



Cyclistic Case Study

How Does a Bike-Share Navigate Speedy Success?

Presented by: Chris Aguirre



Agenda

- Problem/Opportunity Statement
- Business Task
- Data Used for Analysis
- Analysis Plan
- Analysis Results
- Recommendations
- Questions



Problem/Opportunity Statement

Cyclistic's customers include both casual riders and annual members. Annual members are much more profitable and Cyclistic wants to maximize conversion to annual members to help secure the highest possible long-term growth for the business.



Business Task

Determine how casual riders and annual members use Cyclistic bikes differently. Use these insights to provide top 3 recommendations on how to tailor Cyclistic's marketing strategy to maximize annual member conversion.



Data Used for Analysis

- **Cyclistic Historical Trip Data (07/22 – 06/23)**
 - <https://divvy-tripdata.s3.amazonaws.com/index.html>
 - “This data was made available by Motivate International Inc.”



Analysis Plan

Data Cleaning/Wrangling and Feature Engineering

1. Load monthly datasets from JUL 2021 – JUL 2023
 - Create one quarter (3QFY23), one year (JUL 2022 – JUN 2023), and two year dataframes
2. Feature Engineering (new columns):
 - Separate time/date
 - Trip duration
 - Day of week
 - Time of day (early morning, mid-morning, late morning, etc.)
 - Weekend boolean (1 or 0)
 - Start/End Station Name/Id Combo
 - AVG Start/End Lat/Long separate and combo
3. Check for null values
 - Impute null values where possible using dictionaries for start/end station names/ids and end lats/longs
 - Create separate dataset with remaining null values dropped
4. Trim any extra whitespace in dataset
5. Check for misspellings in entries
6. Check for outliers



Analysis Plan (Cont'd)

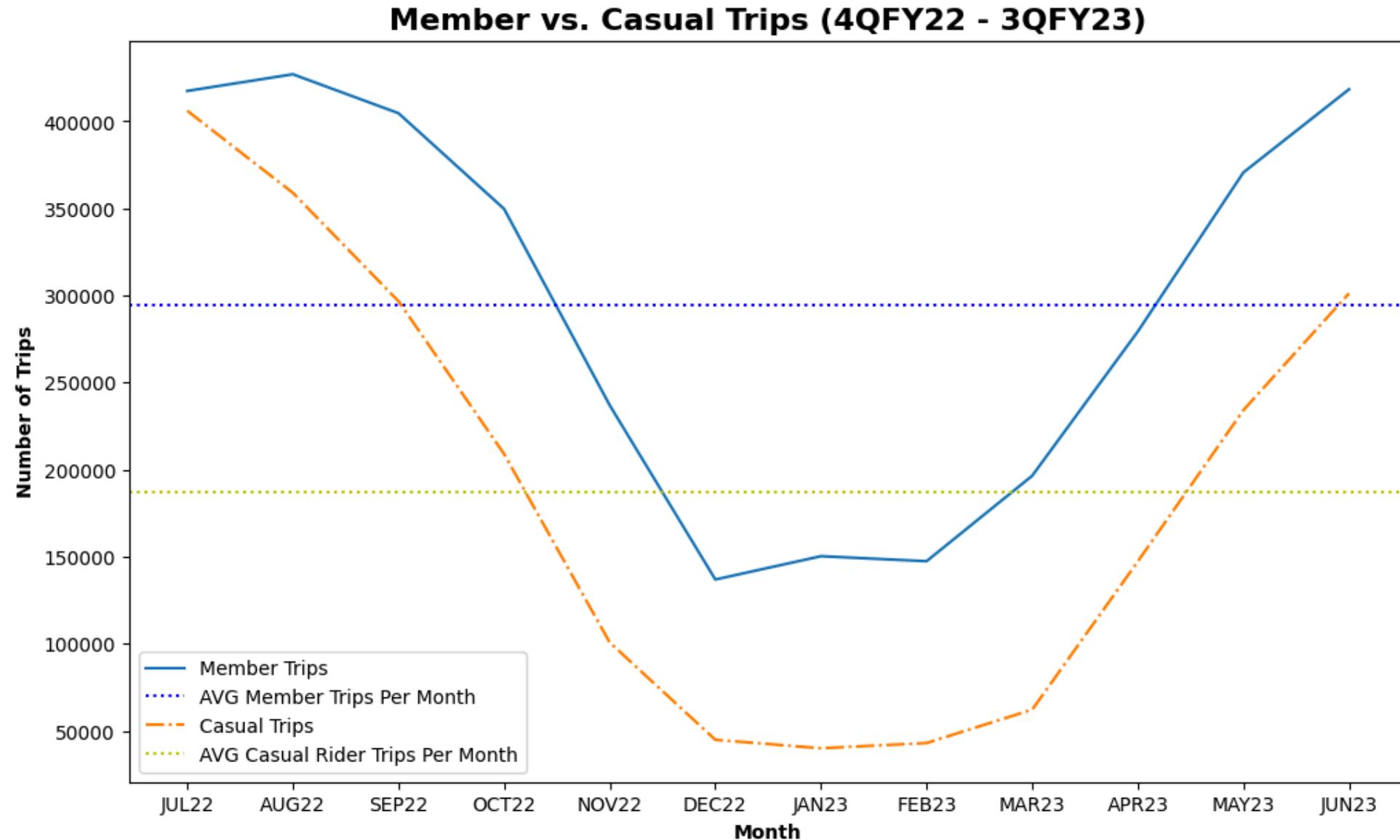
EDA/Machine Learning for Feature Importances

1. Summary Statistics from Overall, Member, Casual Riders dataframes
2. Encode dataframe and plot correlation matrix/heatmap
3. # Trips Overall, Members, Casual Riders over time
4. Member/Casual trips by top start/end stations and station combos
5. One-way vs. round-trips
6. Trip durations Members vs. Casual Riders
7. Long trip durations by start stations
8. Trip durations by bike type
9. Time of day by trip durations
10. Weekday by trip duration
11. Weekday by time of day
12. Weekend vs. Non-weekend by time of day
13. Bike type by top start/end station combos
14. Time of day by start/end station combos and weekday
15. Create geographic scatterplots and spider map using Tableau; implement filters for all relevant variables

Machine Learning Models for Feature Importances (to guide further EDA)

1. Random Forest model
 - Plot feature importances
 - Evaluate model performance
2. XGBoost model
 - Plot feature importances
 - Evaluate model performance

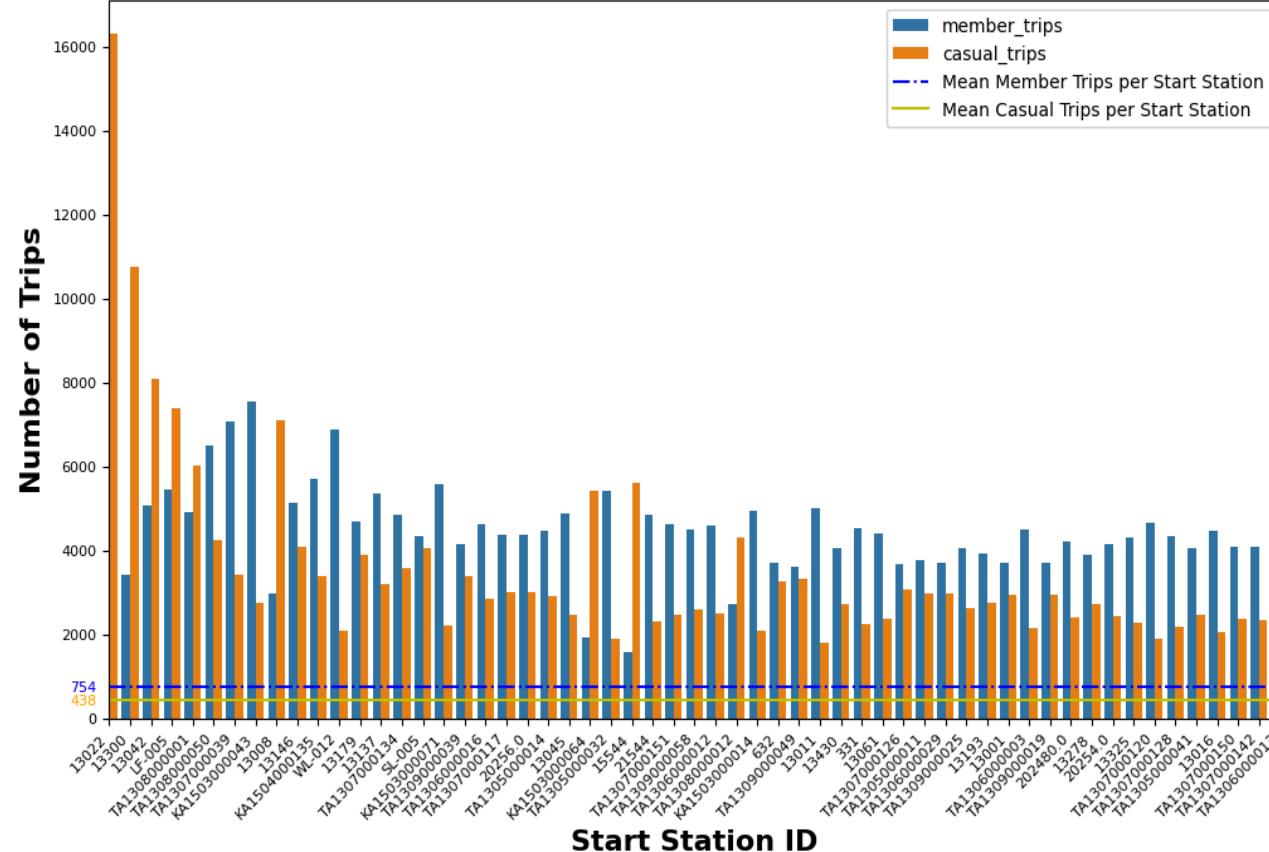
Analysis Results



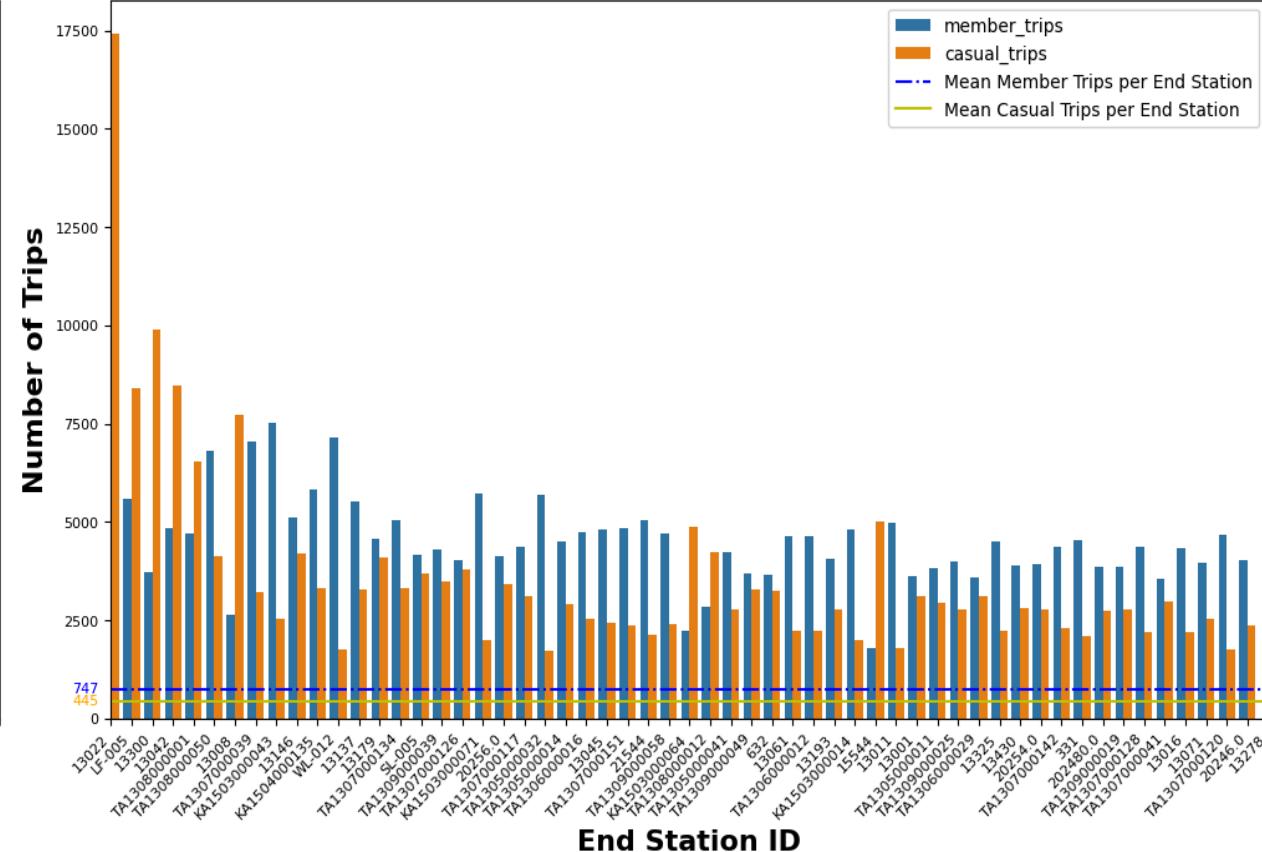
Analysis Results (Cont'd)



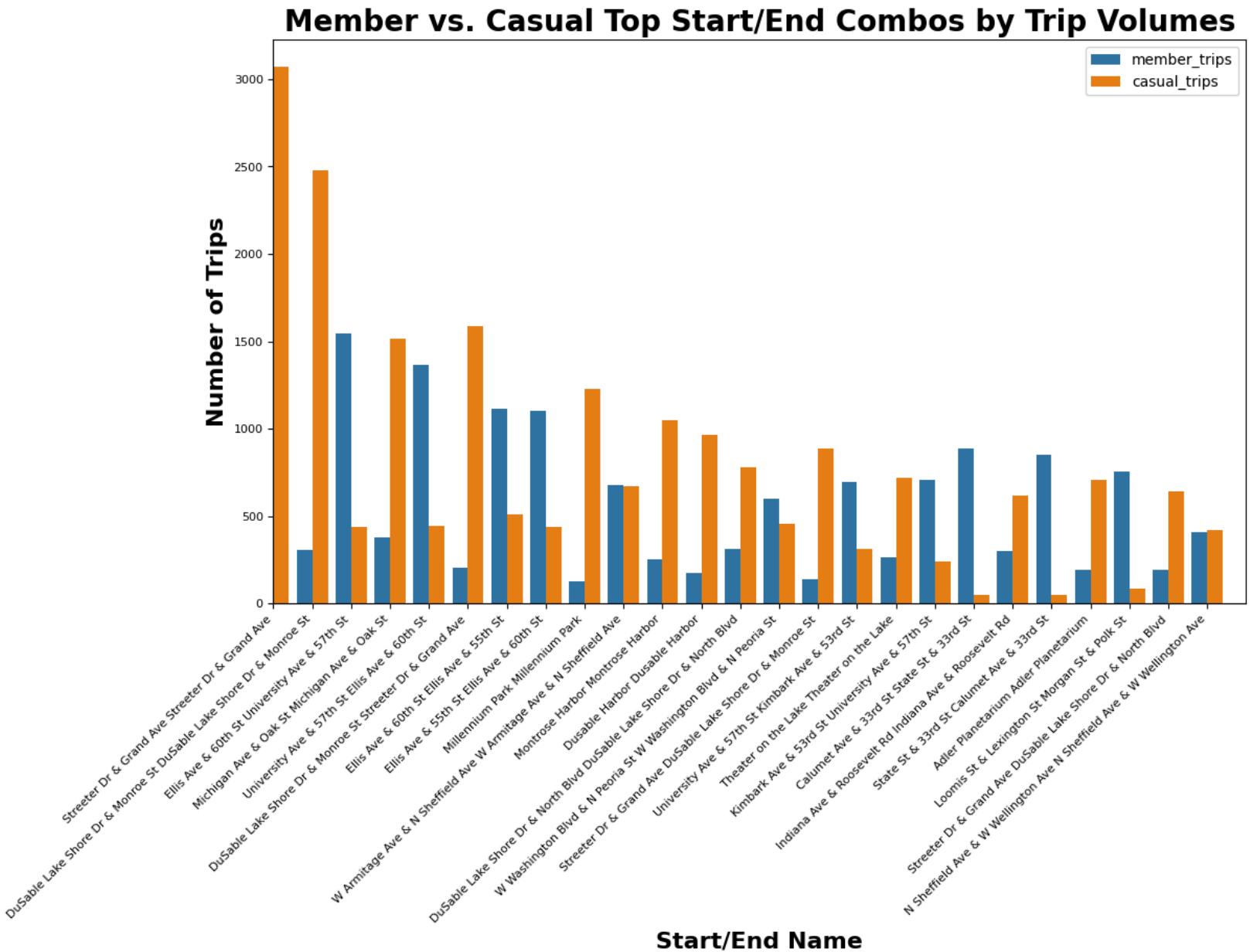
Member vs. Casual Top Start Stations by Trip Volumes



