Casual Riders Vs. Subscription Members: Differences in Behavior

For Cyclistic bicycle sharing company, July 2025

Section 1: Project Overview

This analysis focuses on the behaviors of casual riders and subscription members who rode Cyclistic bikes in Chicago through their bike sharing service. The company would like to convert more casual riders to subscriptions, and tasked me with finding the ways in which the two groups behave differently. The data has a single observation for each ride. The key information that is missing from this data is a way to identify the riders themselves: I will not be able to connect the individual rides to the person riding the bike.

I am using one year of data from divvy-tripdata (divvy-tripdata.s3.amazonaws.com) for the time period from May 2024 through April 2025. This was the most recent data available when I first loaded the data into my table on June 2nd.

It was made available under this license: Data License Agreement | Divvy Bikes (https://divvybikes.com/data-license-agreement)

My goal for this project is to establish ways in which behavior differed between casual riders and subscription members, and to identify patterns that could potentially be relevant to the business goal of increasing the number of subscription members as a share of all riders.

Section 2: Preparation, Data Integrity, and ETL

I began this project by previewing the first and last months of the data in RStudio in order to determine the schema, and whether the columns were the same for the entire year. I then used PostgreSQL (by way of the query tool in pgAdmin GUI) to clean, transform, and analyze the data. First, I created a table for the entire year. I specified all the datatypes for the table, and included a check to ensure the locations specified by the starting and ending latitude and longitude columns were valid. Once the table was specified, I loaded the data into it using SQL. The table is composed of 12 CSV files organized consecutively.

I then checked the data for cleanliness and validity. I used COUNT DISTINCT and GROUP BY statements to determine the number of unique values for the character fields, and searched for null values in every column. The data is very clean, and every observation has a value in the primary key column.

I used a few more basic queries to check for the total number of observations (5,735,884) and to generally explore and get a feel for the data.

Section 3: Data Transformation

The data table contains the following dimensions: ride type, start station, end station, and whether the rider is a subscription member; plus the start time and end time. There was no quantitative data immediately available in the data, and the raw timestamps were not useful in their current form. It was necessary to create some calculated columns.

I started with the quantitative measurements, because these can easily be used in every query without complicated grouping logic. I came up with three:

- **Number of rides**, derived using the COUNT function.
- **Trip duration**, derived by subtracting the start timestamp from the end timestamp and converting the interval to an EPOCH numeric datatype.
- **Absolute distance traveled**, derived through a simple trigonometric formula using the starting and ending longitude and latitude. This formula operated on the assumption of a flat plane of coordinates, which I believe is appropriate for a set of locations concentrated in a single city.

I then ran a series of queries grouping the data with SQL by both membership cohort and each of the qualitative columns, with the three measurements included. This yielded limited insight, so I began transforming the qualitative columns as well. I created the following calculated qualitative columns for my analysis:

- **Month**, extracted from the ride start timestamp
- **Day of week**, extracted from the ride start timestamp
- **Hour**, extracted from the ride start timestamp
- **Trip direction**, derived using a CASE statement with a trigonometric function from the starting and ending longitude and latitude
- **Neighborhood.** This was created by spatial-joining my table to a GIS geometry table that I found on the City of Chicago website (link: Boundaries - Neighborhoods | City of Chicago | Data Portal) (https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9). I chose neighborhoods rather than zip codes or wards because they are more recognizable.

I created summary tables with SQL that featured each of these, and made simple visualizations for each to help with my understanding. For these, I also isolated rides where the rider did not actually go anywhere (likely never used the bike at all), and then updated my queries to exclude these zero-distance rides when I failed to find a useful pattern for when they happened.

I also attempted to load the entire table into Tableau to allow me to more easily manipulate the data on the fly. However, the program would not run smoothly with a data table so large, especially once calculated fields were added. Nearly all data transformation was done through SQL.

Section 4: Analysis

With my relevant columns having been created, I made an overview table in Google Sheets to summarize all of the relationships I was able to measure. I highlighted significant relationships in red, and small but potentially useful relationships in a light pink.

