# **DIVVY’S ROAD TRIP**

**Introduction:**

**About the Company:**

Cyclistic is a bike-sharing company in Chicago, which has since expanded to include a fleet of 5,824 geotracked bicycles stationed at 692 locations across Chicago. The bikes can be unlocked at one station and returned to any other station within the network at any time. Individuals buying single-ride or full-day passes fall into the category of casual riders, while those acquiring annual memberships become recognized as Cyclistic members.

**Tools and Technologies:**

* Tableau/Power BI for dashboard development.
* Python for data analysis

**Phase 1:**

### **About the Dataset:**

The data is publicly available on an [AWS server](https://divvy-tripdata.s3.amazonaws.com/index.html). We were tasked to work with an entire year of data, so I downloaded zipped files (CSV format) containing data from January 2023 to December 2023, one file for each month.

### **Data Structure:**

Each .csv file contains a table with 13 columns with varying data types as shown below. Each column stands for a field that describes how people use Cyclist's bike-sharing service. Each row represents an observation with the details of every ride.

* **ride\_id:** This is a unique identifier assigned to each bike ride. It's like a reference number for the trip.
* **rideable\_type:** This column indicates the type of bike used in the ride. It can be "electric\_bike" or "classic\_bike".
* **started\_at:** This shows the date and time when the ride began. The format is YYYY-MM-DD HH:MM:SS.
* **ended\_at:** This shows the date and time when the ride ended. The format is the same as the started\_at column.
* **start\_station\_name:** This specifies the name of the docking station where the ride started.
* **start\_station\_id:** This is a unique identifier for the starting docking station. It complements the start\_station\_name.
* **start\_lat:** This represents the latitude coordinate of the starting docking station.
* **start\_lng:** This represents the longitude coordinate of the starting docking station. These coordinates might be useful for mapping the station's location.
* **end\_station\_name:** This specifies the name of the docking station where the ride ended.
* **end\_station\_id:** This is a unique identifier for the ending docking station. It complements the end\_station\_name.
* **end\_lat:** This represents the latitude coordinate of the ending docking station.
* **end\_lng:** This represents the longitude coordinate of the ending docking station. These coordinates might be useful for mapping the station's location.
* **member\_casual:** This column indicates whether the rider was a member (member) or a casual user (casual) of the bike-sharing service.

Phase 2:

I used python for data cleaning You can view the Jupyter Notebook for the Process phase [here](https://github.com/K-Jyoshnavi/Capstone-project/blob/main/divvy%20.ipynb)

Here are the steps that I did during this phase

1. Check for null and duplicates
2. Additional columns and data transformation (change the data type, remove trailing or leading spaces, etc.)
3. Extract data for analysis

### **Data Cleaning Result**

Total Row Count before data cleaning: 5745324

Total Row Count before data cleaning: 4268747

**Phase 3:**

**Analyze:**

I used Python in my jupyter notebook to look at the huge data we cleaned earlier. I came up with questions to figure out how casual riders are different from annual members. Then, I made queries to get the answers, helping us understand more and make decisions based on the data.

### **Questions**

Here are the following questions we will answer in this phase:

1. What is the percentage of user types from total users?
2. Is there a bike type preferred by different user types?
3. Which bike type has the longest trip duration between users?
4. What is the average trip duration per user type?
5. What is the average distance traveled per user type?
6. What days are most users active?
7. What months or seasons of the year users tend to use the bike-sharing service?

### **Findings**

* 63% of the total Cyclistic users are annual members while 36% are casual riders.
* Both annual members and casual riders prefer classic bikes. Only casual riders use docked bikes.
* Generally, casual riders have the longest average ride duration (23 minutes) compared with annual members (18 minutes).
* Both annual members and casual riders have almost the same average distance traveled.
* Docked bikes have the longest average ride duration which only casual riders use. Classic bikes have the longest average ride duration for annual members.
* Most trips are recorded on Saturday.
* There are more trips during spring and at least during winter.

