OVERVIEW: This is a case study performed for the capstone project of the Google Data Analytics course on Coursera.

The case study presented information to ask a question and analyze data to provide the findings. The case study was performed using the method of Ask, Prepare, Process, Analyze, Share, Act.

INTRODUCTION: Cyclistic is a bike-share company that has 5,824 bicycles with over 1,500+ docking locations throughout the city of Chicago. They have regular bikes that require to be taken and returned to these docking stations. They also have electric bikes which have an electronic lock and can be taken and left anywhere.

The company currently offers 3 different ways to ride: single-ride pass, full day pass and annual membership. Cyclistic considers single-ride and full day pass riders as casual and annual passes as members. The financial analysis shows that annual memberships generate more revenue and that there is opportunity to convert casual members into annual members.

ASK: The Director of Marketing has asked the data analysis team to review the prior 12 months of data and determine how annual members and casual riders use Cyclistic bikes differently?

PREPARE: Cyclist has its data stored here. The data is in zip files that contain csv for each month. The current time period to be reviewed was January 2023 - December 2023. I

downloaded and extracted all of the files locally. Opening the first dataset in google sheets there were 190,000 records. With just one month of data this was a large set of data to work with within a spreadsheet. Due to the volume of data I imported all 12 sets of data into SQL Server for further review. I then created a new table to consolidate the 12 sets of data into 1 and inserted the records into the new table. The query to create the table and insert the records can be seen here. There were 5,719,877 total records for 2023.

Reviewing the data there were several issues with the data that needed to be addressed. There were rides that were less than a minute that were likely due to a bike issue or maintenance from the Cyclistic team. There were also rides that were over 24 hours long that were determined to be from incorrect registration of the bike being returned. From January - August there were records with rideabale_type of docked_bike which was a legacy term for classic bikes. Some records were missing ending location coordinates but had ending station id which could have the coordinates updated.

PREPARE: To clean the data I had to remove the rides that were less than 1 minute and more than 24 hours. This removed 157,497 records. All records that had a rideable_type as docked_bike were updated to classic_bike, this affected 76,100 records. Any records that did not have ending location coordinates but had an ending station id were updated, this was 117 records. Records that did not have any ending location coordinates or ending station information were removed, 675 records. The query to prepare this data is located here.

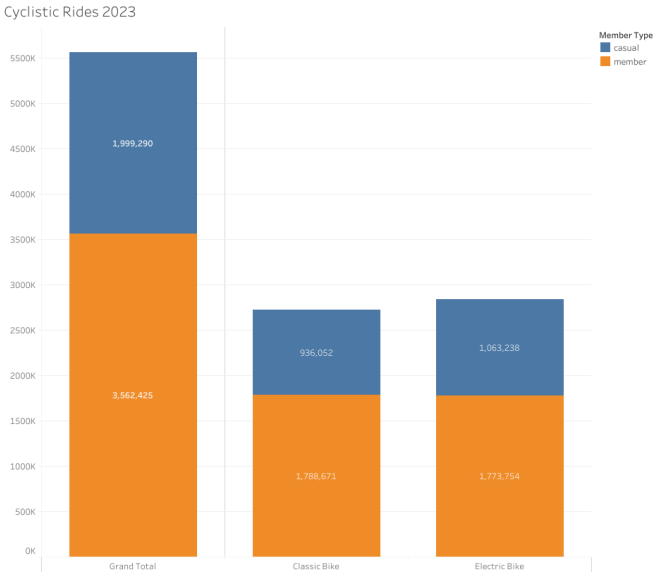
ANALYZE: For analysis the data was transformed to add columns for length of ride in minutes, month of ride and day of the week of each ride. The data was the aggregated as follows:

1. Total number of rides by rider type and bike type
2. Total number of rides each month by rider type
3. Average ride length by month by rider type
4. Total number of rides each day of the week by rider type
5. Average ride length by day of the week by rider type
6. Top 10 destination by rider type

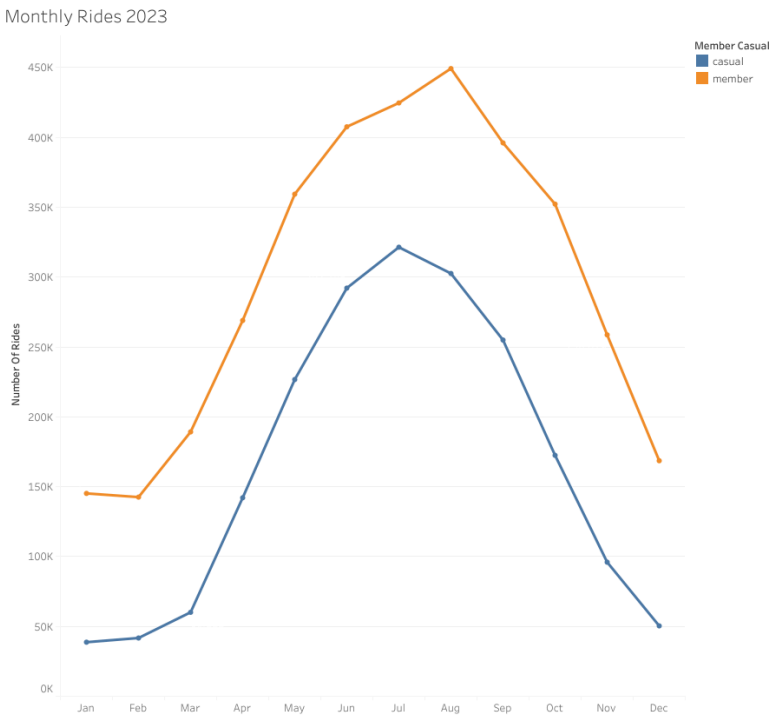
The query to aggregate the data can be located here.

