****

**Google Data Analytics Course Capstone Project – Cyclistic Case Study**

**INTRODUCTION**

I recently completed the Google Data Analytics Professional Certificate course offered and Coursera which had eight different courses in total. In the course, I learnt about the six stages of the data analysis process; A**sk, Prepare, Process, Analyze, Share, & Act,** and how to use tools like**Spreadsheet, SQL, Tableau, R** to explore an ultimately gather insights from data

The last course, represents the Capstone Project which Google recommends that students complete to showcase the skills gained in the course.

**BACKGROUND**

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked 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 very good chance 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.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends.

I will be using the 6 phases of the analysis process (Ask, Prepare, Process, Analyse, Share and Act) to help guide my analysis of the datasets.

**PHASE 1 ASK**

**Key tasks**

1. Identify the business task

2. Consider key stakeholders

**The Business Task:**

Analyse the Cyclistic data set for the last 12 months to understand how annual members and casual riders use Cyclistic bikes differently.

**Stakeholders:**

**Lily Moreno:** The director of marketing. Moreno is responsible for the development of campaigns and initiatives to promote the bike-share program.

**Cyclistic marketing analytics team:** A team of data analysts who are responsible for collecting, analysing, and reporting data that helps guide Cyclistic marketing strategy.

**Cyclistic executive team:** The executive team will decide whether to approve the recommended marketing program.

**PHASE 2 – PREPARE**

**Key tasks**

1. **Identify the data source:**

**Dataset:** Cyclistic’s historical trip data.

The data used for this study was made available by Motivate International Inc. under this [license](https://ride.divvybikes.com/data-license-agreement).

Please click [here](https://divvy-tripdata.s3.amazonaws.com/index.html) to access the data. A total of 12 csv files representing a one-year period were downloaded for the analyses.

The period under review was between May 2021 to April 2022.

1. **Determine the credibility of the data:**

The “ROCCC” system was employed determine the credibility and integrity of the data.

**Reliability:** This data is reliable as it was gathered by Cyclistic and made available by Motivate International Inc. under this [license](https://ride.divvybikes.com/data-license-agreement).

**Originality:** This is an original dataset as it was collected directly from the casual riders and annual members.

**Comprehensiveness:** This data is comprehensive enough for the defined business task. However, it could have been better if information on age of riders where provided to guide on for further exploration and recommendation by age category.

**Current:** This data was collected in 2021 and up to April 2022 which means it is current.

**Cited:** As earlier mentioned, the data was made available by Motivate International Inc. under this [license](https://ride.divvybikes.com/data-license-agreement)

The data integrity and credibility will sufficiently provide reliable and comprehensive insights to Cyclistic. Hence, we can proceed to process and analyze the data.

1. **Sort and filter the data**

This part was completed using **RStudio due to its capacity to handle large datasets.**

The following R packages were loaded and installed for my analysis.

install.packages("tidyverse")

library(tidyverse)

install.packages("lubridate")

library(lubridate)

install.packages("janitor")

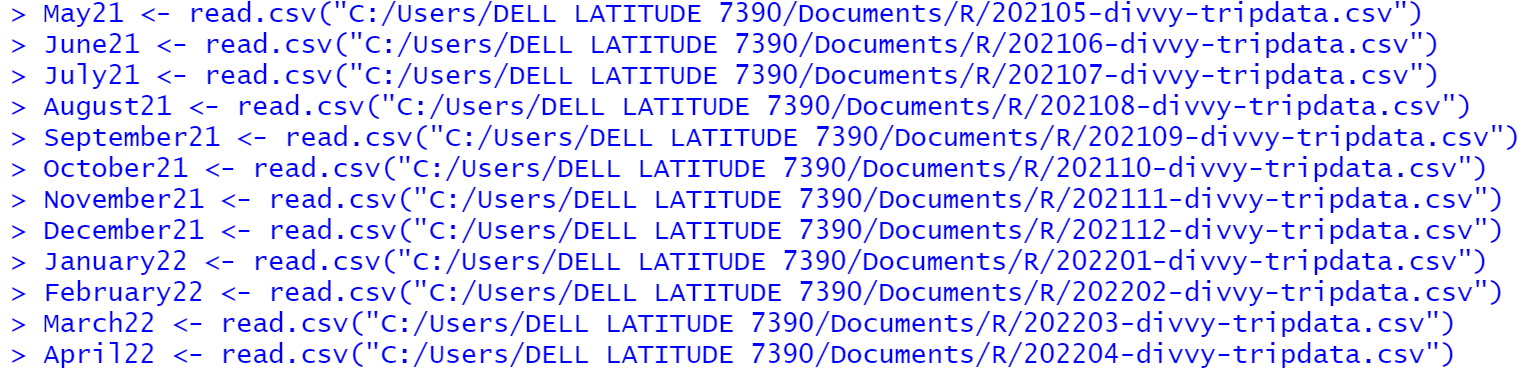
library(janitor)

install.packages("skimr")

library(skimr)

**DATA IMPORTATION**

The data set for each month between May 2021 to April 2022 was imported using the read.csv function

