

ANLY482 Analytics Practicum Interim Report

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
Project Background & Motivation	2
Project Objectives	3
Data Integration and Filtering	4
Data Collection	4
Extracted Tables	4
Challenges	9
Data Cleaning and Exploration	9
Issues	9
Duplicate Values	9
Missing Values	10
Changes in Business Model	11
Findings	12
Users, Customers & Professionals	12
Bookings	15
Services	24
Campaigns	26
Revised Methodology	30
Cluster Analysis	30
Survival Analysis	30
Revised Scope of Work	30
Project Timeline	33
Revised Work Plan	34
References	35

Overview

Project Background & Motivation

Vanitee was officially launched in May 2015, in an attempt to bridge the gap between customers and independent beauty professionals. 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, 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. Hence, 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.

Firstly, to further the success of their application, Vanitee has to place emphasis on attracting new customers as well as retaining their existing customers. Many customers 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, in response to these dormant customers, Vanitee currently has an extensive loyalty program (as shown in Figure 1) 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.

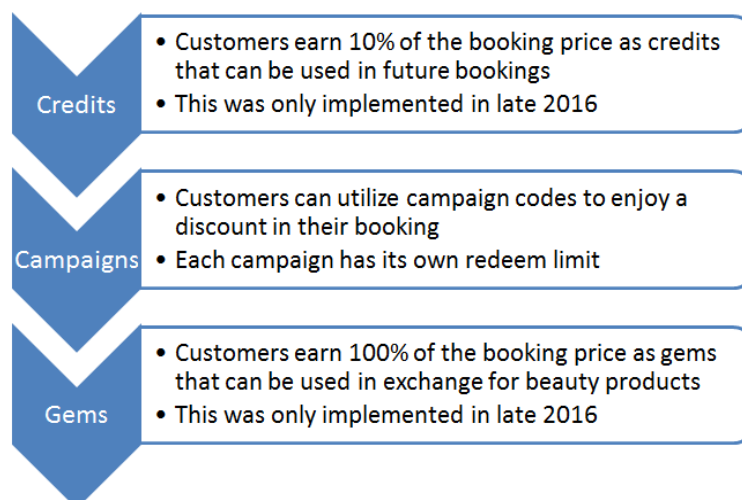


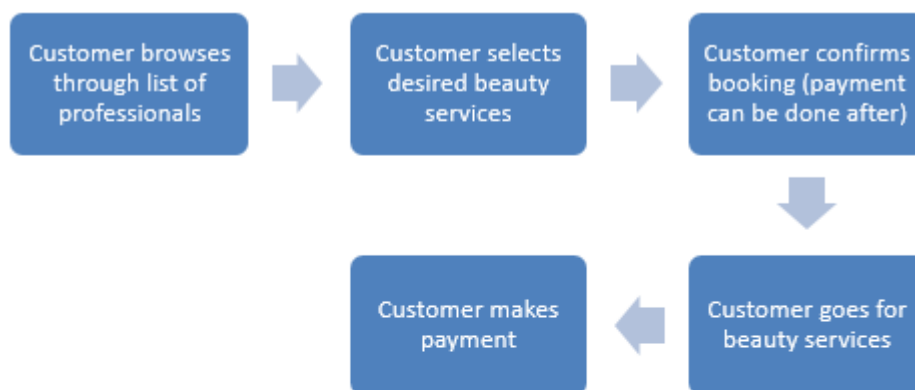
Figure 1 - Vanitee's current loyalty program

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 have set the following objectives:

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 (refer to Figure 2), whether they are accumulating before use or using them in their next booking



- Credits can be used here to reduce the price of booking
- Campaign codes can be used here to reduce the price of booking
- Credits (10%) & Gems (100%) are earned here

Figure 2 - Vanitee customers' interaction with the application

- 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? Do customers respond more to campaigns giving discounts in dollar amounts (e.g. \$20 off) or to percentage amounts (e.g. 20% off)? 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 (based on the following hypothesis).
 - Are beauty professionals more attractive if they have a higher chat response rate?
 - Are beauty professionals more attractive if they have a greater variety of services?
 - Are beauty professionals more attractive if they offer less expensive services compared to other professionals?
 - Are beauty professionals more attractive if they offer services on non-working days or hours?

Data Integration and Filtering

Data Collection

To facilitate our data analysis, Vanitee has provided us with access to their current MongoDB database on the cloud. The database contains numerous tables such as customers, beauty professionals, bookings, campaigns etc. Our team has decided to use two full years' worth of data which ranges from Jan 2015 to Dec 2016.

Extracted Tables

Different types of data are currently represented by different tables in the database. In this case, there were a total of 59 tables for us to utilize. After exploring each table and its suitability for analysis, we eventually narrowed down to the following 7 tables:

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:

Existing Column	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
service_ids	The set of ids belonging to the service(s) purchased
use_credits	Indicates if credits are used in the booking
price_final	The final price chargeable to the customer (after discount)
price_discount	The price of the discount, if any
price_payout	The price of the payout to the professional
price_transaction_fee	The price allocated for payment by credit card

price_customer_cashback_fee	The price that is converted to credits to be obtained by the customer
price_vanitee_transaction_fee	The price that Vanitee receives as profit
campaign_code_id	The campaign code used in the booking, if any
created_at	The date & time that the booking was created at

New Columns Created	Description
created_at_sgt	The date & time that the booking was created at in SGT
created_at_year	The year that the booking was created
created_at_month	The month that the booking was created
created_at_day	The day that the booking was created
is_master	Indicates if the booking is a master booking or not
is_online	Indicates if the booking is an online or manual booking
service_count	The number of services selected in the booking
vanitee_fee	The profit earned by Vanitee (based on the new formula mentioned above)
use_campaign	Indicates if the booking utilized any campaign codes
category_name	The category that the booking falls under

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:

Existing Column	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
-----------------	----------------------------------------------------------

New Columns Created	Description
created_at_sgt	The date & time that the campaign was created at in SGT
start_at_sgt	The date & time that the booking started at in SGT
end_at_sgt	The date & time that the booking ended at in SGT
duration	The duration of the campaign
duration_category	The category that the duration belongs to (e.g. 1 week, 2 weeks etc.)
is_published	Indicates if the campaign has been published or not

Categories

A row in this table represents a specific category that can be used to classify services. The detailed description of the main columns in this table is as follows:

Existing Column	Description
id	The unique id of the category
created_at	The date & time that the category was created at
name	The name of the category
parent_id	The id of the parent category

New Columns Created	Description
created_at_sgt	The date & time that the category was created at in SGT

Users

A row in this table represents a specific user that has an account on the application (either as a customer or professional). The detailed description of the main columns in this table is as follows:

Existing Column	Description
id	The unique id of the user
admin_at	The date & time that the user was made an admin
created_at	The date & time that the user was created at
gender	The gender of the user
date_of_birth	The date of birth of the user
professional_id	The unique id of the professional if the user is one

New Columns Created	Description
created_at_sgt	The date & time that the user was created at in SGT
age	The age of the user
is_professional	Indicates if the user is a professional or customer

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:

Existing Column	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

Professionals

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

Existing Column	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:

Existing Column	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
category_level_1_id	The unique id of the master category tied to this service

New Columns Created	Description
created_at_sgt	The date & time that the service was created at in SGT
is_master	Indicates if the service is a master service or not
is_published	Indicates if the service has been published or not
price_category	The category that the duration belongs to (e.g. <50, 51-100 etc.)

Challenges

As a relatively new startup that is still growing, Vanitee has been doing quite a bit of testing to minimize the number of bugs in their application. However, it has been observed that their testing is not done in a consistent manner which makes it difficult for us to filter out test data accurately. Moreover, some of the tables do not have indicative columns to indicate whether a record is a test data or not. Hence, to overcome this, we have implemented the following measures:

- Narrowed the initial data range from Aug 2014 - Dec 2016 to Jan 2015 - Dec 2016 as Vanitee had mentioned that larger amounts of testing happened in 2014 when they first officially launched
- Remove records that have columns that indicate that the record is a test data (e.g. columns such as `is_test`, `test_at`, `deleted_at`)
- Manually searched for the keyword “test” that appeared in several columns that involve name and description

Another challenge we faced was due to the way Vanitee had structured a few of their tables in the database. For example, the Bookings & Categories table had way more records than expected. The main reason for this was that most of the records were actually related to each other where certain records had a `master_id` column that referenced another record within the same table. Vanitee’s rationale behind the `master_id` column was that they wanted to keep track of the changes made to each record but as new records instead of updating the current record. One specific example is that a single booking can have a parent booking (also known as the master booking). The multiple child bookings represent the different states that the booking has undergone. Hence, to overcome this, we have implemented the following measures:

- For those tables with the column `master_id`, we created a new column called `is_master` to indicate whether that record is a master record or not
- Non-master records are then excluded from analysis and the formulation of graphs

Data Cleaning and Exploration

Issues

Before exploring the data, we faced several issues such as duplication of data, missing values as well as changes in Vanitee’s business model. As these issues may potentially affect the accuracy of our analysis, we have implemented the following measures to overcome these issues prior to performing our analysis.

Duplicate Values

Out of the 7 tables identified above, only the Campaigns table had records with duplicate values. As seen in Figure 3 below, we observed that there were several records that had similar campaign names and almost identical creation date times. However, we realized that only 1 campaign record had campaign codes, while the others did not have. Hence, we made the assumption that only

campaigns that had campaign codes were true campaigns that were carried out. Also, to help reduce such duplicate values, we made use of the column `is_published` to sieve out those campaigns that were actually published over the past 2 years.

Vanitee X Kelynn Nai	vaniteekelynn vtkelynn2016 vtkelynn	0/1	0/1	0	04 May 2016 6:17:48 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:47 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:47 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:46 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:47 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:46 PM
Vanitee X Kelynn Nai		0/1	0/1	0	04 May 2016 6:17:46 PM

Figure 3 - Campaign records that have duplicate values

Missing Values

After examining the tables, we realized that different tables had a varying amount of columns with missing values. Firstly, when we looked at the Bookings table, columns that contained monetary values (e.g. `final_price`, `total_price`, `discount_amount` etc.) mostly had missing values as well. Upon further inspection, we deduced that these missing values actually represented the monetary value of 0. Hence we replaced those missing values with the value of 0. Another column that had a huge amount of missing values was `price_vanitee_transaction_fee` which is the fee that Vanitee profits when an online booking has been successfully checked out. Subsequently, we found out that these missing values were attributed to changes in Vanitee’s business model which we will elaborate further in the next issue.

Another major table that had columns with missing values was the Campaigns table. Columns like `start_at` and `end_at` had several missing values that confused us initially. After analyzing Vanitee’s online dashboard, we learnt that campaigns need not have start and end dates to be created. Campaigns that had missing start dates would use the creation date as replacement as clarified with Vanitee. On the other hand, campaigns that had missing end dates would mean that the campaign codes could be redeemed as long as the start date had passed.

Changes in Business Model

After meetings with Vanitee and email correspondence with their developers, we learnt that Vanitee's current business model was only recently implemented (around Nov to Dec 2016). As seen in Figure 4, there are several changes to their business model. To make matters worse, the Bookings table did not have a column that indicates whether a booking was based on the current or previous business model. In addition, this also partially explains why the column price_vanitee_transaction_fee had missing values (as mentioned above).