Members vs. Casual Riders: Differences in Behavior

		Six Qualities to Compare (in order of how difficult they were to measure)							
2 Primary Cohorts Being Compared	Totals for each measurement for all qualities	Ride Type (Classic Bike, Electric Bike, Scooter)	Month	Day of Week	Time of Day	Direction of Ride (North, South, etc.)	Neighborhood		
Casual Riders Vs. Subscription Members	Members make up approximately 63% of all rides. I dropped rides where no distance was actually traveled from the dataset, and the membership percentage increased to 65%.	Riders used more electric bikes overall. The ratio of classic to electric bikes is similar across both cohorts. Scooters made up a small share of total rides (<1%).	Number of rides followed a similar monthly pattern for both groups, where ridership was much higher in the warmer months. This is expected given Chicago's weather. However, ridership rose much higher for casual	Members and casual riders used the service with **opposite patterns** through the week. Casual riders were more frequent during weekends, while subscription members used the bikes more frequently during weekdays.	The difference in hourly patterns between cohorts is clear. **Members** had the highest clusters of rides **during commute times**, while **casual riders** had fewer rides in the morning, and **higher in midday/afternoon**, with a **peak in the early evening**.	Riders took trips that went North and South somewhat more often than East and West trips. This was more pronounced for casual riders, who took an even higher percentage of trips travelling North and South as opposed to East and West.	**Inner ring neighborhoods** just outside of the city center had the **highest** numbers of **subscription members**, while **tourist destinations** and **outer neighborhoods** had relatively **more casual users**.	Number of Rides	Three Measurements
	After dropping zero-distance rides, I found that casual riders traveled a bit farther overall, but distances traveled were similar on average.	Members traveled a bit farther on electric bikes and casual riders traveled **significantly farther on classic bikes**. This difference may be related to commuting behavior but it is difficult to say.	Distance traveled was slightly higher in warmer months. The pattern was the same for members and casual riders.	Both groups rode a bit farther on weekends, but there was not a lot of weekday variation. Also, subscription members traveled farther on Wednesdays when riding electric bikes.	Riders traveled similar distances throughout the day, with members traveling a bit farther during morning commute hours, and casual riders travelling farther in the middle of the day.	Casual riders and subscription members had a nearly identical pattern of ride distance by direction, where rides that traveled North, Northwest, South and Southeast covered significantly more distance.	Generally, riders took longer trips when starting from North Side neighborhoods. This trend was somewhat more pronounced among subscription members.	Distance	
	Casual riders took significantly longer trips than members did overall. Possible explanations include that casual riders may opt for slower classic bikes, or use them for slow leasure trips more often.	Casual riders rode for longer across **all** ride types, **especially classic bikes**. Again, potentially related to commuting vs. leasure use.	Ride times were somewhat higher in summer compared to in winter. This was true for both cohorts, with both classic and electric bikes.	Ride duration followed a similar pattern throughout the week for both groups. Trips were somewhat longer on weekends, with the shortest trips on Mondays and Tuesdays.	There is an observable hourly pattern in how long rides were by hour. Electric bike rides were similar (~10min) most of the day, but **casual riders took longer rides during midday**. The pattern was similar but **more pronounced for classic bikes**.	Ride duration by direction follows a similar pattern to ride distance.	No obvious pattern here, but it does appear that casual riders took shorter trips when starting from north-central neighborhoods like Sheffield & DePaul and Avondale.	Ride Duration	

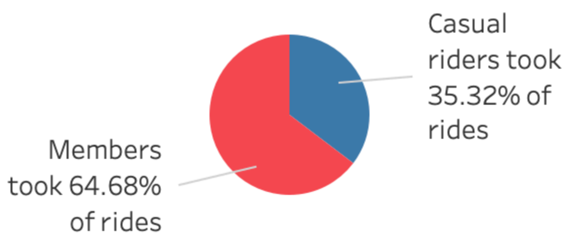
(this table is also included in my repository (https://github.com/Reed-Solomon-9/Google-Data-Certificate-Capstone-Bike-Share) as it may be too difficult to read in this format)

This gave me a somewhat comprehensive look at what is in the dataset. In terms of completeness, the logical next step would be to analyze combinations of multiple qualitative columns, but that could require hundreds of separate queries and charts, and likely a lot of wasted time.

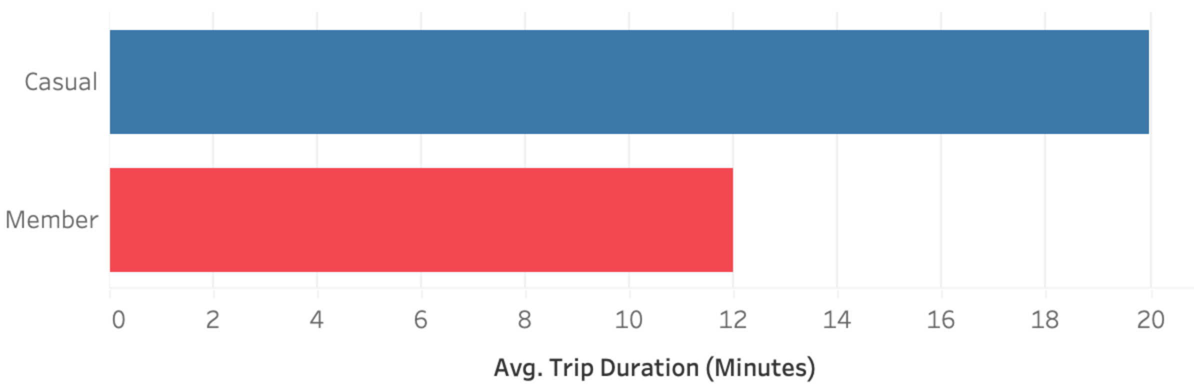
This first analysis yielded some concrete insights about rider behavior, which I outline below, supported with visualizations from Tableau.