Member vs. Casual Top End Stations by Trip Volumes



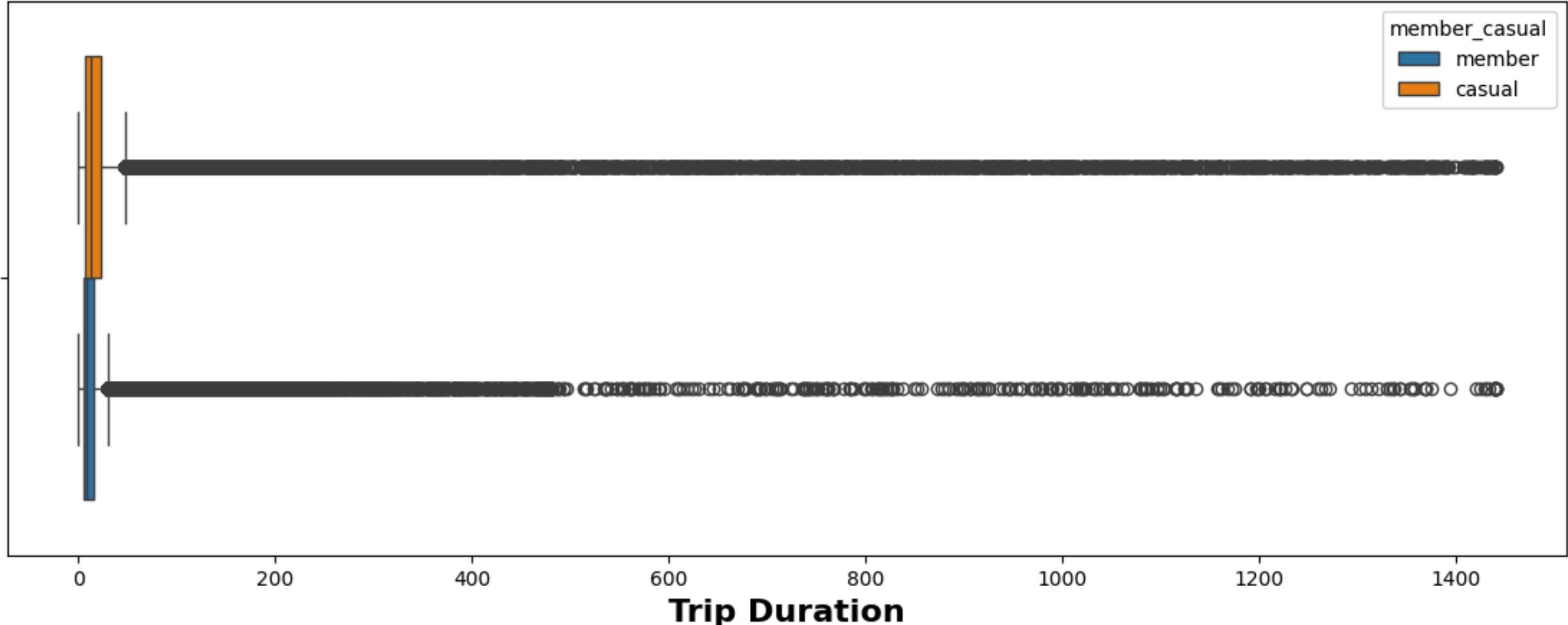
Analysis Results (Cont'd)



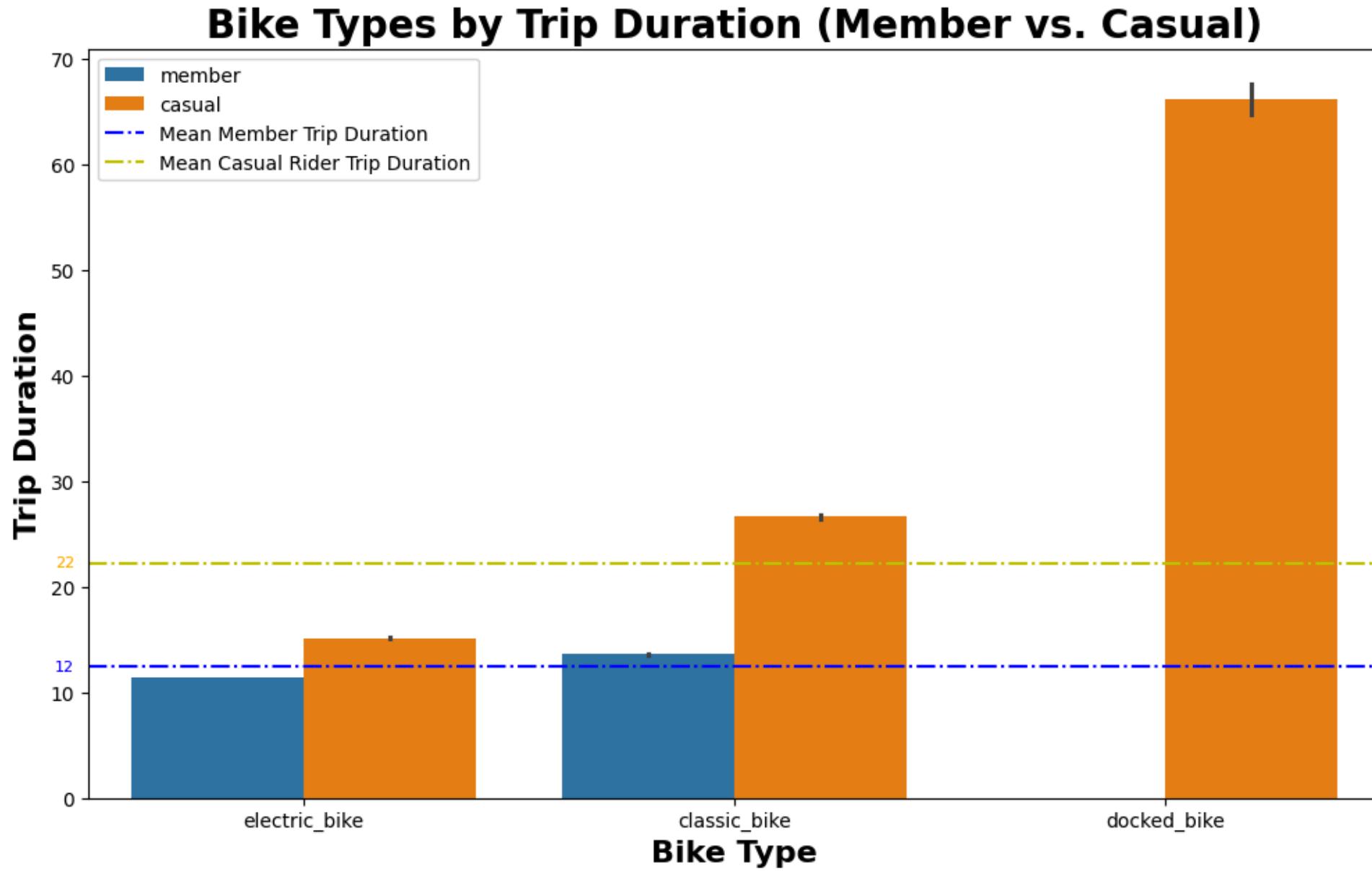
Analysis Results (Cont'd)



Member vs. Casual Rider Trip Durations



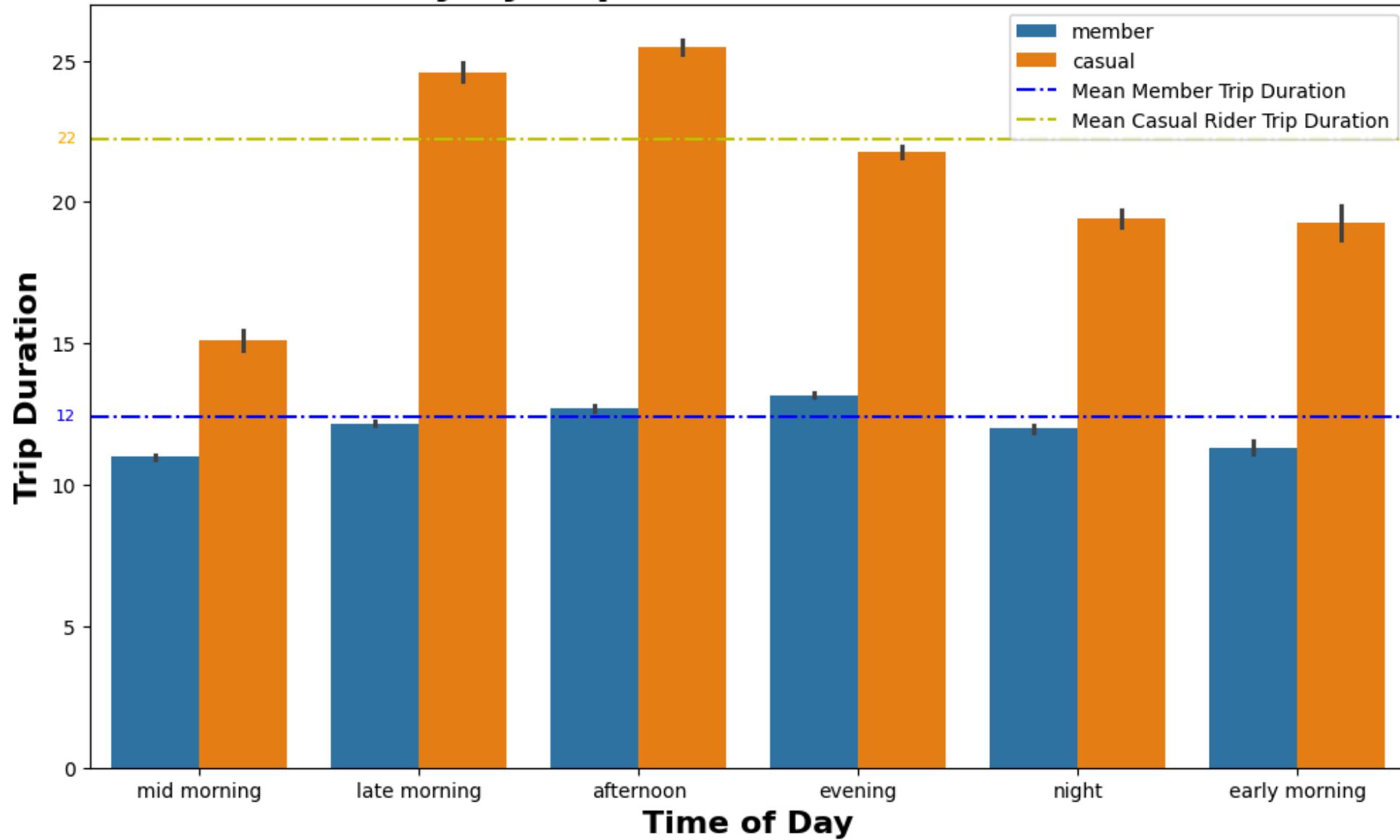
Analysis Results (Cont'd)



Analysis Results (Cont'd)



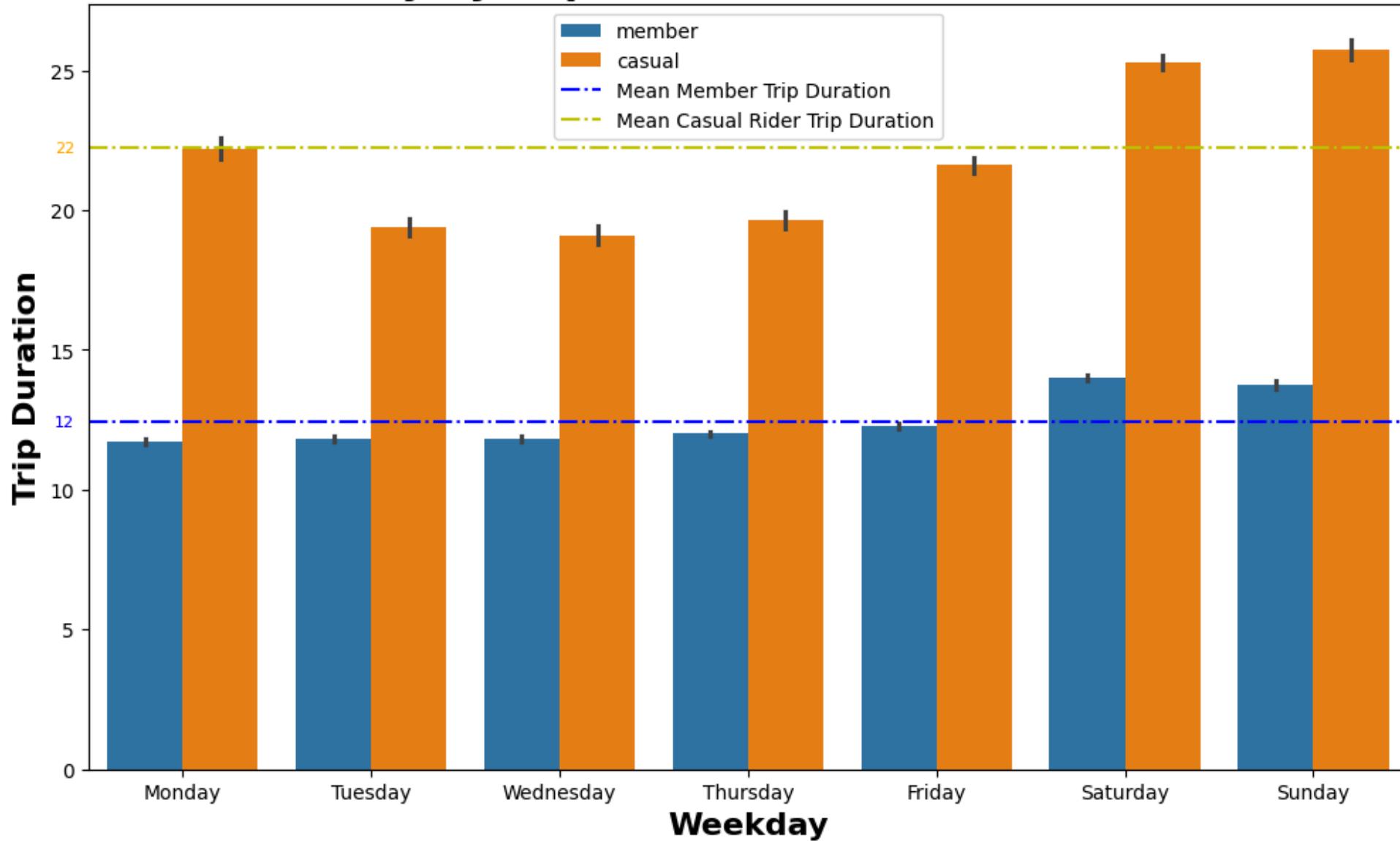
Time of Day by Trip Duration (Member vs. Casual)



Analysis Results (Cont'd)