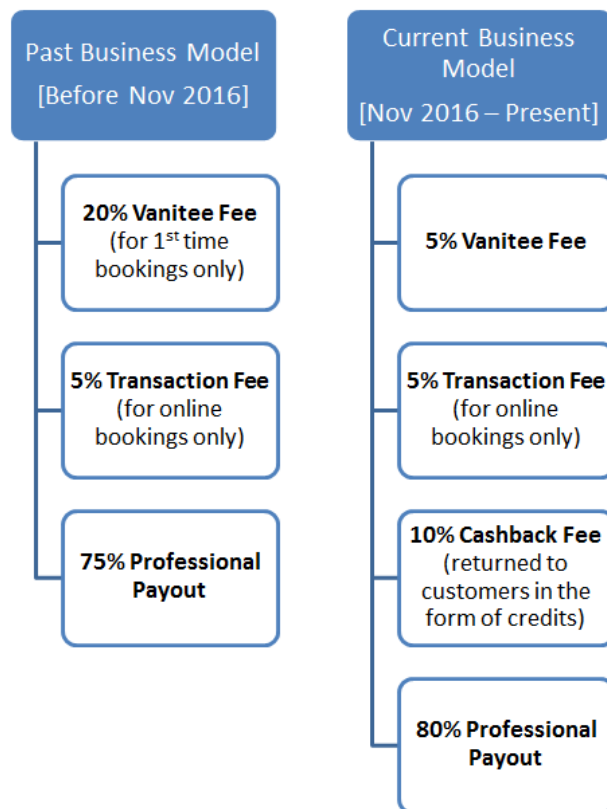


Figure 4 - Comparison of past and current business models (*percentages are based on the price of the specific booking)

Since some of our analysis we were about to do involved calculating the profit or Vanitee fee made per booking, we came up with the formula (as shown in Figure 5) below as an alternative way of calculating that value. To put it simply, the profit per booking is the final price that the customer pays (after any discount) minus away the payout that the professional receives, minus away the transaction fee incurred from payment by credit card and finally minus away the cashback that the customer receives as credits. We felt that the benefit of using this formula was that it ignores the Vanitee fee in its calculation and uses the other columns that do not have much missing values instead. Also, it works for any booking regardless of past or current business model.

$$\textit{profit} = \textit{final price} - \textit{payout} - \textit{transaction fee} - \textit{cashback}$$

Figure 5 - Alternate formula to calculate profit per booking

Findings

Users, Customers & Professionals

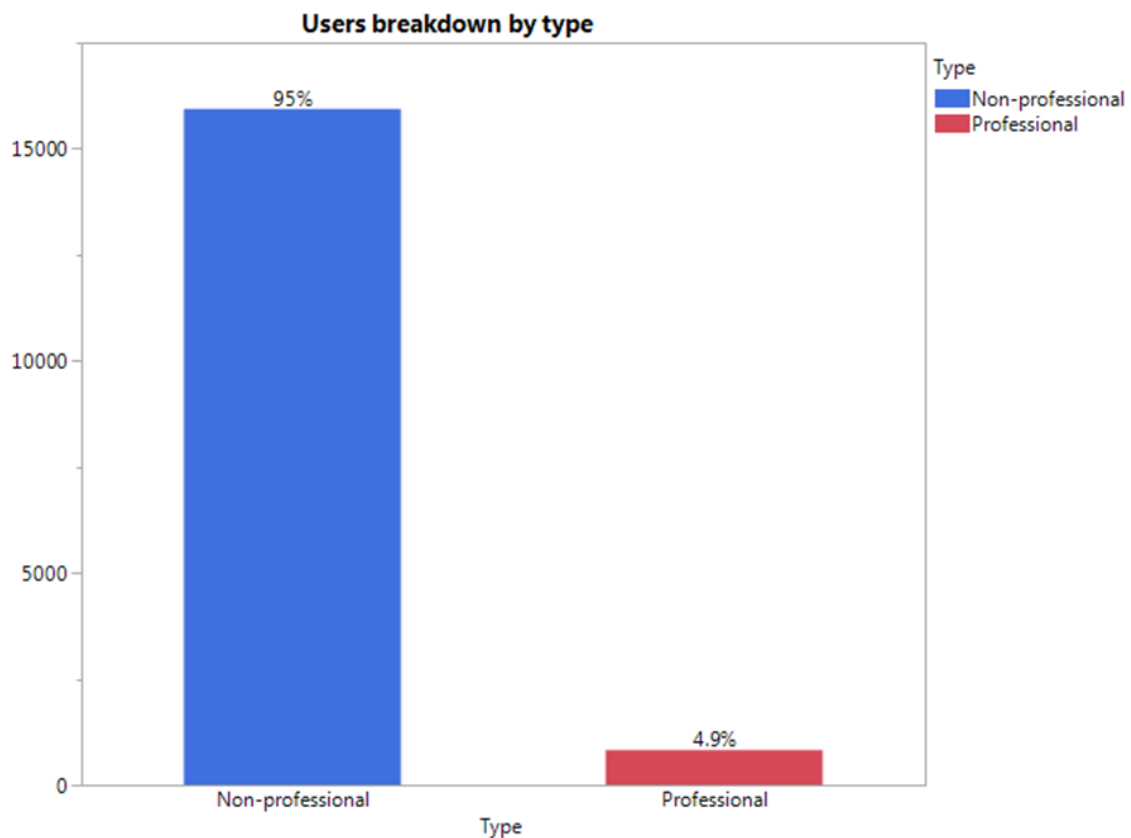


Figure 6 - Users breakdown by type

As we can see from Figure 6, there are much more non-professionals (customers) than professionals who currently have an account on the Vanitee application. This shows that there is indeed a demand for beauty services on this platform as more customers are interested in signing up for an account. However, it is important to note that having an account alone is not indicative of how successful the application is doing in encouraging the customer to book through this platform. We will continue to explore further when we look at the bookings made in the later part of this report.

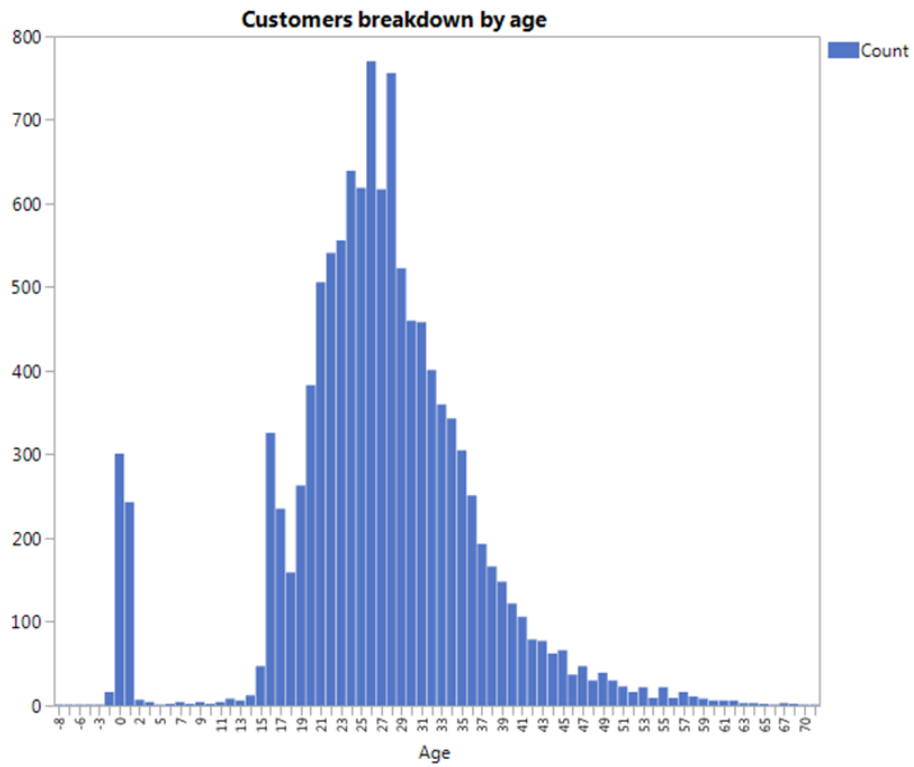


Figure 7 - Customers breakdown by age

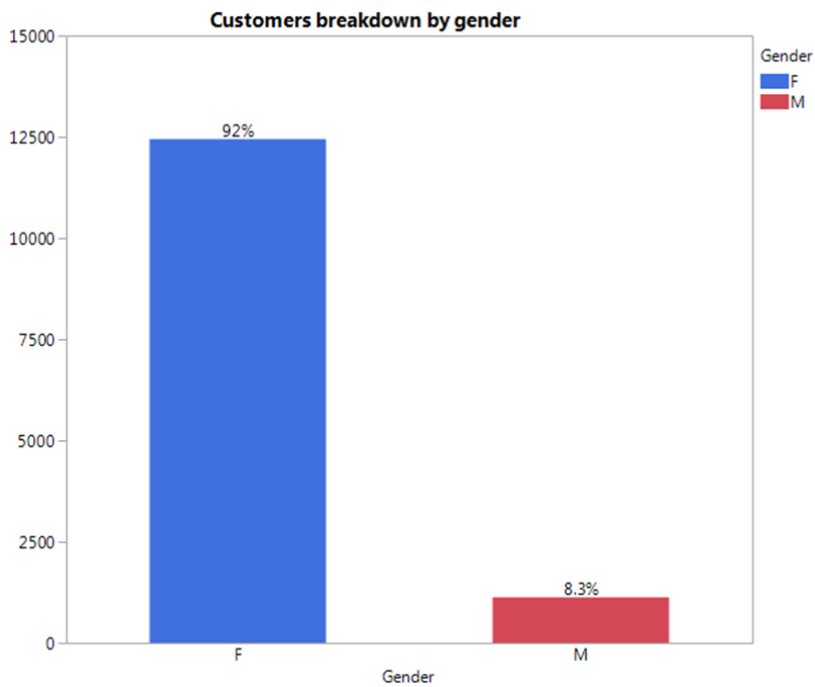


Figure 8 - Customers breakdown by gender

Looking at Figures 7 & 8, we can observe that most customers that use this platform are mainly females aged between 20 to 35.