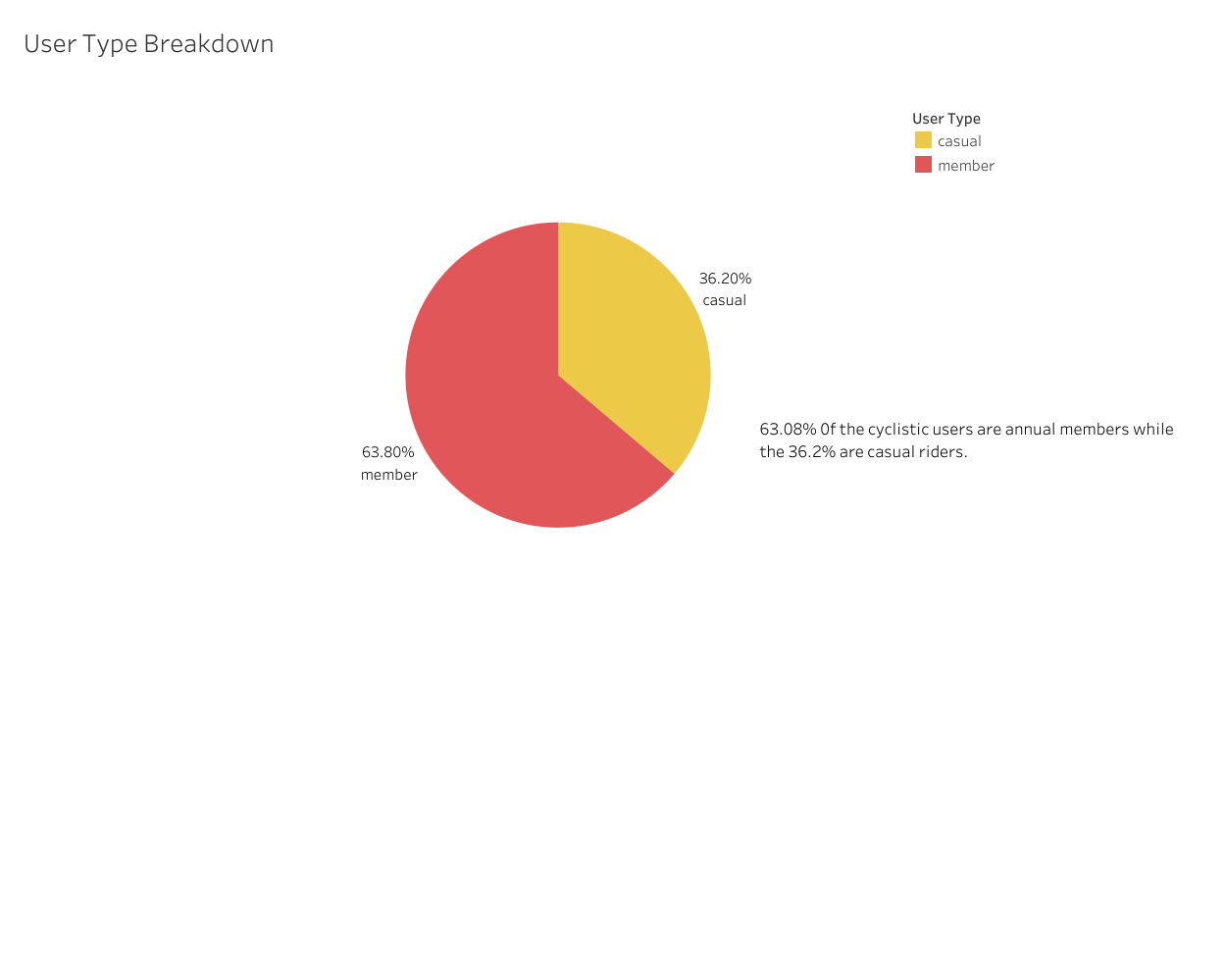
**Phase 4:**

I used Tableau public in making the visualization. You can view the data visualization for the Share phase [here](https://public.tableau.com/app/profile/katabathina.jyoshnavi/viz/divvytripvisualisation/Dashboard7).

