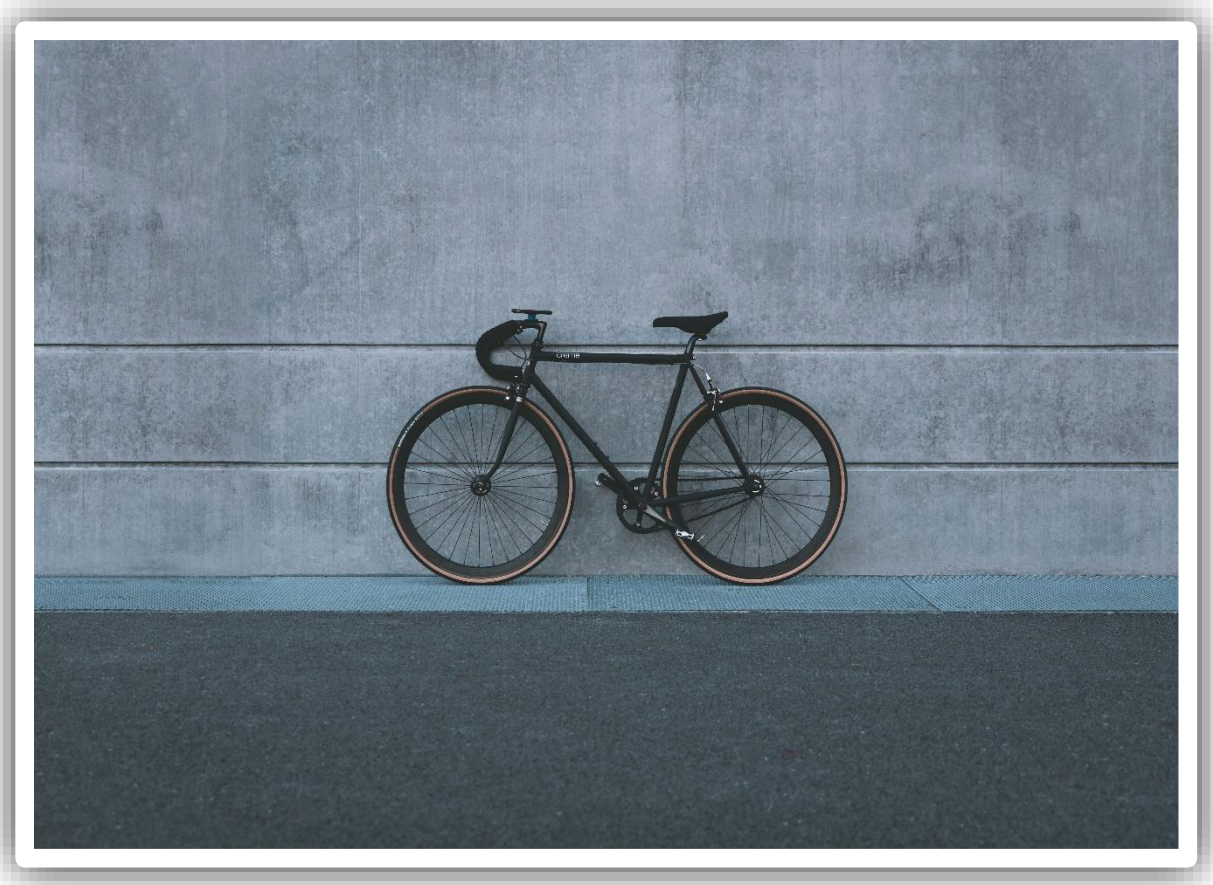
Case study: How do annual members and casual riders use Cyclistic bikes differently?



This Analysis is made By Si Mohamed Abdenour Email: asimohamed@edu.hse.ru

Case description:

a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, We need to undertand how casual riders and annual members use Cyclistic bikes differently. From these insights, my team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must be backed up with compelling data insights and professional data visualizations. But firstly the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership? So this case study is mainly about How do annual members and casual riders use Cyclistic bikes differently?