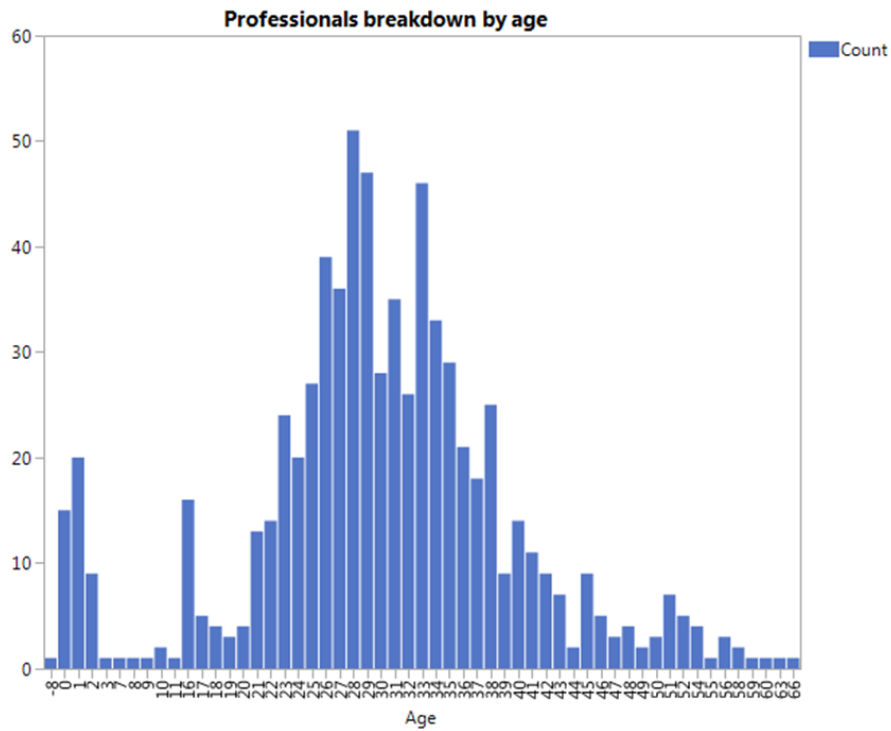


Figure 9 - Professionals breakdown by age

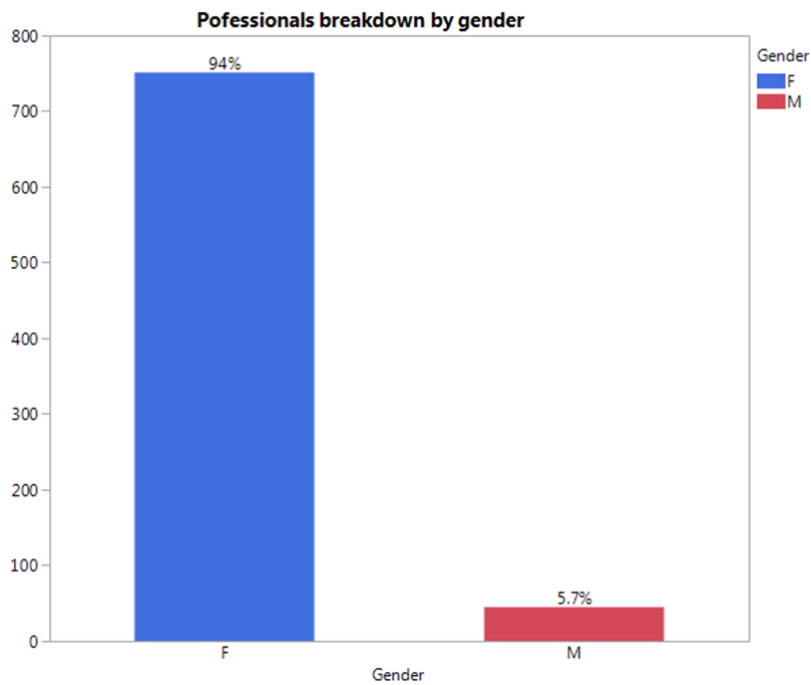


Figure 10 - Professionals breakdown by gender

Similarly from Figures 9 & 10, we can observe that most beauty professionals that use this platform are mainly females aged between 20 to 40.

Bookings

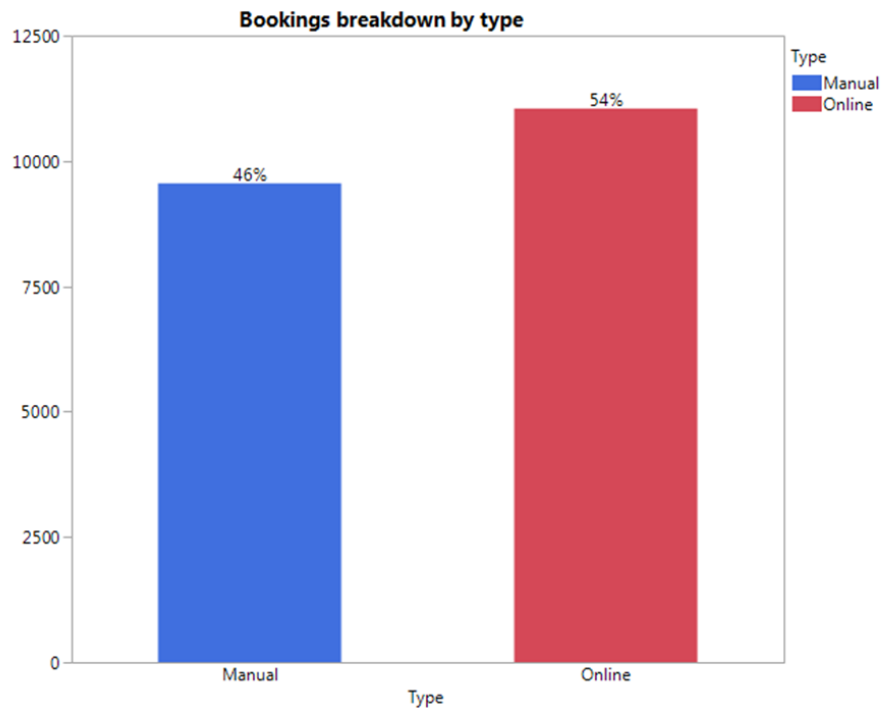


Figure 11 - Bookings breakdown by type

From Figure 11, we can see that there are slightly more online bookings than manual bookings, probably due to greater convenience in using the online platform.

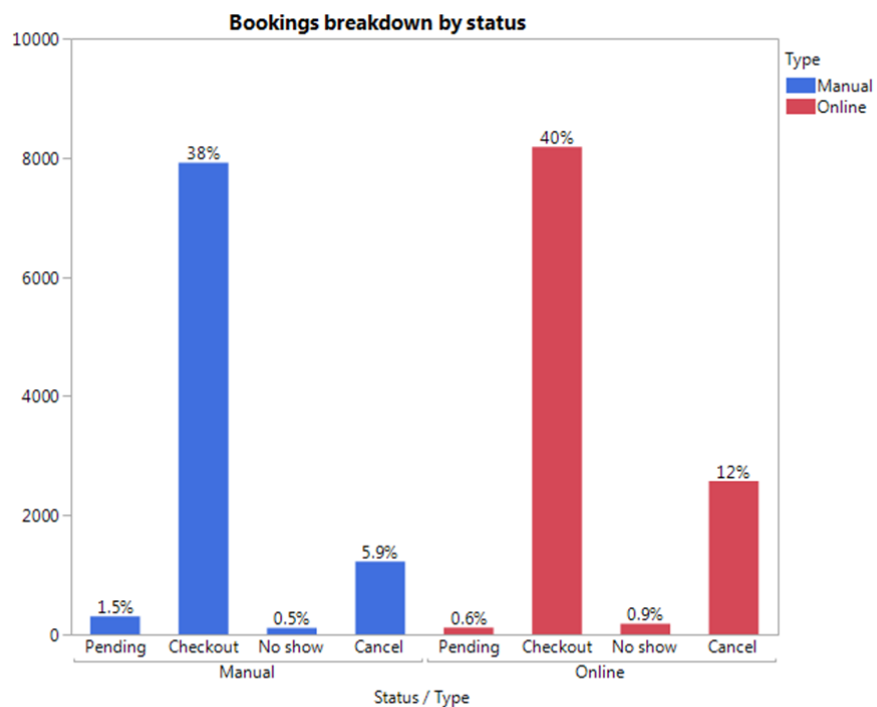


Figure 12 - Bookings breakdown by status

Bookings made through the application can have 4 different statuses namely, Pending, Checkout, No show and Cancel. From Figure 12, we can see how 40% out of the initial 54% of online bookings were successfully checked out. This seems to be a healthy percentage as it can be said that Vanitee only

earns revenue from online bookings that are successfully checked out. Hence, the graphs from this point onward would mainly be based on these online bookings that have the status of check out.

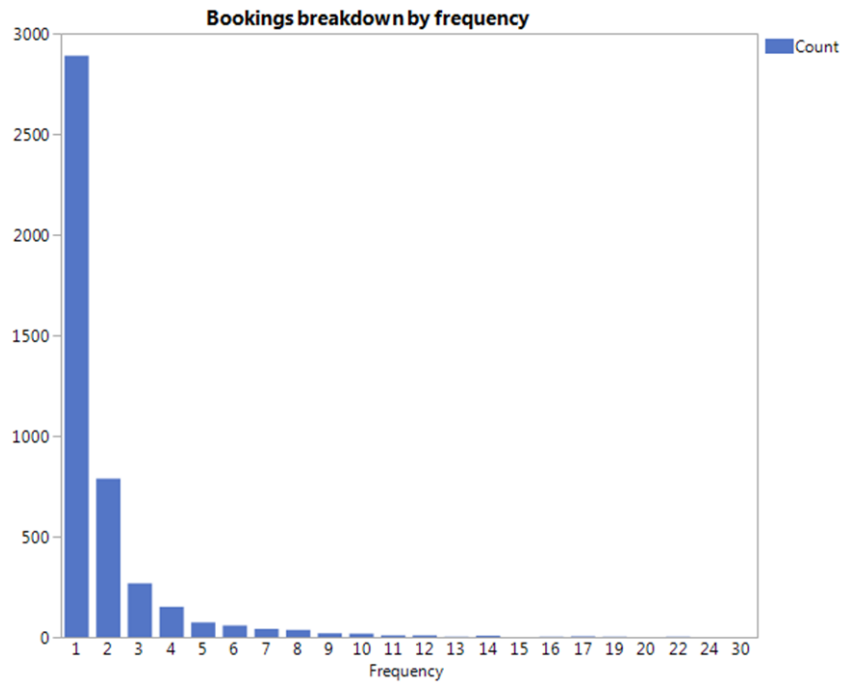


Figure 13 - Bookings breakdown by frequency

The next graph shows the bookings frequency throughout the past 2 years. Surprisingly, there are many users who have only booked once and this number of users actually significantly drops as the bookings frequency increases. There are probably many reasons why a user has only booked once and we hope to be able to identify these reasons in our future analysis as this would greatly help Vanitee improve on their customer retention strategies.

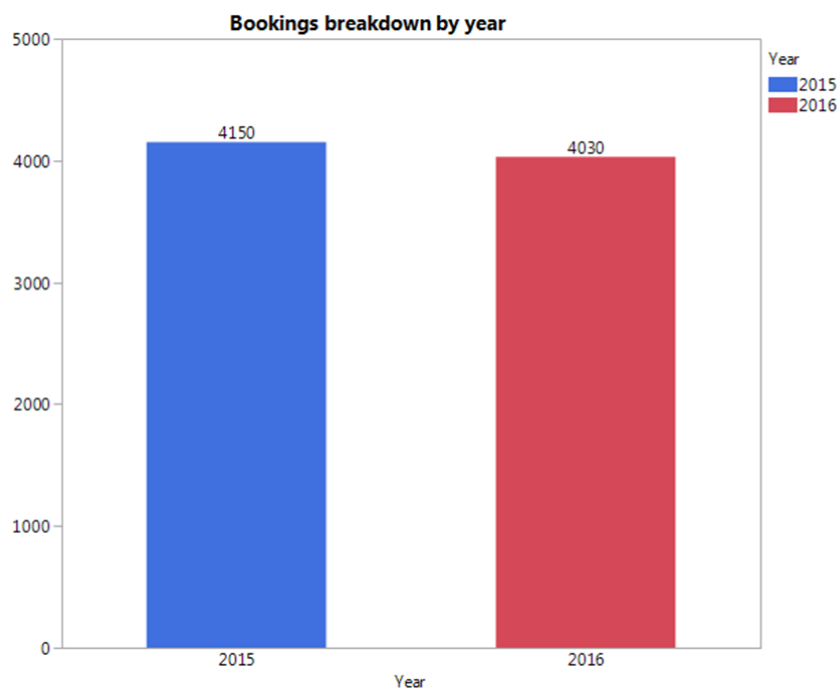


Figure 14 - Bookings breakdown by year

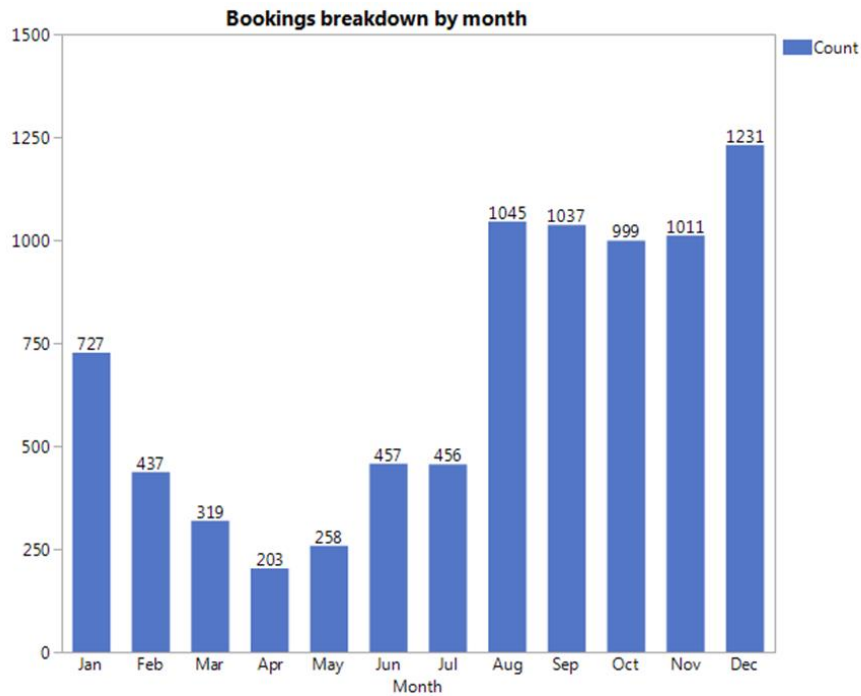


Figure 15 - Bookings breakdown by month

As seen from Figure 14, the number of bookings made in 2015 & 2016 are roughly the same, showing how Vanitee has managed to somehow sustain this over the past 2 years. Moving to Figure 15, it can be deduced that these bookings tend to occur in the last quarter of the year.

