Logo, icon

Description automatically generated

**How does a Bike-Share Navigate Speedy Success?**

Report by: Farrukh Nizam Arain  
Designation: Data Analyst  
Date: 24th February 2022

**Introduction:**

Cyclistic is a cycle company operating from Chicago, USA. In 2016, it launched a successful bike-sharing offering for its residents. Since that year, it has grown to a fleet of 5,824 bicycles that a geographically tracked and locked into a network of 692 stations across the city of Chicago. The bikes can be unlocked from any of the station and then returned to any other station in the system at any time of the day.

Until now, the primary business strategy of the company was to build general awareness and appeal to the broad consumer segments within the city. One approach which was approved earlier to accomplish this strategy was to offer flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships.

Customers who purchase single-ride, and full-day passes are classified as casual riders. Customers who purchase annual memberships are termed as Cyclistic members. Additionally, most riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.

The finance department within the company have concluded that annual members bring much more profits to the company, rather than casual riders. Although the present strategy of the company is to attract more riders, Moreno (Director of Marketing & my manager) believes that maximising the annual members will be the key for the company’s growth. Rather than targeting all potential customers through the marketing campaigns, Moreno believes that it is a good chance to convert casual riders into annual members. She notes that the casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Three questions will guide the future marketing program:

1. How do annual members and casual riders use Cyclistic bikes differently?   
2. Why would casual riders buy Cyclistic annual memberships?   
3. How can Cyclistic use digital media to influence casual riders to become members?

**Problem Statement:**

The purpose of my assessment in this project is to find out how do annual members and casual riders use Cyclistic bikes differently? The result of the analysis will be used to design a new marketing strategy to convert casual riders into annual members.

**Preparation of the Datasets:**

I have downloaded public datasets at the following [link](https://divvy-tripdata.s3.amazonaws.com/index.html) provided by Motivate International Inc. under this [license](https://www.divvybikes.com/data-license-agreement). I have analysed and haven’t found any issues with bias and credibility with the data through the methodology of ROCCC

The data is organized with the tenures of months and quarters from 2013 till 2020 in zipped format. This data will guide me in differentiating casual annual members through comparison or different variables in the dataset.

Following at the columns (variables) and first ten rows (tuples) in the dataset for a sample:

A screenshot of a computer

Description automatically generated with medium confidence

Following are the key schematics of the dataset provided from Motivate International Inc:

* The dataset contains 13 columns in CSV format.
* The names and data types of the columns are:
  + ride\_Id (string)
  + Rideable\_type (string)
  + started\_at (Date and Time)
  + ended\_at (Date and Time)
  + start\_station\_name (string)
  + start\_station\_id (string)
  + end\_station\_name (string)
  + end\_station\_id (string)
  + start\_lat (whole number)
  + start\_lng (whole number)
  + end\_lat (whole number)
  + end\_lng (whole number)
  + member\_casual (string)