Data Preparation:

About data

We will use Cyclistic's historical trip data to analyze and identify trends.We Downloaded the previous 12 months of Cyclistic trip data here. (Note: 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 to answer the business questions. The data has been made available by Motivate International Inc. under this license.) This is public data to use in order to explore how different customer types are using Cyclistic bikes.

Getting prepared

- The Data that was downloaded is for 2019 for the fourth quarters of the year (4 delimiter files) and it was uploaded to Power Query to start processing the data
- The forth files was merged into one file called "Total_trips" through the Dax code
  Table.Combine({Divvy_Trips_2019_Q1, Divvy_Trips_2019_Q2, Divvy_Trips_2019_Q3, Divvy_Trips_2019_Q4})
- Total_trips was sorted by the starting_date

Data Process

All the processes that was applied on the Data was on Power Query using Dax Language

Data Intergity

- **Data Type**
  Changing Trip_duration column from String to Decimal in order to apply some operations on it later on
  = Table.TransformColumnTypes(#"Source", {{"trip_duration", type number}}, "en-US")

  Data Range
- for the trip_duration it should not exceed 24 hours(1440 min) as Max Value and should be more then 00.00 as Min Value
  = Table.SelectRows(#"Source", each [trip_duration] >= 0 and [trip_duration] <= 1440)

- For the Age of of the Bikers They Must be between 6 years and 99 years old, which is between 1920 and 2013 by the birthdate Language
  = Table.SelectRows(Source, each [birthyear] >= 1920 and [birthyear] <= 2019

Data Cleaning

- removing duplicates records
  = Table.Distinct(#"Source")
- Replaced the error values with Null
  = Table.RemoveRowsWithErrors(#"Source")

Data Analysis

- Splitting the starting_time of the trip into two columns Date and Time in order to make our analysis more accurate (the same thing for the arriving time)

  = Table.AddColumn(#"Filtered Rows2", "Time", each DateTime.Time([start_time]), type time)

- Also Adding the day name column is important for the analysis of the trips number by each day of the week
  `week name = FORMAT(Total_Trips[Week Day],"dddd")`

- Adding new column that will be calculated to get the Age of every Biker by subtracting the birthday value from 2019 in order to use the age values in more decent way