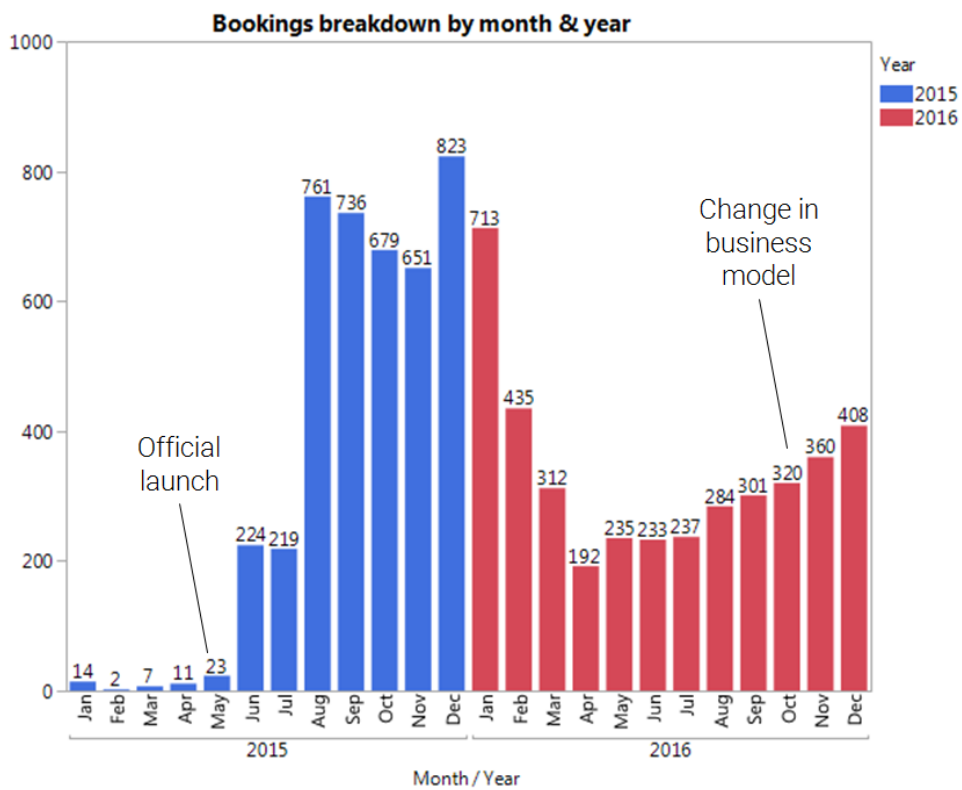


Figure 16 - Bookings breakdown by month & year

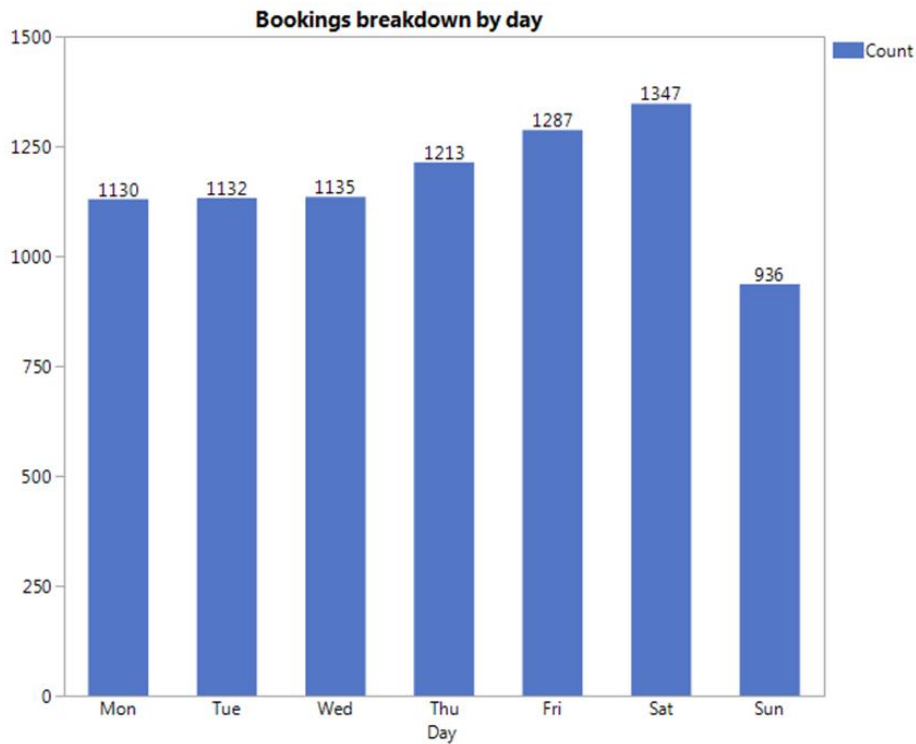


Figure 17 - Bookings breakdown by day

From Figure 16, we get to see a clearer picture of how the bookings vary across the past 2 years. In general, it can be said that Vanitee did much better in the last quarter of 2015 (right after its official launch) as compared to that of 2016. In the earlier part of 2016, it suffered a decline in the number of online bookings that were successfully checked out. However, this number gradually picked up some pace towards the last quarter of 2016. One reason for this could be the change in Vanitee's business model which happened around that period as well. The main difference in the business models is the introduction of customer cashback, in the form of credits, to incentivise customers to make more bookings.

In Figure 17, it can be observed that the frequency of bookings tend to increase nearer to the weekends (Friday, Saturday) where people naturally have more free time to themselves.

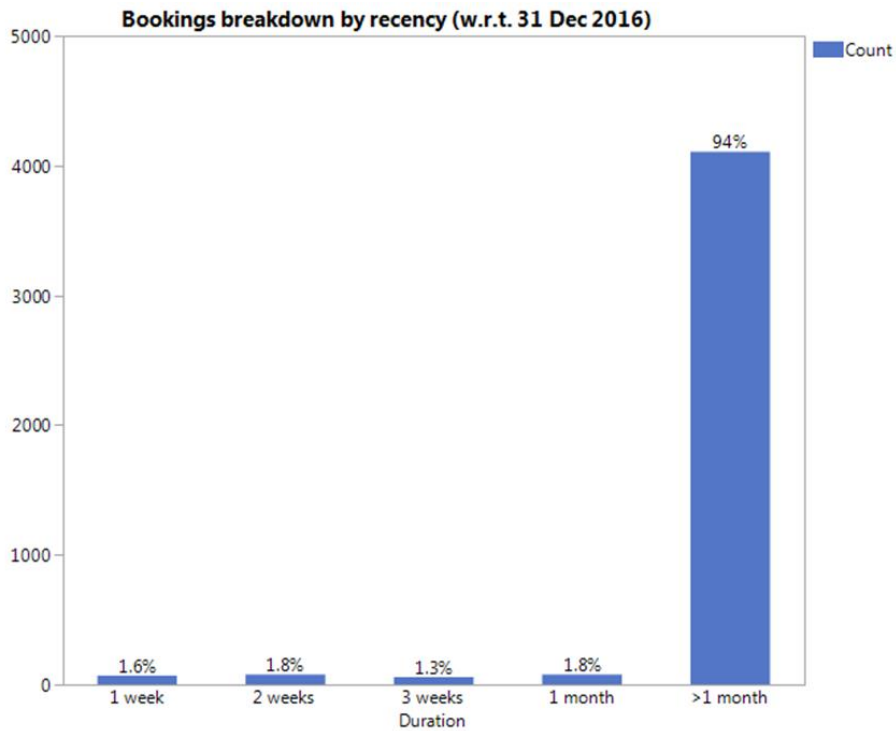


Figure 18 - Bookings breakdown by recency (initial)

The next thing we tried to find out from the Bookings data was whether existing customers have made any recent bookings. This was an important analysis to do as it provides us with a general idea as to how active the customers are in using this platform. As seen from Figure 18, we decided to calculate the duration from the last time that each customer has made a booking to 31 December 2016 which is the latest possible date from the data range we have selected. Initially, we came up with categories in terms of weeks with “>1 month” being the last category. However, upon generating the graph, we realized how skewed the analysis was as approximately 94% of customers fell under the “>1 month” category. Hence, to make this analysis more insightful, we decided to refine the categories to include durations in terms of months.

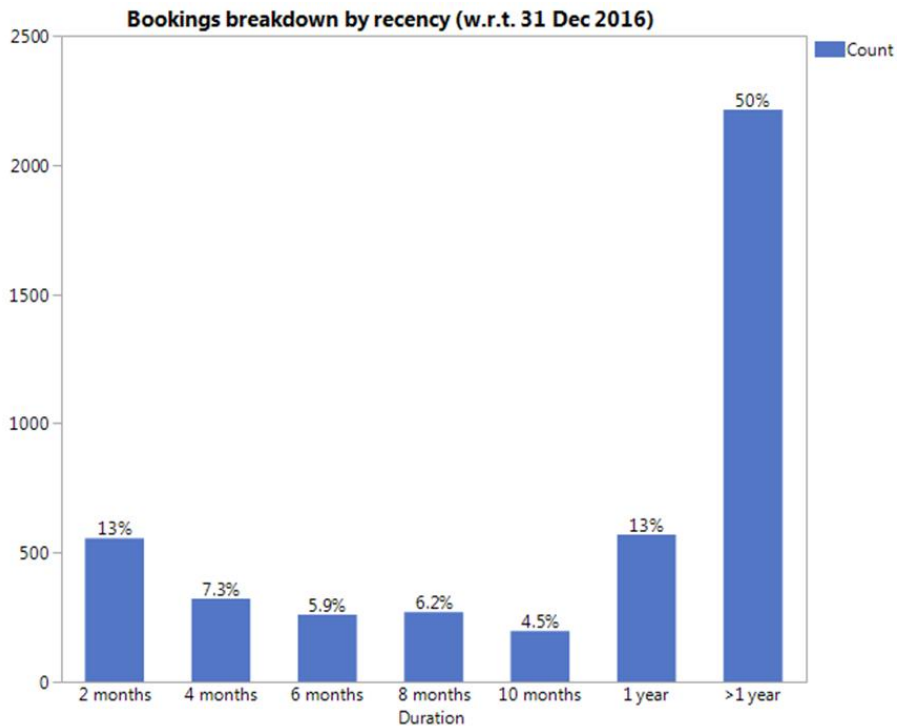


Figure 19 - Bookings breakdown by recency (final)

Figure 19 shows a much clearer picture of the recency analysis where we could conclude that 50% of existing customers have booked within the last year while the remaining half had not.

