CASE STUDY - How does a bike-share navigate speedy success?

Introduction:

This case study is the final part of the Data Analytics Professional Certification course provided by GOOGLE through coursera. Use previously learned skills from the course to solve the case study and experience the real-world problem as a data analyst.

We use six phases of data analysis such as Ask, Prepare, Process, Analyze, Share and Act to solve the business problem in this case study.

Scenario:

Cyclistic a fictional bike-share company in Chicago. The director of marketing believes the company’s future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members.

About Company:

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geo tracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. Until now, Cyclistic’s marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members. Cyclistic’s finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a solid opportunity to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Stakeholders:

Lily Moreno: The director of marketing and our team manager

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members.

Moreno has assigned you the question to answer: How do annual members and casual riders use Cyclistic bikes differently?

Dataset:

Use Cyclistic’s historical trip data to analyze and identify trends. “cycle\_trips\_2019q1.xlsx” and “cycle\_trips\_2020q1.xlsx” are the two datasets used in this case study. Access the dataset use this link <https://github.com/3103203/First-Data-Analysis-Case-Study> .

“cycle\_trips\_2019q1.xlsx” contains the first quarter of 2019 which have trip\_id, start\_time, end\_time, start\_station(name, id), end\_station(name, id), user type, customers gender and birth year columns. Sample data for reference:



“cycle\_trips\_2020q1.xlsx” contains the first quarter of 2020 which have ride\_id, rideable\_type, started\_at, ended\_at, start\_station(name, id, latitude, longitude), end\_station(name, id, latitude, longitude) and user type columns. Sample data for reference:



The datasets have a different name because Cyclistic is a fictional company. For the purposes of this case study, the datasets are appropriate and will enable you to answer the business questions. The data has been made available by Motivate International Inc. under this license. This is public data that you can use to explore how different customer types are using Cyclistic bikes.

Download and store the datasets in device and make a copy of it for further work and keep original dataset unchanged for future reference if we deleted some important data from the working dataset.

Both datasets are in .xlsx format which is excel file and arranged structured in columns and rows. We can use R programming to handle these datasets and also use Excel for simple tasks. If we use SQL we need to convert the format to .csv converting a .xlsx file to .csv some features may loss.

Data Cleaning:

Check missing values:

Check if there is any missing values in the dataset. In excel use filters to find the missing values. Open excel then select all data elements then click” sort and filter” on top right corner of the home menu from the drop down select filter. Now all columns have down arrow next to the headers click it which shows data that are present in that column check for “blank”. if there is blank unselect all other to view only blank cell to locate which row have missing value then same for all columns. Another way to check the blank space or missing values use “find and replace next” to “sort and filter” click it type what values to find in find text box then select find next to check all the values. If missing value found check if it is important data for the analysis. If yes then ask manager or data collection team to get a accurate data if no then remove the column. In this case study, 2019 dataset have missing values in two columns which is gender and customers birthday this information is customers personal information and not important to our analysis so we remove those columns. If the stakeholders are asking which age group of peoples ride most or which gender rides the most then we need to request data in missing cells from manager or data collection team. In 2020 dataset we have one missing row which have ride\_id, start\_time and end\_time all the other values are missing. The start time and end time are same in this row which means the trip may cancel, so we remove this row.

Check duplicate values:

Check for duplicate entries in the dataset. In excel use conditional formatting to highlight the duplicate entries. Click the conditional formatting in the top right of home menu in the drop down select new rules then select format only unique or duplicate values then choose colour and font and click ok. It highlights the duplicate entries then remove it. In our case study, there is no duplicate entries in both datasets.

Check irrelevant values:

Check for irrelevant values in the cell like in customer type column which is texts there mistakenly entered a numeric values. For this we use both filters to manually check, if any column has irrelevant data points or use conditional formatting to check, if any numeric values in text based columns by formula to format then enter ISNUMBER(range). Remove columns that are non relevant to business task and only keep important information for further analysis. In our case study station\_id, latitude, longitude, gender, birth year, ridable type are non relevant or not important to business task so we remove those columns. In 2019 we have total of 12 columns and we use 4 columns(ride id, start trip time, end trip time, user type). In 2020 we have total of 13 and we use 4 columns(trip id, start at, end at, user type).