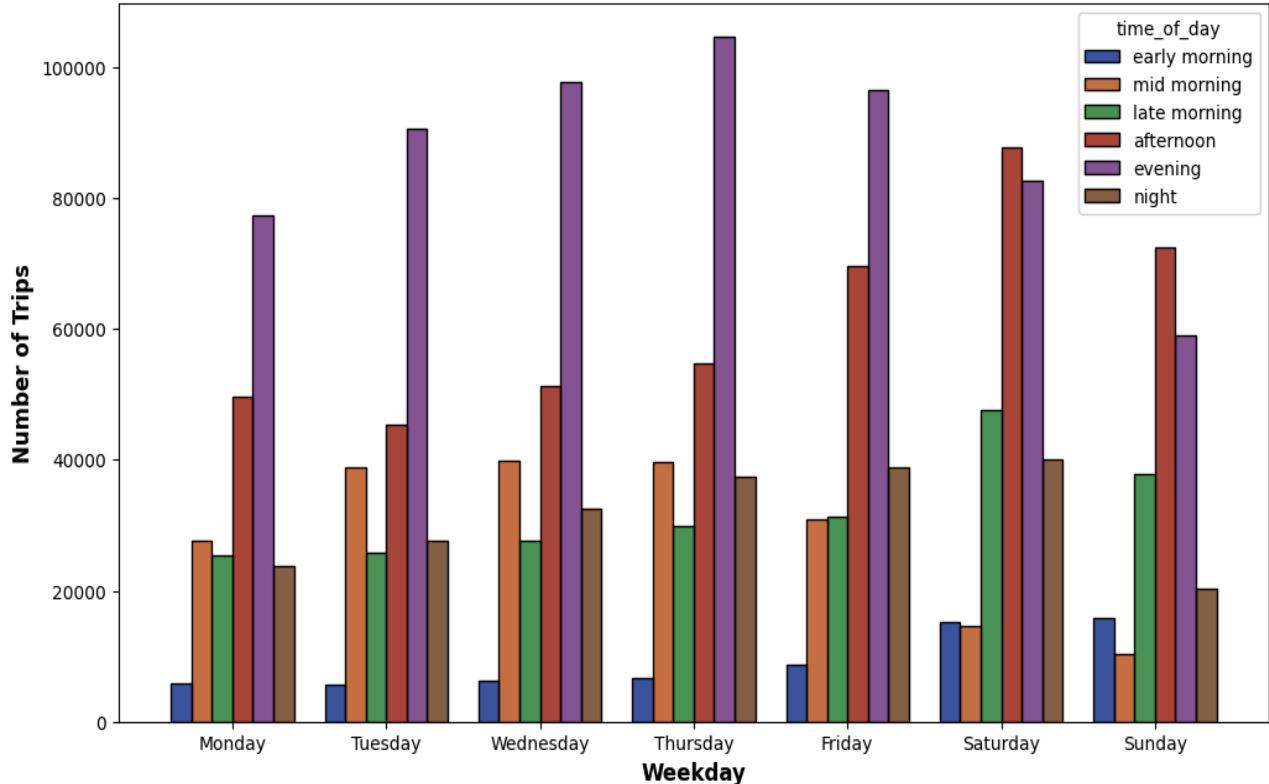
Weekday by Trip Duration (Member vs. Casual)



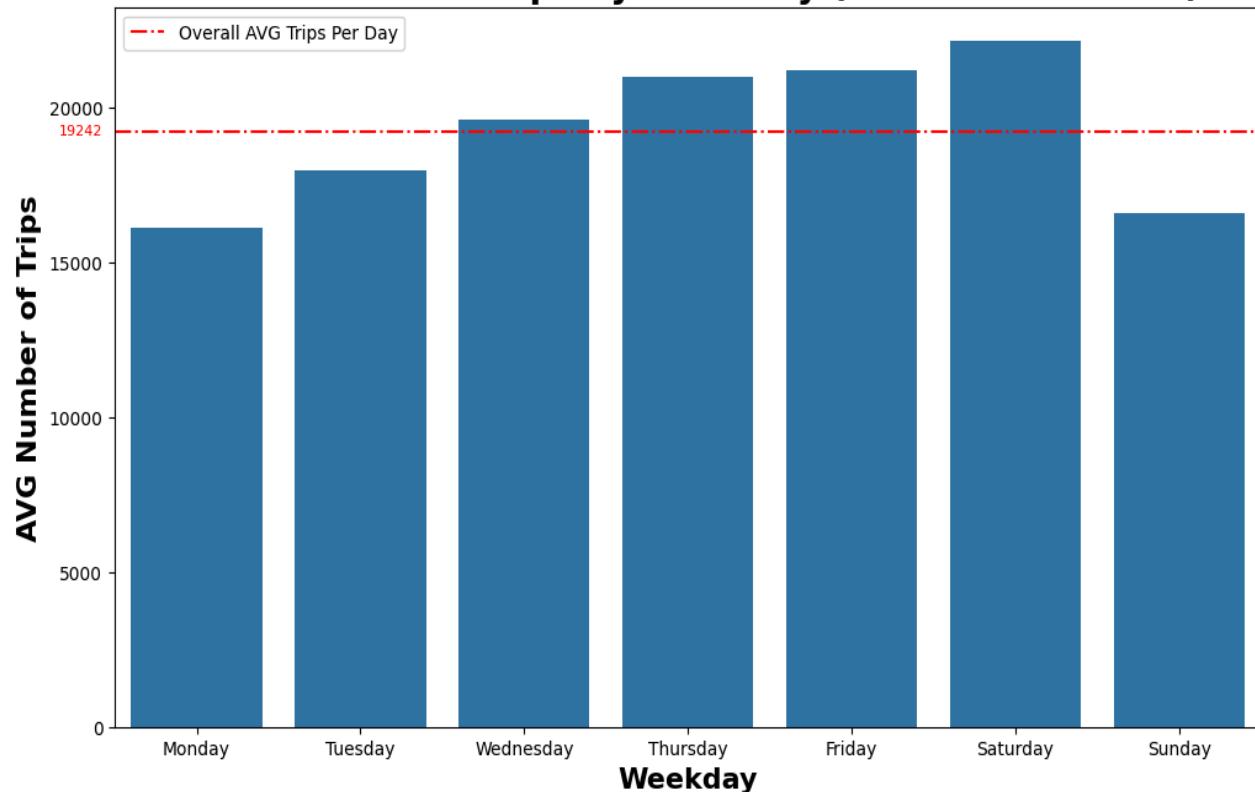
Analysis Results (Cont'd)



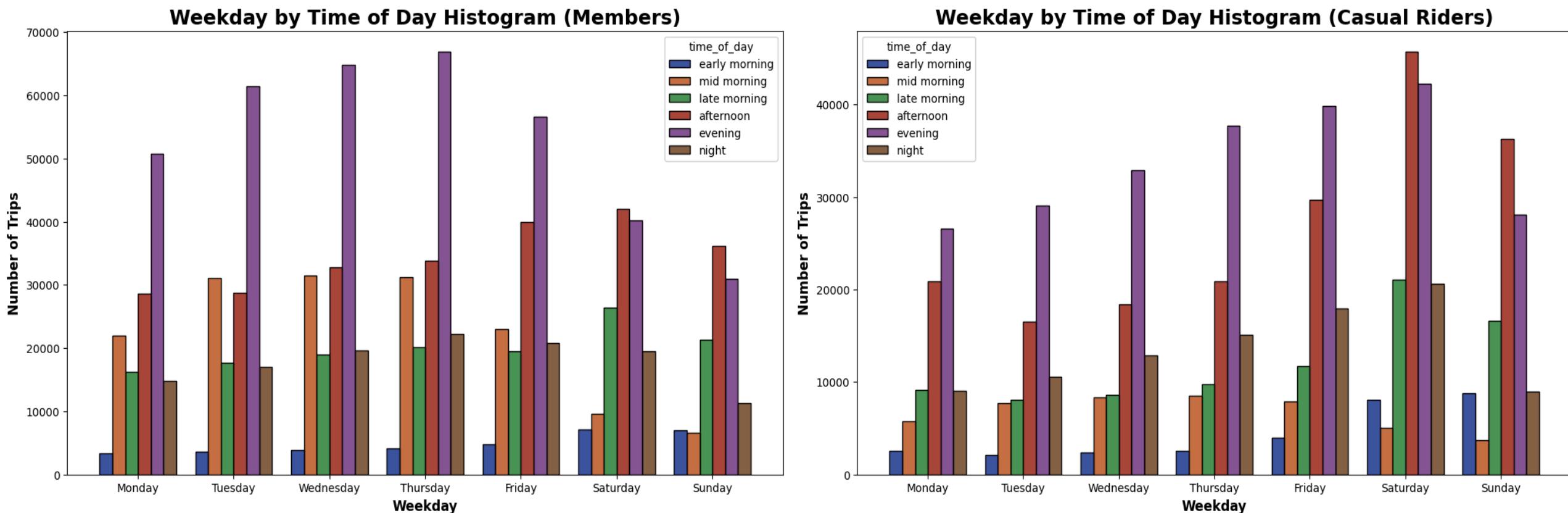
Weekday by Time of Day Histogram (Overall)



AVG Number of Trips by Weekday (Member + Casual)



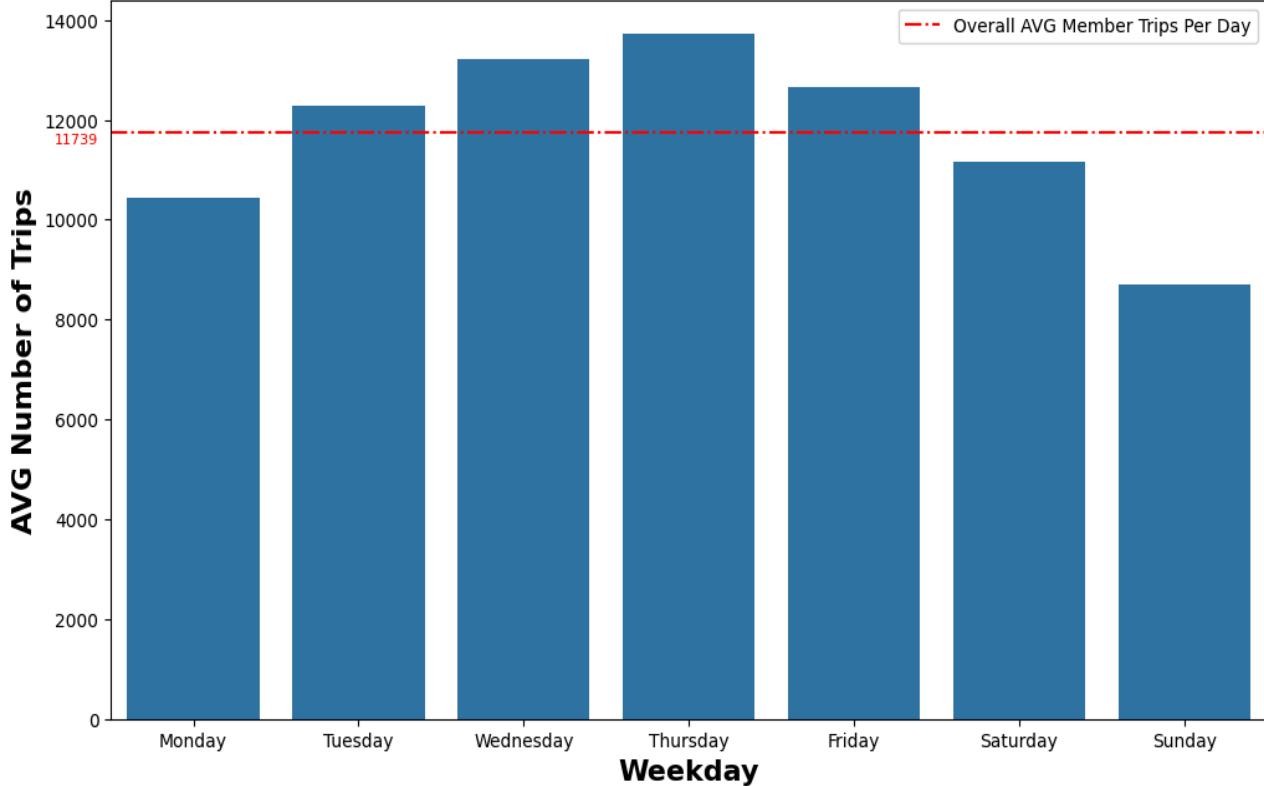
Analysis Results (Cont'd)



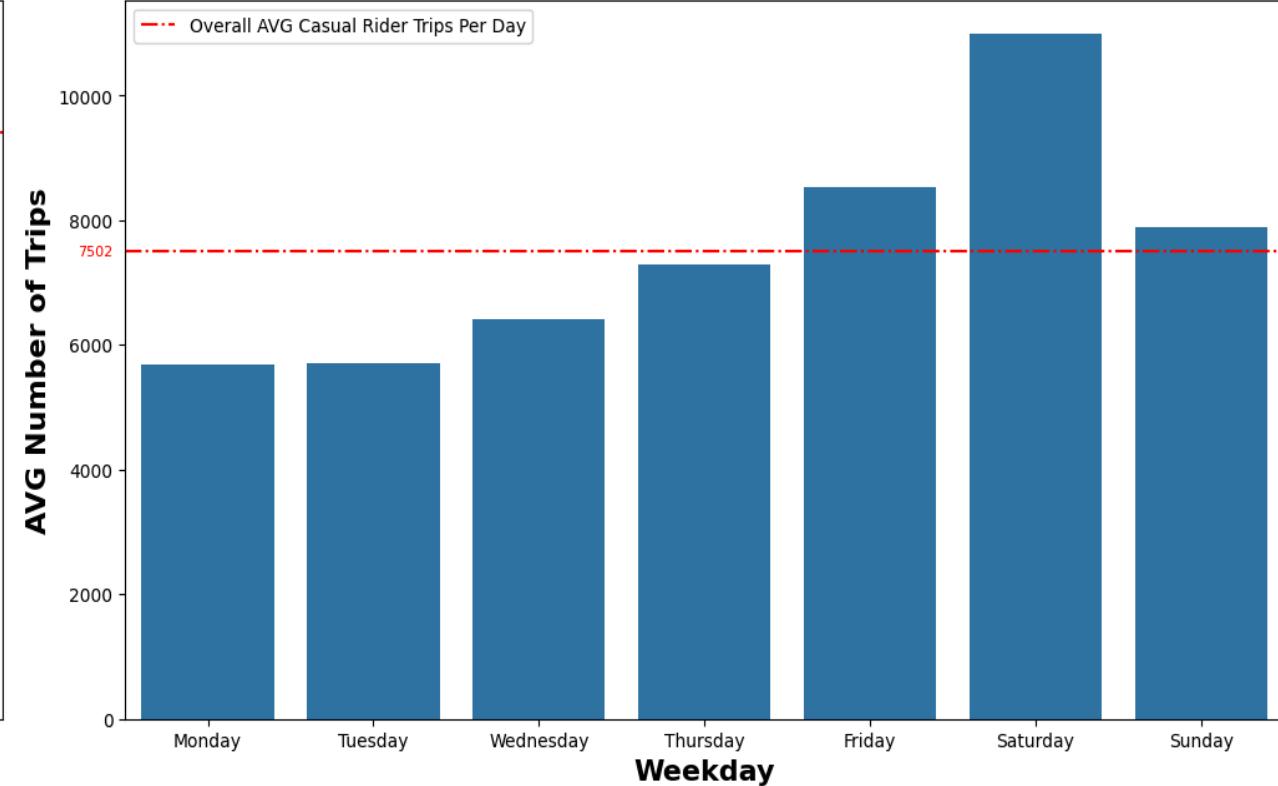
Analysis Results (Cont'd)