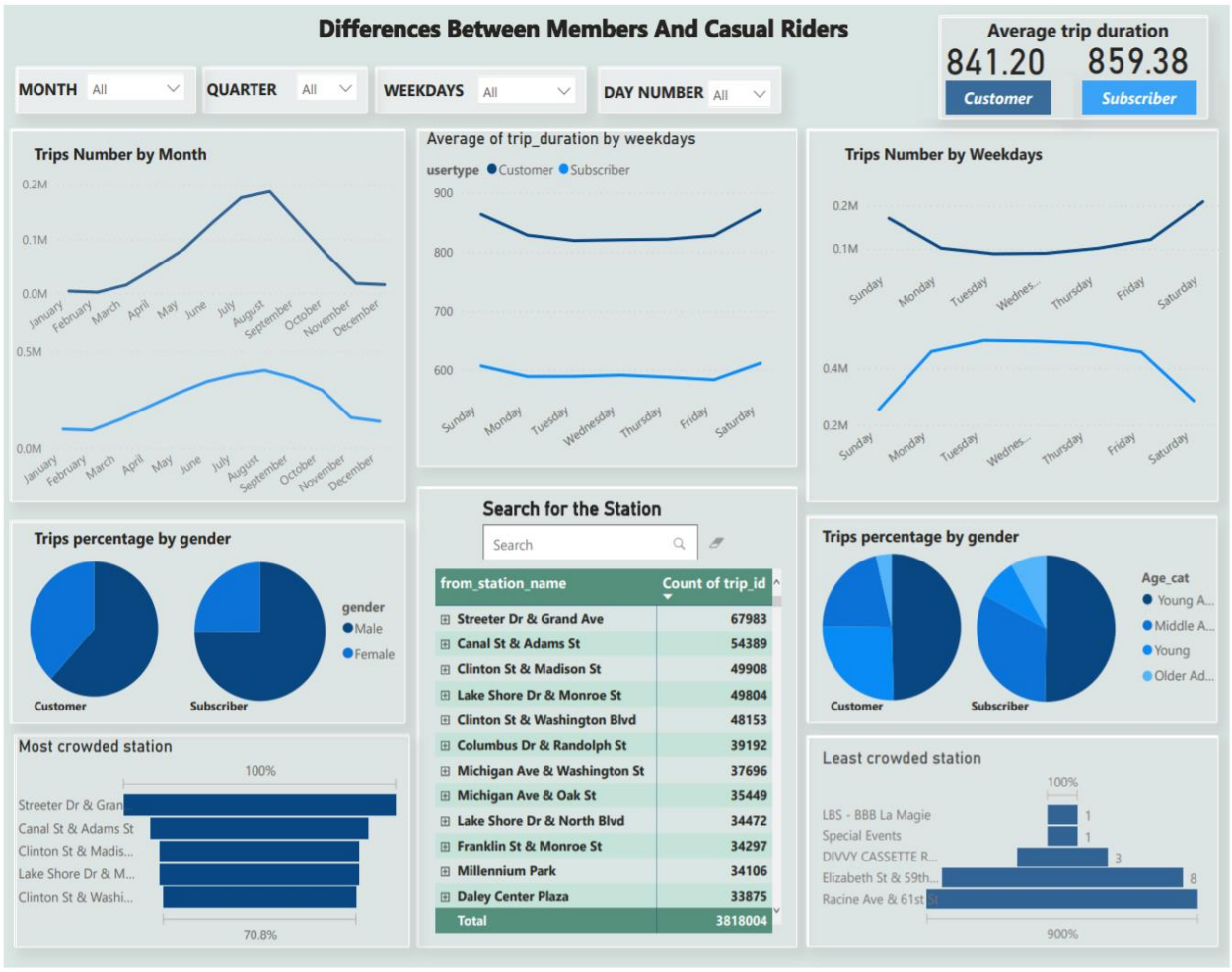
  = Table.TransformColumns(#"Source", {{"Division", each _ + 2019, type number}})

- grouping the age values into 4 Categories to be able to analyse the age factors since it impossible to analyse all the ages values

  = Table.AddColumn(#"Source", "age_category", each if [age] <= 24 then "Young" else if [age] >= 25 then "Young Adult" else if [age] >= 35 then "Middle Age" else if [age] >= 55 then "Older Adult" else null)

Insights

The insights are In PowerBi Dashboard form where it can be controlled by different factors such as Timing Factor



This dashboard include different Data Vizualization and Graph that are:

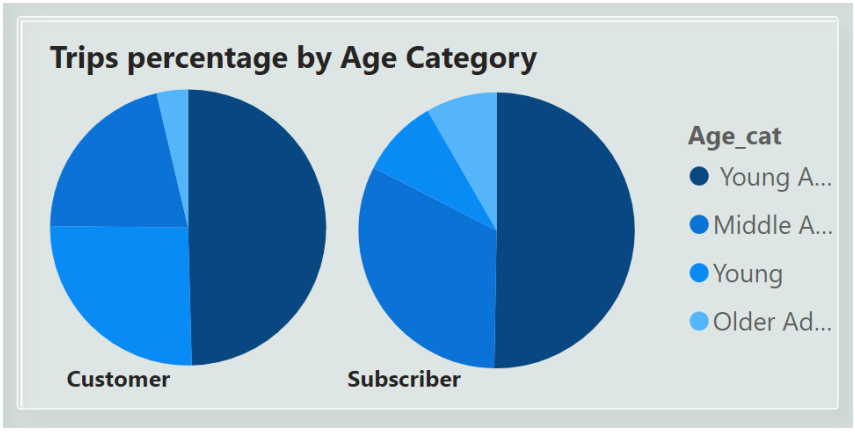
- Timing Bar that include Month, Quarter, Weekdays and Day Number

