SMU
SINGAPORE MANAGEMENT UNIVERSITY

IS434
SOCIAL ANALYTICS & APPLICATIONS

PROJECT PROPOSAL

PPROFESSOR : KYONG JIN SHIM (DR.)
SECTION : G1
TEAM : 6
TEAM MEMBERS : KANCHEN KUMMAR
LAU ZHAO YONG
PANG JIA HAO SHAWN
QUEK CHIEW XIA

What is the client?

a. Jetstar

What is the business problem?

a. Inability to analyse customer feedback based on historical posts for insights on how to improve service offerings & optimise marketing spend (on campaigns / influencers etc.)

What data will we be looking at?

b. Twitter feed - comments, likes, retweets
c. Instagram engagement - comments, likes, shares

What data do we need?

a. Details of your engagement data on various social media platforms
b. Other data that could supplement our analysis (e.g. API keys)

What analysis do we plan to do?

A. Text Analysis
B. Sentiment Analysis
C. Visualization (Word Cloud, Bar Charts)
D. Social network analysis
E. Competitor Analysis

Data Input/Sources

a. Twitter
   We will look at twitter data and crawl tweet text containing words related to "Jetstar" and competitors "Scoot" and "AirAsia" respectively using Twitter's REST API.

b. Instagram
   We will create scripts to crawl Instagram pages of Jetstar Asia, Airasia, and FlyScoot.

Detailed plan

The following steps apply to both Twitter and Instagram, although written in Twitter's context.

1. Cleaning tweet text
   Upon cleaning the tweet text (tokenising, stopping, stemming, spell-checking correction) and removing mentions "@" / retweet sign "RT @", we will do a wordcloud & bar chart to visualise the commonly occurring words in the tweet texts. We will also eye-ball the tweets to spot other positive or negative words that might have been missed.

2. Updating dictionary & Further cleaning of tweets
   For some of these words/phrases (from eye-balling the tweets and from wordcloud/bar chart), we will decide on the magnitude of score to assign to them (unigram &

multigram analysis) and the +/- polarity of the score (POS-Tagging), and add it to our dictionary.

From this analysis, we can identify tweets related to airport quality/services (e.g. airport food), and try to find patterns or words that indicate that the tweet is about the airport and not the budget airlines. For example, if a tweet contains both "airport" and "jetstar", we will take a sample of those tweets and eye-ball them for patterns or indicators of sentiments regarding the airport. Then, we will exclude the parts of a tweet (sentences/phrases) that contain sentiments on the airport.

3. Topic modelling with LDA & popular hashtags identification
   From tweets, we will categorise them into multiple sub-topics regarding flight punctuality? (punctual or delay), after-sales service, in-flight experience (seats & entertainment), in-flight experience (staff), in-flight experience (food), in-flight experience (other customers), pre-sales (check-in), general (for tweets that don't specify reasons), flight cancellations, pricing, online booking usability, pilot skills, etc. These topics can be further categorised into main categories such as "in-flight experience", "online experience", "after-sales", "promotions".

   We can also identify popular hashtags and classify them into topics they are related to, of which we can draw insights on future hashtags to use for maximum engagement.

4. Generating normalised sentiment scores for each topic & comparing with competitors
   Upon performing sentiment analysis and generating a list of sentiment scores, we will normalise the scores by dividing them by the number of text characters in each tweet (excluding stop words), and visualise the distribution of scores using bar chart/boxplot. These scores for each topic will be compared with competitors.

   If competitors' scores are less negative than Jetstar's in a particular aspect, we can find out areas of improvement where Jetstar is lacking in in the eyes of customers. Diving deeper into the topic of interest can give us insights into what kind of campaigns are competitors running - whether competitors' advertising success (in terms of consumer sentiment scores for the topic) is related to certain ways of forming emotional connection with customers, some promotions they're running, or external factors like PR disasters of other competitors.

   Conversely, if competitors' scores are more positive than Jetstar's in a particular aspect, this area becomes an opportunity for Jetstar to improve on. The focus of our recommendations will be on key areas where the magnitude of differences in sentiment scores differ widely between Jetstar and their competitors.

5. Hub identification - from popular tweets & Social network analysis
   From uncleaned tweets, we will identify mentions "@ ", retweets "RT @", likes, and comments. We will then generate a bar-chart to show the top 5 influencers who are the

most mentioned + retweeted from the tweets (containing words related to budget airlines' names).

6. Identifying & ascertaining causes from analysing frequent complainants / mentions
   From all the tweets mentioning any of the 3 budget airlines, we will identify "frequent flyers", those who either tweet about:
   - budget airlines most frequently
   - a particular budget airline (out of the 3) most frequently

   In these cases, we would examine their behaviour and identify if these are genuine reviews or paid endorsements/complaints.

   Depending on our findings, if we find that those who frequently mention a budget airline:
   (I) Are also an identified hub, Jetstar might consider engaging them as partners for paid endorsements, if budget allows for it.
   (II) Are frequently complaining about the budget airline but yet are a frequent flyer i.e. price-sensitive customers, Jetstar could consider sending them promotions more frequently to retain them.

What are the expected outcomes?

Primary Outcomes

A. Overview of Client's Social Media Engagement and Efficacy
   A summary of team's findings on your company's overall and individual social media platforms. It will provide an overview of which social media platform has a higher level of engagement and effectiveness. Additionally, suggestions will be given on how you can better utilise and strategize on the use of their social media platforms to drive a strong media presence and engagement.

B. Competitor's & Other Relevant Stakeholders Analysis

C. Prediction of Popular Topics
   From data obtained through crawling tweets / graph API data / YouTube API, we will draw insights from types of content and engagement - to predict what kind of content would generate more buzz in future. If needed, we could require exported API data from these sources for further analysis.

   For airlines specifically - present findings or information on tourists (e.g. likes and dislikes, what appeals to them etc.) by looking at other channels such as Expedia, which falls within the tourism industry

D. Consumer sentiments on different types of product offerings

With data on what consumers are saying about your brand, we can identify areas for product improvement for your company.

E. Identification of Influential Users to Increase Reach/Engagement for Jetstar
We can help you to answer questions like "Who can make or break our brand?" i.e who are the people who have such huge networks with your potential customers that you are not reaching out to? Jetstar could thus engage these influencers for paid endorsements to increase reach.

F. Identification of frequent flyer & their behaviours
From identifying frequent flyers & their behaviours, we can draw actionable insights on how to retain these groups of customers.

Additional Outcomes (If time permits)

A. Predictive models for identifying customers who are likely to churn
From historical data, we can identify switching behaviour of customers who flew with Jetstar but switched to other low-cost carriers (& vice versa), whether the reason is due to cost differences or not. Training and testing this data can help us identify unhappy tweet users who are likely to churn, and thus Jetstar could look into personalised messages / promotions to appease them if they are of high-value (frequent flyer).

B. Corpora for local languages (Singlish etc.) to increase accuracy of spell-checker corrections

C. Representing social networks in Treemaps