Members Vs. Casual Riders Breakdown

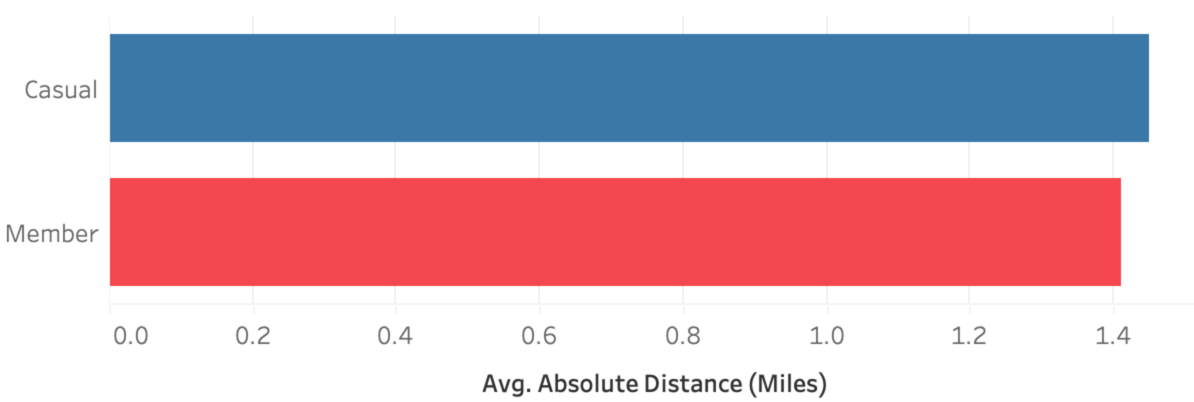
Nearly **two-thirds of rides** were by **members**