- Average Trip Duration for each Customers and Subscribers
- The amount of Trips by whole the Months of the year for Customers and Subscribers
- The Amount of Trips by the weekdays for Customers and Subscribers
- The trip Duration Average by weekdays
- Percentage of the Riders by Age category
- Percentage of the Riders by Gender
- Table of the departure Station Name that include inside the Arrival station Name when you click on the Plus Icon on The Left side (it ordered by the Most crowded to the last

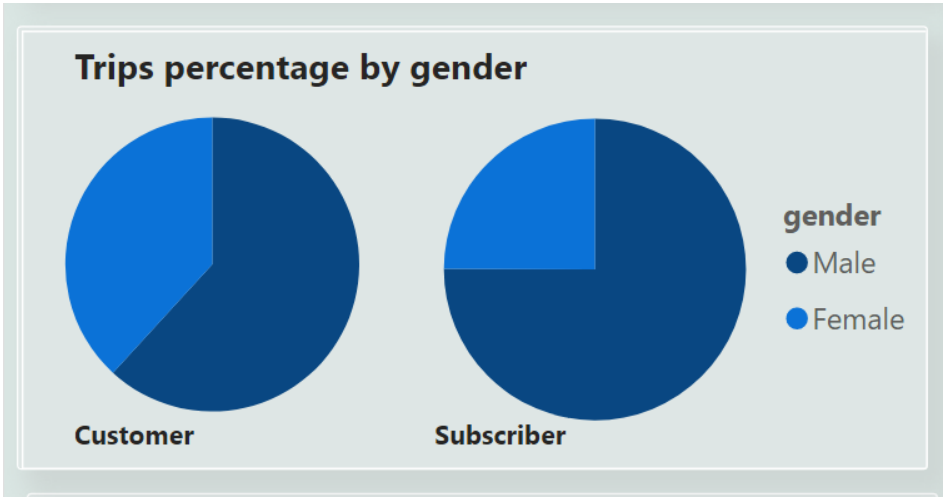
crowded, but when you click on The Arrow Icon inside the Count column, the Order will be reversed

- The 5 Most crowded Station
- The 5 Least Crowded Station

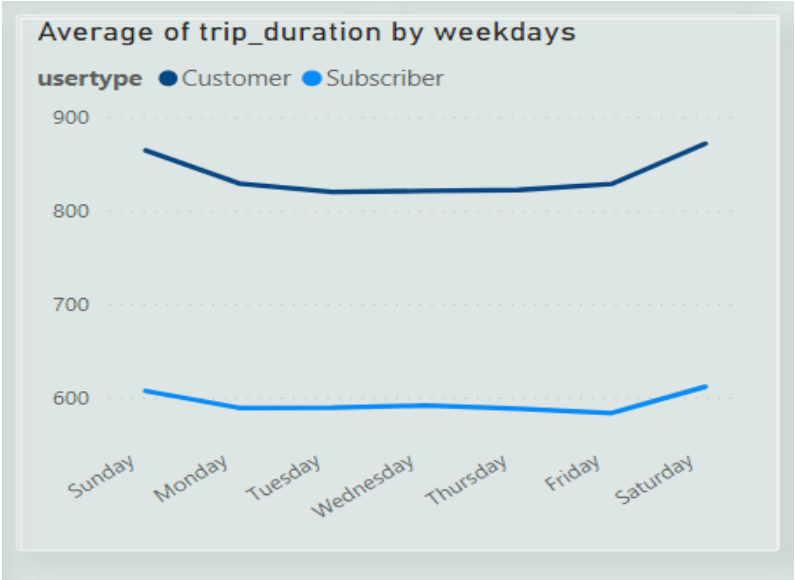
Findings and discoveries from the Dashboards



