

ANLY482 Analytics Practicum Proposal

Vanitee's Customer Retention, Beauty Professional Activeness & Brand Loyalty Analysis

**Lim Chang Jun Andrew | Chow Si Min Sarah
Group 3 - Team V**

Contents

Overview	2
Sponsor Background	2
Project Background and Motivation	3
Team Motivation	3
Project Objectives	4
Data Collection and Description	4
Bookings	4
Campaigns	5
Customers	6
Payouts	6
Professionals	6
Services	7
Wallets	7
Wallet Coupons	8
Challenges	8
Scope of Work	8
Proposed Methodology	9
Data Collection	9
Data Preparation	9
Exploratory Data Analysis	9
Data Cleaning	9
Data Normalisation and Transformation	9
Cluster Analysis	10
Survival Analysis	10
Research	10
Recommendations and Insights	11
Stakeholders	11
Deliverables	11
Project Timeline	12
Project Schedule	13
References	14

Overview

Advancements in technology today have not only brought people closer together, but have also significantly increased our convenience levels. From applications for food delivery to hailing a taxi, these applications have filled in the space between consumers and merchants. Inevitably, the same phenomenon is occurring amongst the beauty scene in Singapore. At Vanitee, the gap between consumers and independent beauty professionals is being bridged. Typically, beauty professionals that are listed on the platform are emerging and independent beauty artists. To put it simply, they are professionals who want to grow their brand and customer base. By providing such a platform, Vanitee is able to help them showcase what they do best.

However, Vanitee does not want to stop at just providing a platform for these beauty professionals and for customers to engage them. Furthermore, with an increasing number of professionals and customers coming on board, evaluating their performance so far becomes much more imperative. Therefore, Vanitee hopes to discover meaningful insights from the data collected thus far to better understand the effectiveness of current efforts in retaining their customers and beauty professionals as well as the potential growth of their business.

Sponsor Background

Vanitee officially launched in May 2015, but it was first created during BattleHack 2014, an annual hacking competition held nationwide. Founders Douglas Gan, Peng Kong Choy, Kuik Xiaoshi, and Meters Ang created Vanitee to bring independent beauty professionals and customers together through a website and a mobile application.

Beauty services range from hairstyling, brows, nails, makeup, massages and facials. As of today, there are approximately more than 1700 beauty professionals and they have served over 20,000 customers so far. As it can be seen, customers and beauty professionals represent Vanitee's two main stakeholders in a two-sided market. Like most two-sided markets, it earns a premium by charging the transactions made through their application between both stakeholders.

Consumers that consume such beauty services on a regular basis have a huge incentive in utilizing such an application out of sheer convenience. With its variety of functions, customers are able to pinpoint the beauty service they need as well as explore new and upcoming beauty professionals with great ease. Likewise for budding beauty professionals, the platform helps them keep proper records for future references in terms of client database, business statistics, and increases their online word-of-mouth via reviews and social connections.

Despite their early success over the past 2 years, Vanitee plans to push their business further forward by focusing on their efforts to grow and retain their customer and beauty professional base.

Project Background and Motivation

As mentioned above, Vanitee was created with the intention of providing a platform for independent beauty artists to provide quality services to customers. This gives them the opportunity to run their own online business store, grow their brand and showcase their works. Vanitee was made to connect customers to these emerging beauty professionals and as of now, the number of beauty professionals and customers are growing steadily.

However, this does not mean that there are no competitors. Competitors include brick and mortar shops in local neighbourhoods and even bigger beauty brands with chain stores such as Jean Yip Group. Even though these are physical stores, they still pose as a threat as customers can still choose to go to these stores instead of using Vanitee to engage a beauty professional.

In order to keep competition at bay and to prevent their website and application from being stagnant, Vanitee needs to constantly improve on their customer retention, activeness of beauty professionals and brand loyalty strategy. This would allow them to differentiate themselves from their competitors and maintain a strong competitive advantage.

Firstly, to further the success of their application, Vanitee needs to grow as much as possible while retaining their existing consumers. Many consumers might have become dormant after just one booking. Hence, analysis can be done to find out why they have turned dormant and identify possible solutions to attract them to make the next booking.

Secondly, as beauty professionals represent the supply of services available on the application, their activeness in using the platform should also be monitored carefully. Some popular beauty professionals may experience uptake on a daily basis while newer beauty professionals have little or no uptake at all. Hence, analysis could be done to find out the main factors of beauty professional attrition.