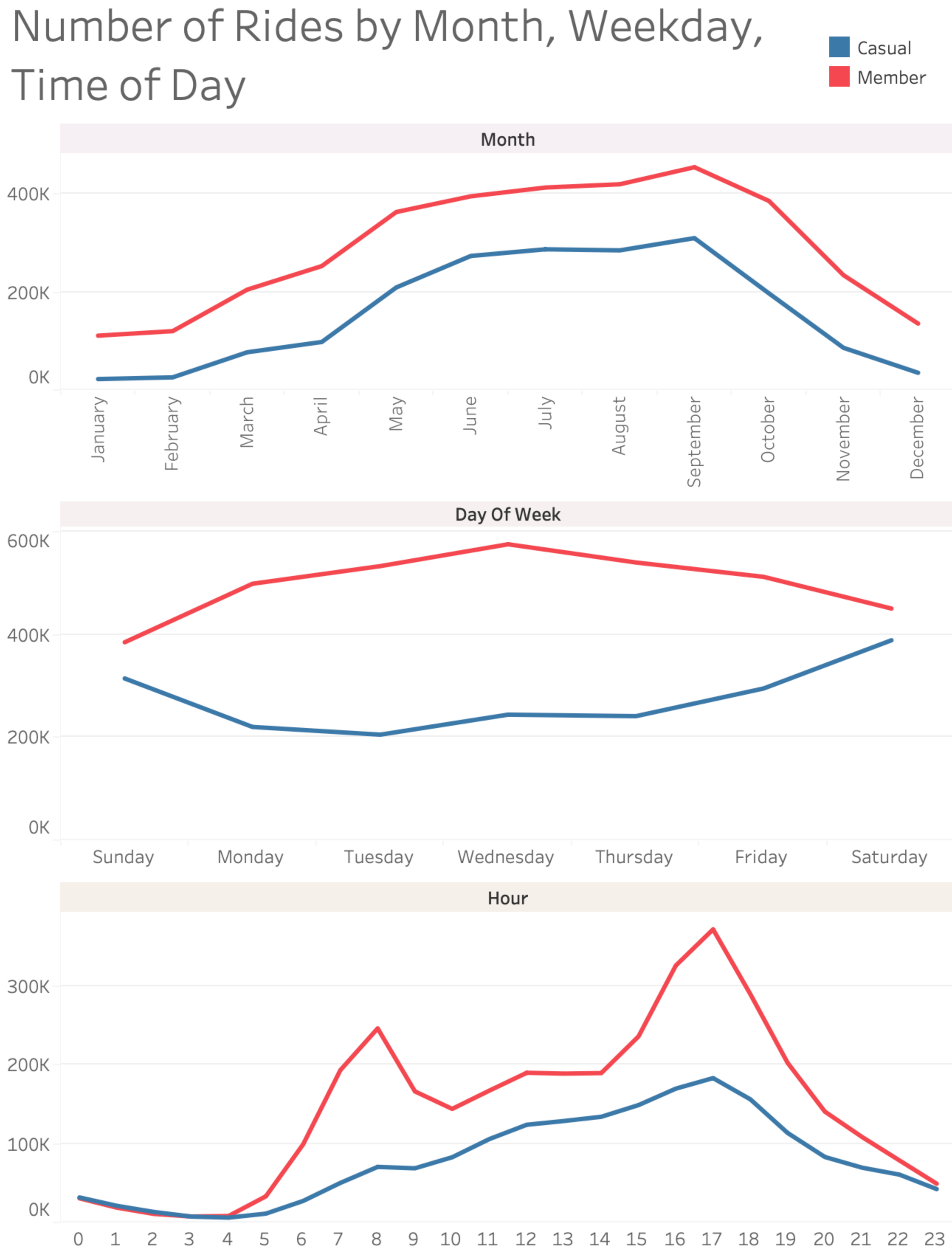
Casual riders **rode for longer**



Both groups traveled **about the same distance**

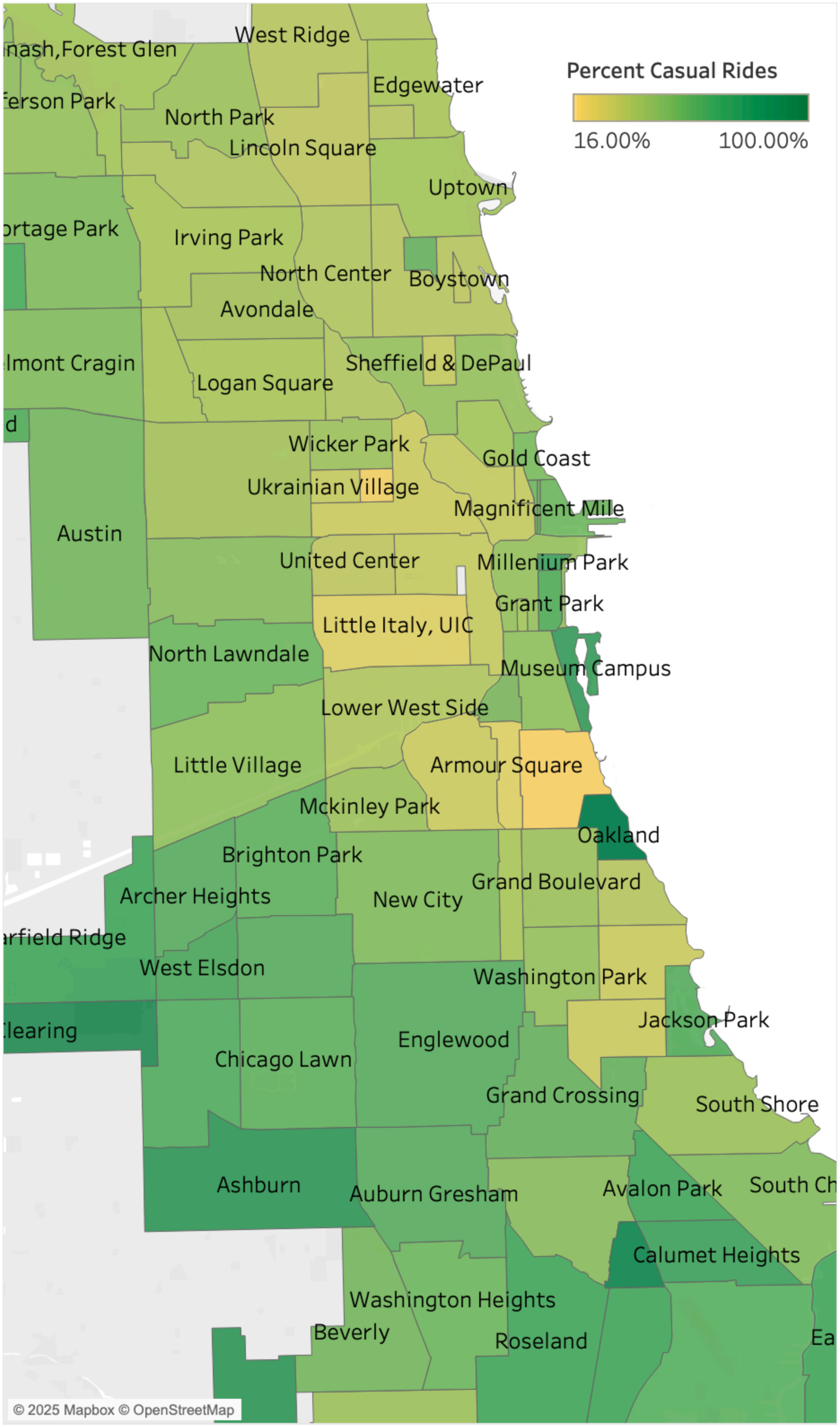


Approximately 65% of trips were taken by subscription members. Casual riders took longer trips (in terms of ride duration) than subscription members. Both groups travelled similar distances.