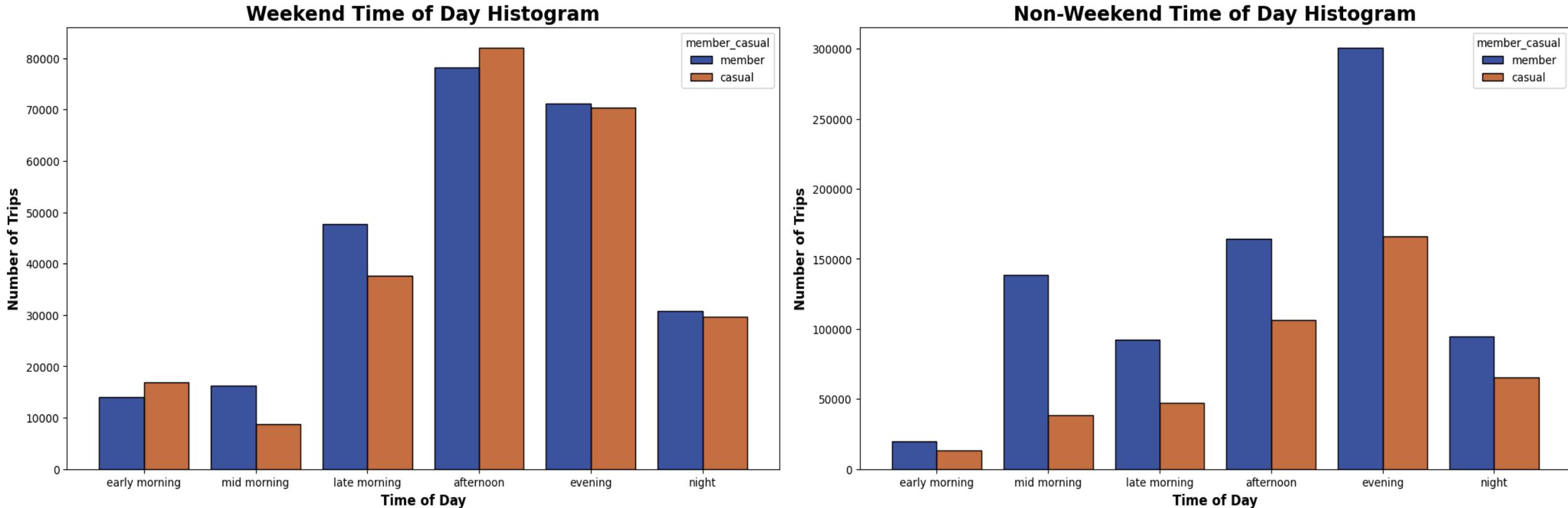
AVG Number of Trips by Weekday (Members)



AVG Number of Trips by Weekday (Casual Riders)



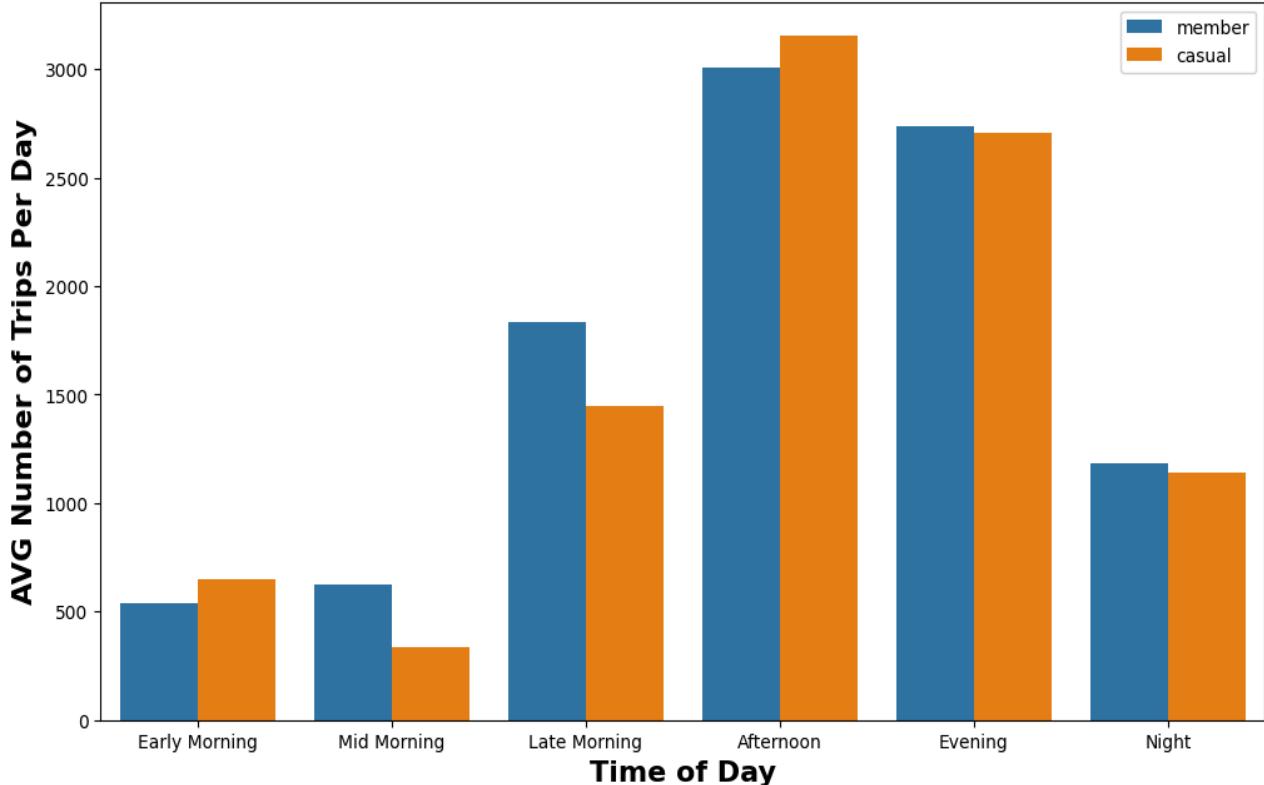
Analysis Results (Cont'd)



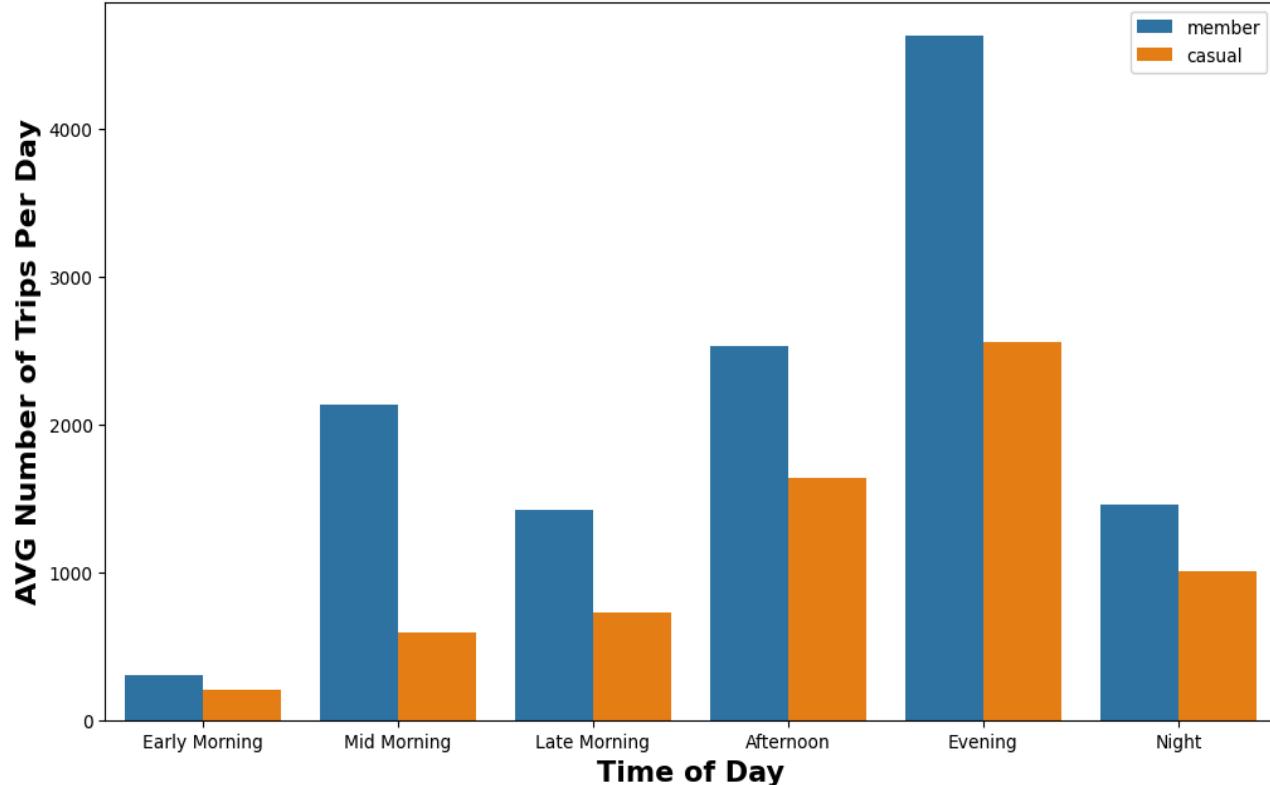
Analysis Results (Cont'd)



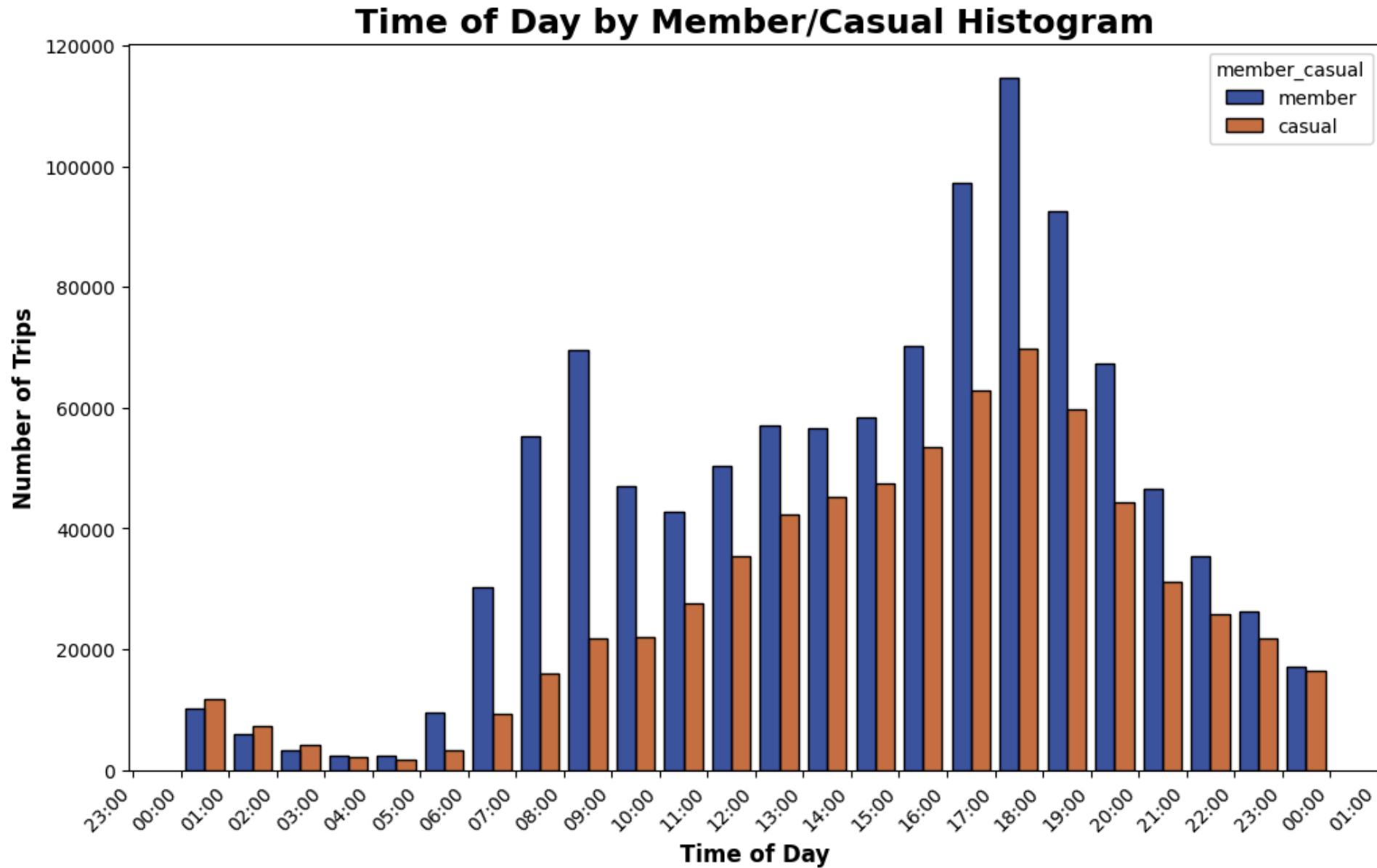
AVG Number of Trips Per Day by Time of Day (Weekend)



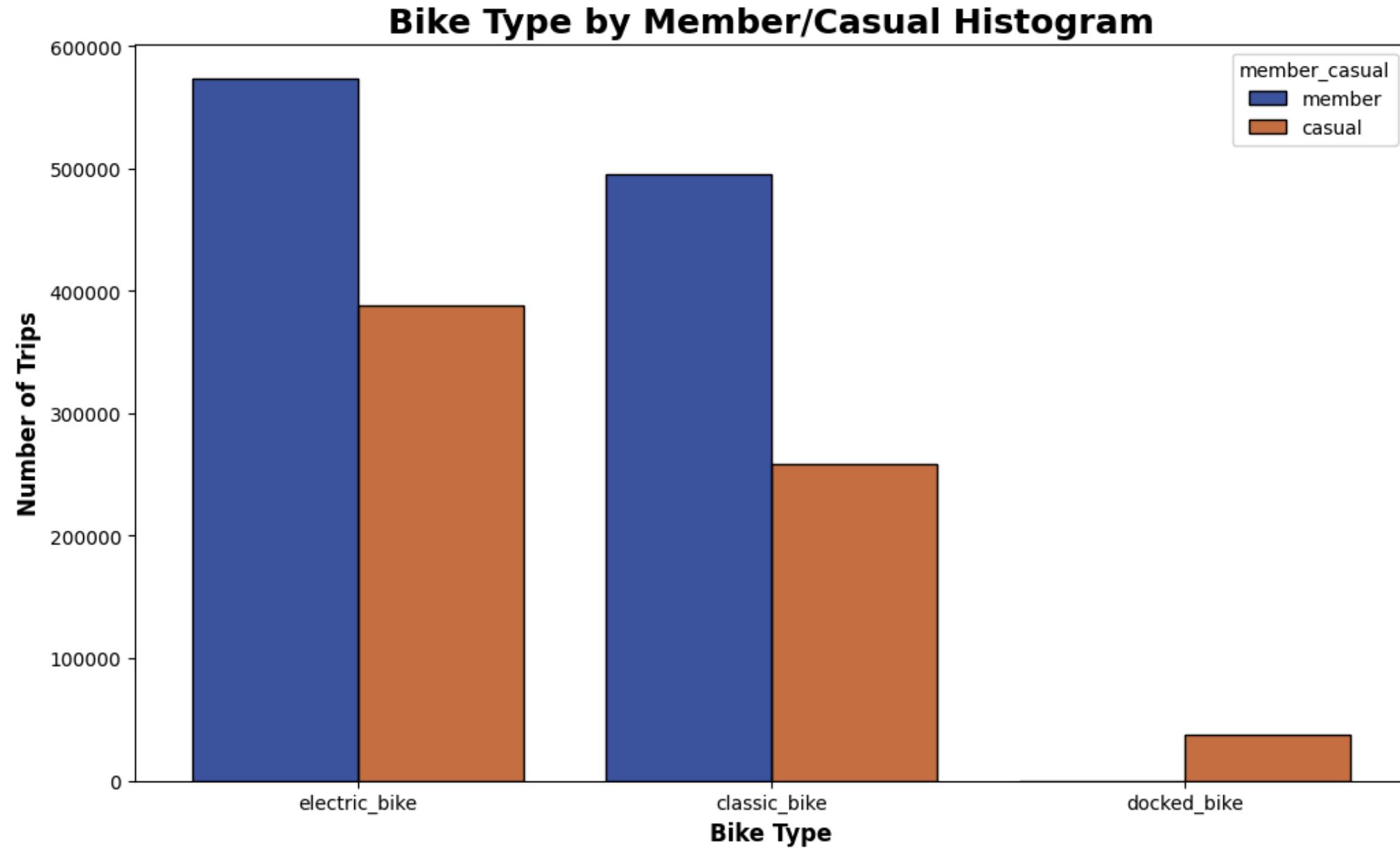
AVG Number of Trips Per Day by Time of Day (Non-Weekend)



