## **Supervisor Meeting Minutes 5**

|  |  |
| --- | --- |
| Date/Time | 27th February 2018 |
| Venue | SIS Meeting Room 4.3 (1 to 2pm) |
| Attendees | Prof Meena, Prof Kam, Eric, Ivan, Shing Hei |
| Agenda | Feedback from Interim Presentation |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Task/Description | Person in Charge | Due Date |
| 1 | **Feedback from Interim Presentation**  Facebook Post EDA  1. The data from KFC is not considered an outlier. It is actually a valid case. It is actually a good example to attract more interest from the Singapore population.  2. slide 20: Use Correlation analysis to understand if they(Likes, share and comment) are related. After which use Principal Correlation Analysis if the variables are the same. If it is the same, then we selectively use those that are unique  3. slide 20: Look by month and look at frequency distribution for each type (Link, Photo, Video). Use mean test for 2 groups. Use (Annova test) analysis of variance for 3 groups. In JMP, aggregate to month and change the representation to a box plot.  4. slide 22:  Analyse the photo group.  Analyse the video group.  Analyse the text group.  5. slide 23: Refer to point 3 above. Do a box plot and a test. The current representation might result in a false illusion.  6. slide 25: Refer to likes/view as a better metrics  Facebook Video EDA  Data Cleaning: F13-17 splitting on Java: can be done on JMP Pro  Slide 32: Number of videos published per day (for one year) Is there anything happening on the data? Esp for the sharp drop for video posted on Saturday. Day of posted -> how to justify the drastic drop in video posts on Sat?   * Wrong hypothesis: Number of views VS. Number published??? Viewer/ number of videos   Slide 35: Null values in Data Cleaning stage: have to decide whether we should go through the posts individually and label it ourselves. If it is too significant, choose not to do it?  Slide 36: Analysis is the same, compare the videos with diff content (how we can articulate it to sponsor)  YouTube EDA  Slide 41: Chi square test (for cross table analysis)  Slide 42: Box plot and do a test to confirm  Slide 43: Rely too much on descriptive and visualization for comparison. Data is skewed. Compare the distribution (mean and the median). 75% of people go more than which mark???  Instagram EDA  Slide 66: JMP has text analysis  Slide70: Quantity the confusion matrix so that we can tell them to pay close attention to confusion between different topics | All | After presentation |

**Upload Tableau Worksheets after presentation**