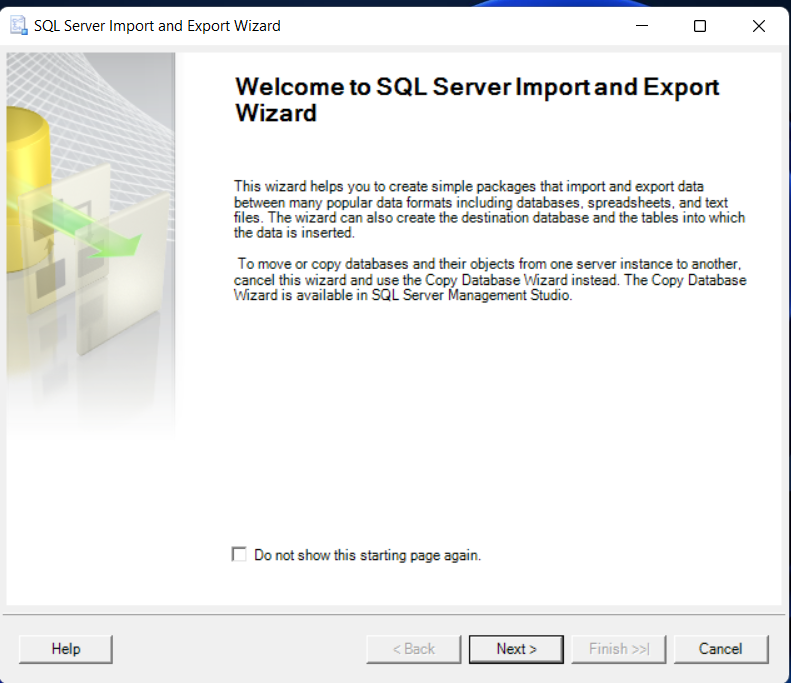
**Processing of the Datasets:**

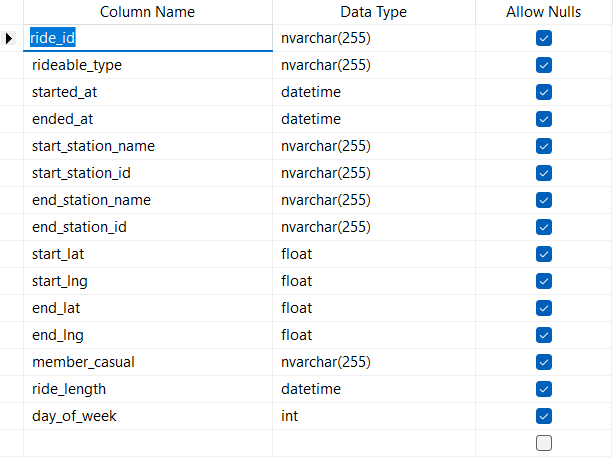
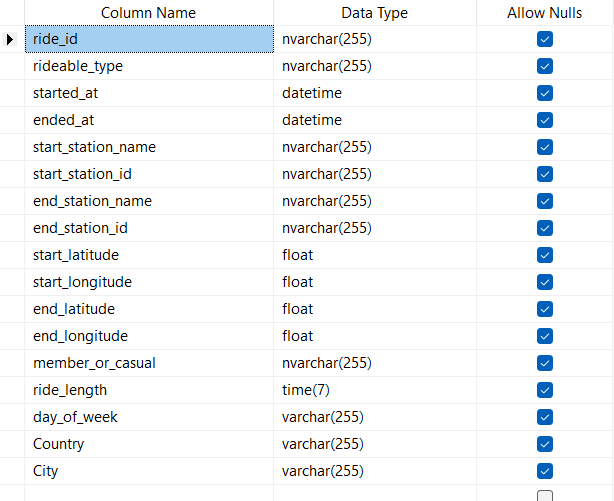
The tools that will be used for the data analytical process are:

* Microsoft Excel 365
* Microsoft Power Query
* Microsoft SQL Server 2019
* Tableau Prep
* Tableau Professional

Following are the steps taken for organization and cleaning of the dataset which was provided:

* Two folders were created for the 2021’s data. One is named as ‘zipped’ and the other is ‘unzipped’. The ‘zipped’ folder is used as a backup of the dataset, while the ‘unzipped’ is used for the actual working for analysis.
* The filenames of the dataset are converted from CSV to XLS to be a bit more compatible with the tools on which I will be working on. Additionally, it will help me to familiarize with the dataset and perform some cleaning activities with the Microsoft Excel’s intuitive GUI as well.
* In Microsoft Excel, a column was created called “ride\_length”. The length of each ride is calculated by subtracting the column “started\_at” from the column “ended\_at” (for example, =D2-C2) and formatting it as HH:MM: SS using Format > Cells > Time > 37:30:55.
* Another column was created, and it was called as “day\_of\_week”. It calculated the day of the week that each ride started using the “WEEKDAY” function (for example, =WEEKDAY(C2,1)) in each row. Additionally, the column was replaced to a string value of the day of the week, noting that 1 = Sunday and 7 = Saturday.
* All the datasets were imported using inbuilt ‘SQL Server Import and Export Wizard’



* After importing, I found out that the data types of the monthly data tables were different from each other. To resolve this, I corrected the data types in all the tables.
* In the next step, I used SQL queries to not only create a NEW table having structure of the existing tables, but also, I append (combined) all the monthly data tables into this new table which I called as ‘Cyclistic\_M\_2021’.
* For changing the column names, I used the same option which I used earlier for changing the data types of columns. Additionally, I created two new columns; ‘Country’ and ‘City’ and added relevant data for better spatial analysis in Tableau further in the project.  
    
  
* For removing the NULLS in the observations, appropriate SQL queries were executed to delete them safely to have a much more clean and accurate analysis in the next step.
* Before visualization and creation of dashboards, quick analysis was carried out to know a bit more about the dataset and how it can further help with the targeted marketing efforts. Following questions were answered using SQL query language:  
  + What are the number of casual and member rides throughout the year?
  + Maximum and Minimum length of rides through our services?
  + Number of rides throughout the week in the previous year?
  + Count of type of bikes used by both members and casual riders throughout the year and the week?
  + What are the top start and end stations used throughout the week?
* In the next step, the SQL Server’s table was connected via Tableau Prep for making it presentable before the actual process of visualization.

**Analysing & Visualization of the Datasets:**

In Tableau Desktop, a series of visualizations were created to answer all the queries pertaining to the previous year, that were asked at the start of the projects and were critically needed before the initialization of next marketing drive that the company is planning:

* Is there a route through which the overall progress can be viewed throughout the previous year?
* What are the top ten start and end stations which our riders preferred?
* What were the peak hours during which casual and members used our bike sharing service?
* During all the months, how many times both the type of riders availed our service?
* What were the average usage (in minutes) of rides by our customers?
* Among the type of bikes, which ones were most of the times opted by our riders?
* During the seven days of the week, which ones were the busiest for us?
* What were the average times for the usage of all types of our bikes?
* Is there a way to interconnect all the findings and trends?