Lastly, Vanitee currently has an extensive loyalty program in place that offers customers credits, gems as well as campaign codes with every booking made. However, one issue they face is the lack of understanding of how consumers utilize these in-app resources. Also, they wish to understand the effectiveness of such a loyalty program in encouraging customers to make repeated bookings in the future.

Team Motivation

We have been consumers our entire life and we will continue to be so for the rest of our lives. It is fascinating that the things we buy, not buy, will buy or consider buying can be studied and analysed by businesses or anyone who is interested. In our search for project sponsors, we found out about Vanitee. Frankly, none of us has ever done a project based on consumers' behaviour. We thought it will be a good opportunity for us to challenge ourselves to do a project that is in an unfamiliar domain. Furthermore, as people who might want to start a business of our own in future, it will be a useful experience and to learn new skills in dealing with business data.

Project Objectives

Hence, by utilizing the data from their current application's database, we would wish to discover meaningful and informative insights which will allow Vanitee to better retain their customers and beauty professionals and understand the effectiveness of their current loyalty program. To achieve the above mentioned, we will perform an in-depth analysis on the the data collected with the following objectives in mind:

Customers

- To determine the customer segmentation (different groups of customers) from the current booking patterns. Which customers are stagnant? Which customers are actively using the app?
- To understand customers' behaviour. When was the last time a customer used the app? How frequent does a customer use the app? How much does a customer spend on average?
- To evaluate the effectiveness of using campaign codes to ensure customers repeat their bookings
- To understand how customers are using credits and gems, whether they are accumulating before use or using them in their next booking
- To determine the Customer Lifetime Value (CLV) by campaign (which promotional campaign drives the highest value customer?) To which campaign do customers react to more? Which customers react and respond more to campaigns, credits and gems?
- Which services generate the most profits?

Beauty Professionals

- To determine if there is any correlation on what makes beauty professionals more attractive to customers?
- To determine if there is any correlation on what makes beauty professionals become active? (bookings/month, chat response rate, new services etc.)

Data Collection and Description

To facilitate the data analysis, Vanitee has provided our team with access to their current MongoDB database (from May 2015 to December 2016) on the cloud. The database contains numerous tables such as customers, beauty professionals, bookings, campaigns etc. To further understand the data given, we will attempt to look at some of the major tables of the database.

Bookings

A row in this table represents a specific booking of a Customer with a Beauty Professional. The detailed description of the main columns in this table is as follows:

Column Name	Description
id	The unique id of the booking

professional_user_id	The unique id of the professional involved in the booking
customer_user_id	The unique id of the customer involved in the booking
start	The start timestamp of the service purchased
end	The end timestamp of the service purchased
service_ids	The set of ids belonging to the service(s) purchased
use_credits	Indicates if credits are used in the booking
price_chargeable	The price chargeable to the customer
price_discount	The price of the discount, if any
price_payout	The price of the payout to the professional
price_transaction_fee	The price that Vanitee receives from the online transaction
price_customer_cashback_fee	The price that is converted to credits to be obtained by the customer
payment_method	The type of payment method used (cash or credit card)
campaign_code_id	The campaign code used in the booking, if any

Campaigns

A row in this table represents a specific campaign (marketing initiative). The detailed description of the main columns in this table is as follows:

Column Name	Description
id	The unique id of the campaign
name	The name of the campaign
description	The detailed information of the campaign
start_at	The start timestamp of the campaign
end_at	The end timestamp of the campaign
usage_limit	The number of times the campaign code can be used per customer
usage_min_spend	The minimum spending required to enjoy this campaign
discount_amount	The discount amount entitled upon usage of campaign code

booking_count	The number of bookings that have utilized the campaign code
---------------	---

Customers

A row in this table represents a specific customer in relation to a specific beauty professional. The detailed description of the main columns in this table is as follows:

Column Name	Description
id	The unique id of the customer
user_id	The unique user id of the customer
professional_user_id	The unique id of the professional that the customer is patronizing
booking_count	The number of bookings made by this customer
revenue_count	The revenue earned by the professional from this customer
discount_count	The discount used by this customer

Payouts

A row in this table represents a specific payout (transfer of funds) for a specific beauty professional for credit card payments. The detailed description of the main columns in this table is as follows:

Column Name	Description
id	The unique id of the payout
admin_user_id	The unique user id of the administrator that performed the payout
professional_user_id	The unique id of the professional receiving the payout
booking_id	The unique id of the booking tied to the payout
amount	The monetary value to be received by the professional
payout_at	The timestamp that the payout was carried out

Professionals

A row in this table represents a specific beauty professional. The detailed description of the main columns in this table is as follows:

Column Name	Description
-------------	-------------

id	The unique id of the professional
user_id	The unique user id of the professional
brand_name	The name of the professional
service_count	The number of services offered by this professional

Services

A row in this table represents a specific service offered by a beauty professional. The detailed description of the main columns in this table is as follows:

Column Name	Description
id	The unique id of the service
professional_user_id	The unique id of the professional offering the service
title	The name of the service
price	The price of the service
duration	The duration of the service
booking_count_total	The total number of bookings that involved this service

Wallets

A row in this table represents a specific wallet belonging to a specific user. The detailed description of the main columns in this table is as follows:

Column Name	Description
id	The unique id of the wallet
user_id	The unique id of the user
credits_total	The total amount of credits available for use
points_total	The total amount of gems available for use
credits_used	The total amount of credits used
points_used	The total amount of gems used

Wallet Coupons

A row in this table represents a specific campaign coupon belonging to a specific user. The detailed description of the main columns in this table is as follows:

Column Name	Description
id	The unique id of the coupon
user_id	The unique id of the user
wallet_id	The unique id of the wallet
points_total	The total amount of gems that this coupon entitles
credits_total	The total amount of credits that this coupon entitles
points_expire_at	The timestamp of when the points expire
credits_expire_at	The timestamp of when the credits expire

Challenges

Firstly, a potential challenge we might face while handling the data is the aggregation of data from the database. As mentioned earlier, Vanitee uses MongoDB as their main storage of data. It is important to note that MongoDB is a type of NoSQL database which means that it allows the insertion of data without a predefined schema. Although this allows for more dynamic schemas to be in place, it also means that each row in a table might not have the same number of columns. Each row is stored in a JSON structure where only columns of data will only be stored if required. Hence, careful selection of what data columns to utilize will be of vital importance.

Secondly, certain tables in the current database contain test data that do not provide accurate insights. Communicating with Vanitee's development team would be crucial in ensuring that we are able to filter out these test data from the live data. Hence, this issue would clearly have to be handled in our data cleaning step.

Scope of Work

Using the data provided by our sponsor from May 2015 to December 2016, we intend to analyse the data and to discover insights to fulfill the above mentioned objectives.

We propose the following scope for our project:

1. Data Collection
2. Data Preparation
3. Exploratory Data Analysis
4. Data Cleaning

5. Data Normalization and Transformation
6. Cluster Analysis
7. Survival Analysis
8. Recommendations and Insights

Proposed Methodology

Data Collection

We will use the data provided to us by Vanitee which through our access to their MongoDB database on the cloud. In particular, we will target data tables that pertain to customers, beauty professionals, bookings and loyalty programmes.

Data Preparation

As mentioned above, data rows within each data table may differ slightly in the number of columns (attributes) they contain. As such, we will attempt to consolidate the data into suitable and consistent formats to be used for analysis.

Additionally, data tables that have relationships with other data tables can be combined into one dataset. Hence, we will attempt to prepare different datasets according to the project objectives. Lastly, data that was created due to testing (by the Vanitee team) will be identified and removed.

Exploratory Data Analysis

We will look into the bookings customers make and also the use of credits and campaign codes when they are making their bookings. From here, we will be able to understand the buying behaviour of customers and analyze the trends in their bookings. Additionally, we will also identify any trends in their usage of gems. As for beauty professionals, we will go into observing the frequency of their bookings, services they put up on the platform as well as their chat responsiveness.

Data Cleaning

Missing values and outliers observed during the exploration of data may invite unnecessary inaccuracy and skewness in our analysis. To handle missing values, we will look at the amount of missing values identified and determine if the value should be estimated or simply removing the entire data row. For outliers, we will attempt to analyze why they exist and decide if they are relevant enough to be included in our analysis.

Data Normalisation and Transformation

As the distributions of values differ amongst different attributes, we will attempt to normalize such attributes before commencing our analysis to prevent these attributes from dominating other attributes. Also, data transformation techniques such as discretization and binarization will be performed to convert the necessary data to categorical and binary form respectively.

Cluster Analysis

Next, cluster analysis will be carried out to determine the existence of clusters amongst Vanitee's customers and beauty professionals. We will attempt to identify the profiles of each cluster according to their booking history and examine the reasons affecting the performance of each cluster. Thereafter, we hope to translate the identified clusters into a form of customer segmentation to help Vanitee better understand its customer base.