### **Data Visualization**

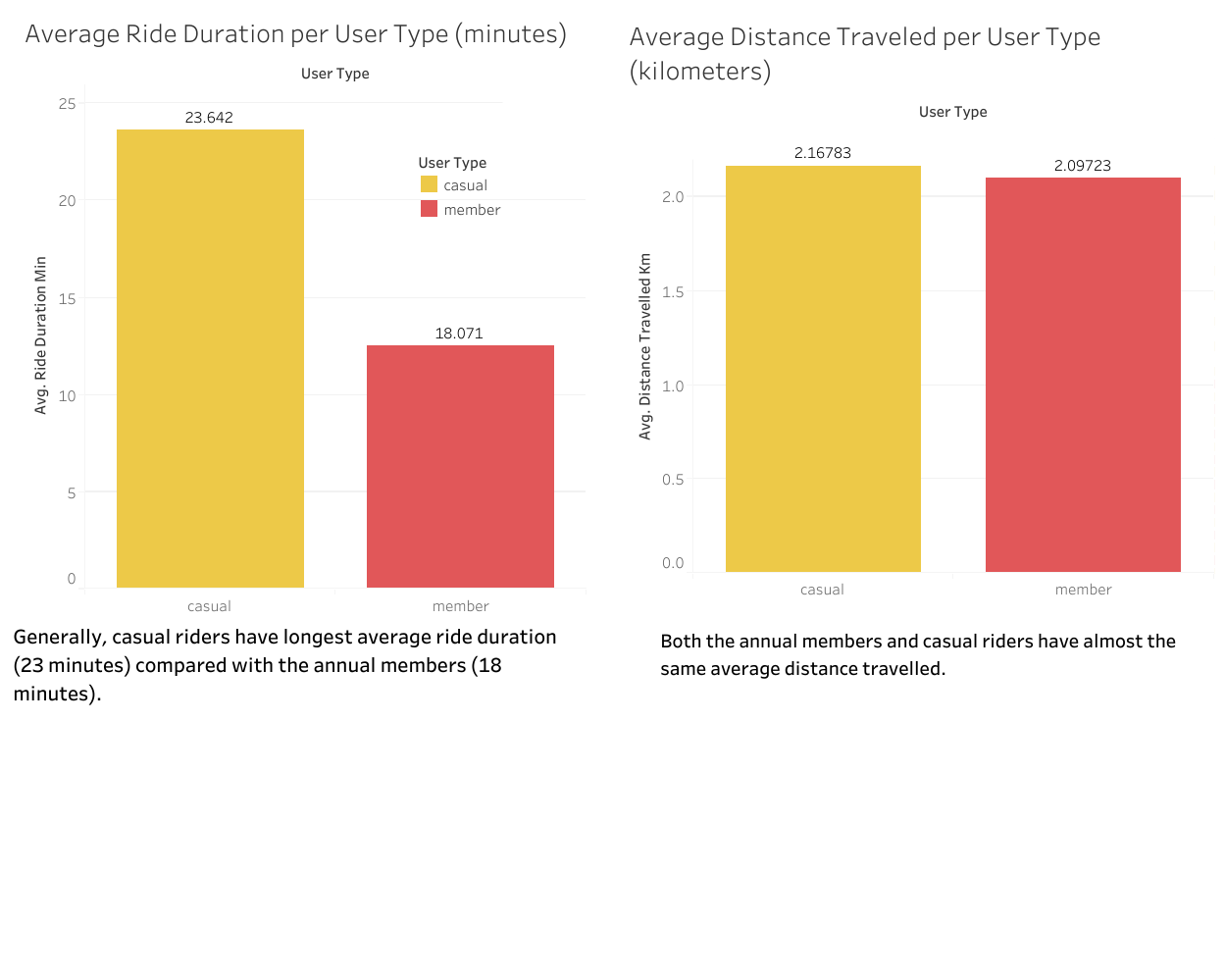
These visualizations are arranged to narrate a compelling story, providing a clear understanding of the analytical process that led to our conclusions.

