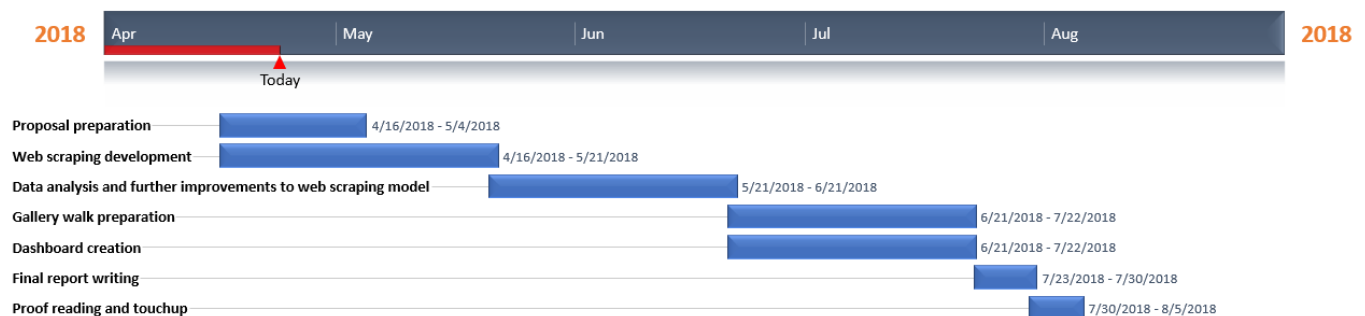
Consultant - Kumar

Role: Consultant on the idea thinking aspect. Does brainstorming for the project for possible inputs.

Secretary - Nurasilah

Role: Secretary for the group that ensure meetings are effectively organized and record keeps minutes and attendance of each member. She jots down key point of the discussion in group meetings and provides intellectual resources such as relevant articles for the team. She also assists Jasper in paperwork and provide technical advice to other fellow consultants.

Project Timeline



As shown in the picture, the timeline has been constructed with consultation from team members to ensure that realistic goals are set. The main phases of the project are web scraping development, data analysis and improvements to the web scraping model as these processes are the backbone of the project. Furthermore, web scraping will be done in Python which is an external module programming language that is new to the team. Thus, there will be more time allocated for the team to get a grasp of the language and then further hone our skills in order to deliver top notch web scraping models. We will mainly be scraping data from reputable food review websites such as Burpple, HungryGoWhere and Google reviews.

After developing the web scraper, we will then draw insights gathered from the data available to us. First, we plan to convert qualitative data into quantitative data by categorizing reviews according to category of complaint/compliment. Take for example, a good review (4 or 5 star) that mentions good service and good ambience will be added to the respective categories classified under “Positive points” and then totaled up to see what is the main value proposition of the cafe. Similarly, the data conversion process will be done for negative reviews (1 or 2 star) and neutral reviews (3 star). This will be done primarily for competitors and (if data is available) for The Caffeine Experience. The comparison allows our client to have a clear view of the most effective value propositions of each company and the biggest hurdle they are facing, which allows The Caffeine Experience to effectively position itself against its competitors in the market.

We will then proceed to prepare for our gallery walk by preparing a draft of our dashboard with key data of the main competitors displayed on big physical dashboards to be placed at our booth. We will also be brainstorming captivating ways to design our booth, promote our project and The Caffeine Experience as well as rehearsing on our speech to the visitors and come up with FAQ on the project.

After the gallery walk, our group will then proceed to write our final report which will consist of the main deliverables, each individual member will also include their reflection on top of providing suggestions on possible improvement to The Caffeine Experience as a whole. The report will then be a guiding tool for us moving forward to prepare for our final presentation where we will explain in detail how our web scrapers work, what is the process of the development, instructions on usage of the web scraper and showcase of the final deliverables such as the functionality of the scraper and also the dashboard that we came up with.

Analysis of Power BI against Tableau

Before our group started performing any operations, we have concluded that it is in the best interest of our client to have an understanding of the various pros and cons of the 2 leading Business Intelligence (BI) softwares that we are choosing between - Power BI or Tableau. We have summarized the pros and cons in bullet points and will then make a concluding statement on which software the group ultimately chose and the underlying reasons behind the choice that our group have come to a consensus.