Analysis Results (Cont'd)



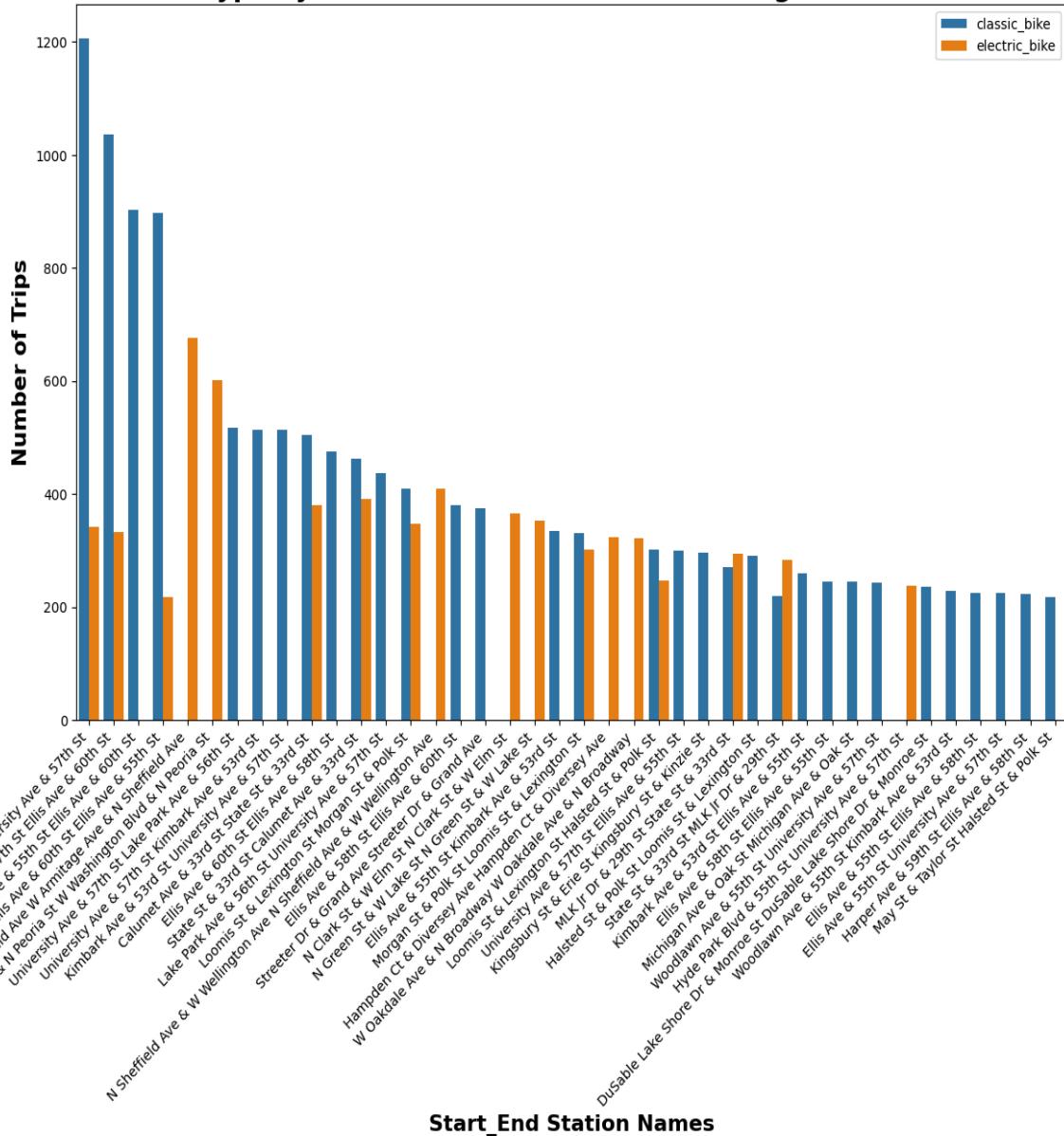
Analysis Results (Cont'd)



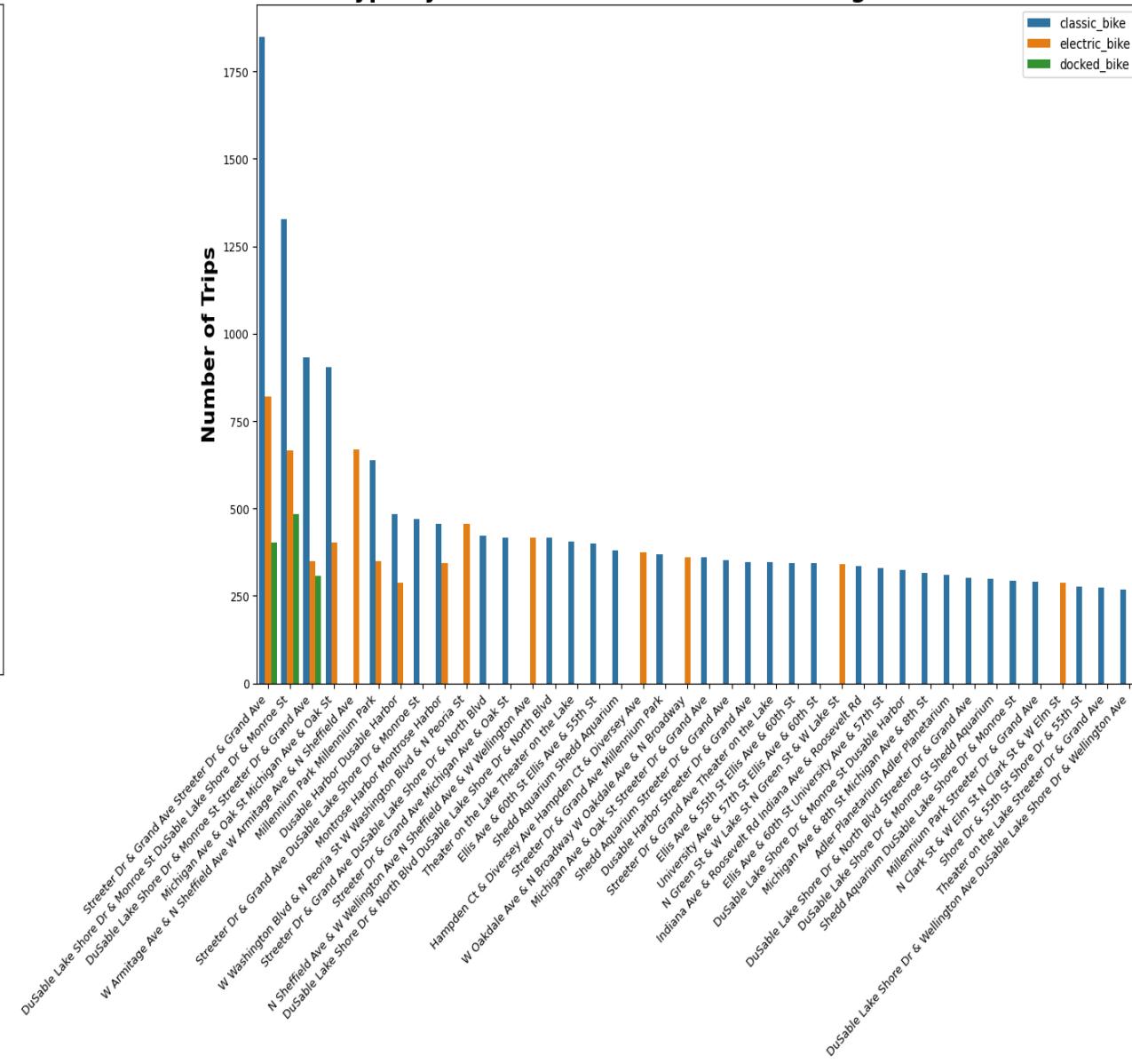
Analysis Results (Cont'd)



Bike Type by Start/End Station Combos Histogram (Members)



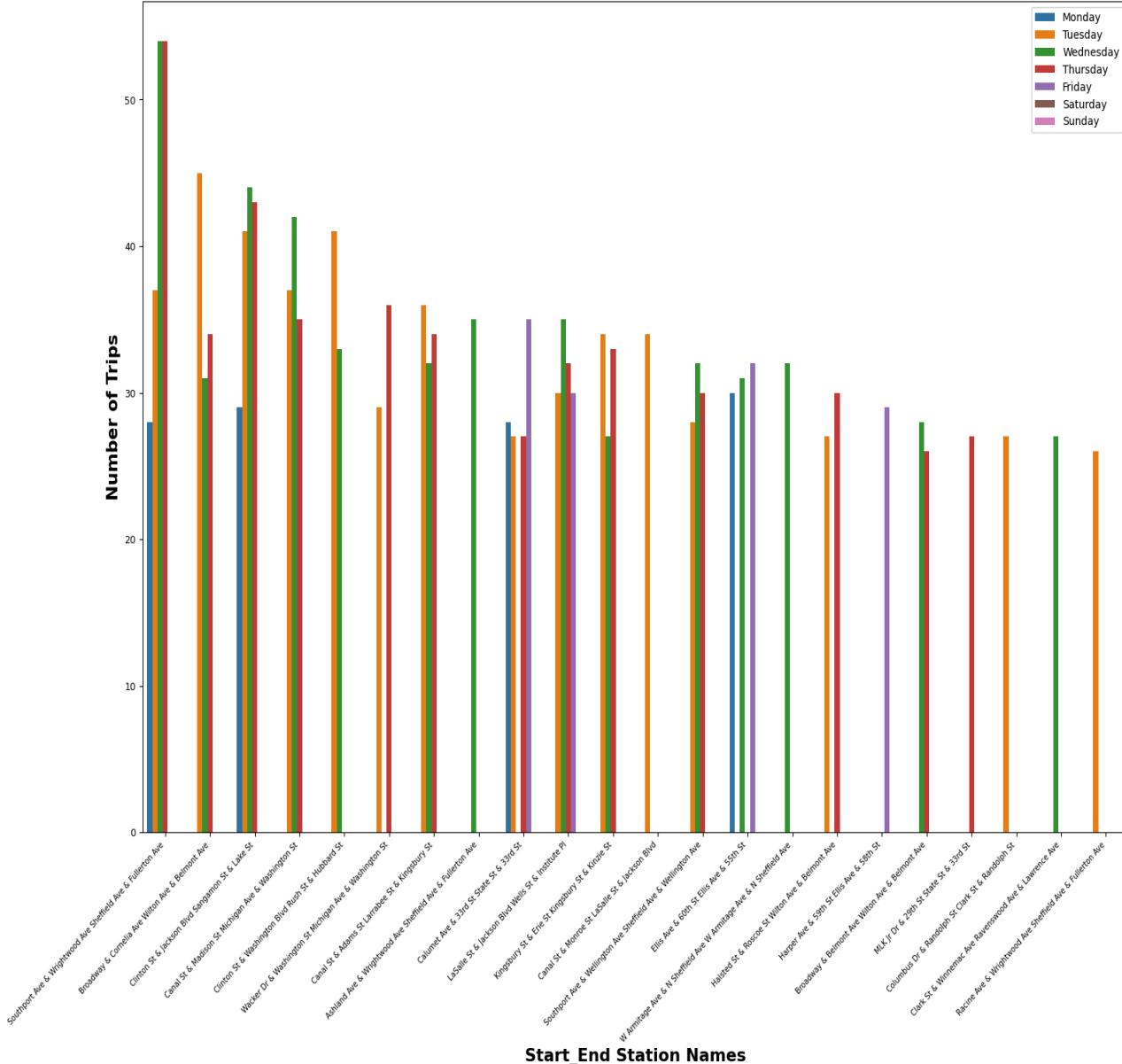
Bike Type by Start/End Station Combos Histogram (Casual Riders)



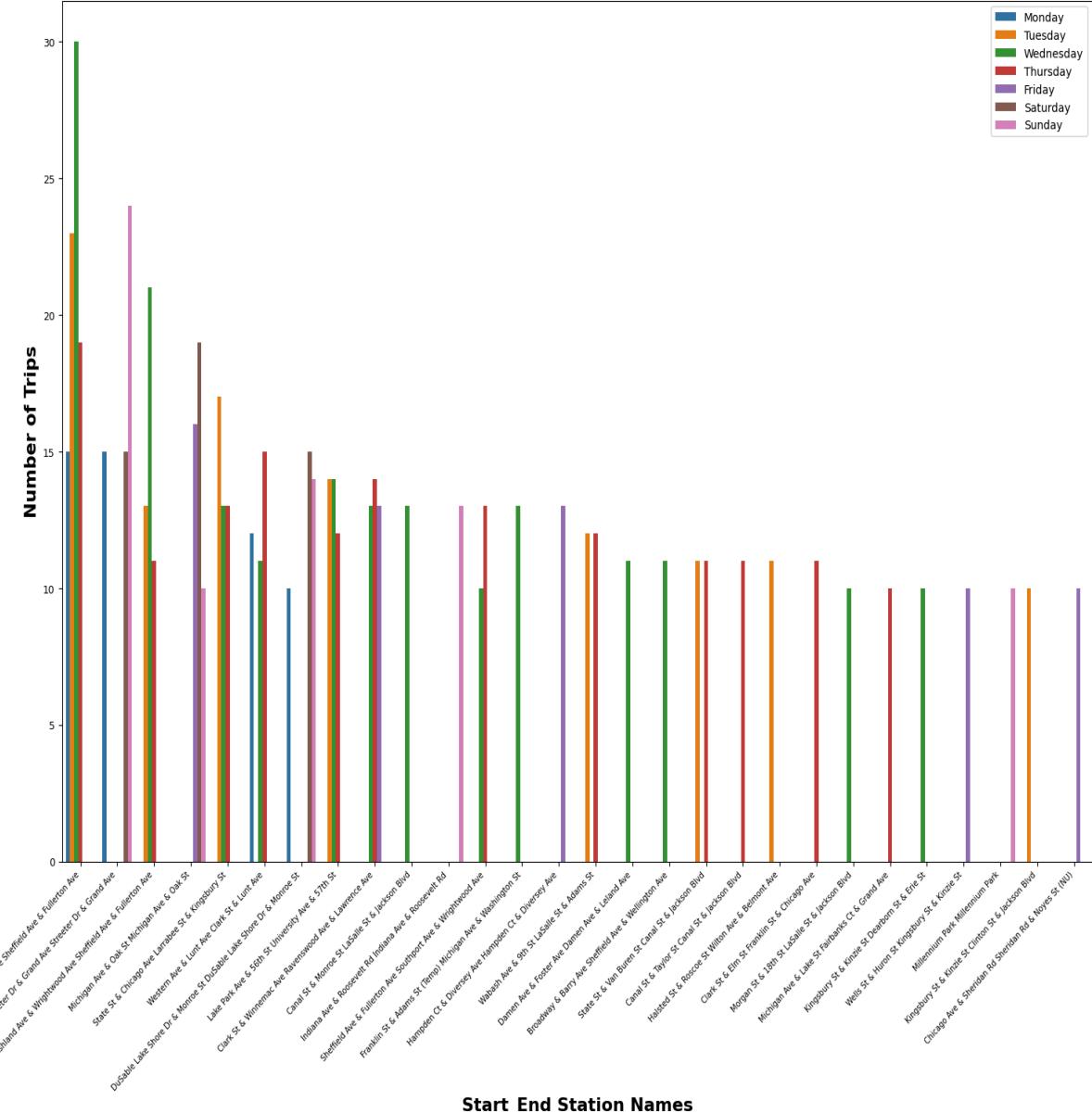
Analysis Results (Cont'd)



