CASE STUDY

HOW DOES A BIKE-SHARE NAVIGATE SPEEDY SUCCESS?

Analysis by Sajith M

INTRODUCTION AND MOTIVATION

My name is Sajith M and I am a graduate in computer science from University Of Calicut and currently I am seeking opportunities to show case my skills at reputed organizations. During my college days I became very Interested in Data science and Data Analysis and decided to make a career in these fields. I really like the concept of data and how much of today's world revolve around them, when I more looked into it I saw these fields of Data Science and Data Analysis as the major areas that studies data.

So, I took upon a course on Data science about 6 months and completed it and also took on a Google certified course at Coursera to further increase my knowledge, I make these case studies to further increase my knowledge as self-improvement is one way of achieving my dreams.

BUSINESS PROBLEM AND DATA

You are a junior data analyst working in the marketing analyst team at Cyclistic, 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, 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. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations

You will use Cyclistic's historical trip data to analyze and identify trends. Download the previous 12 months of Cyclistic trip data here

DATA FEATURE ENGINEERING

- Imported libraries
- Collected the raw data of the previous years and named it according to the year wise
- When I checked the data I saw some of the column names are dissimilar as year wise the formats are changed
- So, I transformed the data by making the column names same, and making the datatypes same

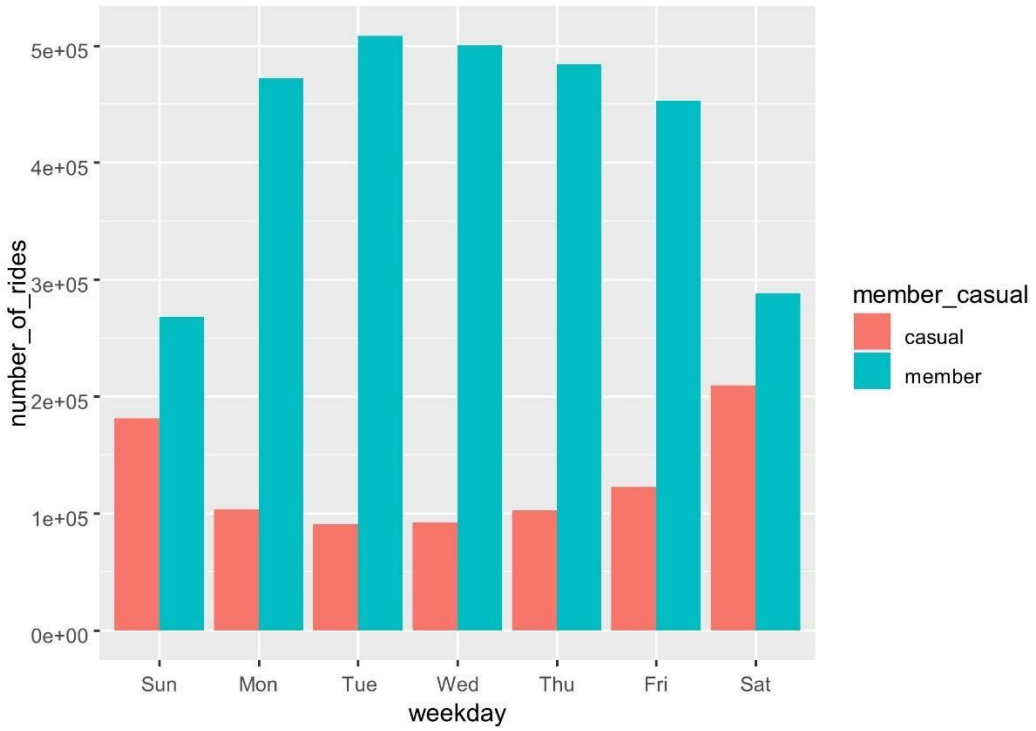
- I also removed some of the column names as they are not consistent across all years
- Now, I combines all the years data into a single dataframe for analysis
- The types of names in member_causal column is different ad there are are names such as “member” and “subscriber” I canged all of therm into “member” and “casual”
- Added a new column names such as date, month, day, year and ride length columns
- Converted the ride length column data type which is in factor to numeric for analysis
- Dropped the rows containing the null values as some of the column values t are not entered by the company