Power BI

Pros :

- Reasonably priced (offers free version, paid version is \$9.99 per user a month)
- Good integration within microsoft tools
- Continual development
- Ease of usage for excel users due to similar tools
- Extensive visualising choices
- Comprehensive database connectivity

Cons :

- Challenging to learn
- Trouble handling large data sources
- Highly complex
- Poor customization
- Lack of data processing capabilities
- Only offers monthly-subscription and not licensing

Tableau

Pros :

- Easy to use
- Visually appealing
- Great performance
- Ability to connect to multiple data sources
- Flourishing community
- Mobile support

Cons :

- Expensive (professional edition \$70 **per** user a month)
- Stringent pricing, not priced to customer's needs
- Poor customer service
- Security concerns
- Difficult to use at deeper levels
- Poor Scalability

- Integration to IT landscape issues
- Steep learning curve for non-analysts
- Data warehouses which are expensive to build and maintain, are usually accompanied with Tableau to achieve its fullest potential

Conclusion

Both Tableau and Power BI are great analytics software with outstanding charting capabilities. However, after considering the needs of The Caffeine Experience being a start-up company with a stringent budget, Power BI being the affordable choice is recommended by our team. It has extensive charting capabilities, great integration with other Office 365 tools such as Microsoft Excel and is continuously developed by Microsoft which means that potential bugs are fixed quickly and the software will receive timely valuable updates.

Major drawbacks of Power BI such as the inability to handle large amounts of data and poor customization such as incapability of generating scheduled reports and personalized views can be overlooked at this price point. The lack of tools to improve data quality will be handled by our web scrapers, which we will first inbuilt data cleaning processes before verifying that the data obtained is of high quality and suitable for providing accurate analysis.

In consideration, Power BI does come with potential hurdles for The Caffeine Experience. Due to the sheer complexity and steep learning curve of Power BI, it is a concern if the user is unable to utilize it without the supervision of an IT-proficient employee. Training on Power BI will also be resource consuming as the user has to attend Power BI courses which translates to a reduction in manpower for operations, trainers have to be compensated and the company may even have to pay for higher-end courses.

Web scraping

For web scraping, our team has decided to build our own scrapers using Python as compared to using software offered in the market. Reasons being: Python is flexible and we could customize our codes to match our utility, unlike utilizing software which we may be limited by the software's capability. Secondly, building a python scraper is the economical choice as several popular web scraping softwares have quoted a market price of ~USD\$100 for a monthly subscription.

The web scraping process involves using a credible framework with extensive guide and community, which is BeautifulSoup (bs4) and Selenium which ensure that the web scraper built is of high quality and is not compromised by the lack of capabilities. The primary language used for web scraping will be Python due to the large community of users which translates to multiple guides on how to troubleshoot the scrapers should any issue arise. However, some knowledge of HTML and CSS is required to analyze the relevant data of the respective food review pages.

After developing the scraper, we will analyze food review websites such as Yelp, Burpple, HungryGoWhere and Google reviews by inspecting their website page element, then scraping out the essential information such as price per pax, number of reviews, review key points, ratings, popular food items etc. The scope of the scraping is on Tanjong Pagar Centre as the main competitors of the business lies there. These information will then be analyzed by our team and illustrated on a dashboard in Power BI. We will also perform drill-down analysis such as filtering out positive reviews, then finding the keywords that appear the most frequent. Eg: Oven baked salmon appears the most frequently in Kraftwich positive reviews.

Data Analysis

Through web scraping, we are able to extract the required data that is needed for data analysis. We believe that each kind of review ranging from negative to positive has its own reasons. As such, our team aim to group the key point(s) of each review into categories. Currently, we plan to group each review into the following categories: Food quality, Ambience, Price, Value, Service, Speed, Variety. These review points are subject to change if the group deem that it is better to add/remove a category.