Create new data for analysis:

Prepare dataset for analyse, create a column called total\_trip\_length for each ride. Calculate the length of each ride by subtracting the column started\_time from the column ended\_time (for example, =D2-C2) and format as HH:MM:SS using Format > Cells > Time > 373055. Then create a column called day\_of\_week, and calculate the day of the week that each ride started using the WEEKDAY command (for example, =WEEKDAY(C2,1)) in that 1 = Sunday and 7 = Saturday.

Analyze:

Identify trends and patterns in the customer type. we use user type, total\_trip\_length and day of the week to find which type of user use and how long they use by each day of the week. For this we use **pivot table** in excel select datasets click insert then click pivot table in top left and enter the range of value for pivot table then enter ok. From pivot table field drag the total\_trip\_length to the value(sum), user type to row and day of the week to column then use the numerical values in the pivot table to understand which user type used to ride most, when and how long. Now alter the table by change value settings to count for both years. A **pivot table** is a data summarization tool in Excel that allows you to quickly organize, analyze, and compare large datasets by grouping and aggregating values.

Pivot tables for 2019:

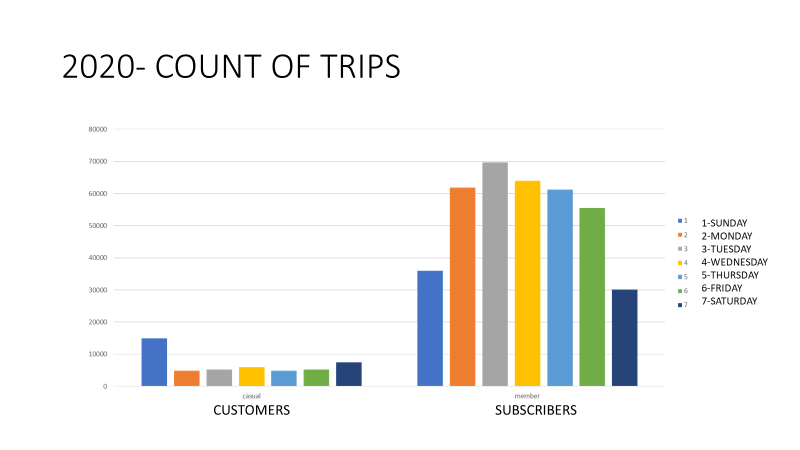
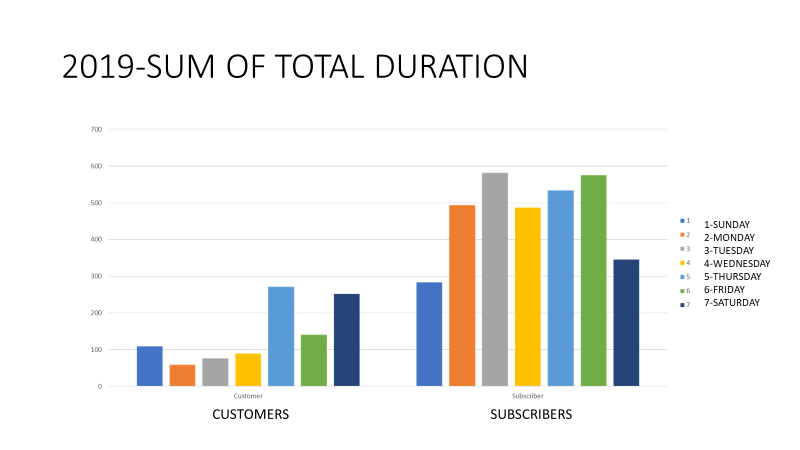
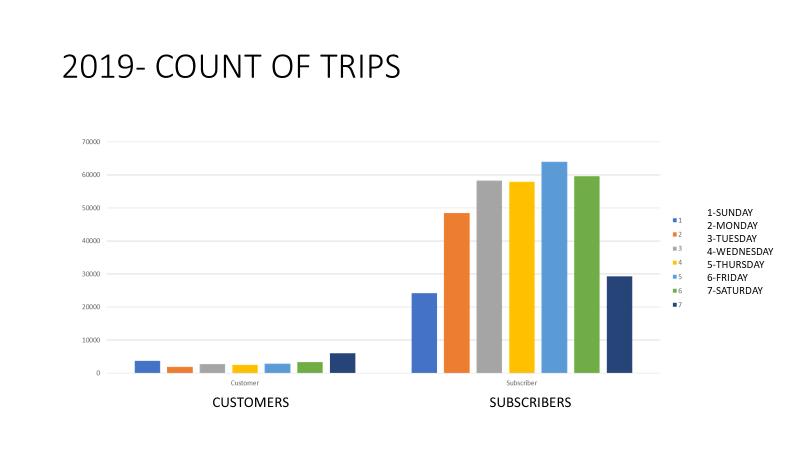
Below tables have total number of rides and total ride duration for both user type across all days of the week.