ANALYZE PHASE

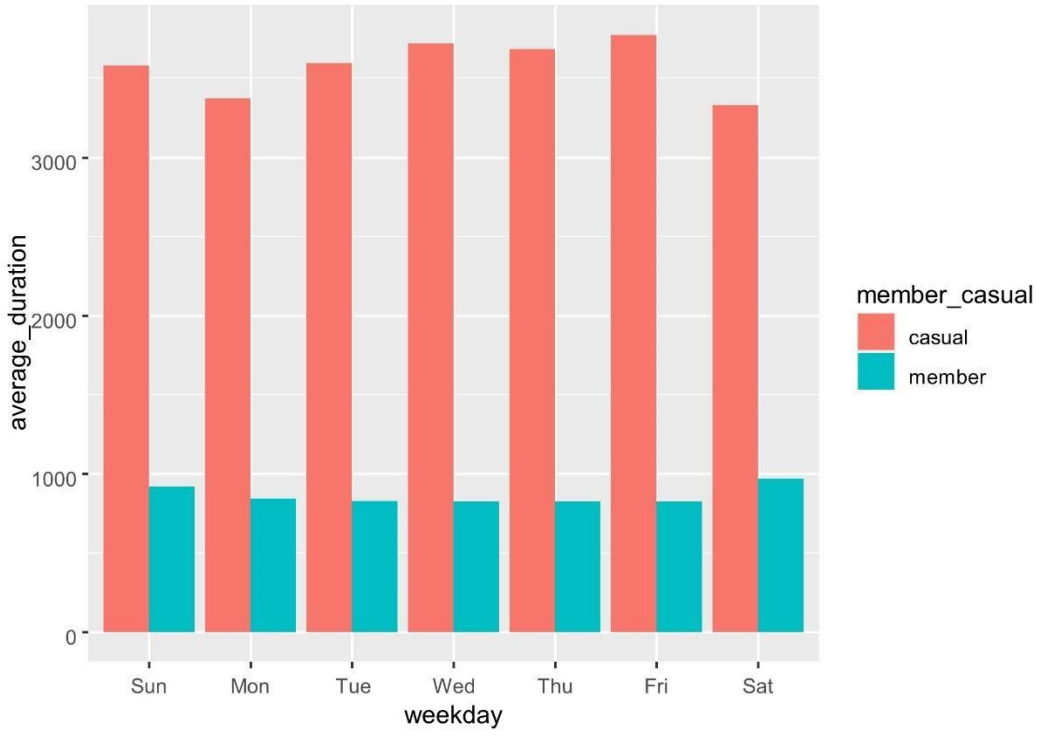
- The mean of the ride length is 1477.691
- The median of the ride length is 712
- The max of the ride length is 9387024
- The min of ride length is 1
- On the casual and members types of customers comparison the mean casual is 3552.7502 and member is 850.0662 aggregating with ride length
- On the casual and members types of customers comparison the median casual is 1546 and member is 589 aggregating with ride length
- On the casual and members types of customers comparison the max casual is 9387025 and member is 9056634 aggregating with ride length
- On the casual and members types of customers comparison the min casual is 2 and member is 1 aggregating with ride length
- Now, aggregating the ride length with day of week and member and casual we see that the order of day of week is different
- By changing the order and aggregating again, we see that on “Friday” the casual rates are high followed “Tuesday” where as in member the rates are high in “Saturday” followed by “Sunday”
- On number of rides and average duration the members are high in “Tuesday” followed by "Wednesday" and casual cyclers are high in “Saturday” followed by “Sunday”

VISUALIZATION

- Let’s visualize the number of rides by rider type



- Let's create a visualization for average duration



- The staff can increase the number of cycles on weekday except Saturday and Sunday as most of the member type cyclists are on these days where as the casual cyclists are somewhat consistent except on Saturday and Sunday

- But in average duration the casual cyclists are high might be because they tend to go on longer rides as they are coming only in some days where as member cyclists tends to

go on same distance everyday, average duration of both type of cyclists are somewhat consistent on each on ever weekday