- Number of rides followed a similar monthly pattern for members and casual riders.
- By week, there was an opposite pattern: **casual** rides were **higher on weekends**, while **member** rides were **higher on weekdays**
- By hour, both groups took the most rides around 5pm (17:00 hour). However, for subscription members there were **two clear peaks**, one around 8am and one around 5pm.

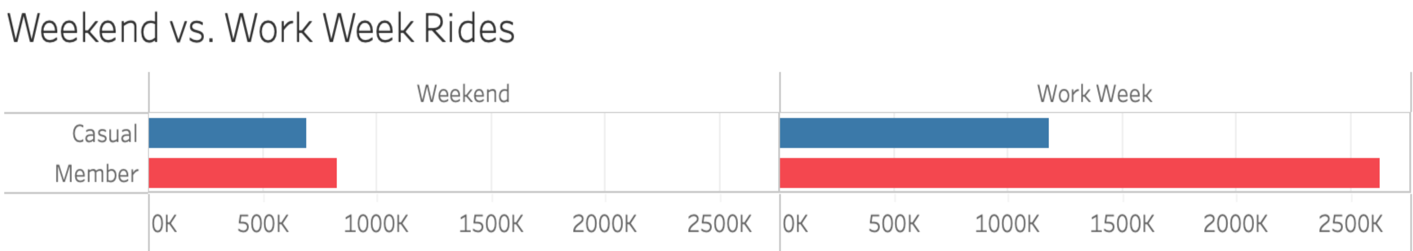
Neighborhood Member Percentage



- Percentage of casual rides varies a lot between neighborhoods.
- There was a high share of casual rides in neighborhoods in the **south** and **west** of the city, and generally far from the highest-density areas.
- **Central neighborhoods** with lots of landmarks also had **relatively high** shares of casual rides.
- **Inner ring neighborhoods** just outside of the city center, particularly to the north, had the **lowest** share of casual riders (highest share of subscribers).

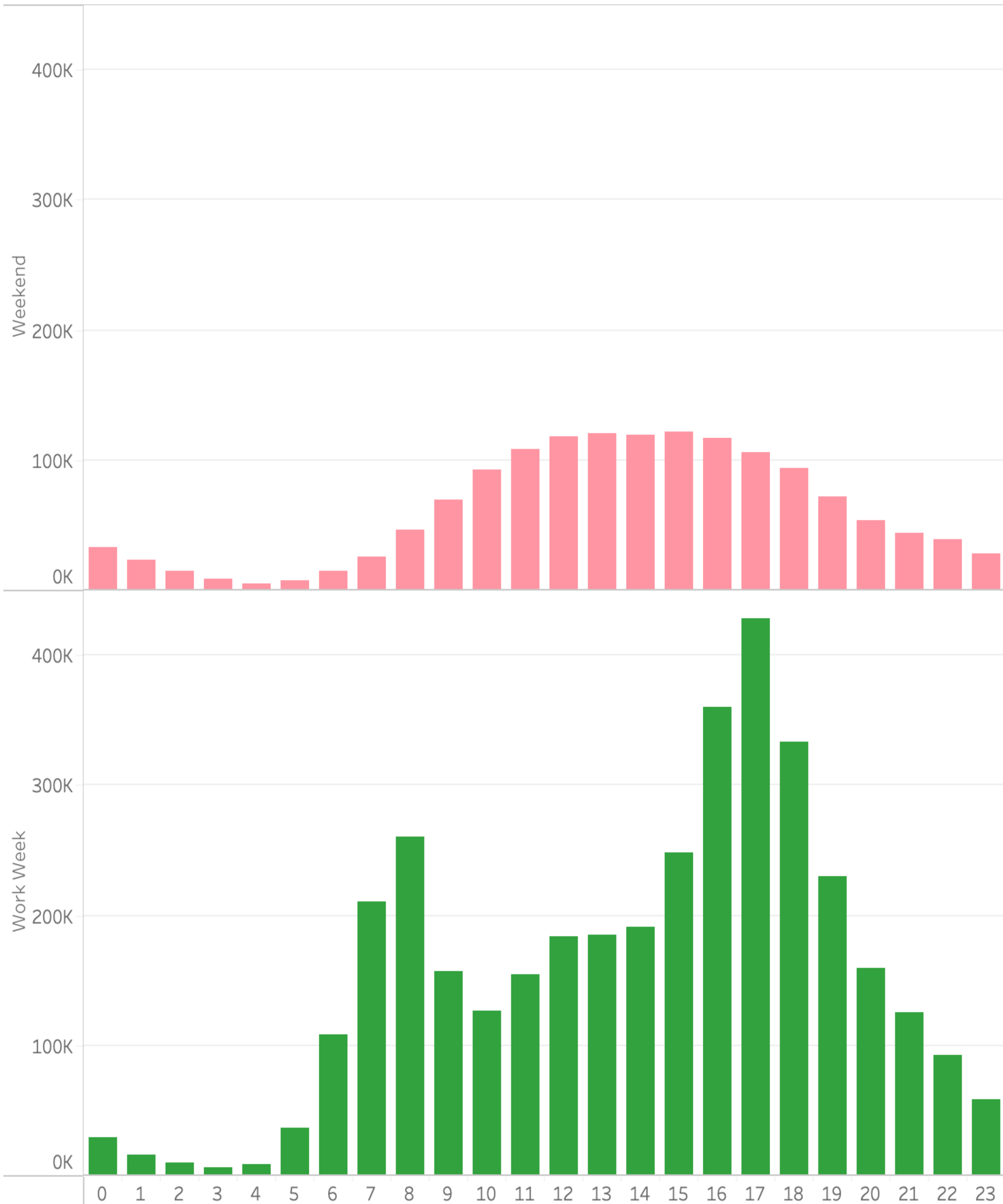
The other relationships between variables were less relevant to this analysis, so I focused on combining the most useful qualitative data.