User Type Breakdown



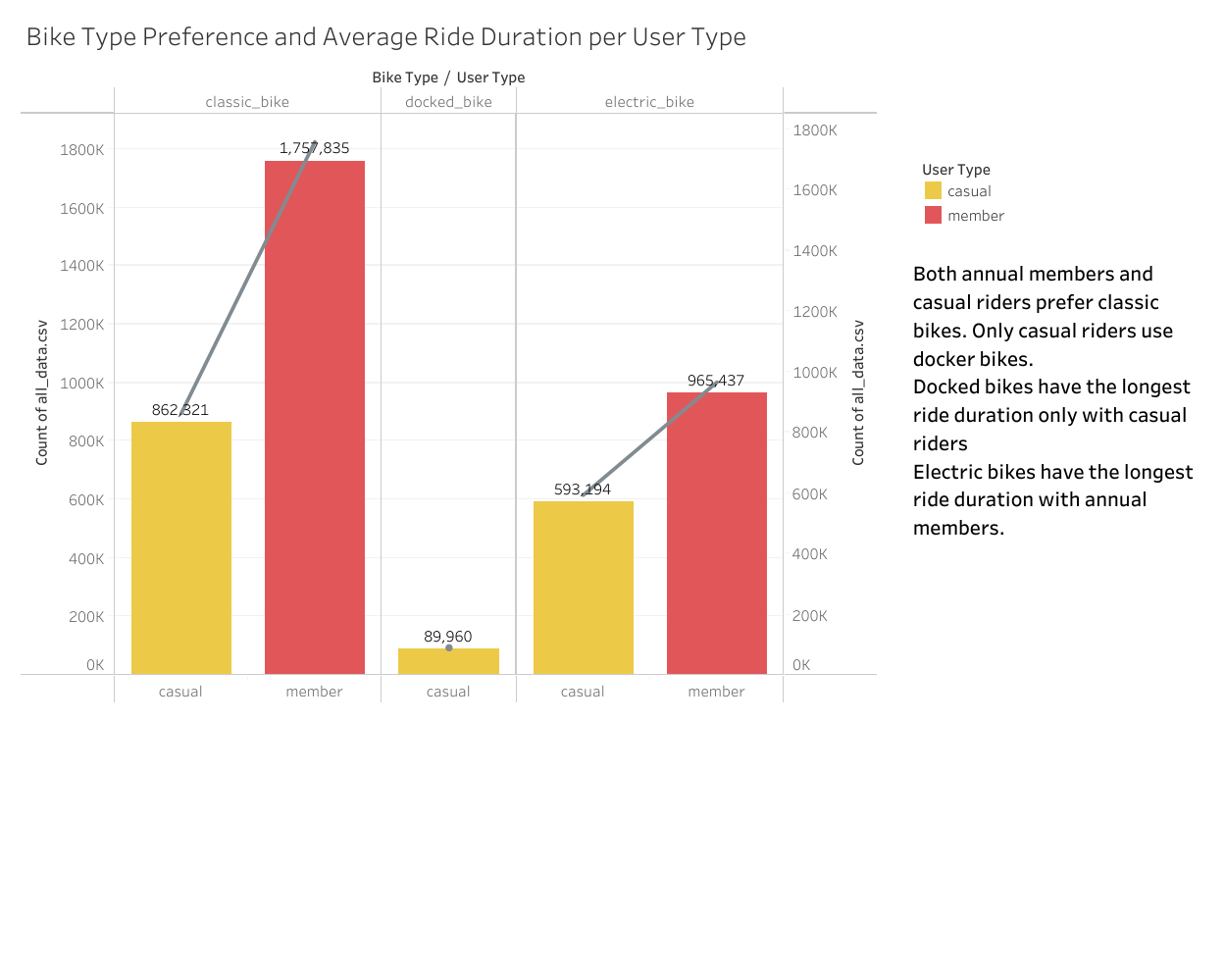
63% of Cyclistic users are annual members, while 36% are casual riders. This suggests that a significant portion of users are already committed annual members.