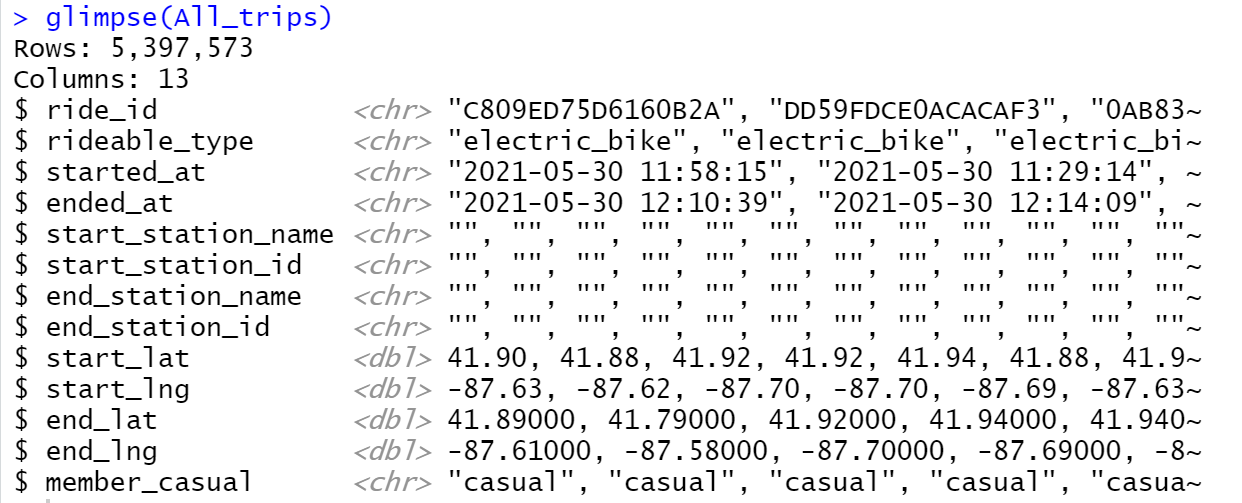


**MERGING THE DATASET**

Since the data for each month all have the same number of columns, I decided to use the *rbind* function to unite all the rows into a single table called *All\_trips*.



I used the *glimpse()* function to review the data summary



The merged dataset, *All\_trips* has 5,397,573 trip records and 13 fields.

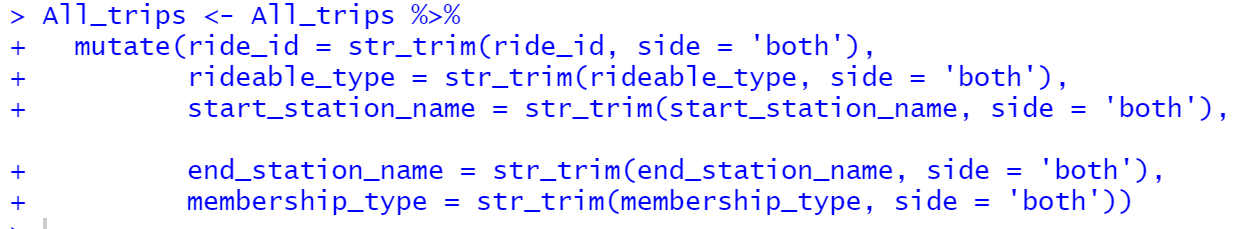
**PHASE 3 – PROCESS**

**Data Cleaning Process:**

1. Renaming of the *“member\_casual”* column to *“membership\_type”*

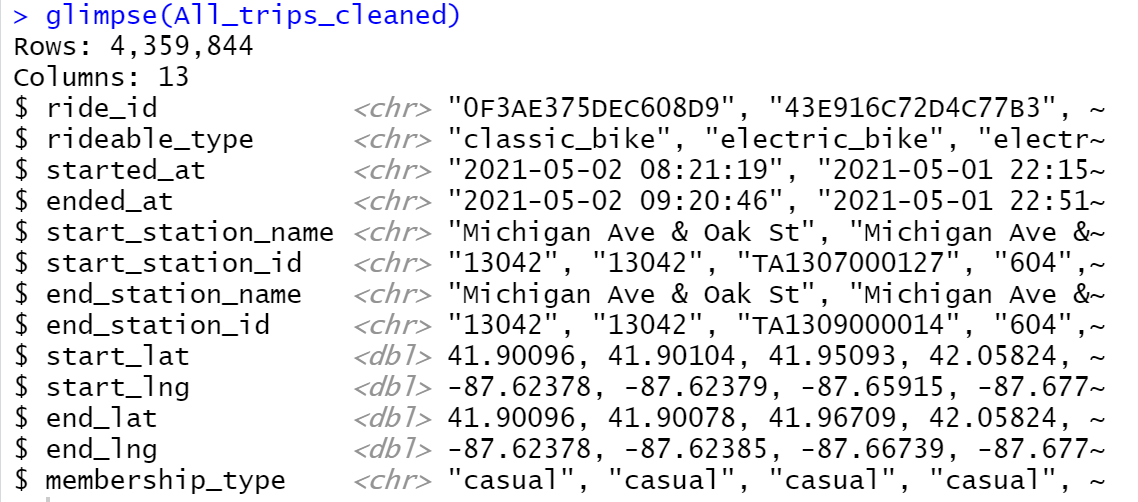


1. Removal of leading and trailing white spaces



1. Expunging records with blank cells in the start\_station\_name and end\_station\_name

All\_trips\_cleaned <- All\_trips %>% filter(!(start\_station\_name == ‘’ | end\_station\_name == ‘’))