I grouped the data by both day of week and time of day. First I determined that most of the difference in the numbers of casual rides and subscription members happened during the work week.



Then I established a distinct difference between rider behavior during the work week (Monday through Friday) and the weekend. There was a clear pattern to rider behavior between these days, and though the pattern was more pronounced for subscription members, it was still clearly observable for casual riders.

Weekend vs. Work Week Rides

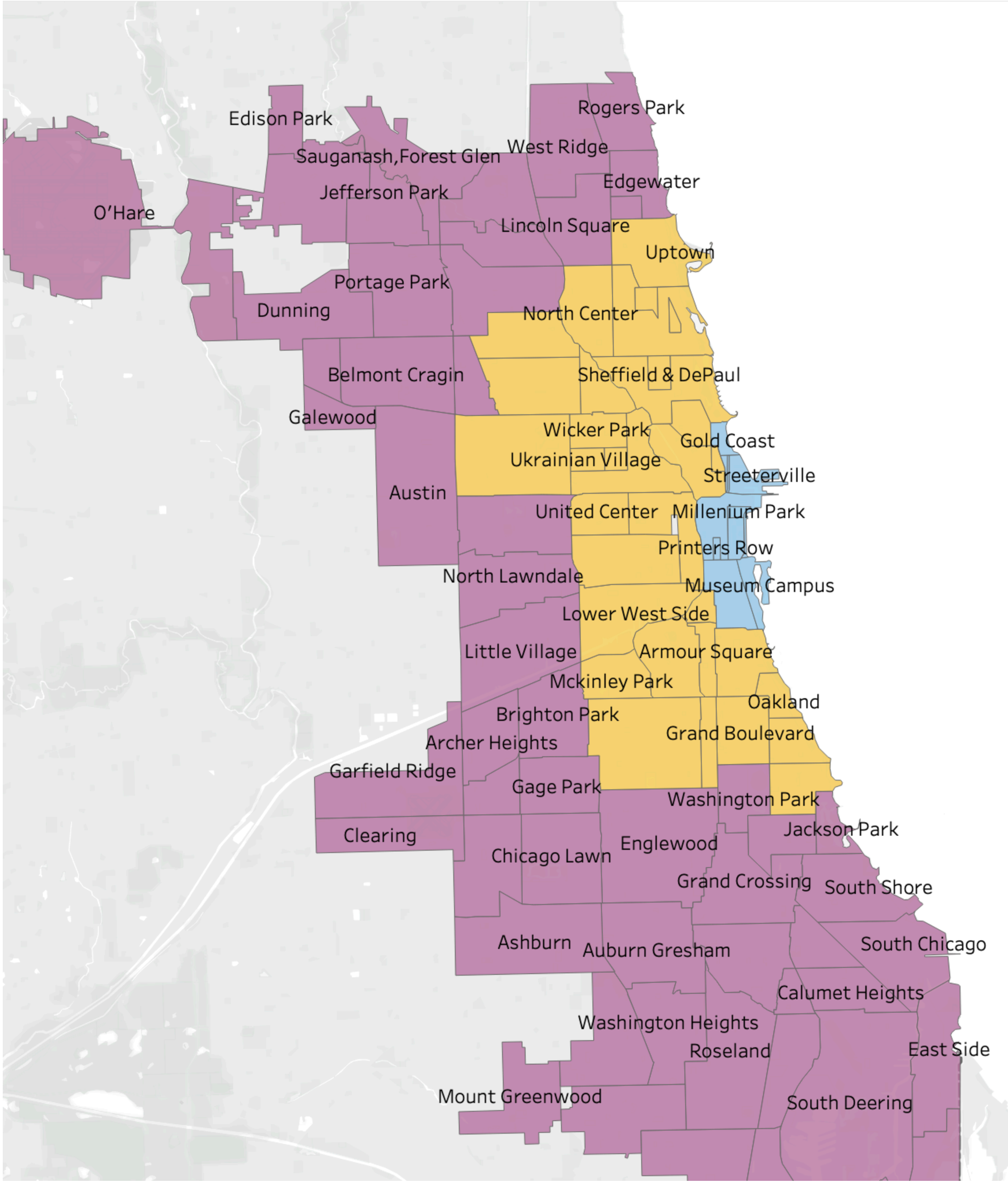