Ride Duration and Distance Traveled



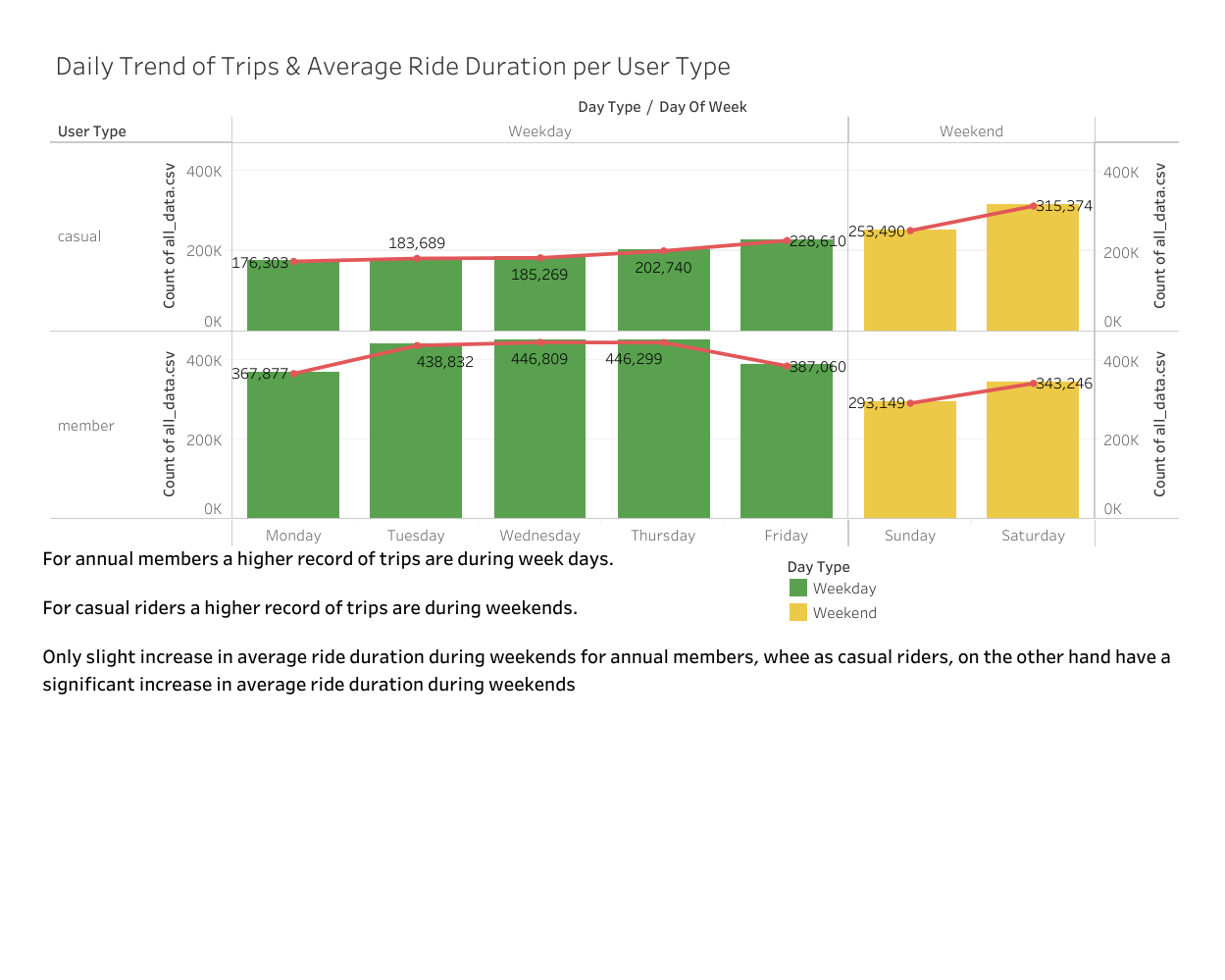
Generally, casual riders have the longest average ride duration (23 minutes) compared to annual members (18 minutes). However, both groups have similar average distances traveled.