Survival Analysis

We will also be attempting to conduct survival analysis to predict the Customer Lifetime Value (CLV) by campaign. Survival analysis is a statistical technique that analyzes the duration to a certain event (e.g. a booking on Vanitee). Hence, such analysis will aim to which campaign drives the highest value customer and in the event of a new campaign, which customer profile will be respond early during a campaign (over a fixed period of time). The effectiveness of campaign codes in ensuring repeat bookings can also be investigated through such an analysis.

Research

Through our consultation with Prof Kam, he suggested that we take a look into Customer Analysis to understand Vanitee's customers and RFM Analysis to analyse a customer's value. As our team is still slightly undecided on the exact type of analysis to use, we decided to heed Prof Kam's advice and do some research on these methods first. Also, these methods can be used in complementary with cluster analysis used in segmenting the customer base.

Starting with segmentation and cluster analysis, it is important that we carefully select the variables we would use to do the analysis. As stated in this paper written by Michael Lynn in 2011, *Segmenting and Targeting Your Market: Strategies and Limitations*, we should keep the number of variables used reasonably small to get good clusters. With each cluster identified, we are then able to look into their specific characteristics to understand their behaviours.

As one of the objectives is to understand the effectiveness of campaign codes and how customers are using their credits and gems, customer analysis can be used here. Moreover, we would also need to find out which customers are making which bookings. After reading articles by Udemy and Marketing 91, it informs us that customer analysis can be useful in identifying the services that are popular and bring in the most profits, the buying patterns of customers and how they make use of Vanitee's loyalty programme.

Articles by Canopy Labs and GainInsights give us the knowledge that, to help identify which customers are more valuable to Vanitee, RFM analysis can be used. RFM analysis is also a segmentation technique that we can use. It will be able to help us identify customers who are more active and who are willing to spend. Further reading into a review on RFM model written by Jo-Ting Wei, Shih-Yen Lin and Hsin-Hung Wu, called *A Review of the Application of RFM Model*, confirms with us that RFM analysis could be a good model for us to use to reach our objectives in finding the most valuable customers for Vanitee. However, it does have its downsides. This model focuses more on the valuable customers hence it gives lesser meaningful scoring for customers who do not buy often. This also means that for smaller firms that are just starting out, most customers may only

have purchased once or a couple of times till date. Furthermore, this could be true for Vanitee since it started slightly before mid-2015.

Recommendations and Insights

We will provide Vanitee with the insights we will generate from the different analysis. Based on the insights collected, we will likewise provide them with recommendations that will help improve their customer retention, activeness of beauty professionals and brand loyalty strategy.

Stakeholders

The primary stakeholders of this project are:

- Project Supervisor: Prof Kam Tin Seong, Associate Professor of Information Systems; Senior Advisor, SIS (Programme in Analytics)
- Sponsor: Kuik Xiao Shi, Co-Founder of Vanitee

Deliverables

The final deliverables of this project are as follows:

- Analytics Practicum Project Proposal
- Interim Practicum Report in Wiki
- Interim Presentation
- Final Report
- Final Presentation
- Project Wiki Page
- Project Poster
- Recommendations to Sponsor

Project Timeline

Tasks		Week -1	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16	
Requirement Gathering	Confirm Project Sponsor																			
	Gather Requirements																			
Data Gathering	Industry and Competitors																			
	Gather Data from Sponsor																			
Project Proposal	Refining Project Scope																			
	Proposal Preparation																			
	Wiki Page Preparation																			
Proposal Deadline - 1 January 2017 - Week 1																				
Data Exploratory	Data Preparation																			
	Data Exploration																			
	Data Cleaning																			
	Generate Findings																			
Interim	Interim Report Preparation																			
	Interim Wiki Update																			
	Interim Presentation Preparation																			
Interim Practicum Presentation - 20 to 24 Feb 2017 - Week 8																				
Model Building	Data Normalisation & Transformation																			
	Cluster Analysis																			
	Survival Analysis																			
	Extrapolation																			
Insights & Recommendation	Create Visualisation from Analysis Results																			
	Generating Insights																			
Final	Formulate Recommendations																			
	Abstract & Full Paper Preparation																			
	Final Wiki Update																			
	Final Presentation Preparation																			
Final Practicum Presentation & Submission - 2 April & 16 April 2017 - Week 14 and 16																				