Weekend rides gradually increased from the early morning to a flat peak from 11am to 5pm, while work week rides showed the hourly pattern where rides had a sharp peak from 7-9am, and then a larger sharp peak from 3-7pm.

I also bucketed the 97 Chicago neighborhoods into three categories: Central, Inner Ring, and Outer Ring. These categories showed noticeable differences in rider behavior.

On the map, this is how they are oriented:

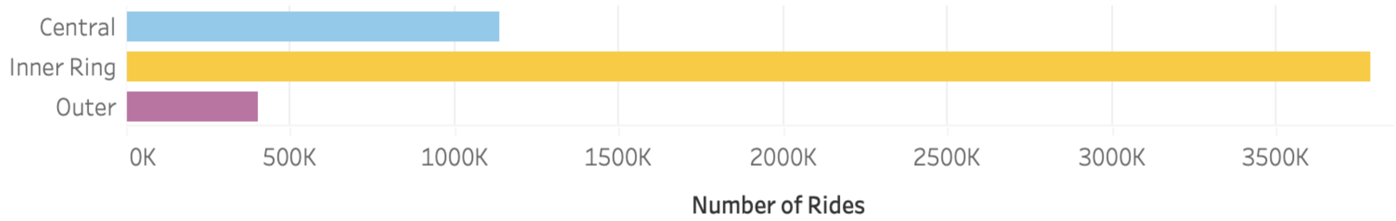
Neighborhood Category Map



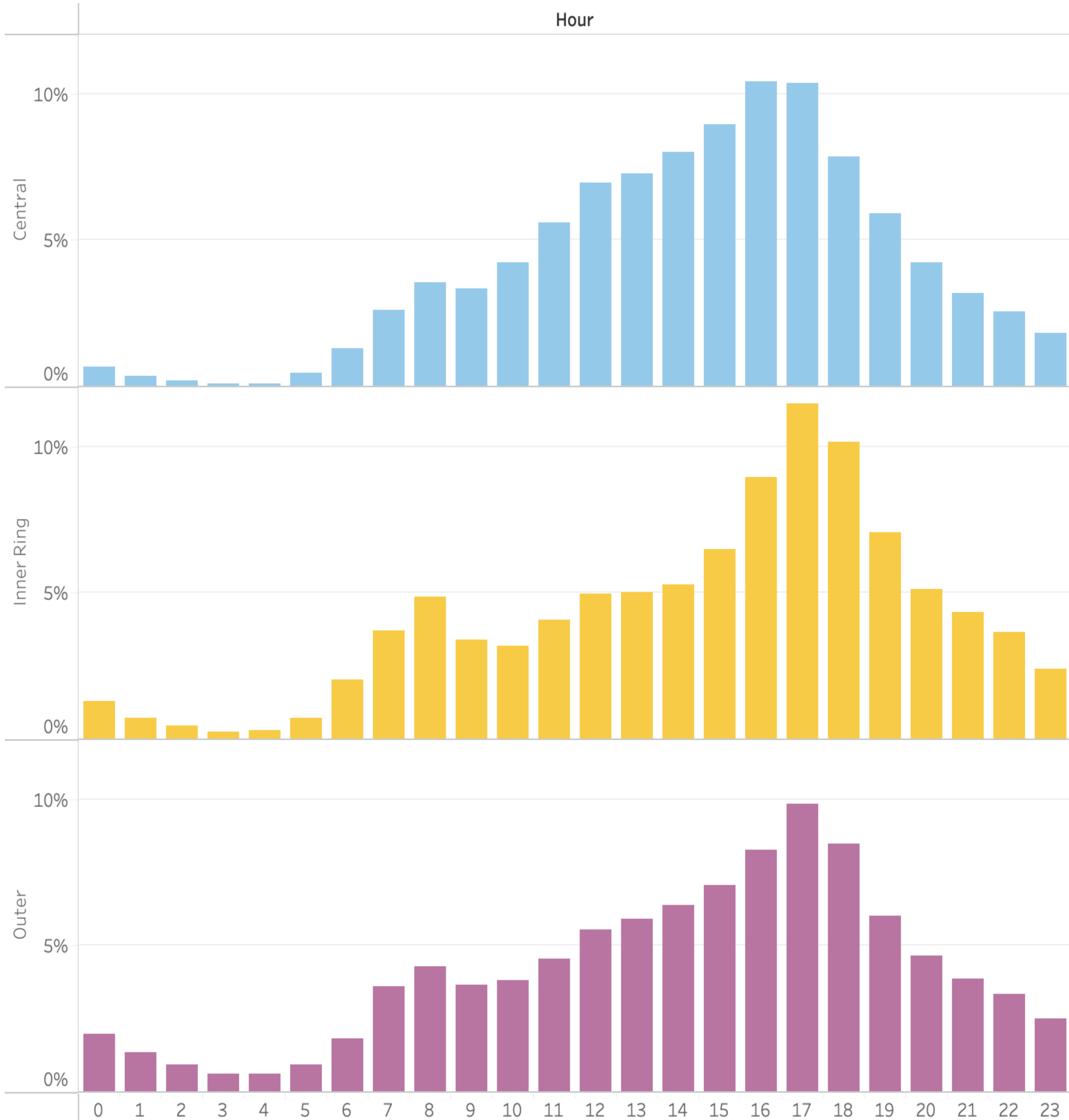
Rides are broken down between these neighborhood categories as shown above.