**Sharing the Findings Found in the Datasets:**

After a thorough process of deep analyzing the data, following findings and trends were found:

* The overall business of Cyclistic Inc. skyrocketed during the tenure during the period of Summer till Autumn, but it did not help much because the only the number of casual riders were increased.
* Top 10 busiest start stations for casual and member riders were:

|  |  |  |  |
| --- | --- | --- | --- |
| Casual | Numbers | Member | Numbers |
| 1. Streeter Dr & Grand Ave 2. Millennium Park 3. Michigan Ave & Oak St 4. Clark St & Lincoln Ave 5. Wells St & Concord Ln 6. Clark St & Armitage Ave 7. Clark St & Elm St 8. Broadway & Barry Ave 9. Dearborn St & Erie St 10. St. Clair St & Erie St | 57,928  29,199  25,545  14,020    13,611  13,332  10,881  10,574  10,320  9,026 | 1. Dearborn St & Erie St 2. St. Clair St & Erie St 3. Clark St & Elm St 4. Wells St & Concord Ln 5. Broadway & Barry Ave 6. Clark St & Armitage Ave 7. Streeter Dr & Grand Ave 8. Clark St & Lincoln Ave 9. Michigan Ave & Oak St 10. Millennium Park | 16,489  16,207  15,550  15,330  14,986  14,065  13,879  13,646  12,177  7,383 |

* Top 10 Ending Stations for the casual and members are:

|  |  |  |  |
| --- | --- | --- | --- |
| Casual | Numbers | Member | Numbers |
| 1. Streeter Dr & Grand Ave 2. Millennium Park 3. Michigan Ave & Oak St 4. Theater on the Lake 5. Clark St & Lincoln Ave 6. Wells St & Concord Ln 7. Clark St & Armitage Ave 8. Wells St & Elm St 9. Clark St & Elm St 10. Dearborn St & Erie St | 60,592  30,513  27,053  15,618  14,130  13,475  13,169  10,768  10,595  10,012 | 1. Clark St & Elm St 2. Wells St & Concord Ln 3. Dearborn St & Erie St 4. Wells St & Elm St 5. Clark St & Armitage Ave 6. Clark St & Lincoln Ave 7. Streeter Dr & Grand Ave 8. Michigan Ave & Oak St 9. Theater on the Lake 10. Millennium Park | 16,656  16,321  15,610  14,484  13,131  12,847  12,285  11,180  9,852  6,797 |

* As far as the peek hours in between the customers are concerned, the peek hours for the casual members were in between **09:00 and 19:00,** with the highest average length of ride in minutes was **20.60 minutes**. As for the members of Cyclitic, the peak hours were between **12:00 and 21:00,** with the highest average length of ride in minutes was **12:80.**Chart, funnel chart

  Description automatically generated
* If we talk about the most rides availed throughout the last year, interestingly August was the peak month for both the type of riders. If we talk about casual and member riders, **341,476** and **332,933** rides were availed respectively.

Chart, funnel chart

Description automatically generated

* A clear difference was noticed when we analyzed the average length of rides from the riders. The highest peak for the members (13.26 minutes) was just **62.84 percent** of the peak for the casual riders (21.10).

Chart, line chart

Description automatically generated

* If we talk about the type of rides preferred by casual and member riders, both of them preferred classic bikes, rather than the modern electric ones.

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

* Usage statistics in a week differed between the casual and member riders significantly. Weekend was the preferred choice for the casual riders, while working / business days were preferred by the member riders throughout the year.

**Top Recommendations Via the Analysis:**

Following are the recommendation which will help the marketing department after analysis of the historical data provided:

* Marketing team needs to initiate a marketing campaign at the top 10 start and end stations to help them convert casual to member riders.
* As per the analysis of the dataset, the marketing campaign should be in between the working hours (08:00 to 18:00) for generation of new subscribers and to have a much more cost-effective and targeted use of resources.
* As per the August monthly data, almost **50 percent** of the total customers were casual members. So, it will be much more feasible to start the marketing campaign **from July till November** because during these months the number of users is higher than the rest of the months.
* The gap of average length of time for using our bike sharing service is huge in between the casual and member riders. So, some discount packages should be for the casual members during the peak times.
* Almost **7 percent** of the bikes were docked by the casual members, leading to wastage of operational costs. A time limit with a penalty should be introduced to cut down the number of docked bikes to **less than 2 percent.**
* As we know that casual members majorly use our bike sharing program during the weekends (Saturday and Sunday), so if the marketing department is short on budget or resources, then marketing campaign can also be run during those days as well.

**Conclusion of the Analysis:**

It was a great learning experience for me during this professional project, as it helped me not only to learn new things, but also to find interesting patterns and findings to help the stakeholders for planning the strategies.