SHARE: To create the visualizations to share with the stakeholders, each aggregated dataset was uploaded to Tableau. The dashboard with the data is located here.

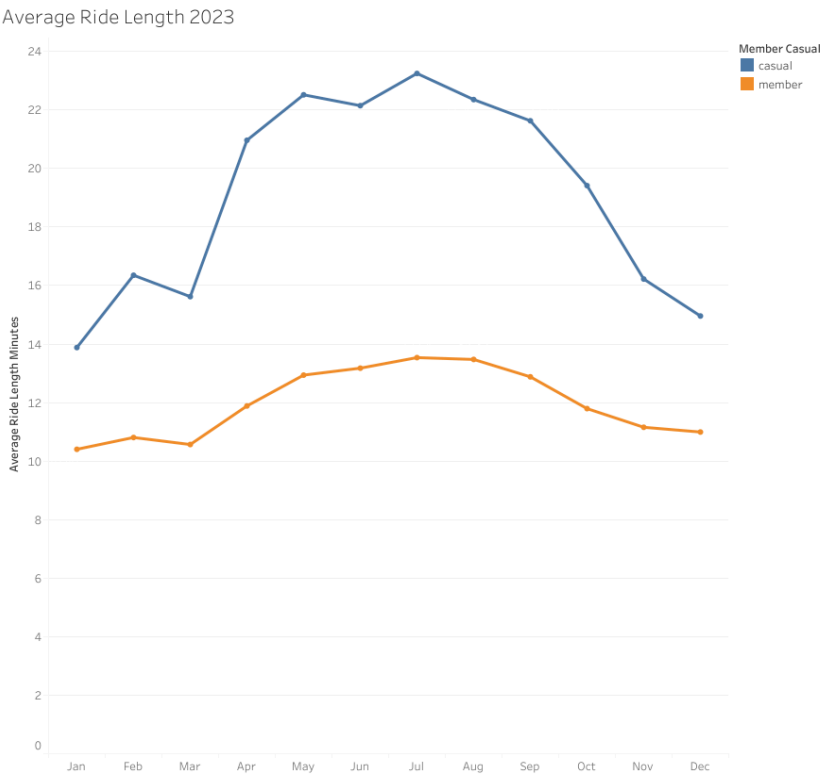
The data shows that 64% of total rides in 2023 were taken by members. There was no significant difference in the type of bikes used between casual riders and member riders.



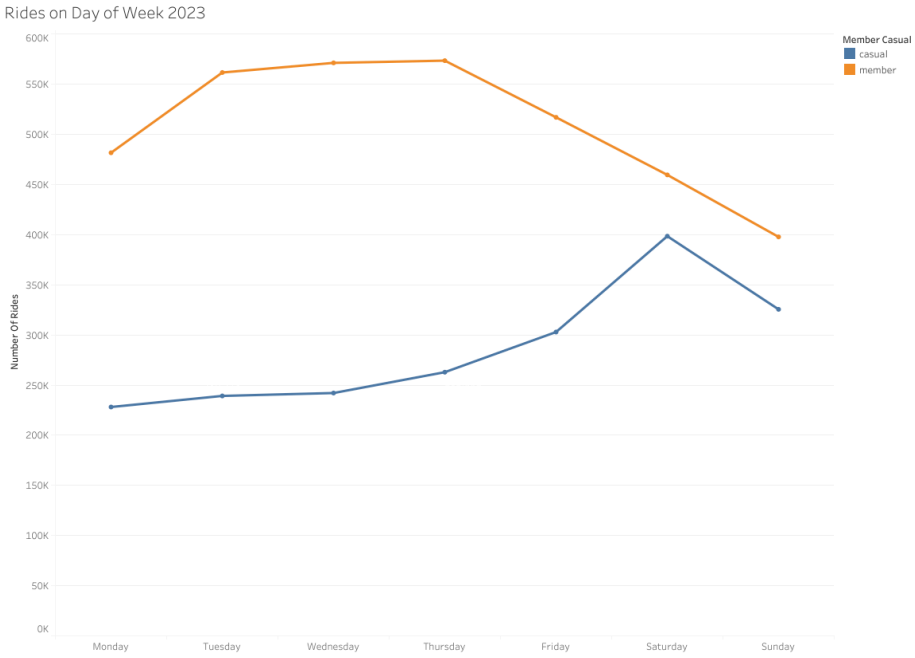
The number of rides taken increased in April and began to decrease in October. The proportion of rides taken by casual riders and member riders remains relatively the same.