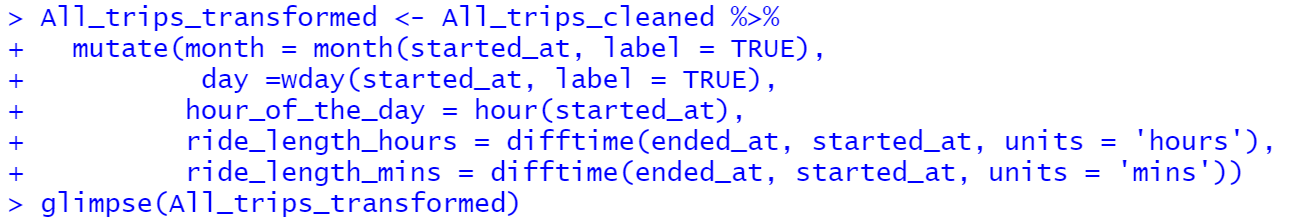


**Data Transformation:**

1. A new column name trip\_route is formed from the start and end station name

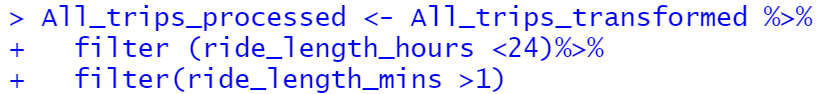


1. Creating new columns that show the months, days of the week, hour of the day, ride\_length in hours and in minutes.

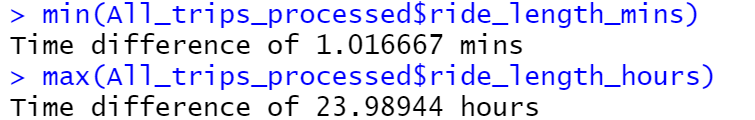


1. Some of the ride\_lengths are less than a minute while other are more than 24 hours.

This will be all be removed for consistency and integrity of the data.



For confirmation, I ran the *min* and *max* functions on the ride\_length columns.



This ends the Data Processing steps in R. Next, I’ll export the data into single CSV file for Analysis and Visualization in Power BI.

write.csv(All\_trips\_transformed, 'C:/Users/DELL LATITUDE 7390/Desktop/Google Data Analytics/Capstone/Bike\_trips.csv')

**PHASE 5 – ANALYZE**

Now we have 19 columns and 4,359,844 rows.



To commence the analysis phase, I took the following steps below for discovery of possible trends and patterns

* I categorized the days of the week as either Weekend or Weekday using DAX in power BI



* I also categorized the months of the year into seasons to understand each season’s peculiarity:
* [Spring](https://www.metoffice.gov.uk/learning/seasons/spring) (March, April, May)
* [Summer](https://www.metoffice.gov.uk/learning/seasons/summer) (June, July, August)
* [Autumn](https://www.metoffice.gov.uk/learning/seasons/autumn) (September, October, November)
* [Winter](https://www.metoffice.gov.uk/learning/seasons/winter) (December, January, February)



**Further cleaning**

* Changed the column *rideable\_type* to *bike\_type*
* Changed the column *started\_at* to *start\_time*
* Changed the column *ended\_at* to *end\_time*
* Rounded off *ride\_length\_hours*
* Rounded off *ride\_length\_mins*