Bike Preference



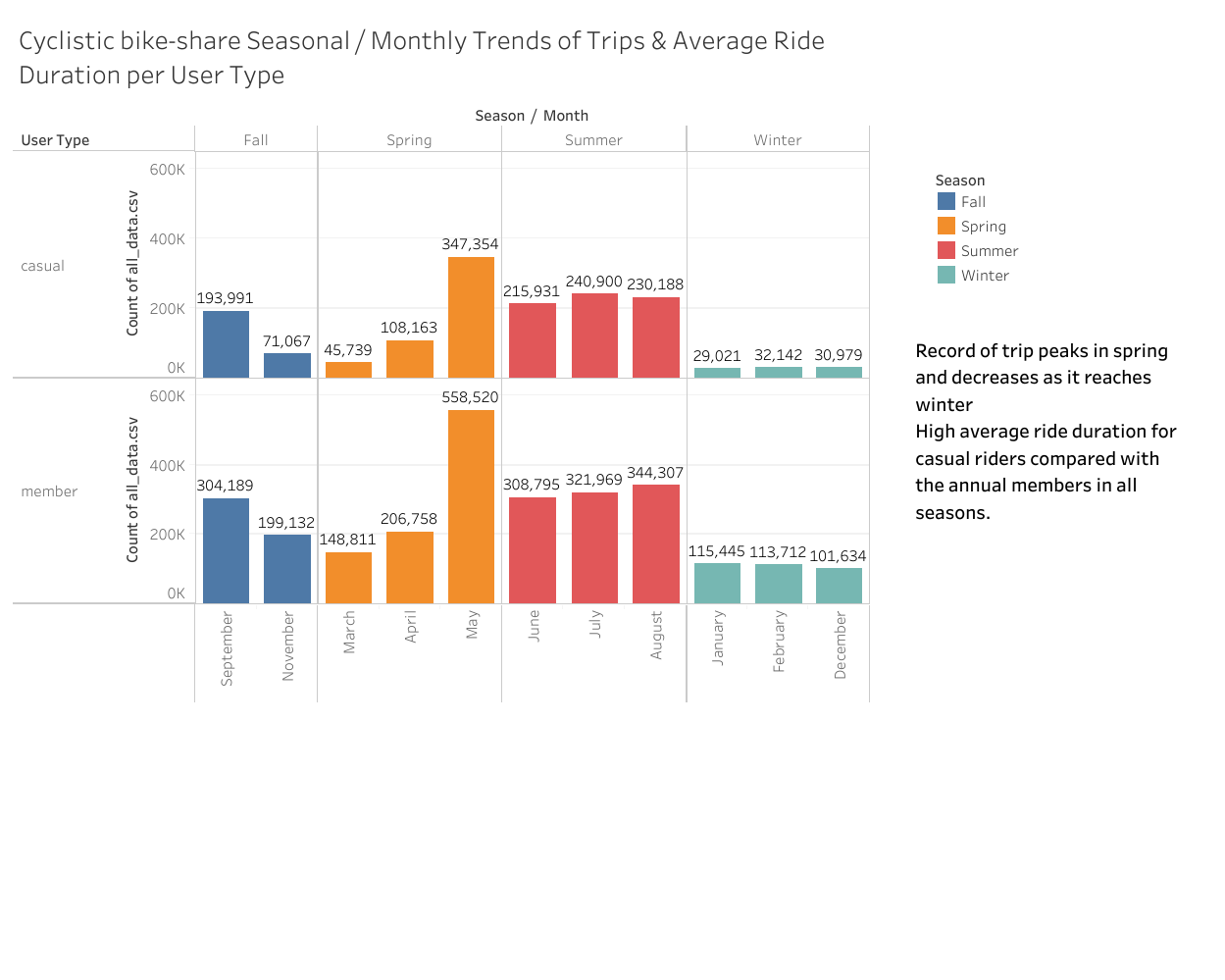
Classic bikes are preferred by both annual members and casual riders. Docked bikes are exclusively used by casual riders and have the longest average ride duration.