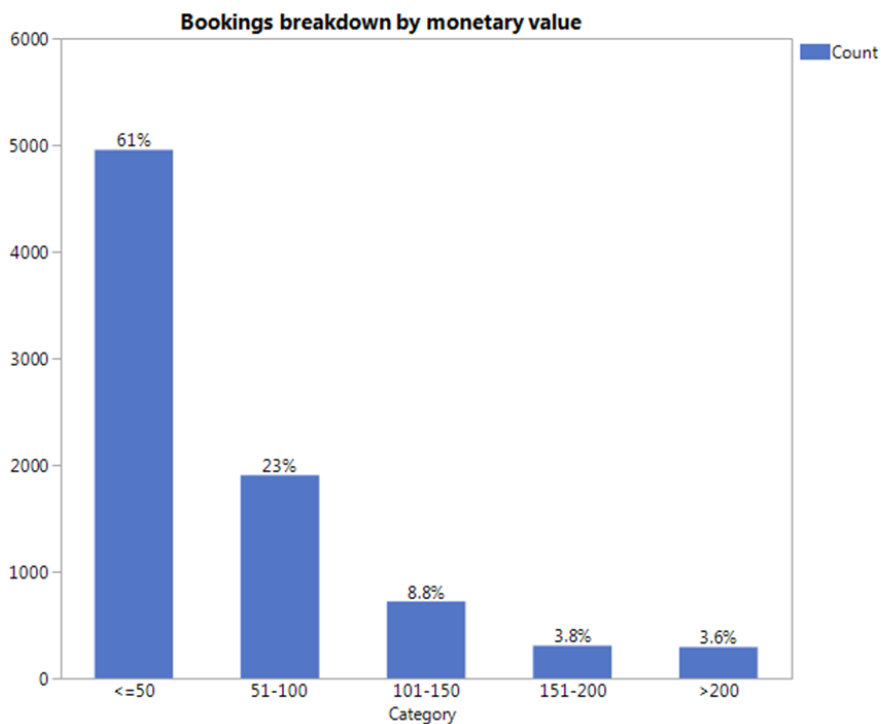


Figure 20 - Bookings breakdown by monetary value

The next figure above shows that 61% of online bookings made have a monetary value of less than \$50. This possibly shows how customers are generally willing to pay for beauty services that cost around \$50. Exceeding this mark shows a huge decrease in the number of bookings made.

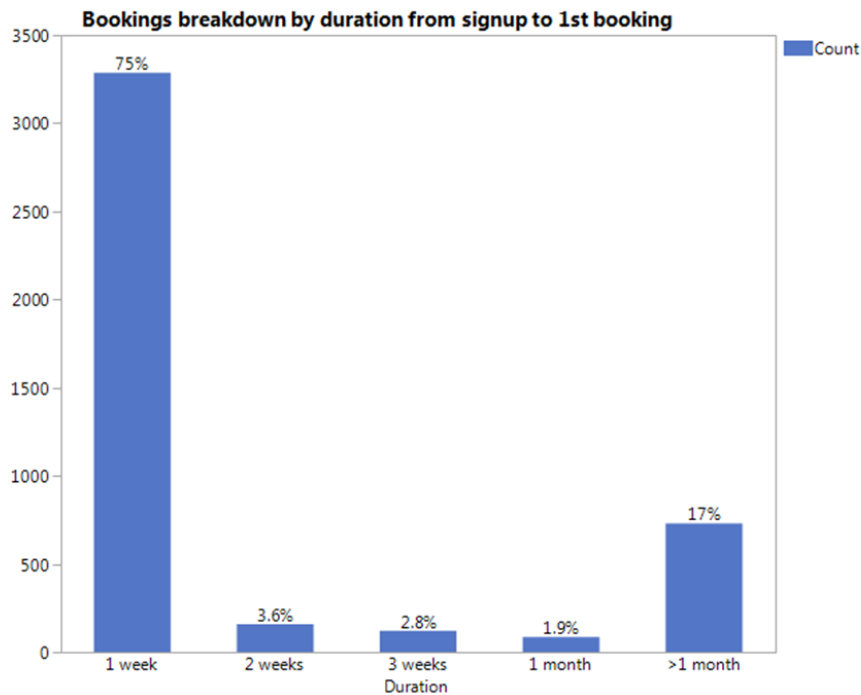


Figure 21 - Bookings breakdown by duration from signup to 1st booking (initial)

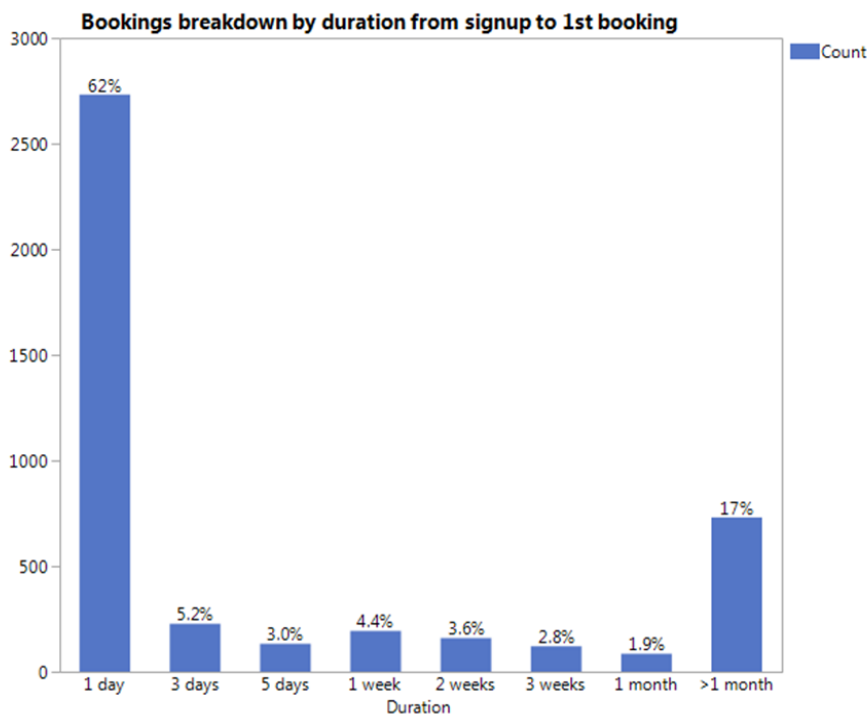


Figure 22 - Bookings breakdown by duration from signup to 1st booking (final)

Figure 21 & 22 shows the bookings breakdown by the duration from when customers signup to their very first successful online booking. While generating this analysis, we faced a similar issue of having an overly skewed results, in this case 75% of customers make their first booking within 1 week of their signup. As this percentage was pretty large, we decided to take the same action and break it down to smaller durations in terms of days as seen in Figure 22. The final result ended up to be very positive showing that 62% of customers make their first bookings within a day of their signup.

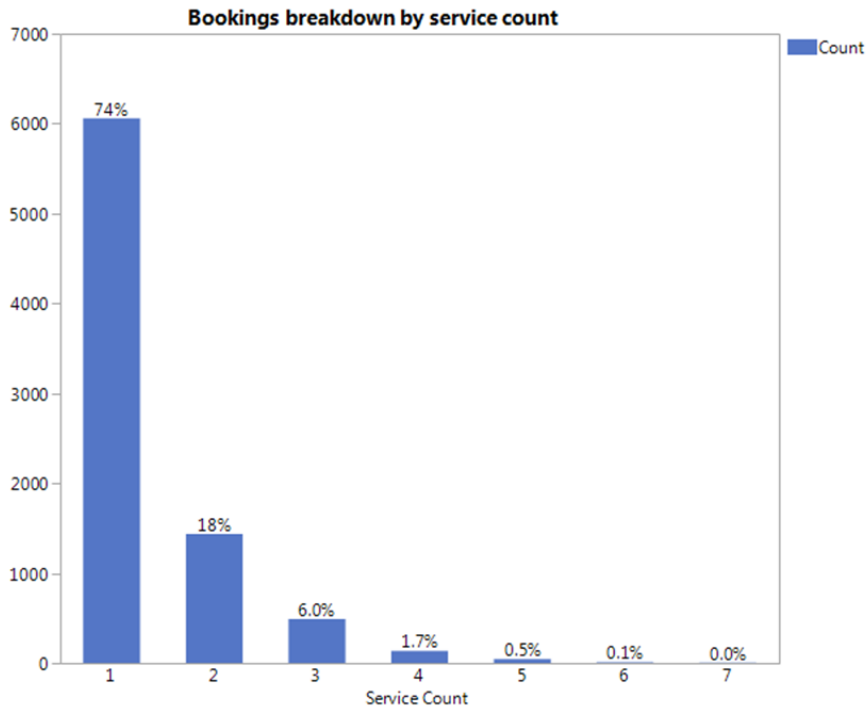


Figure 23 - Bookings breakdown by service count

Figure 23 shows us the bookings breakdown by service count. In this case, 74% of bookings only involve 1 beauty service. This shows that most customers are specific in targeting the main beauty service that they want to engage in, be it nails or brows or other types of services.

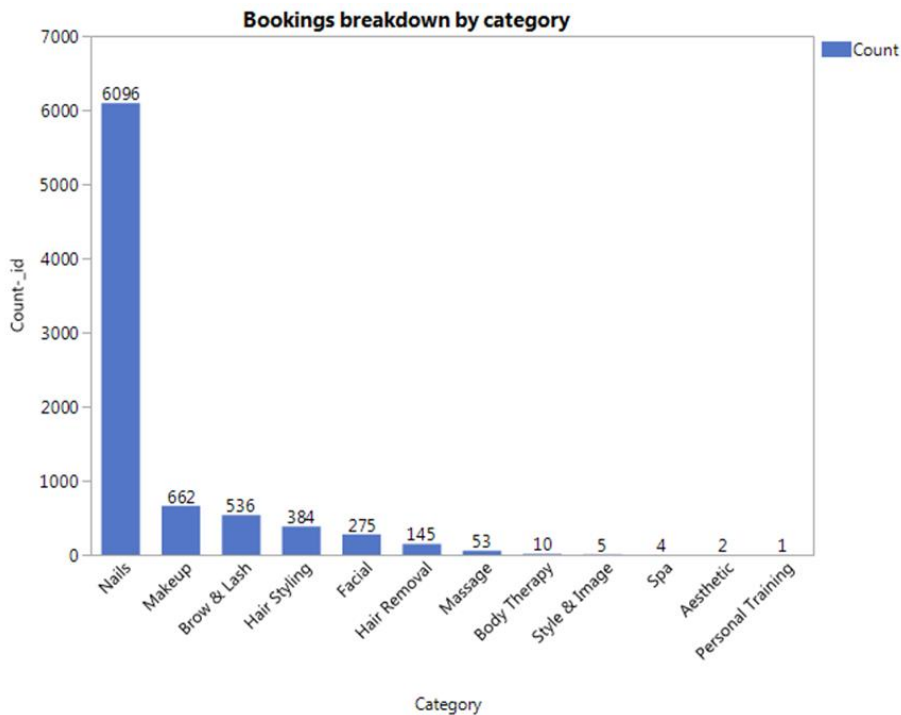


Figure 24 - Bookings breakdown by category

This figure simply shows that most bookings involve nail, makeup and brow services.