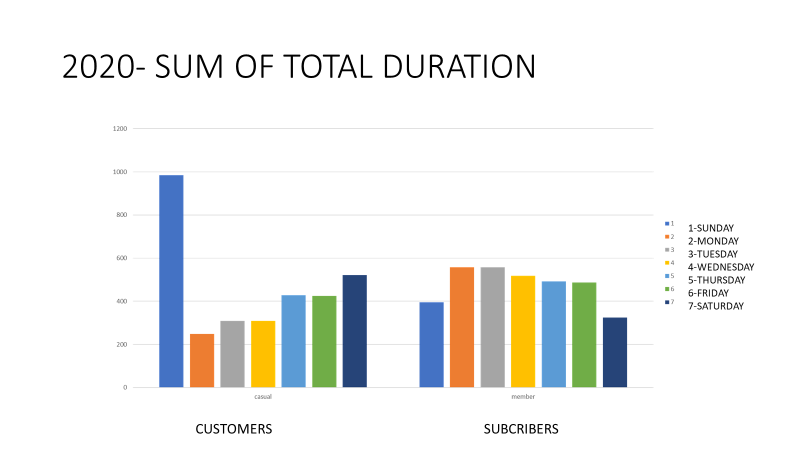


Pivot tables for 2020:



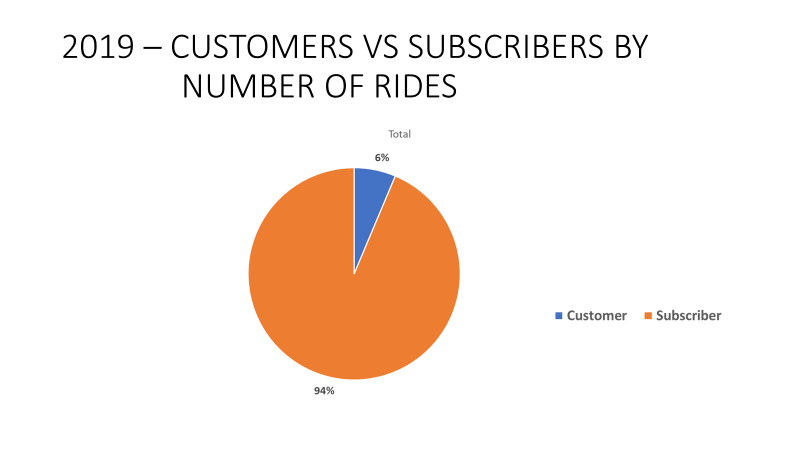
Charts created from these tables:

