**CASE STUDY NOTES**

**Background Information:**

* **Shareholder Target:** 
  + Maximizing the number of annual memberships
  + Understand differences between casual riders and annual members
  + Design a new marketing strategy to convert casual riders into annual members
* **Rider’s Information:**
  + 8% riders – assistive options
  + 30% riders – commute to work
* **Customer Type:**
  + Casual Riders: single ride pass, full day pass
  + Members: annual pass

|  |
| --- |
| **Ask** |
| **Questions:**   * What is the problem you are trying to solve? * How do annual members and casual riders use Cyclistic differently? * How can your insights drive business decisions? * Understand customers’ behaviors could help us better our service according to the needs of casual members, which could increase conversion rate. * Why would casual riders buy Cyclistic annual memberships? * How can Cyclistic use digital media to influence casual riders to become members? |
| **Key tasks:**   1. Business task  * Data source: internal, previous 12 months of Cyclistic customer data. * Scope of work: analyze the customer historical data to identify the patterns or trends of membership riders and casual riders, and then identify their differences.  1. Key stakeholders  * Lily Moreno, the director of marketing * Cyclistic marketing team |
| **Prepare** |
| Questions:   * Where is your data located?   + Company internal data * How is the data organized?   + Monthly customer data stored in CSV files. * How did you verify the data integrity?   + Check how the data was generated and who is responsible for the data collection and management.   + Identify if there are bias when collecting the data |
| **Key tasks:**   1. Download the data 2. Identify how it’s organized, check the data features and fields. 3. Sort and filter the data 4. Determine the credibility of the data |
| **Process** |
| **Questions:**   * What tools are you choosing?   + Because the data is more than 800MB, it is very slow to process in spreadsheets, and thus I choose BigQuery to process the data. * What steps have you taken to ensure that your data is clean?   + I checked the unique value of each column and make sure there is no null value.   + I calculated the ride length column and make sure it’s not less than 0.   + Some other outliers (i.e. large ride length number to be discussed with operation team) * Have you documented your cleaning process?   + Yes. I have noted down all the SQL sentences in one document.   **Key Tasks:**   1. Check the data for errors 2. Delete the data which the end time is prior to the start time 3. Calculate the ride\_length column 4. Calculate the day\_of\_week column 5. Document all the cleaning and transforming process |
| **Analyze** |
| **Questions:**   * How should you organize your data to perform analysis on it?   + Combine the monthly data into one table and only keep the useful columns (ride\_id, rideable\_type, started\_at, ended\_at, member\_casual, ride\_length, day\_of\_work). * Has your data been properly formatted?   + Yes. Make sure the ‘started\_at’ and ‘ended\_at’ are stored in TIMESTAMP * What surprises did you discover in the data?   + The average ride length of casual members is longer than the member riders. * What trends or relationships did you find in the data?   + The ride length is longer at the weekends than in the workdays.   Key Tasks:   * Perform calculations * Identify trends and relationships using SQL and R |
| **Share** |
| **Questions:**   * Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?   + Yes. Casual riders tend to ride longer than the member riders.   + During the spring (Feb, Mar, and Apr), casual members ride more than the members.   + Casual riders use our service more during the weekend while member riders ride more during the workdays. * How do your findings relate to your original questions?   + There are some differences between casual riders and member riders. We can cater to the needs of casual riders to convert them to our long-term users. * Who is your audience? What is the best way to communicate with them?   + My audience is the executive team. I will use PowerPoint slides to present my results given the visualization is not very complicated. * Can data visualization help you share your findings?   + Yes. It can clearly show the trends and different users’ preferences.   **Key Tasks:**   * Use R ggplot2 for data visualization * Use Tableau to generate a dashboard * Use MS PowerPoint to present the findings |
| Act |
| Questions:   * What is your conclusion based on your analysis?   + Launch a season pass targeting casual riders using our service frequently during Spring   + More discount or promotion during the weekend for membership   + Change the single ride to 30-minute pass given the casual member’s average ride is more than 30mins so they can buy longer passes * How could your team and business apply your insights?   + Discuss the recommendation during the high-level meetings * What next steps would you or your stakeholders take based on your findings?   + Maybe run a small project to test whether the recommendations are feasible   Key Tasks:   * Wrap up the case study and upload to my portfolio * Practice the presentation |