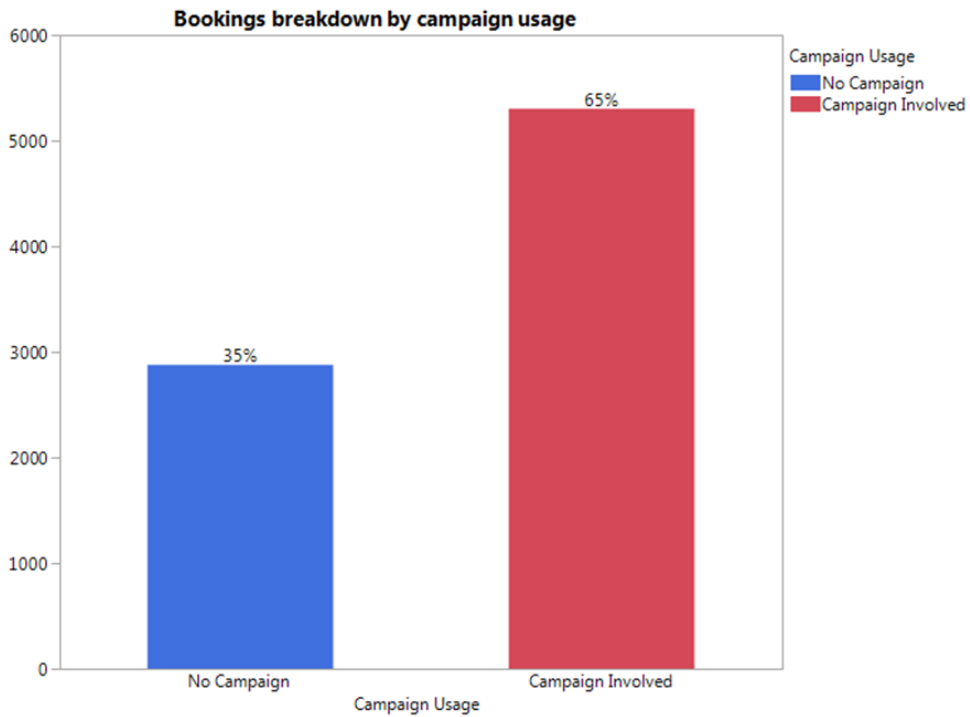


Figure 25 - Bookings breakdown by campaign usage

As seen from Figure 25, about two thirds of the bookings made involve some form of campaign which allows customers to enjoy booking discounts.

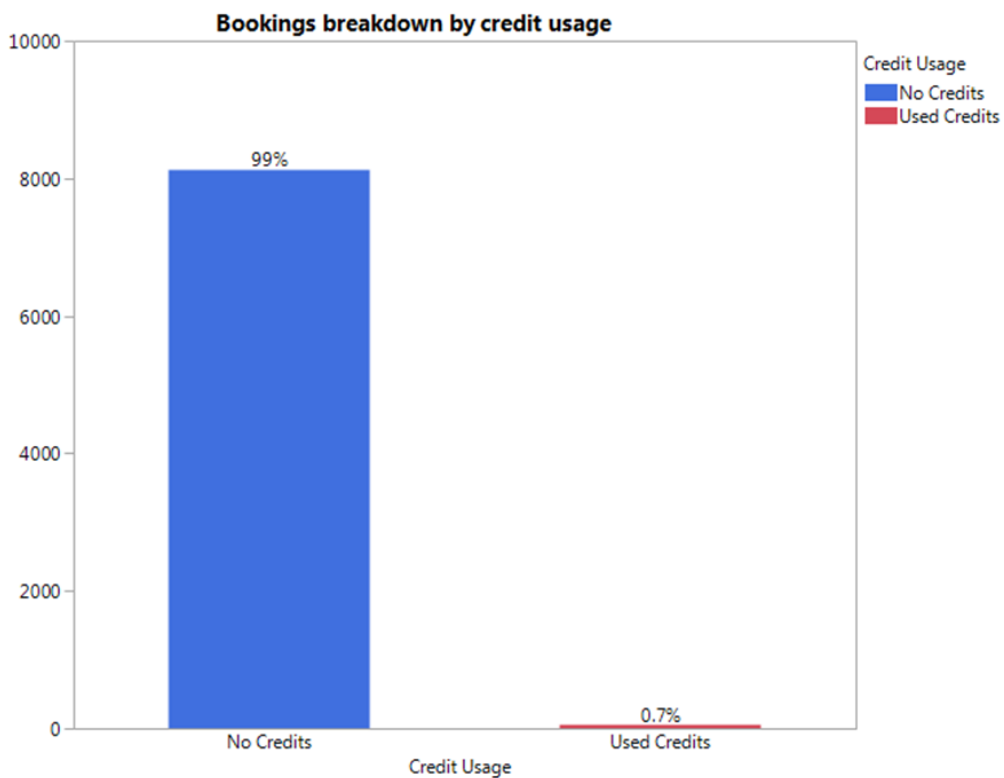


Figure 26 - Bookings breakdown by credit usage

However, when we tried looking at the percentage of bookings that utilized credits, we found out that this percentage only amounts to approximately 1%. One possible reason for this extremely low number is that most customers have only made 1 booking, which means that they have not even utilized the credits earned from their 1st booking. Another reason is that the current business model that involves customer cashback in the form of credits was only introduced in the last quarter of 2016.

$$profit = final\ price - payout - transaction\ fee - cashback$$

$$vanitee\ fee = final\ price - \frac{payout}{total\ fee} * final\ price - \frac{transaction\ fee}{total\ fee} * final\ price - \frac{cashback}{total\ fee} * final\ price$$

Figure 27 - Profit per booking formula comparison

The last thing that we wanted to find out was the total profit that Vanitee had made over the last 2 years. Utilizing the 1st formula in Figure 27, we calculated the profit to stand at a surprising value of -\$98k. One of the main reasons for this is that their current business model doesn't allow for much profits to be made, especially when there are Vanitee discounts involved in bookings. As compared to bookings that have no discounts or have professional discounts, we noted that Vanitee discounts are naturally absorbed by Vanitee itself and may be part of their marketing budget to try to incentivise more people to use the platform. Also, in this specific scenario where Vanitee discount is present, other fees such as payout, transaction fee & customer cashback are still calculated based on the initial price before the discount.

As we felt that such calculations may put Vanitee at a disadvantage, we came up with a revised formula as seen in Figure 27 that calculates the profit based on the scenario where the above mentioned fees are calculated based on the final price after the discount. As expected, the final profit turned out to be a less negative amount of -\$85k. A simple change in formula actually resulted in \$13k being saved. However, we do understand that there may be a possibility that the calculated losses may very well fall within Vanitee's budget.

Services

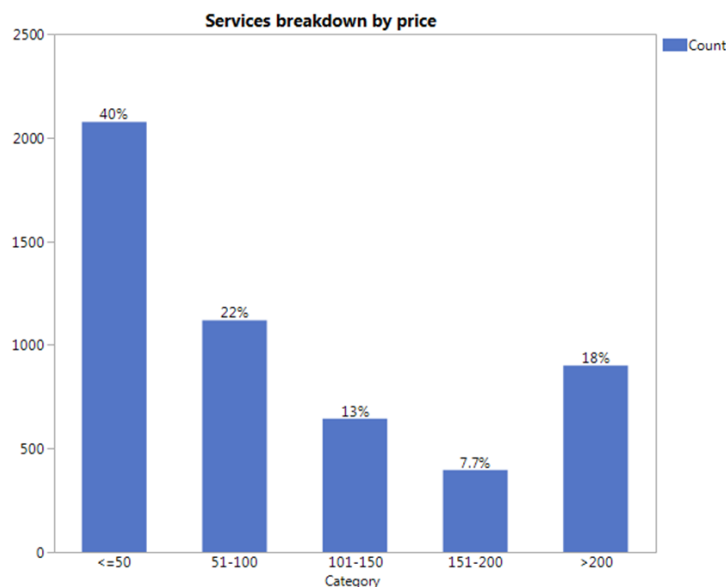


Figure 28 - Services breakdown by price

Next, we shifted our focus to the services table, which includes the services that professionals have created through the application. For the following few analysis, we have only looked at services that are published, meaning only those active services that are currently visible to customers for their selection.

To start off, Figure 28 shows how majority of services are priced at \$50 or less. This shows that professionals have a rough idea that services that are priced at \$50 or less may be more attractive to customers. The previous graph showing how majority of bookings have a monetary value of \$50 or less reinforces this point.

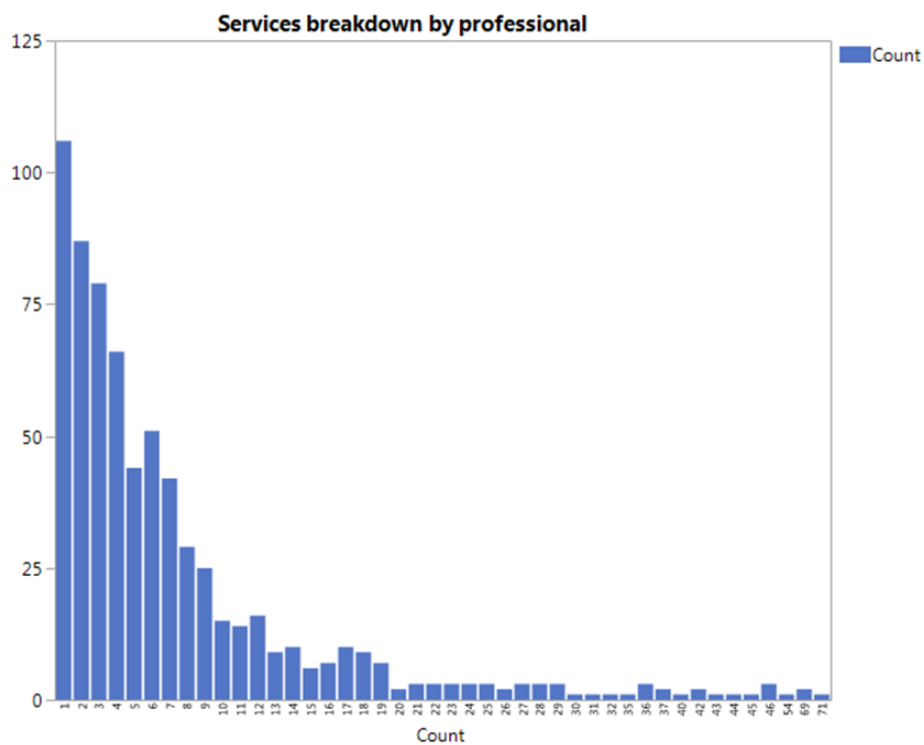


Figure 29 - Services breakdown by professional

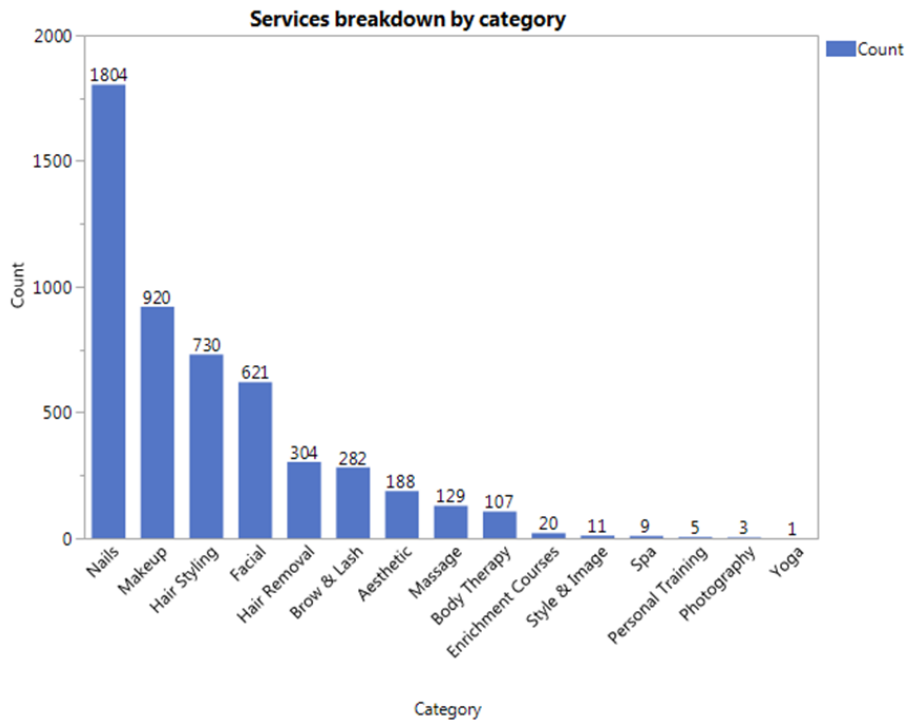


Figure 30 - Services breakdown by category

Next, Figures 29 & 30 show the number of services that most professionals have as well as the category the most services belong to respectively. We can see that most professionals only have 1 service that customers can select from. A possible explanation for this could be the fact that some of these services could be over generalized such that the customer gets to only customize his or her service on the actual day of the booking itself. Also, as expected, majority of services offered involve nails, which corresponds the the graph earlier where most bookings are categorized under nails as well.

