TEAM V
ANLY482 SUPERVISOR MEETING
MINUTES 7

Date	2 March 2017
Time	17:00 - 18:00
Venue	SIS Meeting Room 4.5
Attendees	Team V: Andrew, Sarah Supervisor: Prof Kam
Agenda	1. Discuss some findings in EDA with supervisor 2. Update supervisor of revised EDA progress 3. Update supervisor of project progress

S/N	Item Discussed	Remarks
1	Discuss Findings - Campaigns	- Andrew showed findings about campaigns to Prof Kam. Every published campaign has at least 1 booking. - To see how the effectiveness of the campaigns, Prof Kam suggested the team to look into periods where there are campaigns versus the periods where there are no campaigns and see the difference in bookings. - If in the case where there are various campaigns running all the time, we will need some sort of control. - For this, Prof Kam suggested that we can look at periods where there are at least 3 campaigns versus periods with more than 3 campaigns. - Andrew then showed Prof Kam the chart on campaigns that are used and not used, however, the results were strange. - Hence, Prof Kam ask the team to clarify with the sponsor to get a clear idea of what is happening.
2	Revised EDA Progress	- Andrew showed Prof Kam the revised EDA for bookings. In the previous bookings EDA, we used the creation date of the bookings. But our

			sponsor wanted to see the bookings using the dates the bookings will be carried out on. Prof Kam acknowledges the charts.
			- Next, Andrew showed Prof Kam the booking by monetary value. They are broken down into intervals of $10 and $50. From $0 to $50, they are in intervals of $10 but above $50, they are in intervals of $50.
			- To address this, Prof Kam said that it is dangerous to do so. He said that the intervals should remain the same or it will be difficult to read the graph. He suggested to split the graph into two parts. We can do this in intervals of $50 first then breakdown into intervals of $10.
			- Next, andrew showed the charts for bookings by monetary value is split by services such as nails, makeup etc.
			- Prof Kam suggested that we can use a boxplot for each service.
			- Prof Kam directed us to plot a histogram using te nails data. In the histogram, we see some outlier values of above $700. Prof Kam directed Andrew to select on those values to find out what these data represent.
			- In the data table, we can right click on "selected" and click on "data view"
			- We also generated a distribution table from here.
			- Prof Kam and us found out that there are people who are willing to pay $40 to $300 for nail services.
			- Prof Kam mentioned that these people belong to the niche market, but we will have to look into these further to see that these are not nail packages and if they are legitimate bookings.
			- This is the top 25% (niche market) in which there are willing to spend.
			- Prof Kam suggested we can look into the niche market and segment the mass market which is the rest of the 75%.
			- Next, we looked at makeup service, and similarly, 25% Prof Kam said is the niche market, the rest belongs to the mass market.
			- He advised us to use the data filter the generate the graphs instead of saving it separately.
			- The charts for services breakdown by price are shown next by Andrew and Prof Kam suggested the same so that we can compare it with the

		customers. - As such, we can look to see if professionals that charge a high price will correlate to bookings made in the niche market.
3	Project Progress - Analysis	- Andrew raised some question on how we can do this, especially choosing the right attributes to use. - Prof Kam suggested we use the attributes from the RFM model and he said for now, we have chosen the the F and M parts but not the R (recency) part, so we can include it in. - We can also look at the services that are being used. - Prof Kam said that we will never know when a variable is suitable until we get the clustering results. As such, we will need to try several times. - There is a variable in the output of the analysis which will tell us if it played a dominant role in the clustering. - Andrew mentioned the two type of clustering that we can look into. K-means and Hierarchical. - Prof Kam said for K-means, the convergence will improve when more iteration are run. - Prof Kam mentioned that we can run both to compare the results. - Andrew mentioned to Prof Kam about our intention to try out Survival Analysis on campaigns effectiveness. However, our team maybe not have sufficient data to do this. Prof Kam said that it is more important to do a good job for our clustering segmentation. So, in this case, we may be able to remove doing survival analysis. - Instead Prof Kam said that we can look into Association Rule / Market Basket Analysis in which we can see what different service a customer will tend to book upon booking a service. (i.e. when a customer books a nail service, she will book a hairdressing service) - Through this, Vanitee could come up with packages for promotional efforts, or give a discount to customers buying these services. - Prof Kam said, based on current price, if 25% of customers are willing to pay, should we expand the category or the price? - Andrew brought up the topic about looking at

			factors that affect the customers from dropping off the app.
			- However, Prof Kam said it is quite impossible to do that at this point in time because of the lack of specific data. He said that as Vanitee does not have any complaints, feedbacks input or a measure to measure customer satisfaction, we are not able to build a model out of this.
			- Andrew suggested we can look at the patterns of customers booking using the app then moving on to manual. However, Prof Kam said that there is most probably insufficient time for us to do so. He would prefer us to focus on the clustering analysis.
			- For this, Prof Kam mentioned that we can propose to Vanitee on the specific type of data they can collect to do this.

S/N	Action Item	Action By	Deadline
1	Continue with EDA revision	Andrew, Sarah	By 21 Mar 2017
2	Attempt cluster analysis - pinpoint attributes	Andrew, Sarah	By 21 Mar 2017