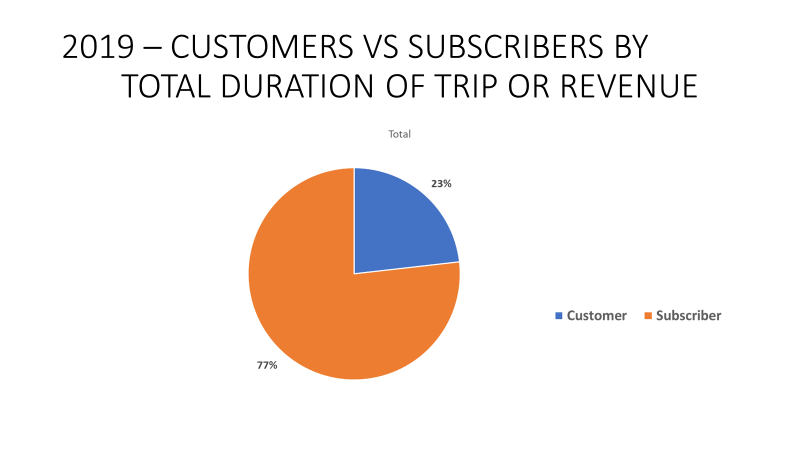


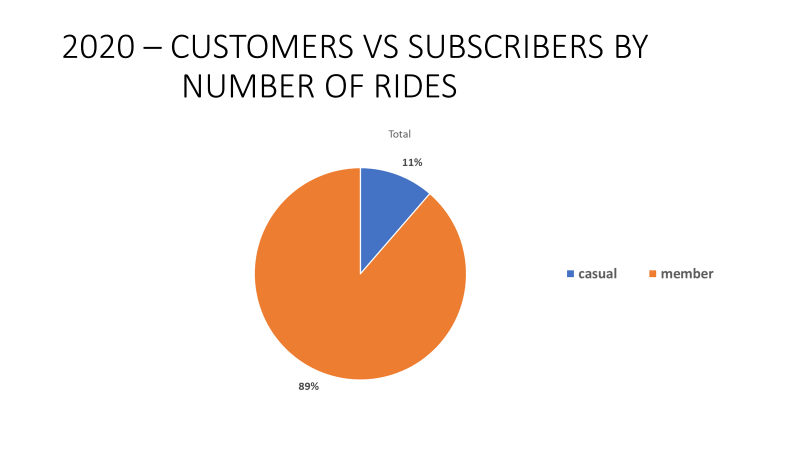


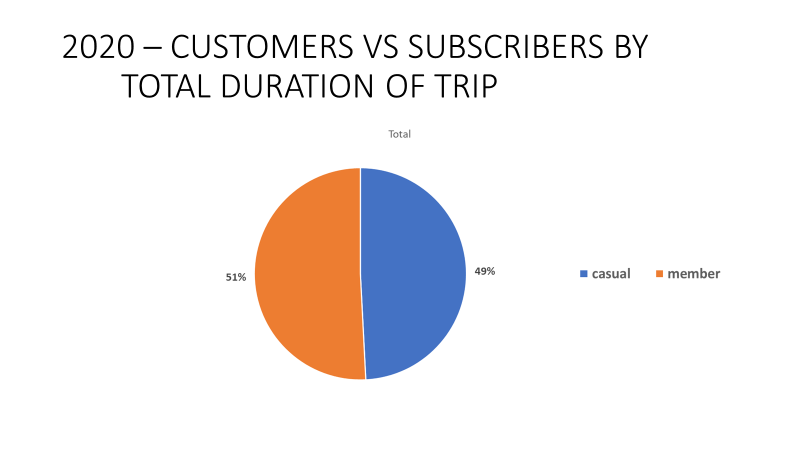
Then create a charts for the created pivot tables to understand the patterns visually. For our case study, we use clustered column chart to compare both type users. Casual riders are mostly ride in weekend and members are mostly ride in weekdays. Which means casual riders are occasional or leisure riders and members are mostly commute to works.

Then create charts for each year calculate how many casuals ride and members ride and how long the total trip duration for each category using bar charts. For this create pivot table drag customer type to row and drag total trip duration or ride Id to value and select “count” in the value setting to get a total number of casual and member rides. Then change the value settings to “sum” to get the total trip duration by each category.