Below is how they were distributed:

Over 71% of all rides were in inner ring neighborhoods.



The commute-time pattern is visible for casual riders in all categories, but most pronounced in inner ring neighborhoods.



Section 5: Conclusions

- Behavior differs significantly between casual riders and subscription members by **weekday.** Many more midweek/work week riders are subscription members, while casual riders are more frequent users on weekends. This implies that subscription members use the bikes to commute to work.
- Behavior also differs by **neighborhood.** Casual rides are most frequent as a share of all rides in the most **high-traffic and tourist friendly neighborhoods** like Millennium Park, Wrigleyville, Grant Park, Museum

Campus, Streeterville, Gold Coast; as well as **outer neighborhoods** with far lower ride frequency like Lawndale, Brighton Park, and Ashburn.
- Ride behavior follows a consistent pattern by hour throughout the day. Rides are highest on **weekdays** (Monday through Friday) **during morning and especially afternoon commute hours**, while the daily distribution is a smooth curve on weekends (Saturday and Sunday) with the peak occurring from **11am to 5pm**.
- Subscription rides are most frequent as a share of all rides in **inner ring neighborhoods** like Little Italy, West Loop, East Village, and Douglas.
- A **large majority (over 71%)** of all bike share users reside in these **inner ring neighborhoods**. These neighborhoods have strong subscription uptake, but they still contain a solid majority of casual riders.
- I've pinpointed a few neighborhoods, particularly **Lakeview** and **Lincoln Park**, that have a combination of large usage and relatively low subscription rate that may be worth targeting with neighborhood-specific promotions.
- This dataset does not have any information on the riders themselves- only the information about each individual ride. Because of this, it is necessary to infer group differences between members and casual riders based on each group's average behavior.
- Subscription uptake is quite strong, meaning that gains will have to be made on the margins- either to increase already high subscription use in the inner ring neighborhoods, to convert a relatively small population in outer neighborhoods, or to convert the casual weekend users in central neighborhoods who are not just temporarily visiting the city.
- Riders starting from **outer neighborhoods** make up **less than 8%** of total rides. They have generally higher percentages of casual users, so it looks like there could be some opportunity for subscriber growth, but I think it's important to note that from these locations, it is hard to position bike share rides as a viable alternative to the rail system for commuters.

Section 6: Business Recommendations

After reviewing this dataset in a comprehensive manner, I propose the following business decisions. These may rely on other factors, but they are rooted in insights from this data.

- Target the **highest traffic inner ring neighborhoods** with online targeted advertisements and neighborhood-specific promotions, and perhaps physical advertisements such as billboards/wallscapes/mural advertising. Chicago residents take pride in their neighborhoods, and the neighborhood names are highly recognizable. I believe they will respond well to neighborhood-specific appeals.
- Target **early morning weekday riders who are not subscribed**. Casual riders ought to be shown how they can benefit from a subscription when they are making the decision to purchase a ride. Giving them this option for early rides will target the riders who would be most likely to want a subscription.
- **Present subscriptions differently on weekends versus weekdays**. Weekend subscriptions could be offered as a way to enjoy the city, while the practicality of weekday commute hour subscriptions could be emphasized.
- Consider a **"nights and weekends" subscription option**. This could be usable outside of commute hours, to segment the market and allow the offerings to be priced more optimally.
- Consider a **discount for longer rides in outer neighborhoods**. Riders took longer rides from these neighborhoods, and they have significantly lower rates of subscription. If financially feasible, a discount could allow the bikes to be competitive with rail transportation. This could potentially be more efficient than advertising due to the lower concentration of riders.