Daily Trip Trends



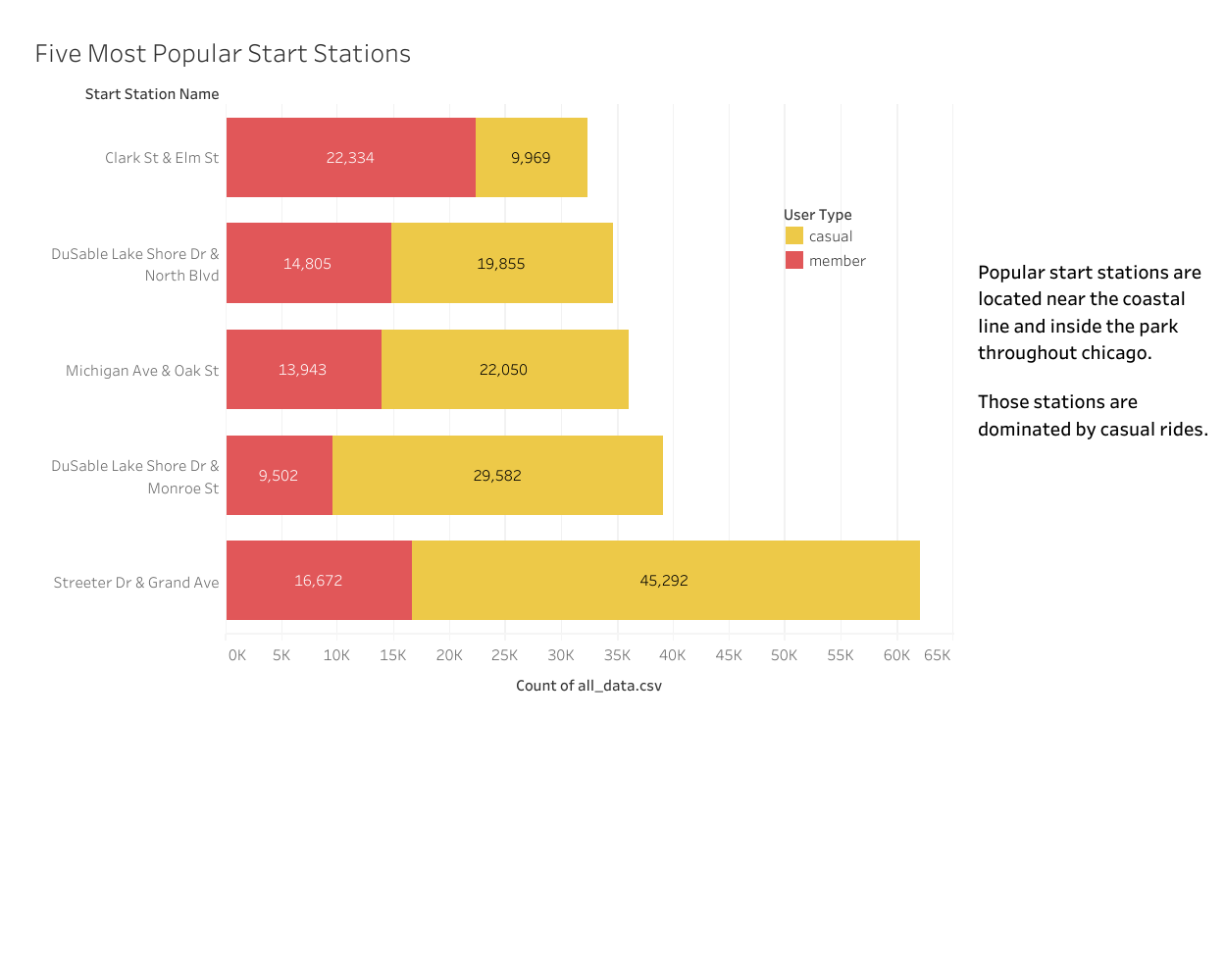
Annual members have a higher record of trips during weekdays, indicating a likely use for daily commuting. Casual riders, on the other hand, prefer weekends for bike rides.