Then I created some measures

1. Total trips



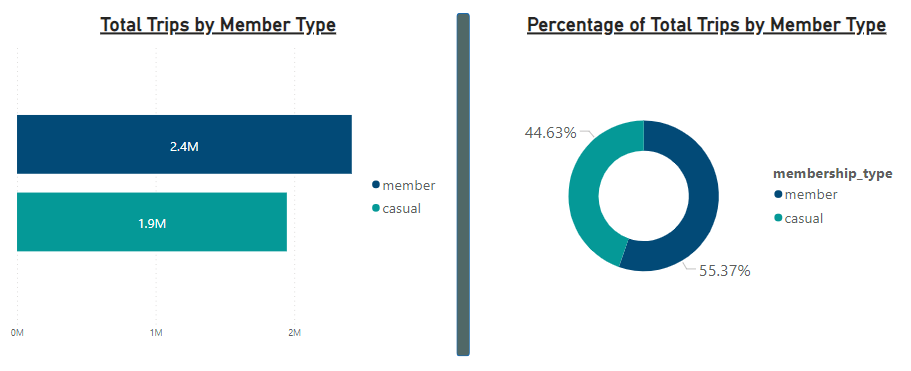
1. Average ride length in minutes



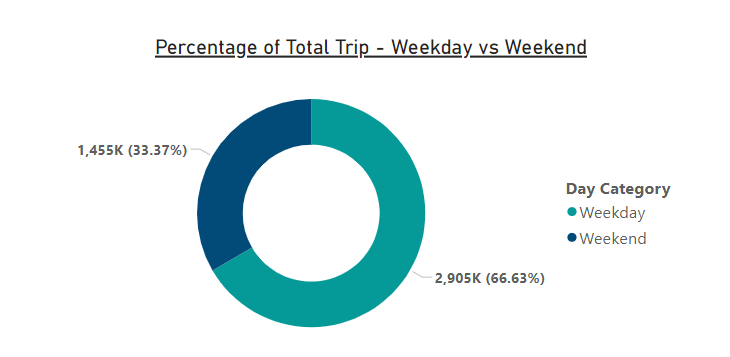
**Analyzing the Data**

**Trip Count**

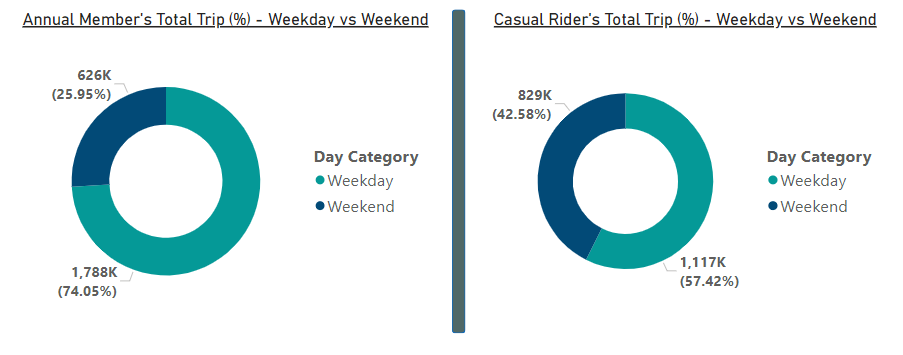
Over the 12-month period, Annual Members took a total of 2.4Million trips representing a little over 55% of the total trips taken while Casual Riders took around 1.9Million trips which is was about 45% of the total trips taken within the period under review.



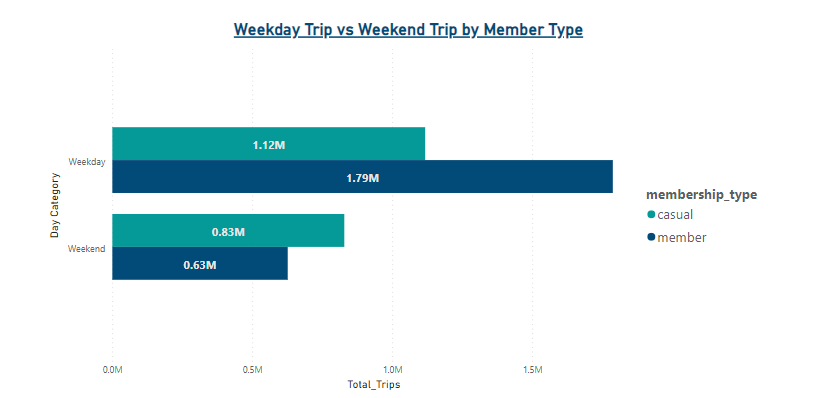
From the Donut Chart, about 67% of the total trips were taken during the weekday while a little over 33% was taken on weekends.



Drilling further down and segmenting the weekday-weekend trips by member type, around 74% of the trips by Annual members was made during the weekdays unlike Casual Riders whose weekday trip accounted for a little over 57%. The percentage of the weekend trips for Annual Members and Casual Riders are 25.95% and 42.58% respectively.



*Please see below bar chart for direct comparison of the total trips taken by both member types segmented into Weekday and Weekdays*

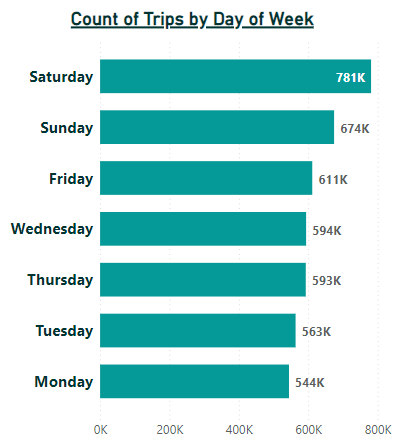


Looking at the chart above, one needs to ask the following questions:

* Why do Annual Members use Cyclistic’s services more during the week?
* And why do the trips taken by Casual Riders outnumber those of Annual Members on weekends.

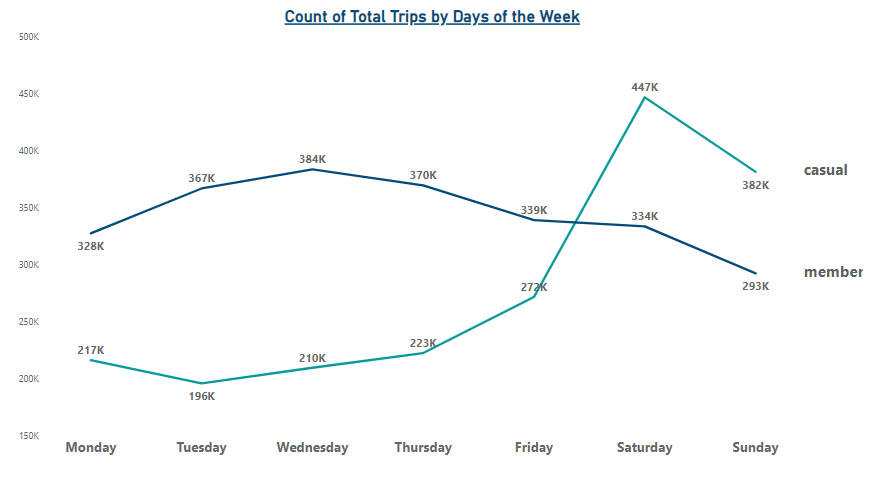
Before we answer the questions raised, let us breakdown the count of total trips taken by day of the week.