The data scrapped will then be cleaned and processed. Firstly, the data scrapped will be converted into a normalized and structured format i.e. converting the Python output into excel format. Afterwards, the data will be cleaned such as excluding cafes in the price analysis of which price per pax information is unavailable due to lack of data. This helps with the data analysis process as it ensures that only quality data that is

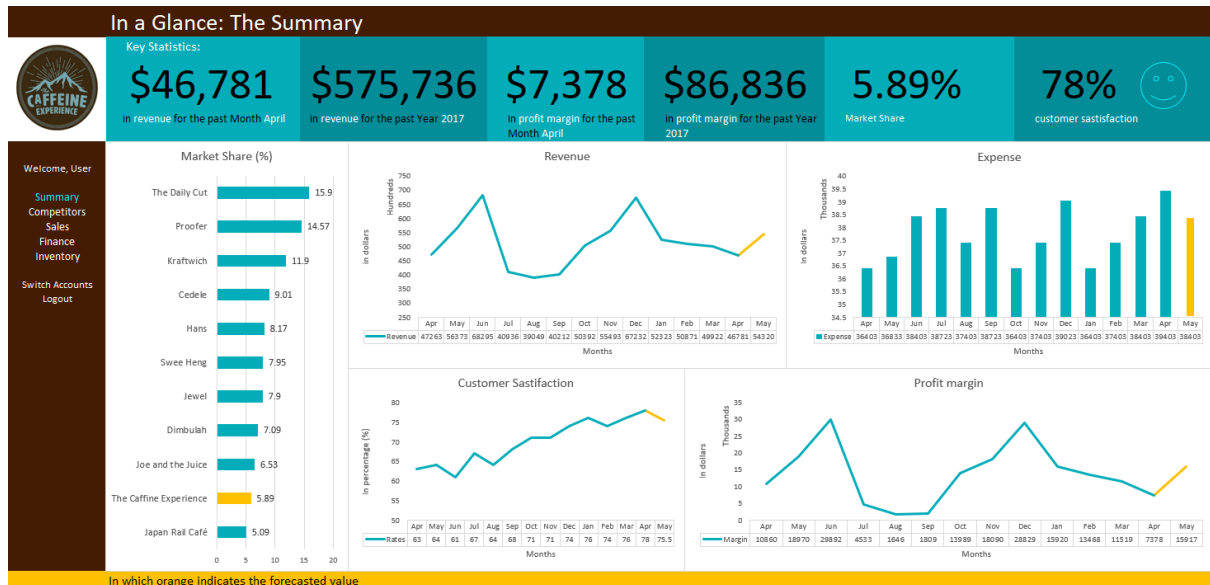
properly structured is utilized during, which improves data accuracy and reduces potential hurdles with data ambiguity moving forward with the project. After the dataset is structured and cleaned, we can then proceed to load them into Power BI to design our dashboard and analyze data to draw significant insights of key competitors.

Afterwards, our group will be analysing the data based on categories such as location, ambience, variety, service. Take for example a review quoted from HungryGoWhere (just an example): “ $\frac{4}{5}$ stars. The bar next door is full or deniers, loud and noisy but Jacob's is a quiet place. Easy drinking was what I was here for and I got that. Nice beer selection.” would add points into the category of Ambience and Variety. After which, the group will total up the points of each kind of review (Positive- 4 to 5 stars, Negative- 1 to 2 star(s), and Neutral 3-4 stars) and uncover where is the main strength and weakness of each and every relevant competitor to The Caffeine Experience. From the insights gathered through the data gathered and analyzed, we will allow executives of The Caffeine Experience to make quality business decisions that will position themselves adequately in the market to compete against the various competitors in Tanjong Pagar Centre.