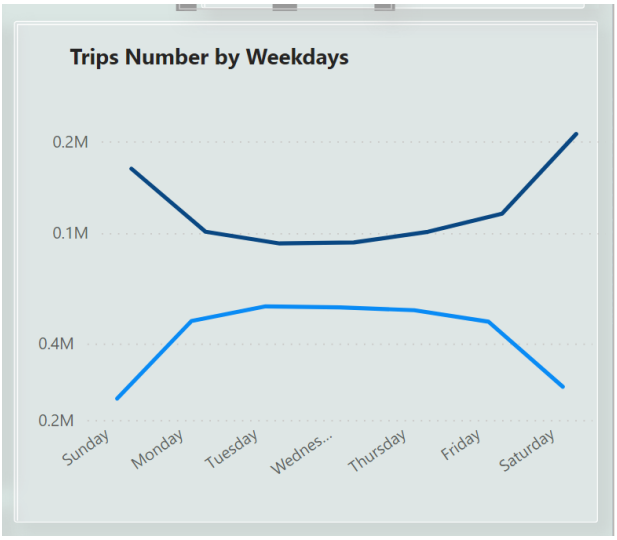
- Almost 50% for both Customers and Subscribers for Young Adult
- Young Category is Lower in Subscribers (9%), in the other side it 26% for Customers
- Middle Age is Higher for Subscribers (32%), Where it 21% for Customers
- Older Age is relatively Low percentage for Both Riders Type, 8% for Sub and 4% for Cus.



the Sub Males are more then the Cus Males with 75% Sub Males and 62% Cus Males, however in both types Male Riders are relatively more then the Female Riders



We can notice from this Graph that Average Trip Duration is Higher on Weekends, however it is the relatively the same on Working Days for both Rider Types.

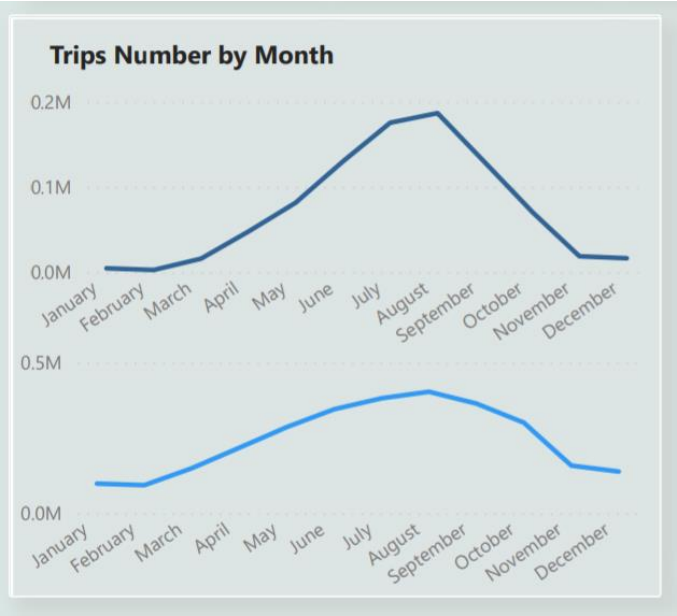


- The Trips number is remarkably higher on weekends then on working days for Customers
- Trips number is lower on weekends then on working days for Subscribers

Conclusion:

Subscribers tend to use Bicycles more on working days which give a sign that they use it to move to their work, study or any Activity that person can practice on Daily basis during working days