**Trip Count by Day of Week**



Saturday has the highest number of trips taken followed by Sunday. While the day with the least number of trips taken is Monday.

Now, let us check out the contribution of each member type to the trips taken on each day of the week.



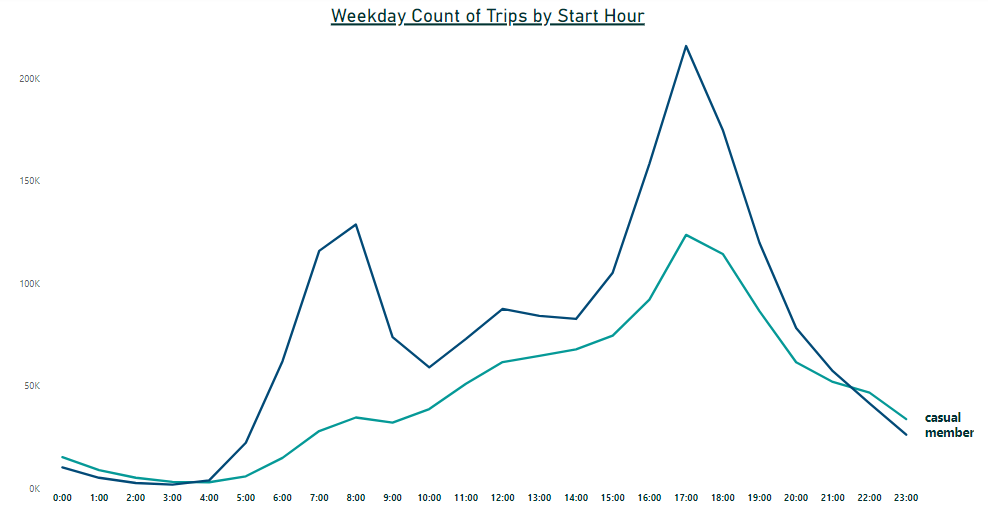
From the line graph above, it is observed that Annual members maintain a steady trip count during the weekdays before seeing a gradual drop going into the weekends; whereas, Casual riders recorded a comparatively low trip count during the weekdays before experiencing a sharp rise in trip count rise during the weekend.

**We can hypothesize from the findings that Annual Members use Cyclistic’s bikes for work related commuting purposes as they maintain a steady trip count during the weekdays; whereas, Casual riders use the services for leisure which correlates with the sharp increase in the trip count on the weekends.**

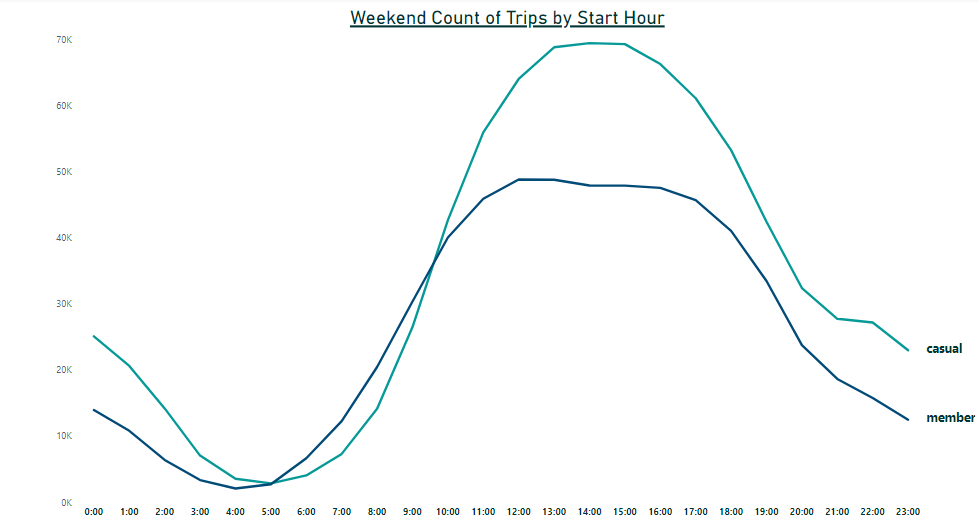
**To further confirm this hypothesis, let us explore the count of trips of both members by start hour.**