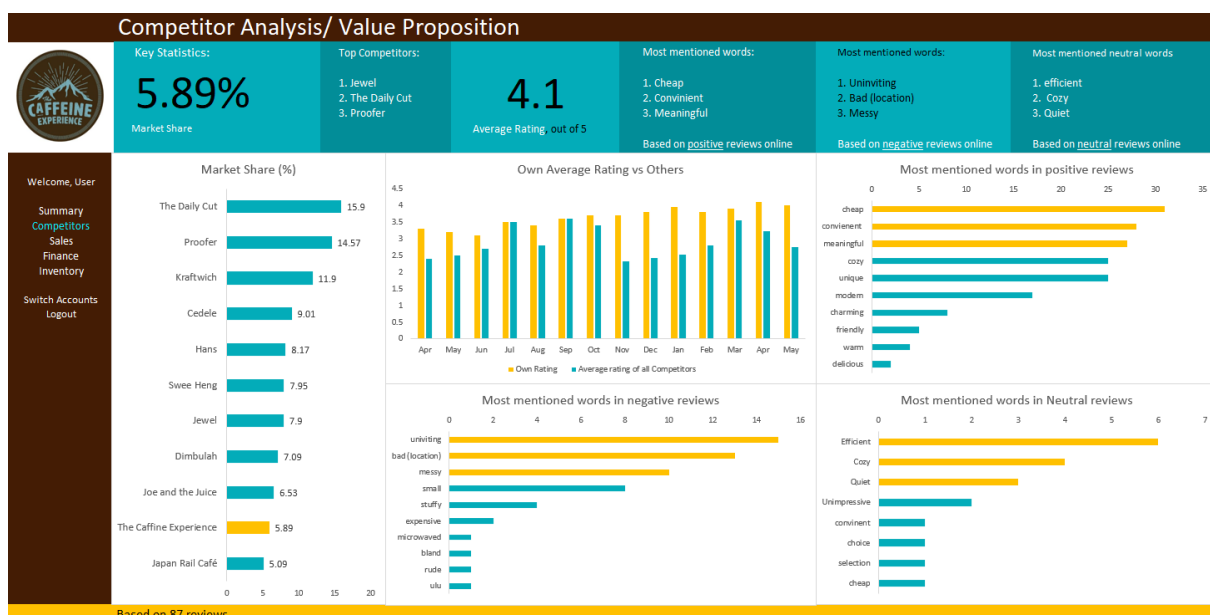
Our team went to Tanjong Pagar Centre at 12-1pm on a weekday for observations and through the process, we noticed that there are many competitors in the mall which offer salad, tea and coffee. The Caffeine Experience is a viable business at Tanjong Pagar mall as there are numerous patrons at salad shops during the lunch time, and quite a handful of also purchased beverages. The main competitors that we have identified are: The Daily Cut, Kraftwiche and Jewel Coffee. As such, our competitor analysis will mainly focus on these 3 competitors.

Database Prototype

Disclaimer: all the features and numbers mentioned in the database prototype are subjected to change and are for demonstration purposes only.



When the client first load the dashboard, they will see the summary page which will include all the important statistics, such as the revenue and the profit margin made on each month. The top bar contains all the important figures the client will need to know at a glance. The sidebar acts as the navigation bar, where the client is able to access different sections of the dashboard, such as the competitors and the inventory sections, all of which will give an in depth detail of the topic.



To showcase what some of the other sections might look like, this is the competitor/ value proposition page, where there will be statistics on where the client's cafe stands in their competitors' landscape. Some statistics might include the market share, the most popular description of the cafe based on the different type of reviews (positive, neutral and negative), so the client can work on building on their proposition and know what to work on to improve their cafe. We have also planned to include statistics that will further breakdown the competitors into direct and indirect competitors, so that the client can build their value proposition around what their competitors lack which will draw a loyal customer base for The Caffeine Experience.

References

GoWhere, H. (2018). *Jacob's Cafe - Nice mix of local and Western fare at Changi Village*. [online] Hungrygowhere.com. Available at:

https://www.hungrygowhere.com/singapore/jacob_s_cafe/ [Accessed 3 May 2018].

Molag, T. (2017). *Power BI vs Tableau: What to Consider | Encore Business Solutions*. [online] Encore Business Solutions. Available at:

<https://www.encorebusiness.com/blog/power-bi-vs-tableau/> [Accessed 3 May 2018].

Post, T. (2017). *Straight Talk Review: Strengths and Weaknesses of Microsoft Power BI*. [online] Yurbi. Available at: <https://www.yurbi.com/blog/straight-talk-review-strengths-and-weaknesses-of-microsoft-power-bi/>

[Accessed 3 May 2018].

Sam-solutions.com. (2016). *Pros and Cons of Visualization Software | SaM Solutions*. [online] Available at: <https://www.sam-solutions.com/blog/tableau-software-review-pros-and-cons-of-a-bi-solution-for-data-visualization/>

[Accessed 3 May 2018].