Mid Morning Trips by Start/End Stations and Day of Week (Members)



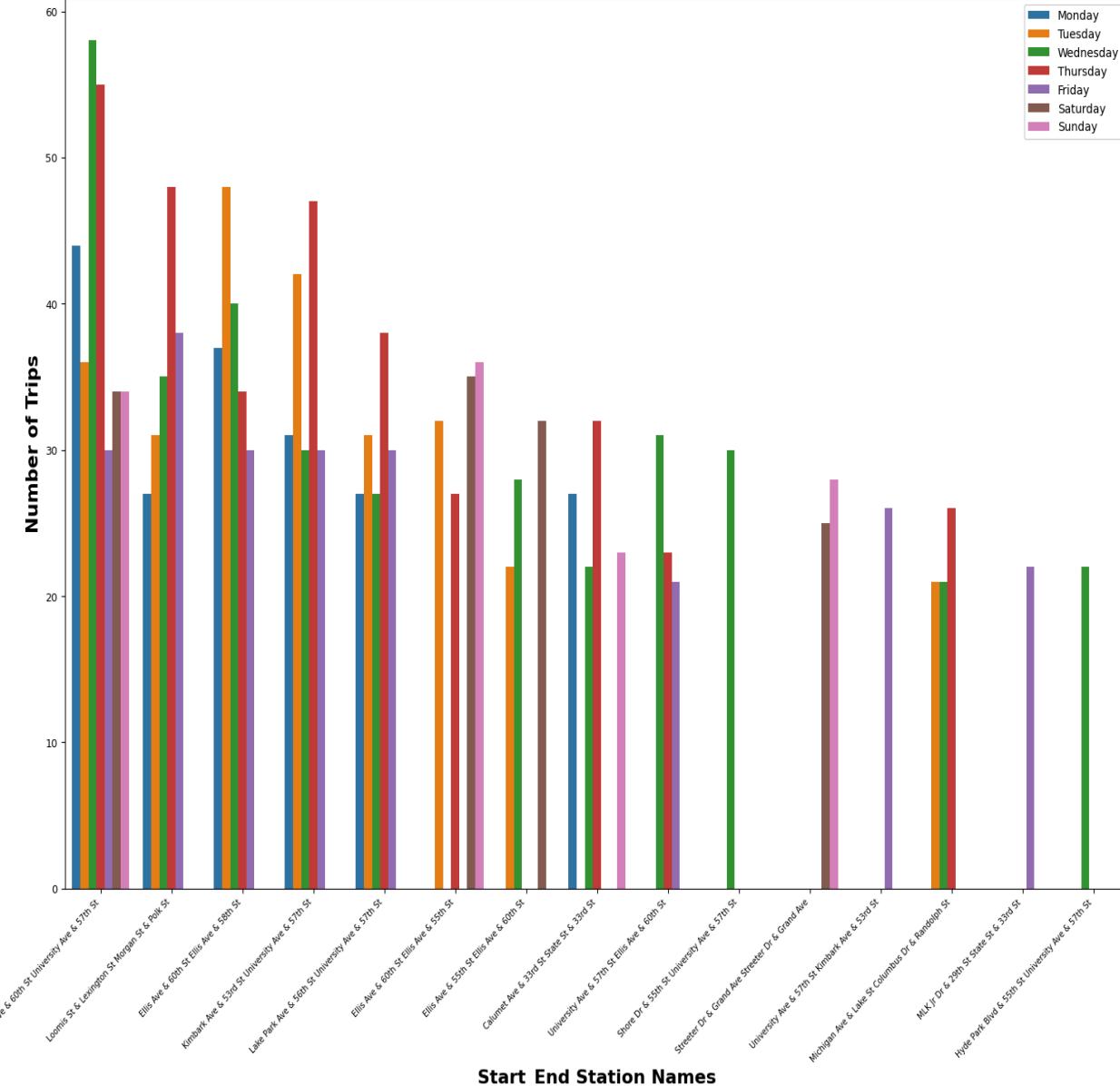
Mid Morning Trips by Start/End Stations and Day of Week (Casual Riders)



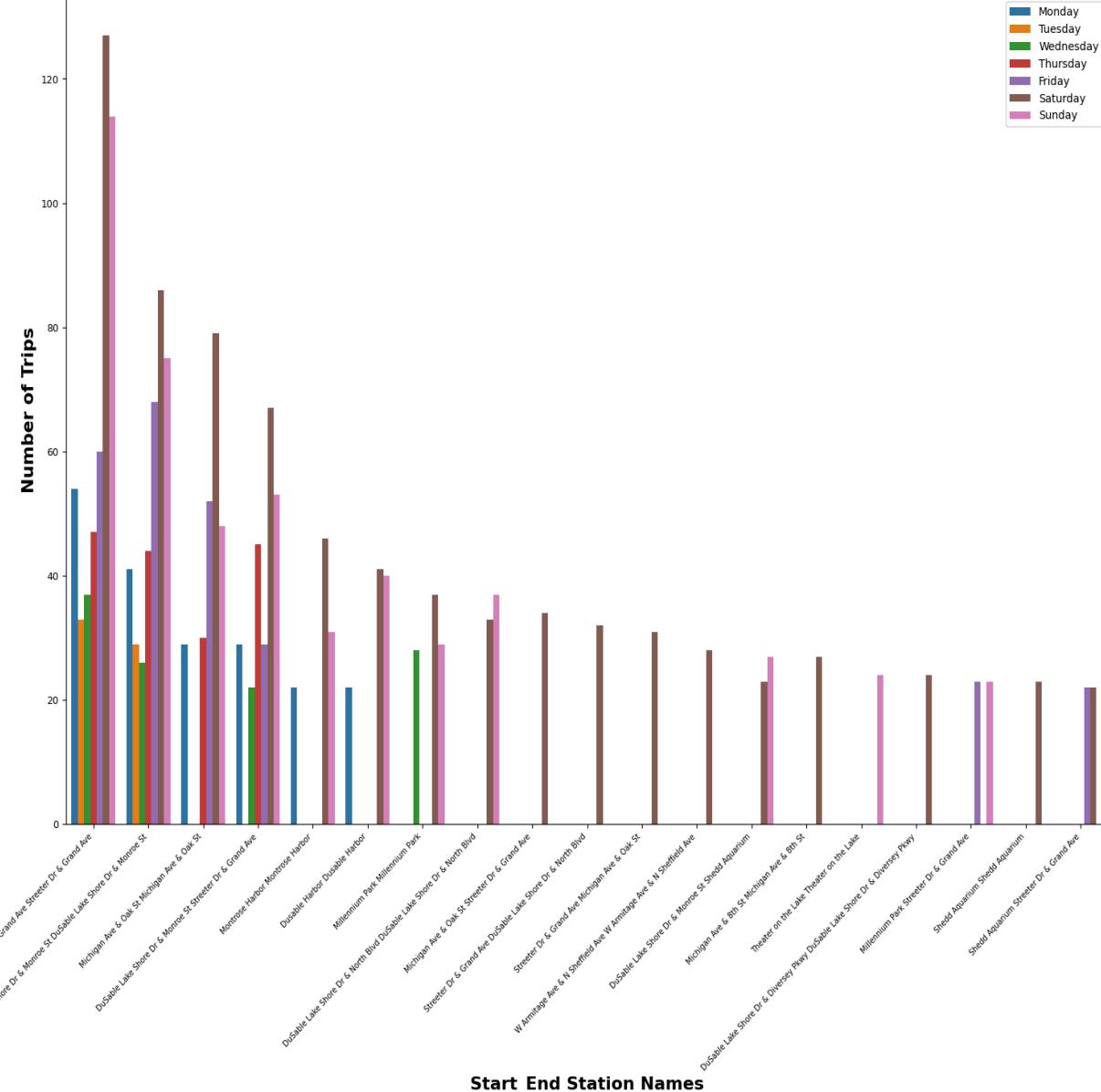
Analysis Results (Cont'd)



Late Morning Trips by Start/End Stations and Day of Week (Members)



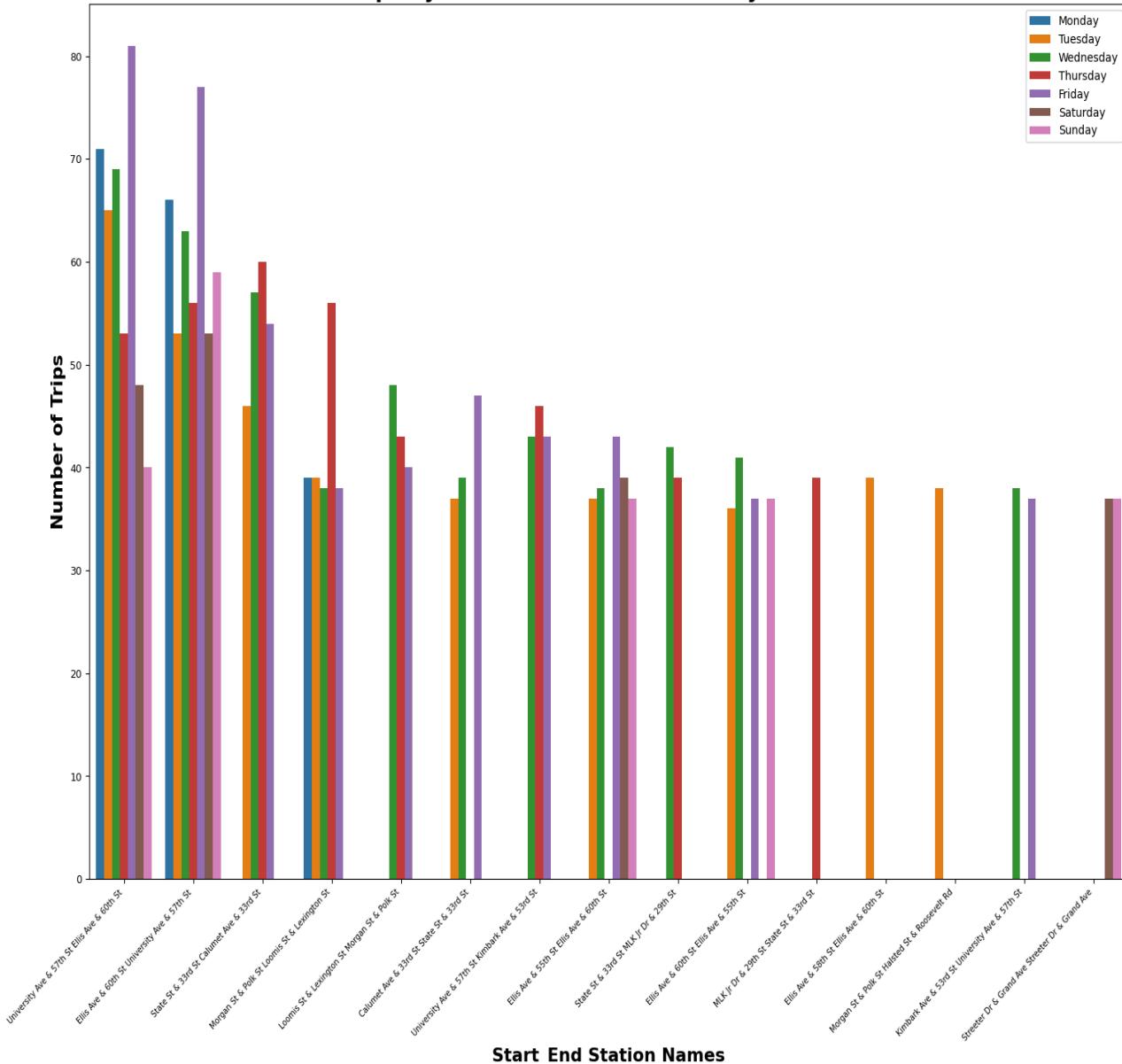
Late Morning Trips by Start/End Stations and Day of Week (Casual Riders)



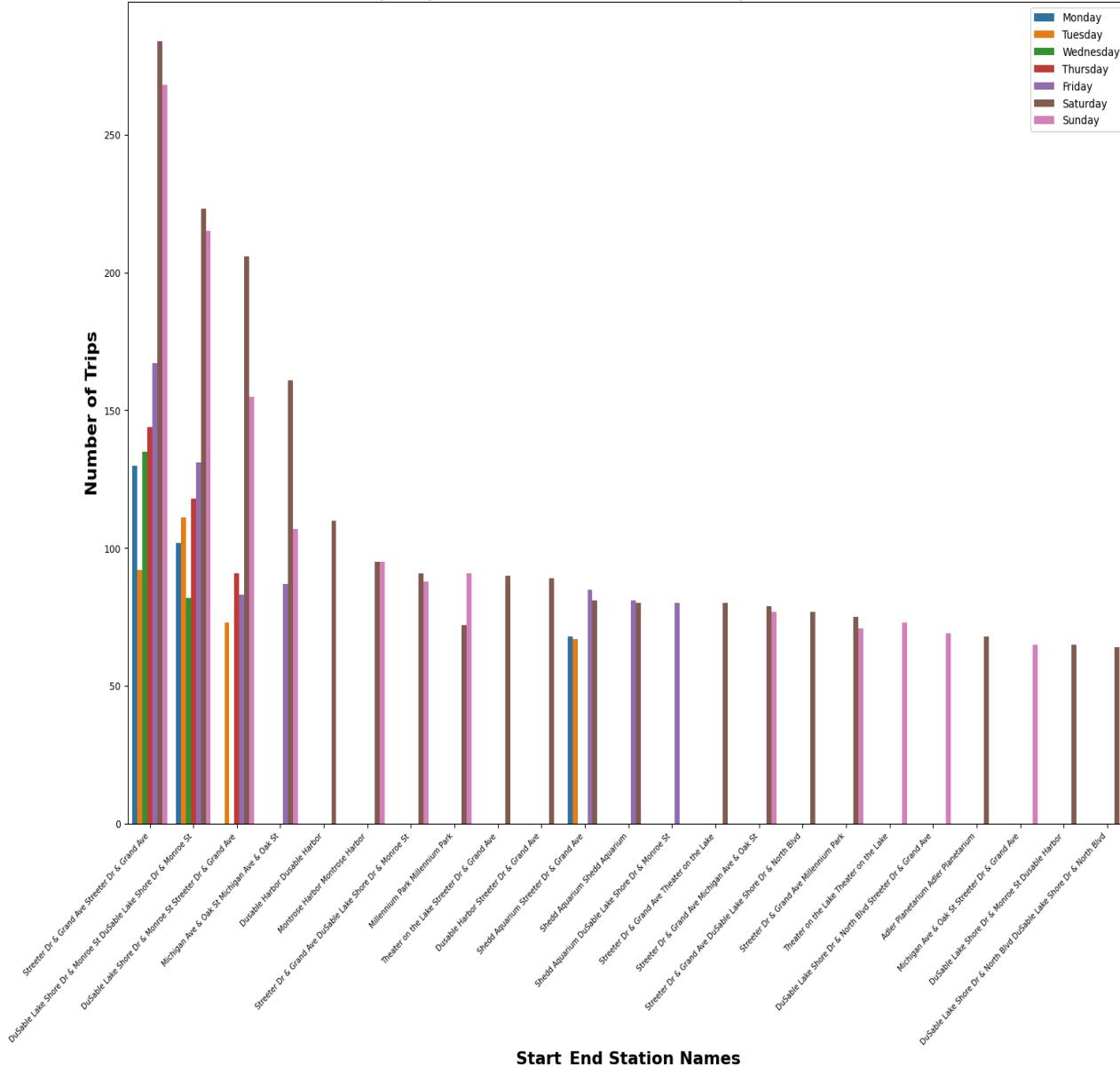
Analysis Results (Cont'd)



Afternoon Trips by Start/End Stations and Day of Week (Members)