**Trip count by Start Hour**

**Weekday Count of Trips by Start Hour**



**Weekend Count of Trips by Start Hour**



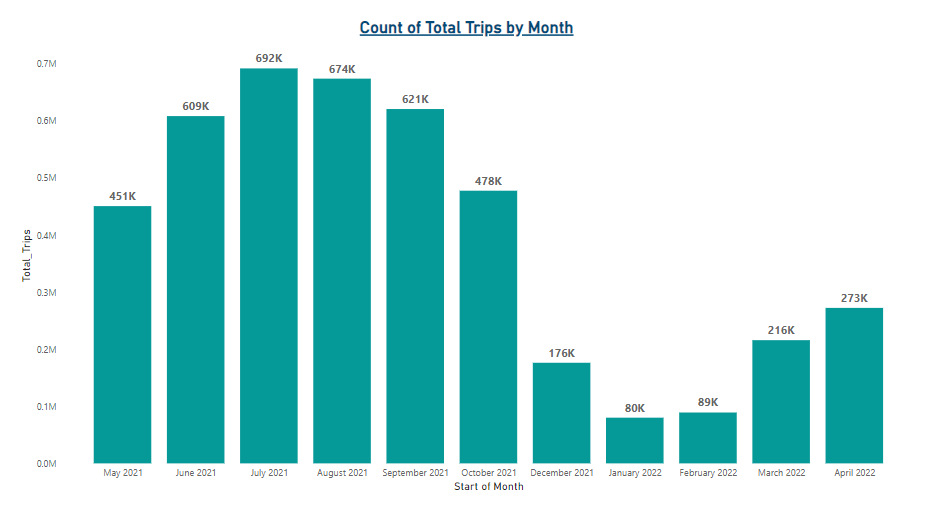
The line graphs shows that on the weekdays, Annual members’ trip count rose significantly between 6:00hrs and 8:00hrs and peaked between 16:00hrs and 18:00hrs, before experiencing a considerably sharp drop after 19:00hrs. On the other hand, Casual riders had a somewhat steadier trip count across the start hours of the day with an abrupt spike in levels of trip count around 17:00hrs.

The start hour pattern for both members are similar on the weekend except that the trip count levels of Casual riders are significantly higher than those of the Annual members.

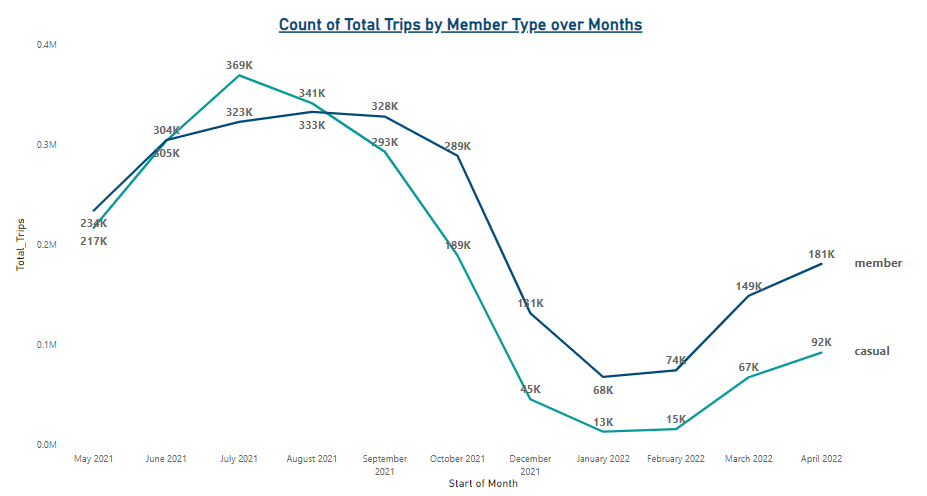
Hence, from these findings, we can further substantiate our hypothesis that **annual members use Cyclistic bikes for work related commuting purposes** as their trip count levels peak during the typical work commuting times (8:00hrs / 17:00hrs) and drop outside of popular office hours.

**Trip Count by Months**

How about we check out the monthly comparison of total trips taken.



From the column chart above July and August recorded the highest number of trips taken and January and December had the lowest trip count. Let us go further by segmenting the total trips by Member Type to see what insights we would uncover.

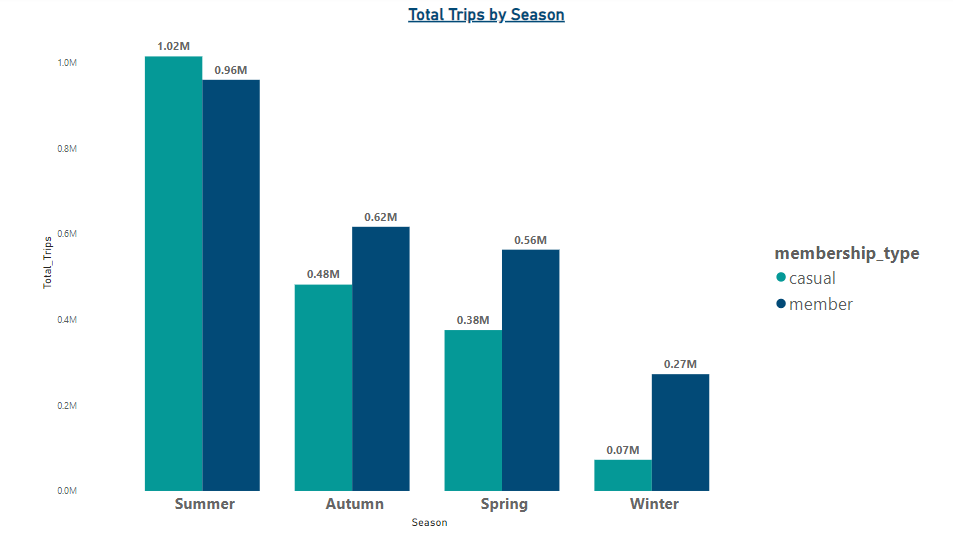


Generally, both Member Types recorded a high number of trips between July and August and their lowest between December and February.