Campaigns

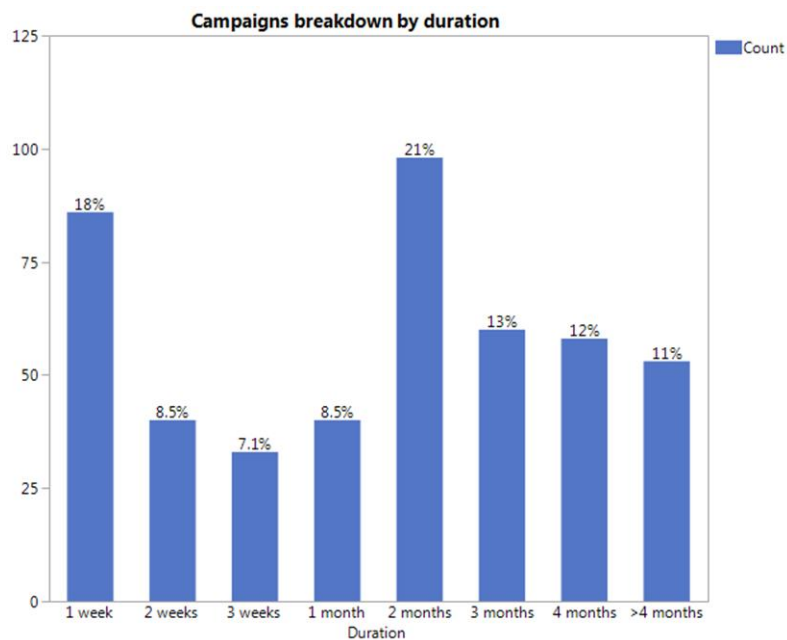


Figure 31 - Campaigns breakdown by duration

The final data table that we have explored and analyzed is the Campaigns table. Figure 31 shows that most campaigns last for either 1 week or 2 months. This duration is calculated by taking into account the start and end date of the campaign. However, this graph alone does not tell us why the above stated durations are more prevalent.

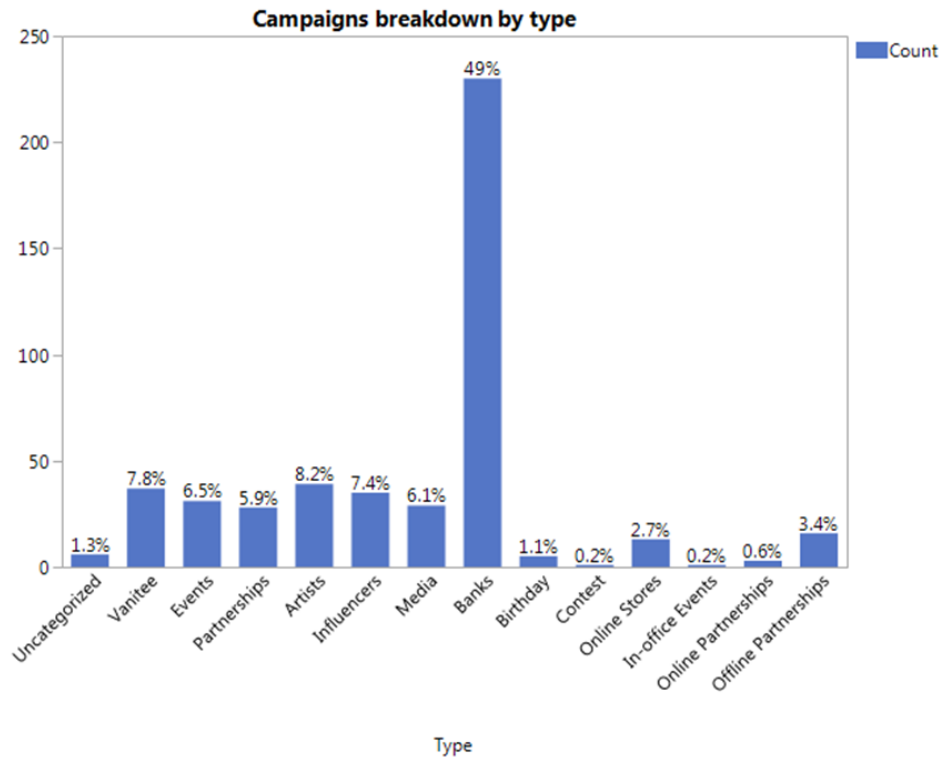


Figure 32 - Campaigns breakdown by type

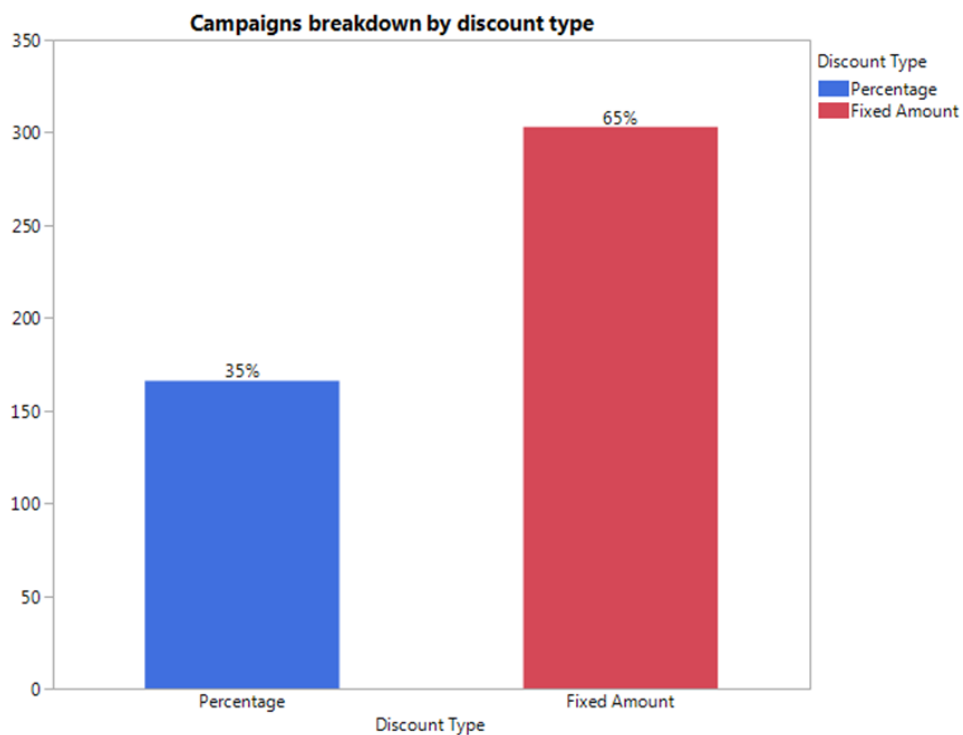


Figure 33 - Campaigns breakdown by discount type

Figures 32 & 33 show the breakdown of campaigns by type and discount type respectively. We can see that majority of the campaigns involves partnering banks such as DBS etc. Also, these campaigns often offer a fixed amount of discount.

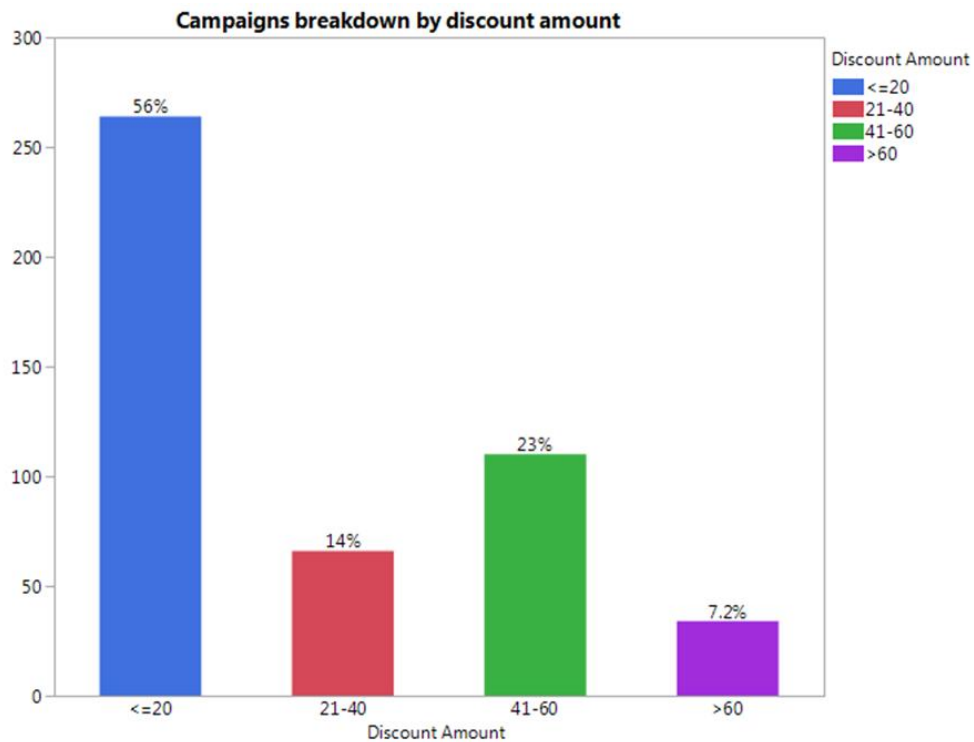


Figure 34 - Campaigns breakdown by discount amount

This discount normally amounts to less than \$20 per booking which we feel is already a substantial amount for customers to utilize.

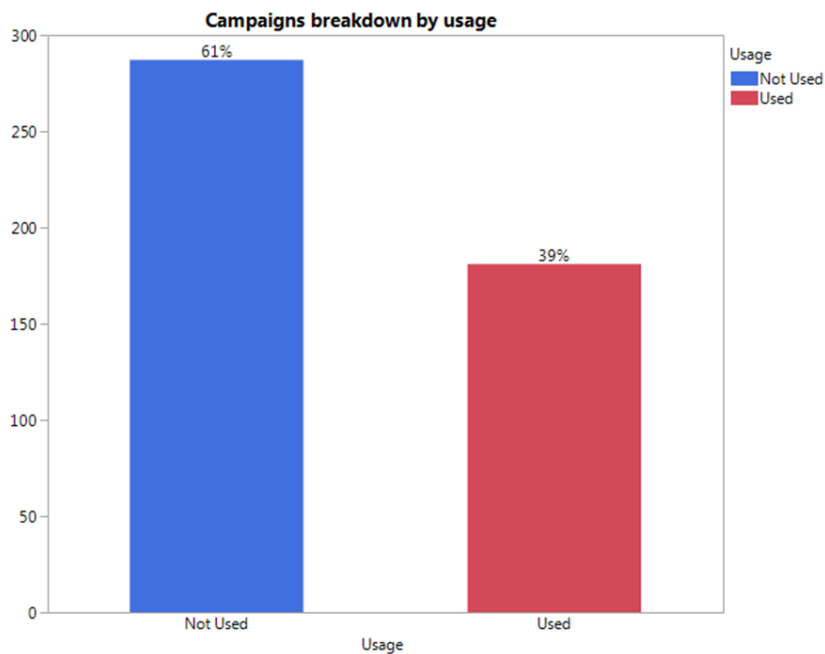


Figure 35 - Campaigns breakdown by usage

Next, we have also analyzed the campaigns by their usage. The way we calculated the usage basically whether a particular campaign has any booking that utilized the campaign codes belonging to that campaign. However, based on our previous meeting with Vanitee, we understand that such a graph may not be as accurate as it seems due to the presence of test data that we currently have no way of filtering out.