Legend	
	Planned
	Actual
	Milestone

Project Schedule

Tasks		Start Week	End Week	Planned Man Hours	Actual Man Hours	Team Members Involved	Status
Requirement Gathering	Confirm Project Sponsor	-1	0	4	4	All	Completed
	Gather Requirements	-1	0	2	2	Andrew	Completed
Data Gathering	Gather Data from Sponsor	-1	0	2	2	Andrew	Completed
Project Proposal	Refining Project Scope	0	2	3	5	All	Completed
	Proposal Preparation	0	2	20	20	All	Completed
	Wiki Page Preparation	0	2	10	10	Sarah	Completed
Research	Literature Research	1	3	5	8	All	In Progress
Data Exploratory	Data Preparation	1	-	15	-	All	In Progress
	Data Cleaning	1	-	15	-	All	In Progress
	Data Exploration	2	-	25	-	All	In Progress
Proposal Deadline - 15 January 2017 - Week 3							
Data Exploratory	Generate Findings	3	-	20	-	All	Not Started
Interim	Sponsor Review	4	-	5	-	All	Not Started
	Interim Report Preparation	4	-	30	-	All	Not Started
	Interim Presentation Preparation	5	-	20	-	Andrew	Not Started
	Interim Wiki Update	6	-	20	-	Sarah	Not Started
Interim Practicum Presentation - 20 to 24 Feb 2017 - Week 8							
Model Building	Data Normalisation & Transformation	8	-	15	-	All	Not Started
	Cluster Analysis	8	-	20	-	Sarah	Not Started
	Survival Analysis	9	-	20	-	Andrew	Not Started
Insights & Recommendation	Create Visualisation from Analysis Results	10	-	30	-	All	Not Started
	Generating Insights	11	-	30	-	All	Not Started
	Formulate Recommendations	12	-	25	-	All	Not Started
Final	Sponsor Review	12	-	8	-	All	Not Started
	Abstract & Full Paper Preparation	12	-	45	-	All	Not Started
	Final Wiki Update	13	-	10	-	Sarah	Not Started
	Final Presentation and Poster Preparation	14	-	20	-	All	Not Started
	Final Paper Preparation & Submission	14	-	30	-	All	Not Started
Final Practicum Presentation & Submission - 2 April & 16 April 2017 - Week 14 and 16							
Conference	Undergraduate Conference on Data Analytics	16	-	8	-	All	Not Started

References

- The Straits Times. (2015, Sep 13). The brain behind beauty start-up. Retrieved from <http://www.straitstimes.com/business/invest/the-brain-behind-beauty-start-up>
- The PE Hub Network. (2015, Dec 4). Singapore beauty tech startup Vanitee takes in funding. Retrieved from <https://www.pehub.com/2015/12/singapore-beauty-tech-startup-vanitee-takes-in-funding/#>
- TechCrunch. (2015, Dec 3). Vanitee Lands \$3.5M To Expand Its Beauty Booking Service Across Asia. Retrieved from <https://techcrunch.com/2015/12/03/vanitee-lands-3-5m-to-expand-its-beauty-booking-service-across-asia/>
- TechAsia. (2016, Jun 6). Singapore's Vanitee acquires startup to get beauty salons on its marketplace. Retrieved from <https://www.techinasia.com/vanitee-mefitted-acquisition>
- NCBI. (2010, Sep). A Practical Guide to Understanding Kaplan-Meier Curves. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3932959/>
- Thebmj. (n.d.) Survival Analysis. Retrieved from <http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/12-survival-analysis>
- Statistics Solutions. (n.d.). Conduct and Interpret a Cluster Analysis. Retrieved from <http://www.statisticssolutions.com/cluster-analysis-2/>
- Michael Lynn (2011). Segmenting and Targeting Your Market: Strategies and Limitations. Retrieved from <http://scholarship.sha.cornell.edu/cgi/viewcontent.cgi?article=1238&context=articles>
- Udemy Blog. (2014, May 1). Customer Analysis: How to Effectively Target the Market. Retrieved from <https://blog.udemy.com/customer-analysis/>
- Marketing 91. (2016, Dec 1) Customer Analysis. Retrieved from <http://www.marketing91.com/customer-analysis/>
- Canopy Labs. (n.d.) Intro Guide to the RFM Model. Retrieved from <https://canopylabs.com/resources/an-introduction-to-the-rfm-model>
- Gain Insights. (2014, Nov 26). Customer Segmentation Using RFM Analysis. Retrieved from <http://gain-insights.com/solutions/retail-analytics/customer-segmentation-using-rfm-analysis/>
- Jo-Ting Wei, Shih-Yen Lin, Hsin-Hung Wu. (2010, Dec 9). A review of the application of RFM model. Retrieved from <http://www.academicjournals.org/journal/AJBM/article-full-text-pdf/EB3418D18198>