The trips taken by Annual Members outnumber those taken by the Casual Riders across all the months except in July and August

**Trip Count by Seasons**

Now we go even further by grouping the months into seasons to discover more insights from our data. *Allons-y!*

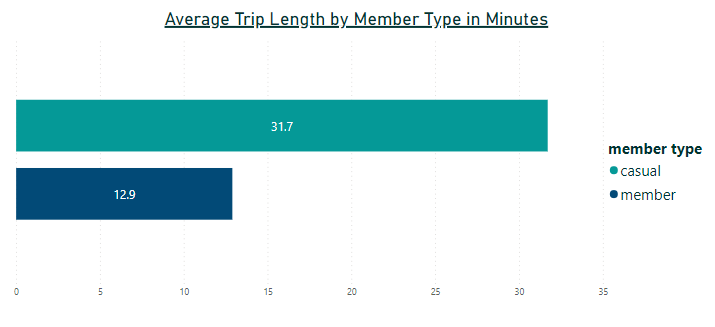


The Annual Members took more trips than the Casual Riders across all seasons except in the Summer where Casual Riders took over 55,000 more trips than Annual Members. One can infer that tourists who visit mostly in the Summer months for holidays due to the warmth account for the increase in number of Casual Riders for that period. In addition, Annual members’ trip count did not drop as much as the Casual rider’s in the winter months. Thus, this further strengthens our hypothesis that **Casual riders use Cyclistic bikes more for leisure purposes.**

**Average Trip Length**

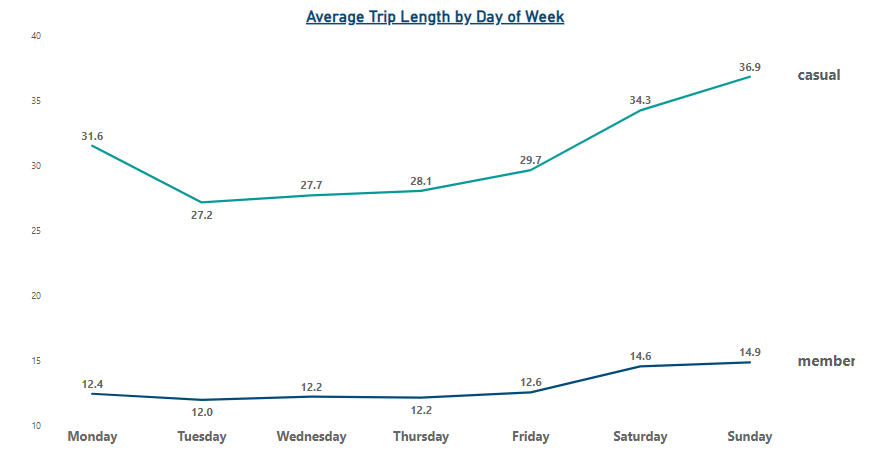
**The average trip length for all recorded trips is 21.29 Minutes.**

**Let us go ahead to breakdown the Average Trip Length by Member type.**

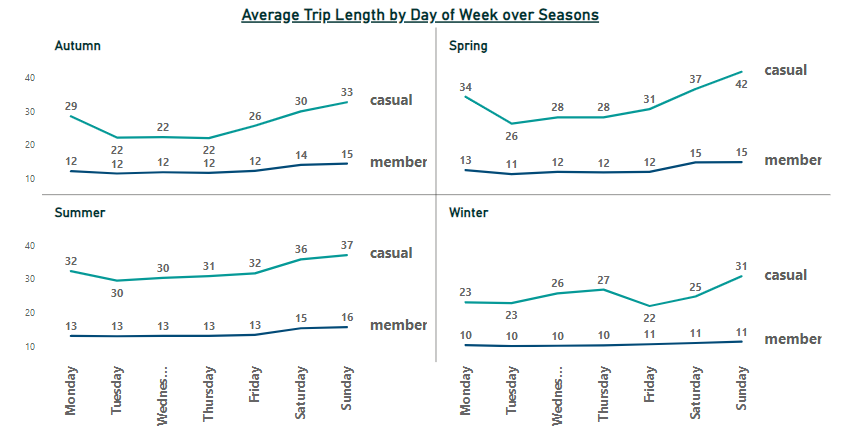


**Annual members tend to take shorter trips compared to Casual riders whose average trip length is almost 2.5 times greater.**

**Let us go ahead to see the average trip length for each day of the week**



**The weekends have the highest Average Ride Length for both member types and this seems to be the same for all seasons as shown below:**

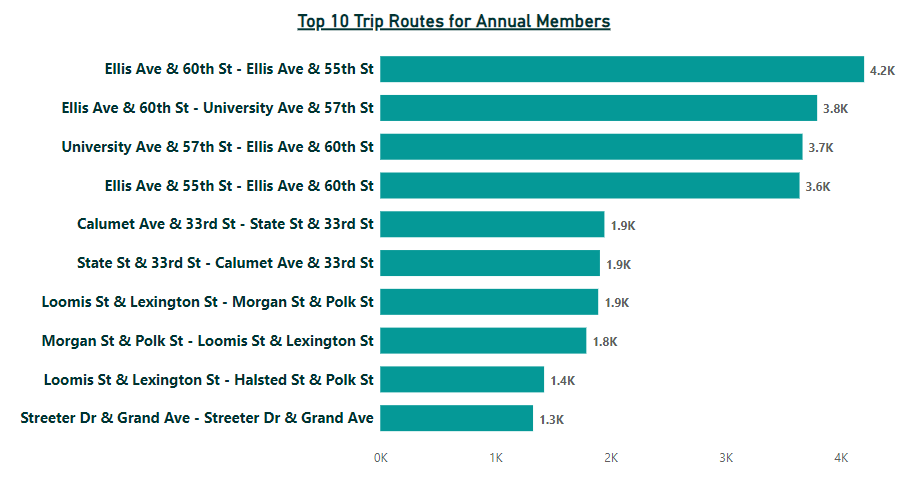


**Spring has the highest Average Trip Length followed Summer, while Winter recorded the lowest Average Trip Length. Obviously, people will remain indoors and take trips only when necessary in the Winter due to the cold. Across all seasons, Sunday is the most preferred day for a ride by both Annual Members and Casual Riders.**