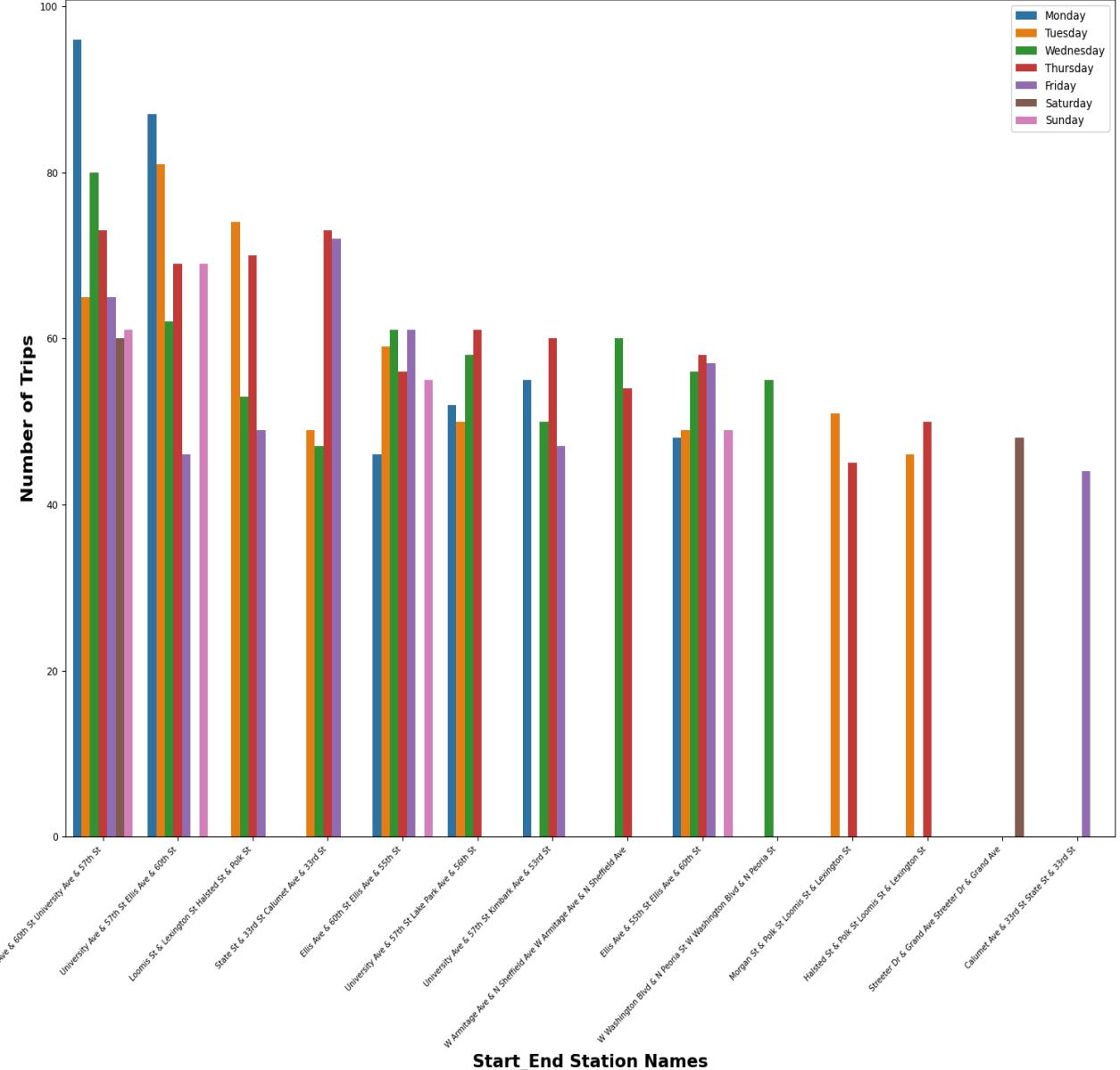
Afternoon Trips by Start/End Stations and Day of Week (Casual Riders)



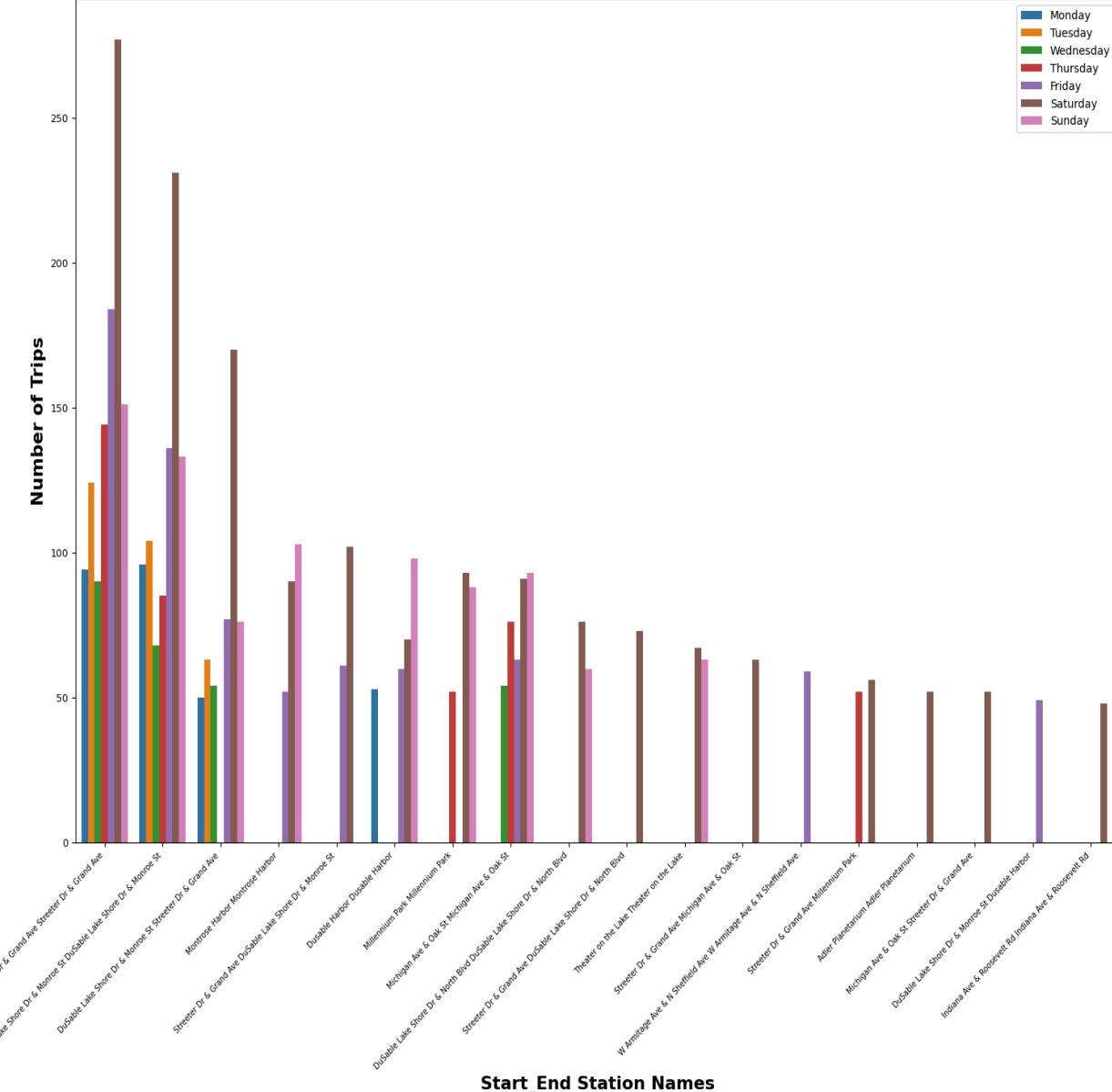
Analysis Results (Cont'd)



Evening Trips by Start/End Stations and Day of Week (Members)



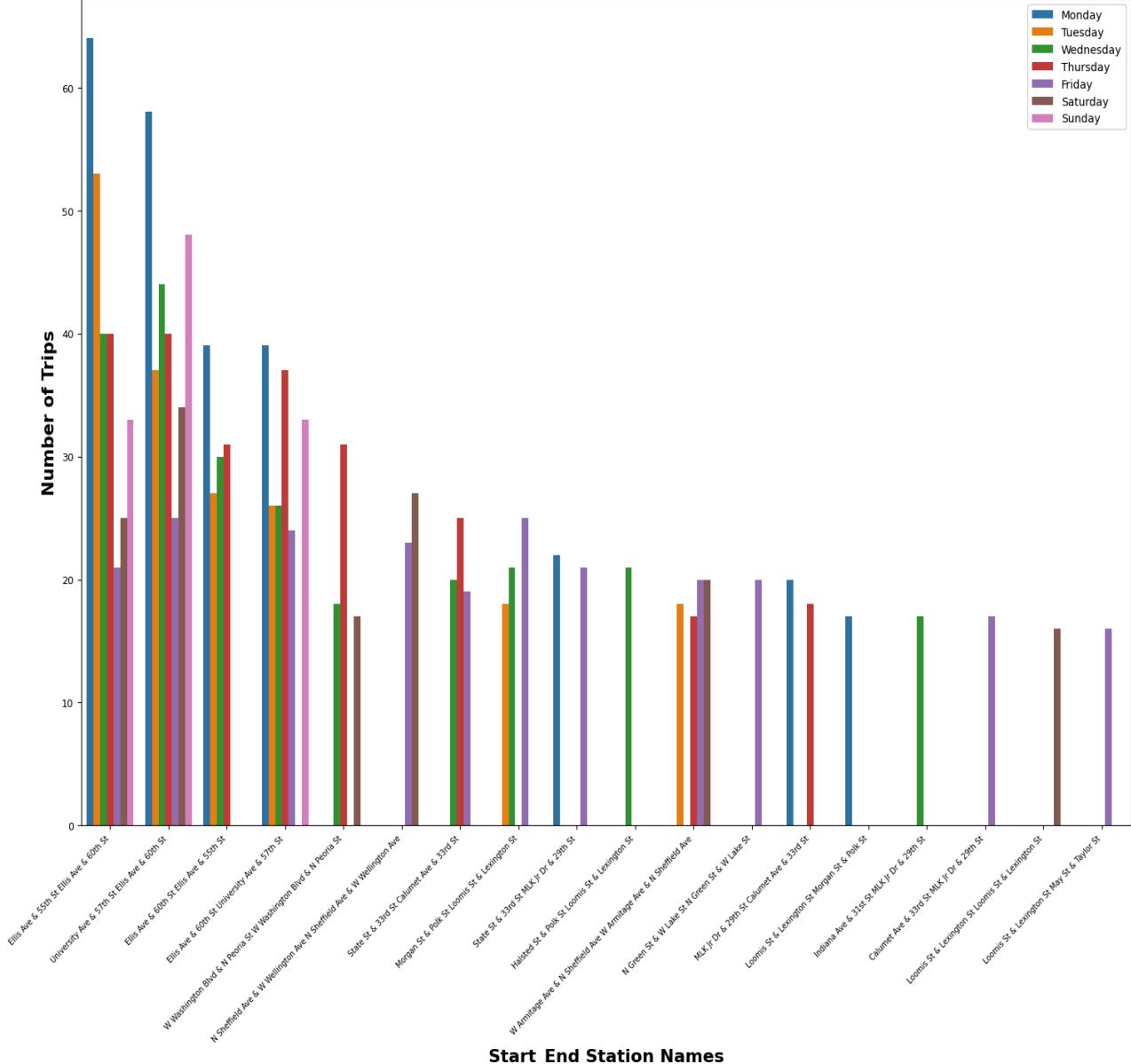
Evening Trips by Start/End Stations and Day of Week (Casual Riders)



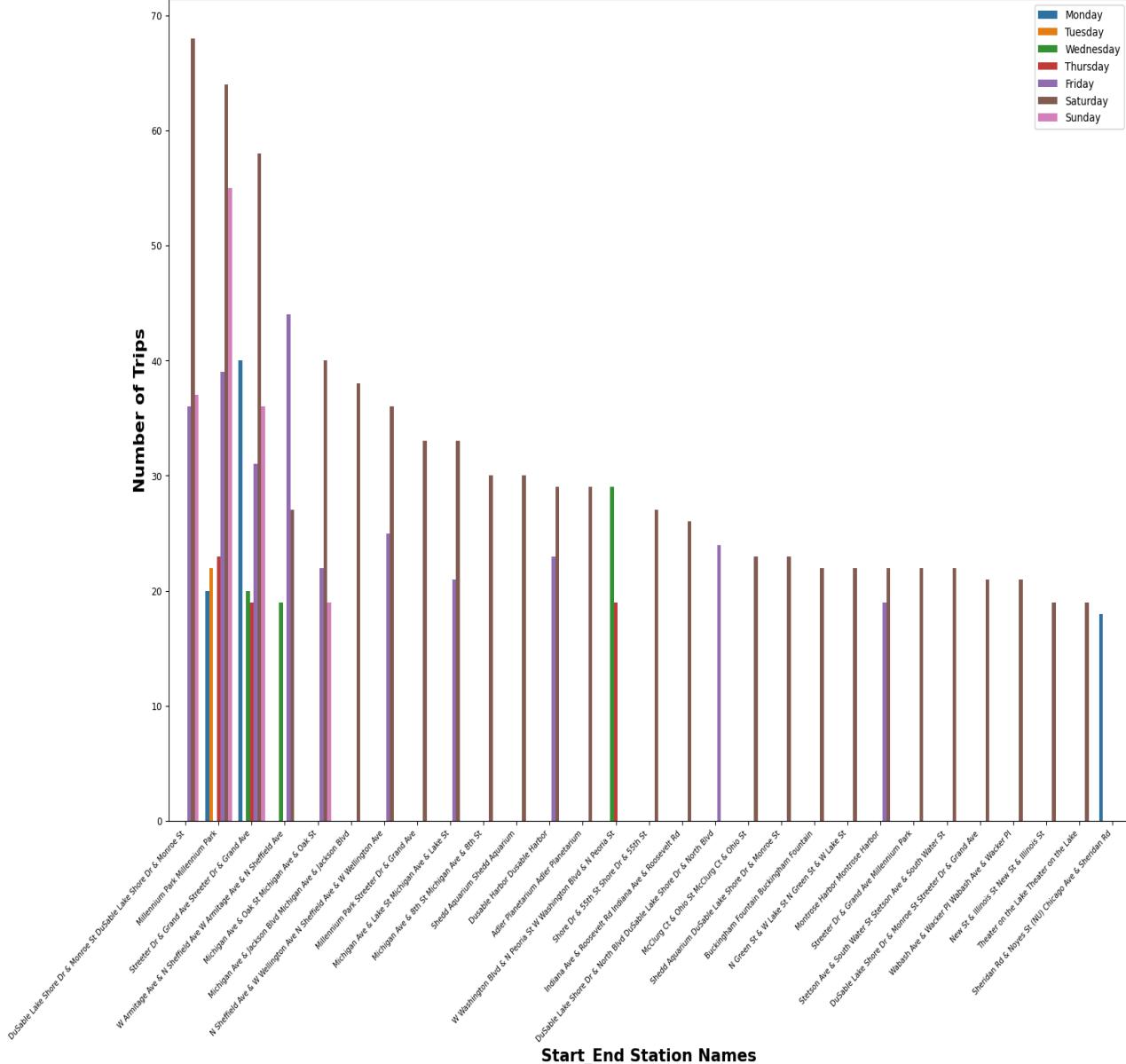
Analysis Results (Cont'd)



Night Trips by Start/End Stations and Day of Week (Members)



