Case Study : How do annual members and casual riders use Cyclistic bikes differently?

|  |
| --- |
| Case Study Roadmap - Ask |
| Guiding questions  ● How Does a Bike-Share Navigate Speedy Success ? |
| Key tasks  1. How do annual members and casual riders use Cyclistic bikes differently?  2. Stakeholders : Lily Moreno - manager |
| Deliverable  1. A clear statement of the business task  2. A description of all data sources used  3. Documentation of any cleaning or manipulation of data  4. A summary of your analysis  5. Supporting visualizations and key findings  6. Your top three recommendations based on your analysis |

|  |
| --- |
| Case Study Roadmap - Prepare |
| Guiding questions  ● Where is your data located?  Public data website/server  ● How is the data organized?  Data is organised chronologically, in csv format  ● Are there issues with bias or credibility in this data? Does your data ROCCC?  Data is not biased as it has all bike types and members types, as it is from the organization database , its credible. Yes Data is ROCCC  ● How are you addressing licensing, privacy, security, and accessibility?  Its under license and it’s from Motivate International Inc, different privacy levels are available in sheets and google sheets, google cloud provides security to data , whole data set has different level of accessibility , ex. Only one team can access as of now  ● How did you verify the data’s integrity?  As its downloaded directly from company’s server and stored in local and then in google sheets . so, there is no chance of data being altered.  ● How does it help you answer your question?  Provided data has sufficient data like start trip and end trip information, number of riders, time , date and membership type . So, we can calculate riding time, riding date using different columns (variables)  ● Are there any problems with the data?  Some values are missing/blank, columns are not consistent |
| Key tasks  1. Download data and store it appropriately.  2. Identify how it’s organized.  3. Sort and filter the data.  4. Determine the credibility of the data. |
| Deliverable  A description of all data sources used : CSV format data , saved again as EXCEL sheet |
|  |

|  |
| --- |
| Case Study Roadmap - Process |
| Guiding questions  ● What tools are you choosing and why?  Excel , looking at the size of data SQL/R is sufficient to prepare , process, Analyse, share  ● Have you ensured your data’s integrity?  Yes, since its directly from company’s database there is no change on data or update .  ● What steps have you taken to ensure that your data is clean?  Checked naming conventions, column names, removed incomplete data, checked for blanks, made column names consistent in all tables, added some new columns  ● How can you verify that your data is clean and ready to analyze?  By following checklist and reviewing change logs  ● Have you documented your cleaning process so you can review and share those results?  Yes |
| Key tasks  1. Check the data for errors.  2. Choose your tools.  3. Transform the data so you can work with it effectively.  4. Document the cleaning process. |
| Deliverable  Documentation of any cleaning or manipulation of data |

Steps of cleaning –

1. Renamed columns in q4\_2019, q3\_2019,q2\_2019 to the following

,ride\_id = trip\_id

,rideable\_type = bikeid

,started\_at = start\_time

,ended\_at = end\_time

,start\_station\_name = from\_station\_name

,start\_station\_id = from\_station\_id

,end\_station\_name = to\_station\_name

,end\_station\_id = to\_station\_id

,member\_casual = usertype

1. Converted ride\_id , rideale\_type to character type
2. Combined all data frames into one big data frames.
3. Removed columns start\_lat, start\_lng, end\_lat, end\_lng, birthyear, gender, "01 - Rental Details Duration In Seconds Uncapped", "05 - Member Details Member Birthday Year", "Member Gender", "tripduration
4. (a) In the "member\_casual" column, there are two names for members ("member" and "Subscriber") and two names for casual riders ("Customer" and "casual"). We will need to consolidate that from four to two labels.

(b) The data can only be aggregated at the ride-level, which is too granular. We will want to add some additional columns of data -- such as day, month, year -- that provide additional opportunities to aggregate the data.

(c) We will want to add a calculated field for length of ride since the 2020Q1 data did not have the "tripduration" column. We will add "ride\_length" to the entire dataframe for consistency.

(d) There are some rides where tripduration shows up as negative, including several hundred rides where Divvy took bikes out of circulation for Quality Control reasons. We will want to delete these rides.

|  |
| --- |
| Case Study Roadmap - Analyze |
| Guiding questions  ● How should you organize your data to perform analysis on it?  Aggregate data  ● Has your data been properly formatted?  Yes  ● What surprises did you discover in the data?  Casual riders are having mean, median, max, min of ride length more than Members, even on every day of the year but number of rides is less for casual compared to Members .  Even sum of ride length/duration is more for casual riders  ● What trends or relationships did you find in the data?  Member riders use frequently but for less duration whereas Casual riders use less but for longer duration. For example : Member riders used for 10 times and for 1 minute on each ride whereas Casual riders used for 7 times and each time for 3 minutes  ● How will these insights help answer your business questions ?  As I can see from data :  Member riders use for (less duration + a greater number of times) in a year ,  Casual riders use for (more duration + a smaller number of times) in a year. |
| Key tasks  1. Aggregate your data so it’s useful and accessible.  2. Organize and format your data.  3. Perform calculations.  4. Identify trends and relationships. |
| Deliverable  A summary of your analysis |

|  |
| --- |
| Case Study Roadmap - Share |
| Guiding questions  ● Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently? YES  ● What story does your data tell? Annual Members use frequently for short duration whereas Casual use for long duration but not frequently.  ● How do your findings relate to your original question? Findings shoe difference between casual and annual members in terms of ride duration and ride numbers  ● Who is your audience? What is the best way to communicate with them? My audience are stakeholders like immediate manager and other team members.  ● Can data visualization help you share your findings?  off course  ● Is your presentation accessible to your audience? Yes |
| Key tasks  1. Determine the best way to share your findings.  2. Create effective data visualizations.  3. Present your findings.  4. Ensure your work is accessible |
| Deliverable  Supporting visualizations and key findings |

|  |
| --- |
| Case Study Roadmap - Act |
| Guiding questions  ● What is your final conclusion based on your analysis?  Annual members ride for less duration but frequently whereas Casual Riders ride for long duration  But not frequently.  ● How could your team and business apply your insights?  Team and Business can come up with some new annual membership plans for these types of riders (casual riders)who ride less but for more duration.  So, if ***we target customers who ride for less duration but frequently*** and ***convert them*** into ***Annual members*** then it will be ***profitable***.  ● What next steps would you or your stakeholders take based on your findings?  Next step would be come up with a membership plan which attracts casual riders and profitable for Cyclistic  ● Is there additional data you could use to expand on your findings ? |
| Key tasks  1. Create your portfolio.  2. Add your case study.  3. Practice presenting your case study to a friend or family member |
| Deliverable  Your top three recommendations based on your analysis   1. Convert Casual riders to Annual rider : Come up with a membership plan to attract casual members, which is profitable for Cyclistic as well. 2. Target new customers : Members are those who ride less duration , less distance but frequently or on daily basis . so, we can connect win new audience who ride less duration but frequently like office goers and try to make them members. 3. We can target casual members with a campaign and attractive membership plans in the month of July , Aug, Sept that is 3rd Quarter |