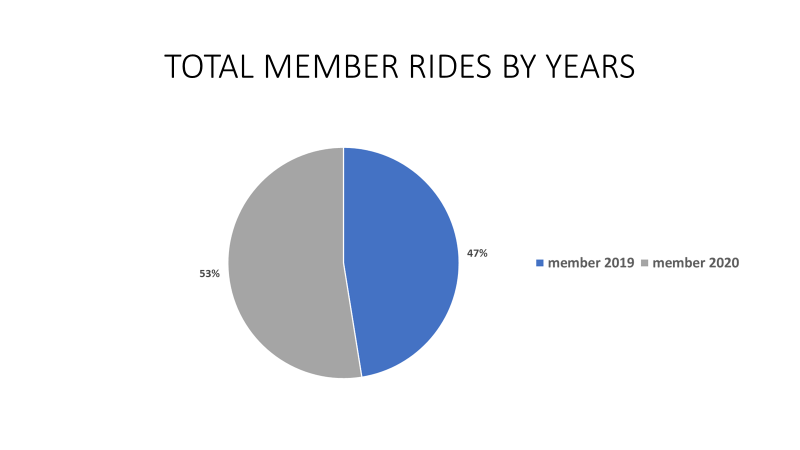


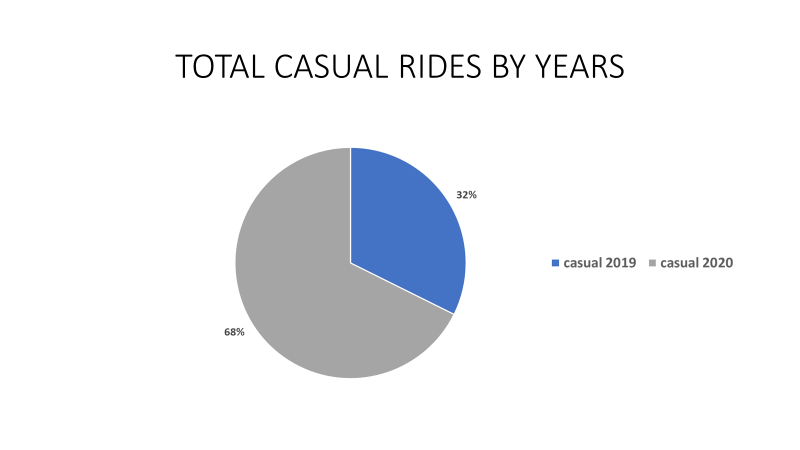






Analysis of ride data from the first quarter of 2019 and 2020 reveals a **6% increase in member rides** and a **36% increase in casual rides**. This growth highlights the rising popularity of Cyclistic bikes, especially among casual users. Pie chart for reference:





Share Findings:

To share the insights gained from the analysis, a **PowerPoint presentation** was created using the charts and graphs from the above Analysis. This presentation is designed to communicate the key findings and strategic recommendations clearly to Cyclistic stakeholders and the marketing team.

Github link for ppt: <https://github.com/3103203/First-Data-Analysis-Case-Study>

Conclusion:

**Capitalize on Casual Rider Growth**

The significant 36% increase in casual rides suggests a growing pool of potential members. Cyclistic should focus on converting these casual riders into annual members through strategic initiatives. These could include personalized ride summaries, in-app recommendations, and cost-saving comparisons that highlight how much users could save with a membership.

**Tailor Marketing by User Behaviour**

Members show more consistent weekday usage, indicating they often use the service for commuting. In contrast, casual riders dominate on weekends, likely using the service for leisure. Marketing strategies should reflect these patterns—offering commuter incentives or loyalty rewards to members to promote long-term engagement and boost retention, and introducing recreational packages, weekend promotions, and trial discounts or benefits for frequent casual users to encourage them to switch to annual memberships.

**Optimize Operations with Data-Driven Planning**

Cyclistic should adjust bike availability and station stocking using historical usage trends. This ensures bikes are reliably available during peak hours and popular days, improving the riding experience for both user types.