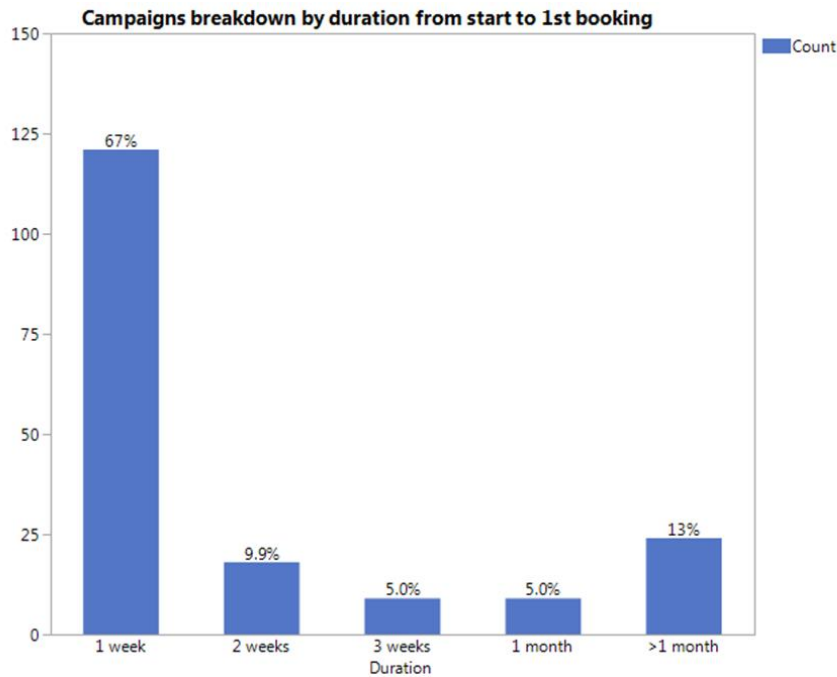


Figure 36 - Campaigns breakdown by duration from start to 1st booking (initial)

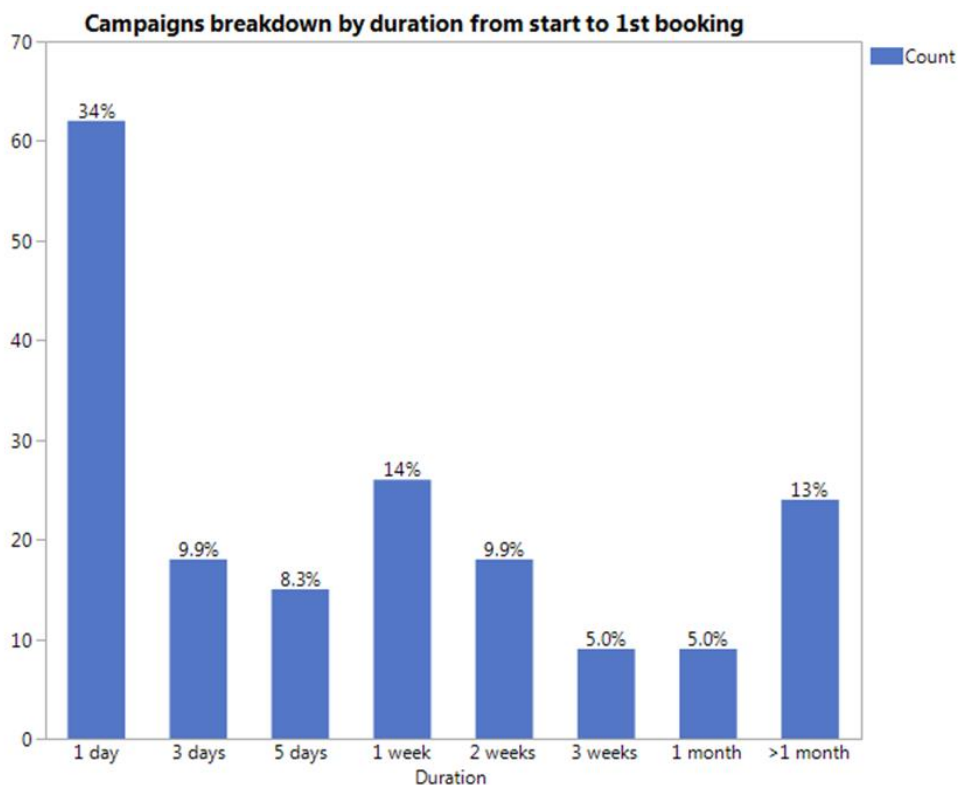


Figure 37 - Campaigns breakdown by duration from start to 1st booking (final)

Due to the uncertainty of the previous graph on whether campaigns created are actually being used, we decided to zoom into those current active campaigns that have at least one booking tied to it. From there, we came up with the above 2 graphs that tells us the breakdown of campaigns by duration from the start of the campaign to the very 1st booking made. Initially, 67% of the campaigns received their very first booking within 1 week of launch. After looking into further, we learnt that a positive percentage of 34% of campaigns actually had their 1st booking within 1 day. This shows how customers are quick to react to any new campaigns created.

Revised Methodology

Cluster Analysis

Firstly, 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.

Revised Scope of Work

Our scope of work mainly remains the same as before as stated in our project proposal. However, based on the comments given by our sponsor and project supervisor, our team has decided to revise some of the exploratory data analysis. For certain charts, we will break down further to derive more specific information that will help Vanitee understand the situation on hand. The charts we will supplement are as follows:

1. A chart to display the breakdown of bookings based on the day it is being checked out, meaning the day the service is being booked for and carried out. This chart will complement the already existing chart that displays the breakdown of bookings based on the day the booking was created, meaning the day the customer made the booking.
2. A chart that will display the breakdown of individual services by price. For now, the current bar chart shows the breakdown of of the services by price.
3. A chart that will display the breakdown of campaigns that have the most redemptions. This chart will facilitate our analysis to find out the ideal duration and discount a campaign should have.

From the feedback we got from Prof. Kam, we have decided to focus our analysis more on customers and bookings (which includes the usage of campaigns and credits) as compared to the beauty professionals. Also, our sponsor feels that it is more beneficial to them to understand more about their customers. For this reason, we will not be looking at the activeness of a beauty professional but we will continue the analysis on the attractiveness of a beauty professional as the attractiveness of a beauty professional affects the way customers make their booking whereas the activeness of the beauty professional is less likely to do so.

Moving forward, our analysis will allow us discover insights based on Vanitee’s revenue and bookings. For instance, for the service(s) that have the most bookings made, Vanitee would like to know what the ideal price of the service should be. Also, we will look into successful bookings that used campaign codes to find out the ideal type of discount and for what duration.

The table below will illustrate better our updated work scope:

Completed ✓	Removed ✕	In-Progress and New Tasks +
<ol style="list-style-type: none"> 1. Data cleaning on all 7 tables by replacing missing values and filtering out duplicates. 2. Identifying the preferred type of booking. 3. The following completed tasks are based on bookings made online: 4. Identifying the status at which most bookings fall under. 5. Identifying the number of bookings made typically. 6. Identifying the number of bookings made within the range of data used for each year, month, a combination of both year and month, and lastly week. 7. Identifying the amount of bookings that were made recently. 8. Identifying the bookings within a certain price range. 9. Identifying the duration a booking is usually made from the time of account sign up to the first booking. 10. Identifying the number of service(s) bookings typically have. 11. Identifying the service that most bookings are made for. 12. Identifying the amount of booking that utilise a campaign code and credits 	<ol style="list-style-type: none"> 1. Conducting EDA on activeness of beauty professional 	<ol style="list-style-type: none"> 1. Analyse the breakdown of bookings based on the day it is checked out. 2. Analyse the breakdown of individual services and the prices. 3. Analyse the campaigns that have the most redemptions to follow up on finding the ideal duration and discount a campaign code should have. 4. Analysis on the attractiveness of a professional. 5. Transforming and normalising the data to prepare it for further analysis. 6. Perform cluster analysis for customer segmentation. 7. Conduct survival analysis to find out the value of a customer and the effectiveness of a

<p>separately.</p> <ol style="list-style-type: none"> 13. Identifying the number of users for each user type. 14. Identifying the number of customers for each age and gender. 15. Identifying the number of professionals for each age and gender. 16. Identifying the number of services under different price ranges. 17. Identifying the number of services a professional typically offers. 18. Identifying the number of services provided in different categories. 19. Identifying number of campaigns for various range of duration, type, discount type, discount amount, usage and duration from start of a campaigns to the first booking made. 		<p>campaign.</p> <ol style="list-style-type: none"> 8. Written report, final presentation, poster, insights and recommendations to Vanitee.
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	------------------------------------------------------------------------------------------------------------------------------------------------------------

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	Planned	Actual																
	Gather Requirements	Planned	Actual																
Data Gathering	Gather Data from Sponsor	Planned	Actual																
Project Proposal	Refining Project Scope		Planned	Actual															
	Proposal Preparation		Planned	Actual															
	Wiki Page Preparation		Planned	Actual															
Research	Literature Research			Planned	Actual														
	Data Preparation			Planned	Actual														
Data Exploratory	Data Cleaning			Planned	Actual														
	Data Exploration			Planned	Actual														
Proposal Deadline - 15 January 2017 - Week 3																			
Data Exploratory	Generate Findings					Planned	Actual		Actual	Actual									
Interim	Sponsor Review and Preparation					Planned			Actual	Actual									
	Interim Report Preparation					Planned	Actual	Actual	Actual	Actual									
	Interim Presentation Preparation						Planned	Actual	Actual	Actual									
	Interim Wiki Update							Planned	Actual	Actual									
Interim Practicum Presentation - 20 to 24 Feb 2017 - Week 8																			
Model Building	Data Normalisation & Transformation									Planned									
	Cluster Analysis									Planned	Actual								
	Survival Analysis										Planned	Actual							
Insights & Recommendation	Create Visualisation from Analysis Results										Planned	Actual							
	Generating Insights										Planned	Actual							
	Formulate Recommendations											Planned	Actual						
Final	Sponsor Review																		
	Abstract & Full Paper Preparation														Planned	Actual	Actual		
	Final Wiki Update														Planned	Actual	Actual		
	Final Presentation Preparation															Planned	Actual	Actual	
	Final Paper Preparation & Submission																Planned	Actual	Actual
Final Practicum Presentation & Submission - 2 April & 16 April 2017 - Week 14 and 16																			
Conference	Undergraduate Conference on Data Analytics																		Planned

Legend	
	Planned
	Actual
	Milestone

Revised Work Plan

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	Completed
Data Exploratory	Data Preparation	1	4	15	15	All	Completed
	Data Cleaning	1	5	15	20	All	Completed
	Data Exploration	2	5	25	35	All	Completed
Proposal Deadline - 15 January 2017 - Week 3							
Data Exploratory	Generate Findings	3	8	20	30	All	Completed
Interim	Sponsor Review	4	8	5	6	All	Completed
	Interim Report Preparation	4	8	30	25	All	Completed
	Interim Presentation Preparation	5	8	20	20	Andrew	Completed
	Interim Wiki Update	6	8	20	15	Sarah	Completed
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

Our team will continue working on the project in the next few weeks by looking into revising our exploratory data analysis based on the comments our sponsor and project supervisor gave. Moving forward, we will then attempt to normalise and transform the data to prepare it for the building of our models.

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/>

TechInAsia. (2016, Jun 6). Singapore's Vanitee acquires startup to get beauty salons on its marketplace. Retrieved from <https://www.techinasia.com/vanitee-mefitted-acquisition>