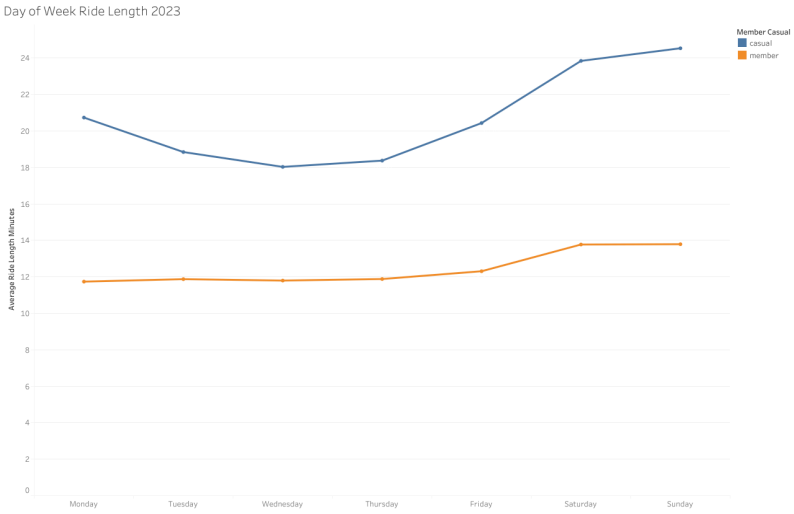


On average casual riders take longer rides than members. The ride length for casual riders increases significantly in the months from April to October while member riders only increase slightly.

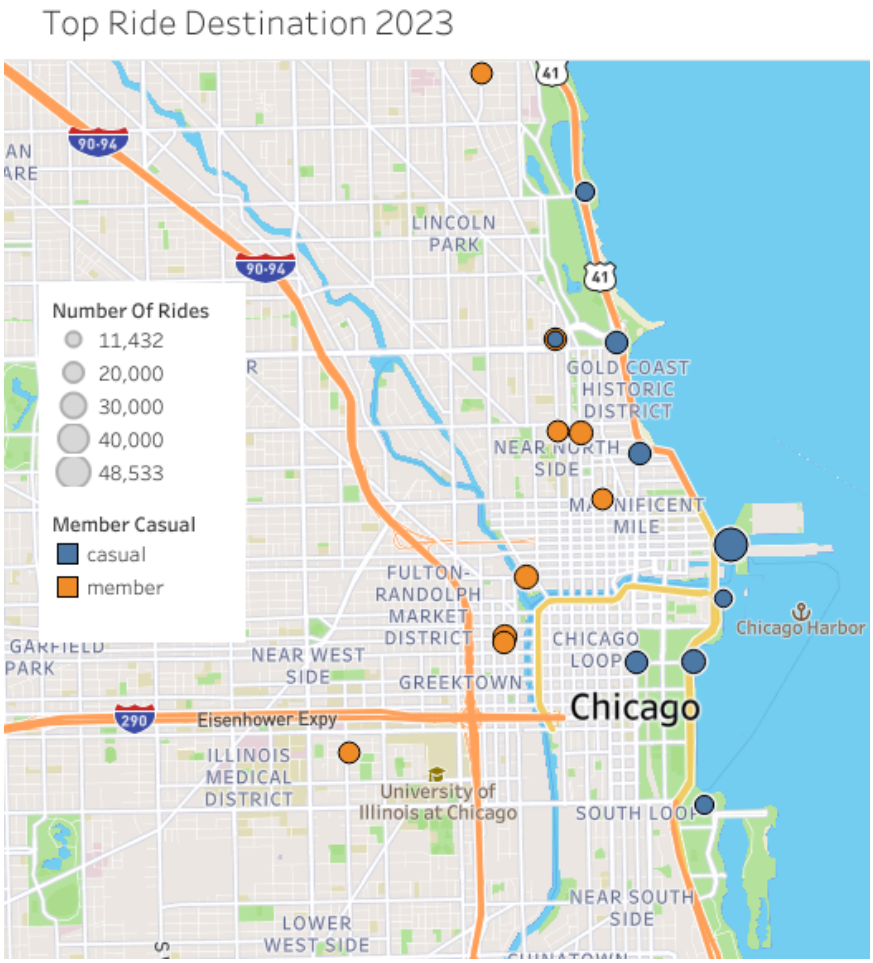


Casual riders take more frequent and longer rides Friday - Sunday while member riders take less rides during this time period.





Casual riders' top destinations are parks, beaches and Navy Pier and member riders' top destinations are located more in places with more buildings.



How do casual and member riders use the bike-share service differently?

Casual riders use the bike-share service for leisure and member riders use the bike-share service as transportation.

Casual riders take longer rides and use the service more on weekends when they have more time available. They take the bikes to locations with more outdoor activities, such as parks and beaches. This also suggests that there may be an increase in activity for casual riders due to tourism.

ACT: Some suggestions to convert some of the casual riders to member riders is to offer different types of memberships.

1. Weekend rider pass that would offer unlimited rides for a weekend
2. Summer rider pass that would offer unlimited riders during the months of April - September