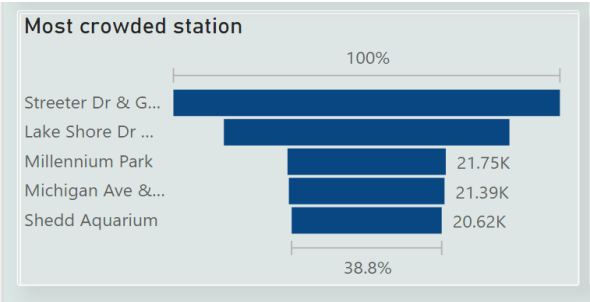
Consumers tend to use Bicycles more on weekends which can give a big possibility that they use it in their Free time and as sort of hobby as any other hobbies that person do it on weekends and Freetime

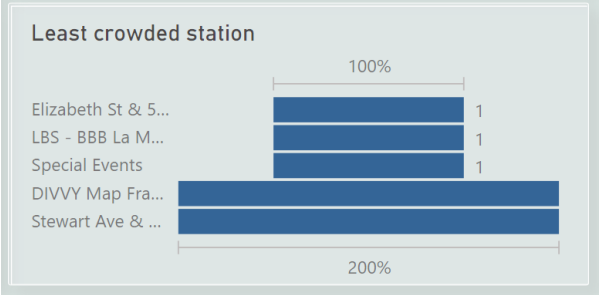


- Both Rider Types use the bicycle during Summer as highest peak, with less degree for the Spring and Autumn
- Where we see the using is on lowest level during the Winter
- This show that the use is more when the weather is tend to be warm and vice versa
- The curve tend to be show big difference between the Summer the other seasons for the Customers comparing with the Subscribers Where it relatively Higher, However the Gap is not that much big!
- That confirm the previous Hypothesis that I said where I supposed that Consumers Ride usually in their Free time where in the summer it also considered as a Holiday that usually employees and Students.. take their Annually Holiday during this season

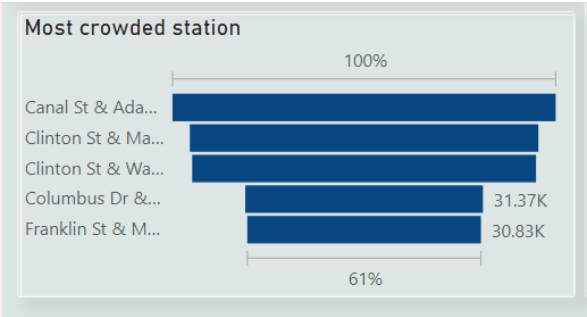
In the end let’s take a look of the most and the least crowded departure stations for both Cus and Sub

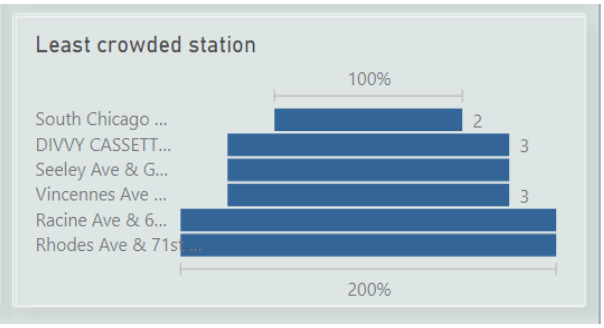
Let’s first start with Customers





And for Subscribers





If we try to compare the stations between the two types we Actually do not find any connection or Link between them which give a sense that the location of the station does not play a role to make a difference between Consumers and Subscribers

Final Conclusion

- Young Subscriber percentage (From 6 to 24 years) tend to use bicycle more then Young Consumer percentage, while it the opposite for middle age where it more percentage for the Cus and less for Sub, however The Aged Adult and Young Adult tend to have close proportions for both Cus and Sub
- Sub use the Bicycle for his Daily Activties (except weekends) while Cus use it on his Free time which is during Weekends
- Both Sub and Cuz tend to Bike when the weather is clear and warm and they avoid using the bicycle when it cold!
- Typically the Male Gender use the Bicycle more then the Female, However the proportion is Higher for the Male Subscriber then the Male Consumer