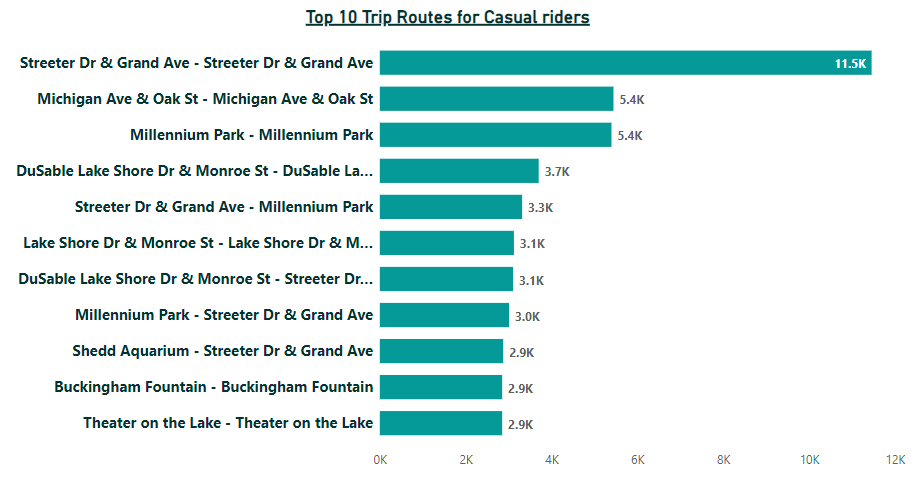
**These insights strongly substantiate our hypothesis and we can reach the conclusion that Annual Members are mainly commuters using Cyclistic’s bikes for short length trips during the weekdays to get to work and casual riders typically use Cyclistic’s bikes on the weekend for leisure.**

**Top 10 Routes among Annual Members and Casual Riders**

Annual Members



Casual Riders:



Taking a look at the top 10 routes of Casual riders and Annual members, we would observe that the preferred routes for Casual riders involve parks, lakes, lake shores and fountains which are typical tourist destinations; on the contrary, Annual members’ trips are more inland, usually and on major streets which may probably be office and apartment locations.

**PHASE 5 – SHARE**

**SUMMARY OF FINDINGS**

**Annual Members**

* Annual members use the Cyclistic bikes mostly for **work related commuting purposes as they maintain a steady trip count during the weekdays** all through the year. They however, experience a drop in trip count in the winter months.
* Their peak times are 8am and 5pm on weekdays which coincides with work opening and closing commuting times and between 12pm and 1pm on weekends.
* 74% of their trips are taken on the weekdays and about 26% during weekends further supporting our point that the bikes are used mostly for work related commuting purposes.
* Their average trip length is 12.9 minutes which is less than half the average trip length of casual riders (31.7 minutes).
* Their top 10 preferred trip routes start and end on major streets where typical of office and apartment locations. Their preferred route is Ellis Ave & 60th St - Ellis Ave & 55th St (4.2k trips in total)

**Casual Riders**

* Casual riders use the Cyclistic bikes mostly for leisure outnumbering the Annual members in trip count in the summer month and dropping drastically in the winter months.
* Their peak times are between 1pm and 3pm on weekends and 5pm on weekdays.
* 57% of their trips are taken on the weekdays and about 43% during weekends.
* Their average trip length is 31.7 minutes which is almost 2.5 times the average trip length of Annual members (12.9 minutes).
* Most of their preferred routes involve parks, lakes, lake shores and fountains which are typical tourist destinations. Their preferred route is Streeter Dr & Grand Ave - Streeter Dr & Grand Ave (11.5k trips in total)

**PHASE 6 – ACT**

Based on the findings from the analyses. I would like to give the following recommendations to the Cyclistic marketing team for designing new strategies to aid in the conversion of Casual riders to Annual members:

* Since Casual riders have peak time of 5pm on weekdays, the company should design annual membership plans targeted at fitness and wellness enthusiasts that take part in after-office-hours sessions. There will be a need for more bikes to be made available at this hour as 5pm is also the major peak time for existing annual members. This can be achieved by partnering with major fitness and wellness companies.
* Digital Screens for adverts should be installed in the Casual riders’ most popular stations. These screens should be explored for marketing campaigns for special discounted annual membership packages in the spring, through the summer, and on weekends. The screens can also be used to advertise their strategic partnership with fitness and wellness companies for weekend and after-office-hours activities.

**RECOMMENDATION FROM DESIGN THINKING**

* Cyclistic should invest in heat-generating bike clothing for riders to improve the low patronage in the winter months.