Seasonal / Monthly Trip Trends

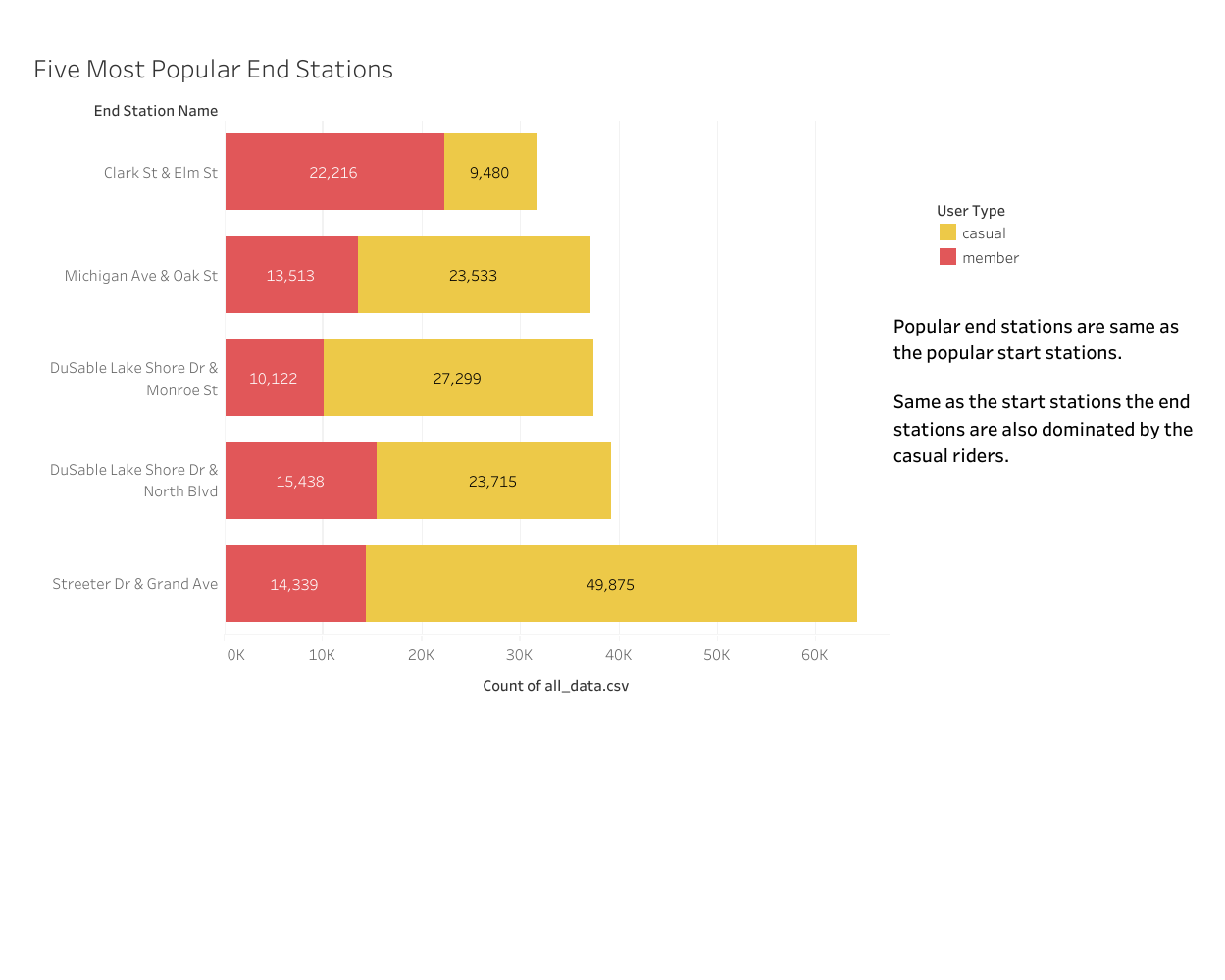


Trips peak during spring and decrease as winter approaches possibly due to good weather & road conditions. Casual riders consistently have a higher average ride duration in all seasons.

Popular Stations



Popular start stations are located near the coastlines and inside the parks throughout Chicago. Those stations are dominated by casual riders.



The popular end stations are the same as popular start stations. Same as the start stations, popular end stations are also dominated by casual riders.

### **Conclusions**

* 63% of the total Cyclistic users are annual members while 36% are casual riders.
* Both annual members and casual riders prefer classic bikes. Only casual riders use docked bikes.
* Generally, casual riders have the longest average ride duration (23 minutes) compared with annual members (18 minutes).
* Both annual members and casual riders have almost the same average distance traveled.
* Docked bikes have the longest average ride duration which only casual riders use. Classic bikes have the longest average ride duration for annual members.
* Most trips are recorded on Saturday.
* There are more trips during spring and at least during winter.