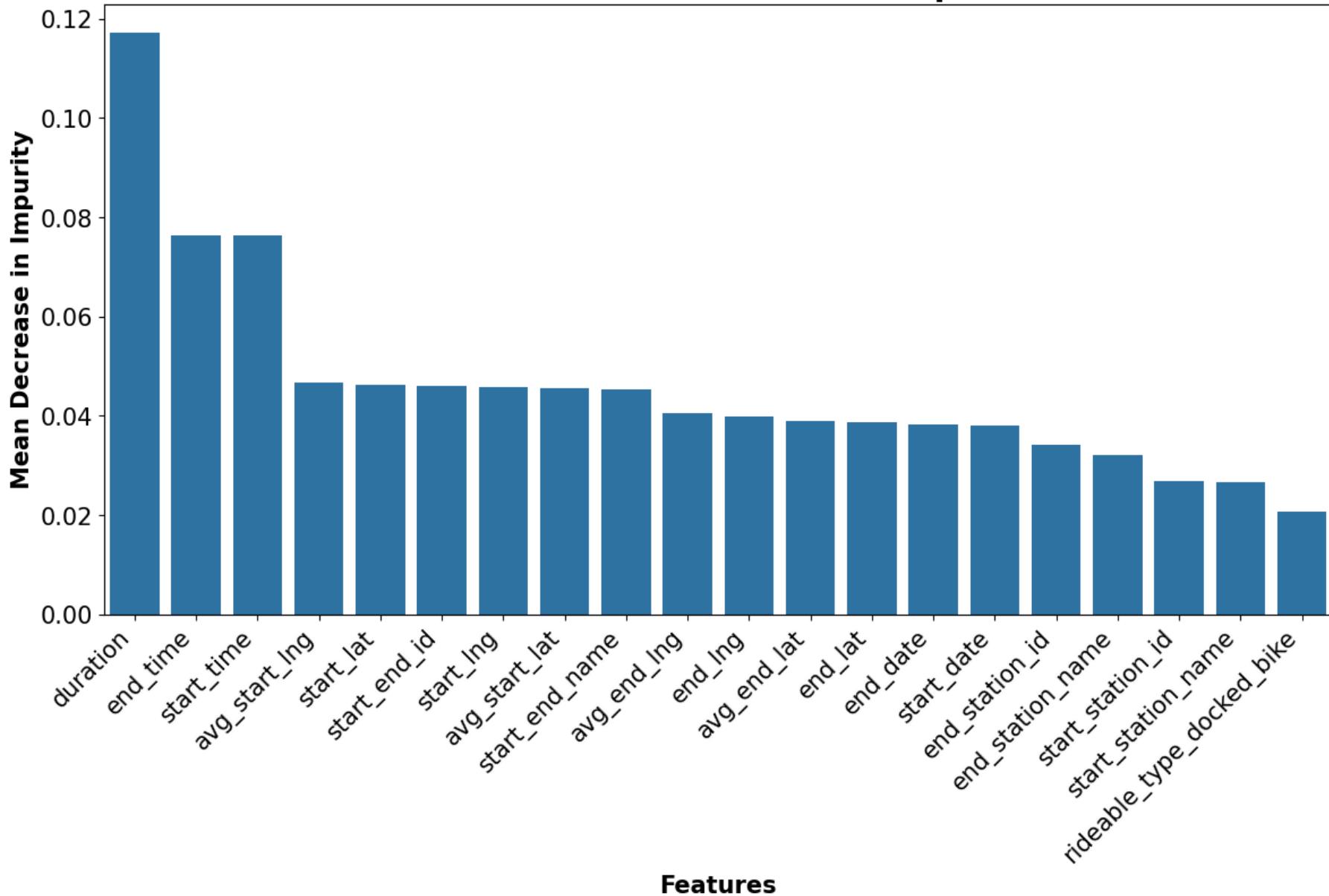
Night Trips by Start/End Stations and Day of Week (Casual Riders)



Analysis Results (Cont'd)



Random Forest Model Feature Importance



Analysis Results (Cont'd)



Member Casual

(All)

Trip Count

- 1
.000
.000
.000
.000
.000
.106

Member Casual

-  casual
 -  member

Weekday

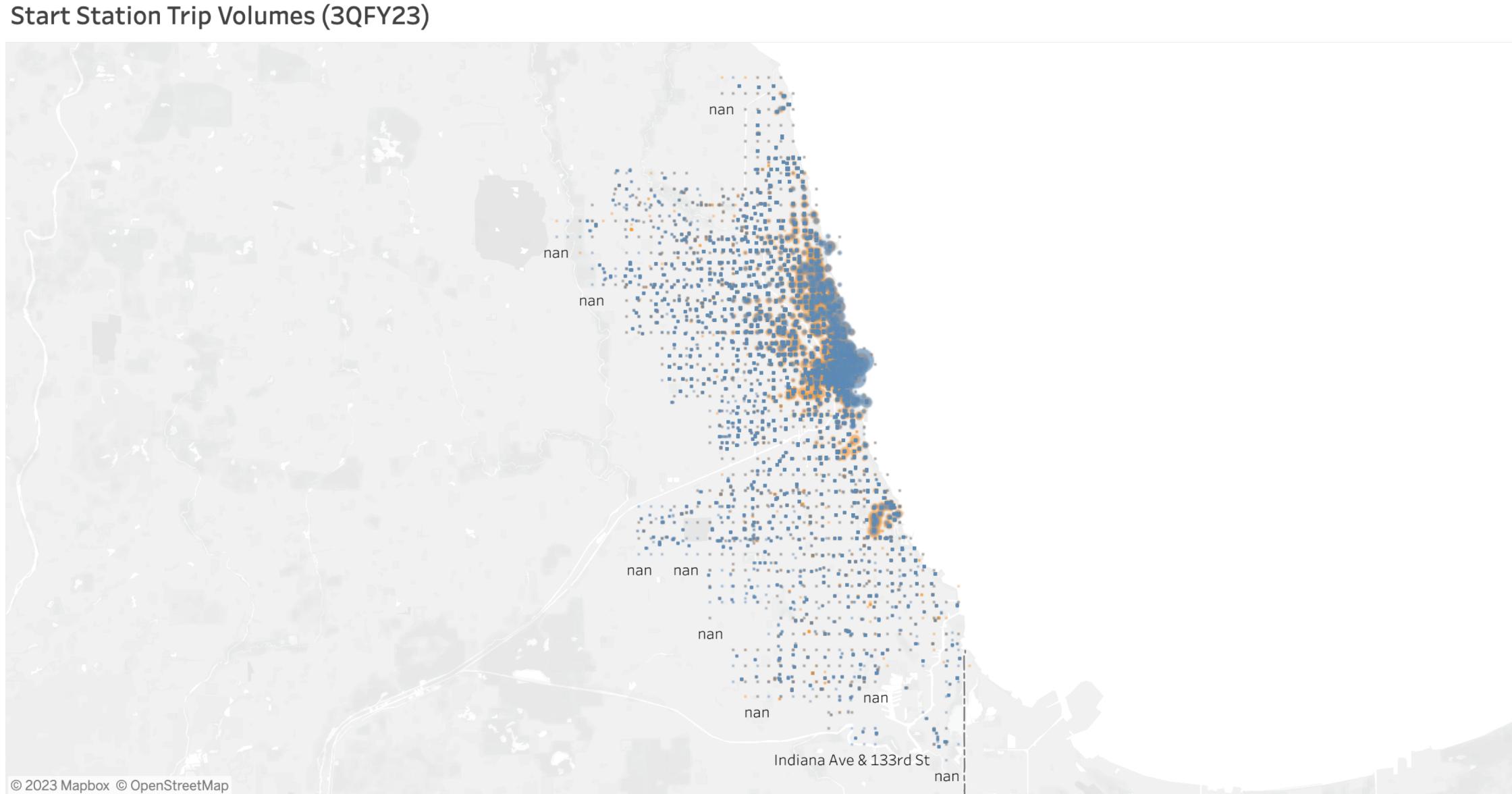
(All)

Time Of Day

(All)

Rideable Type

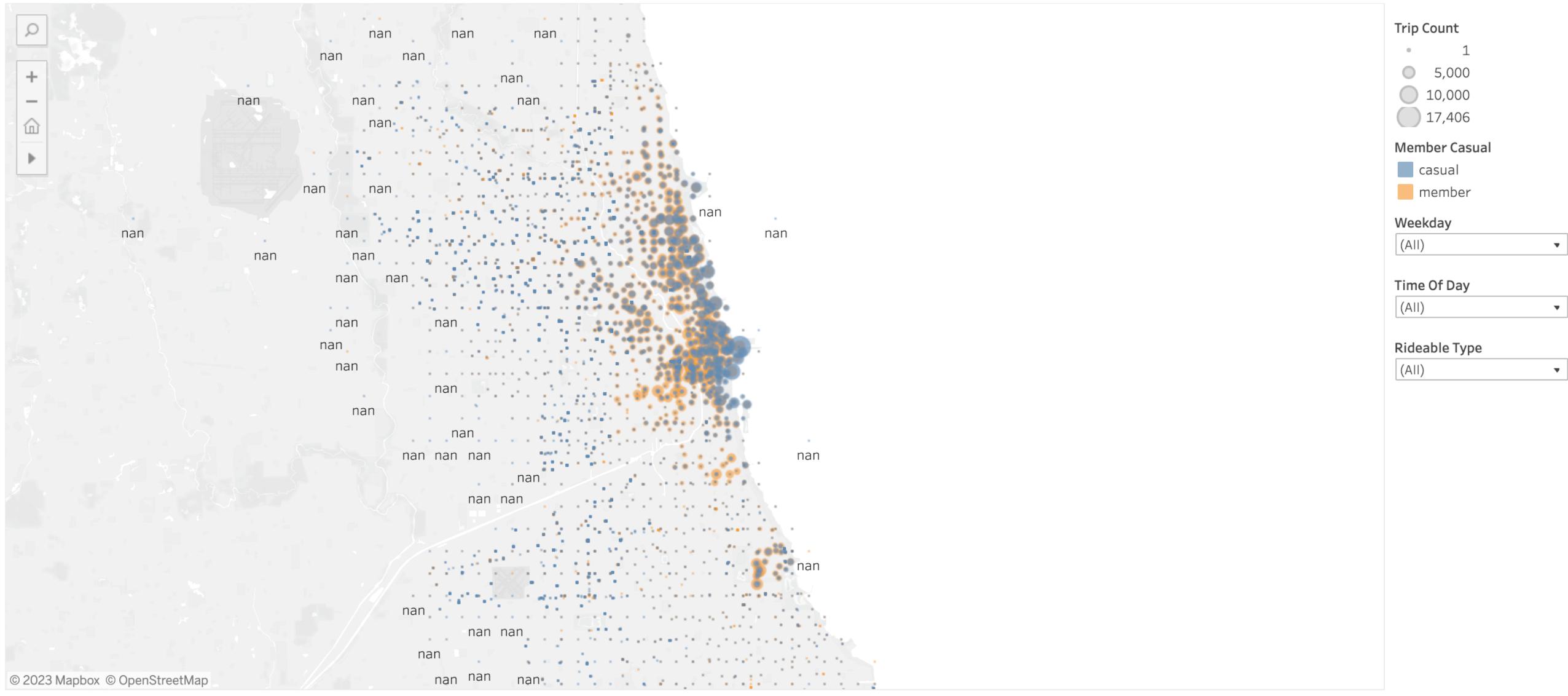
(All)



Analysis Results (Cont'd)



End Station Trip Volumes (3QFY23)



Analysis Results (Cont'd)



Member Casual

casual

member

1

Weekday

Sunday

Time Of Day

afternoon

Rideable Type

electric_bike

Start End Station Names

(AII)

Route Trip Count Legend

1

50

100

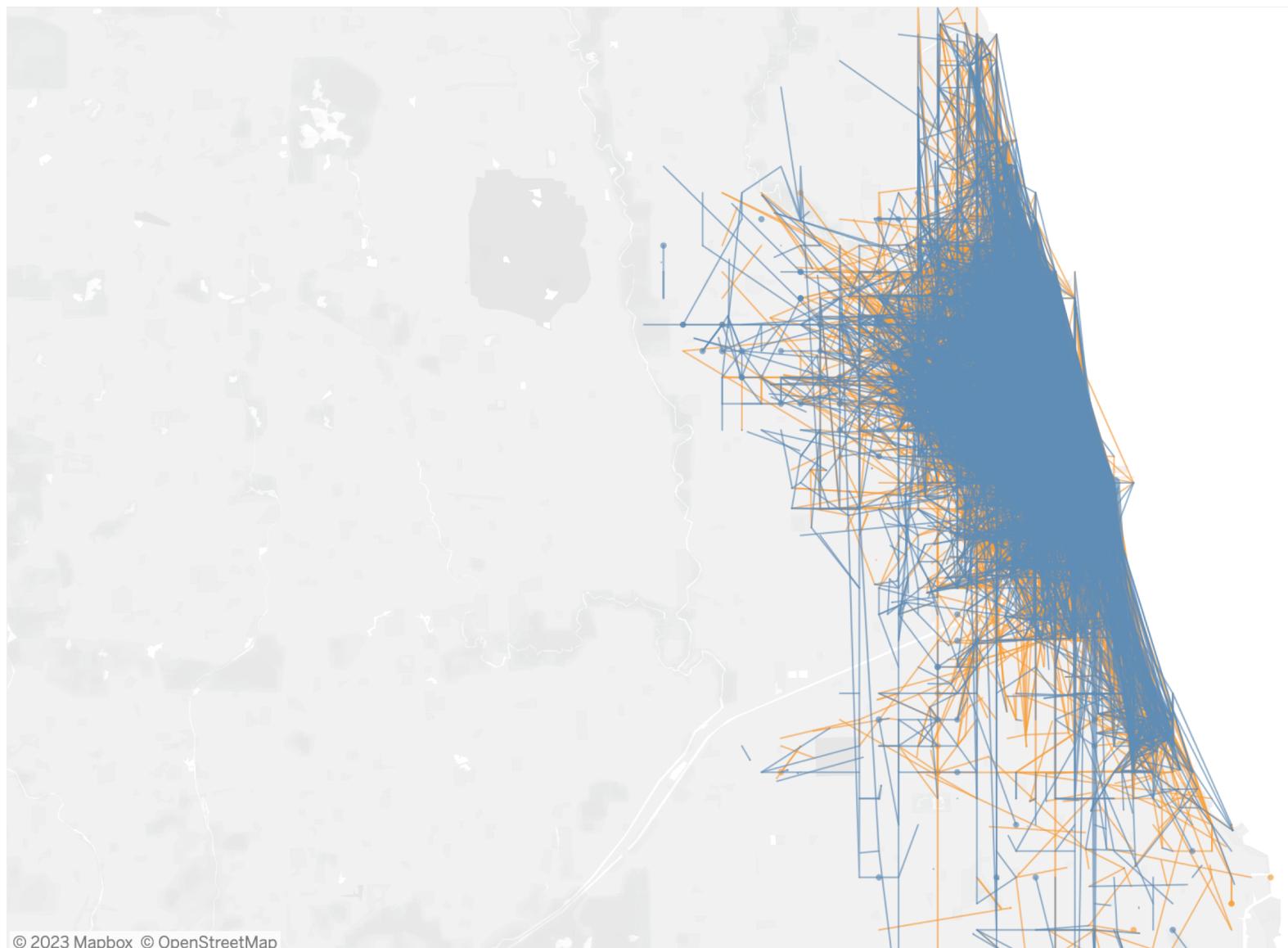
Route Trip Count

Month of Start Date

(All)

Hour of Start Time

(All)





Customer Profiles (Members)

Members

- **Likely city residents and University of Chicago students**
- More likely to take trips to non-tourist destinations
- More likely to have shorter trip durations
- More likely to travel Monday - Friday instead of on weekends
- Most traveled day of the week: Thursday
- Top Start/End station: Kingsbury St & Kinzie St (id: KA1503000043), a major business district
- Top Round Trip station: W Armitage Ave & N Sheffield Ave (id: 20254.0), near subway and bus stations
- Slightly more likely than Casual Riders to take trips earlier in the day
- Trip volumes peak in the evenings during the week and in the afternoons on the weekends
- Make up higher percentage of trips overall in non-Summer months than Summer months
- Less likely to visit more different stations
- More likely to travel more different routes (start/end station combos), except in the early mornings
- Trips per station likely to vary more widely
- Trips per route likely to vary less widely
- Not as likely as Casual Riders to take round trips
- Less likely to take round trips from more different stations
- More likely to use electric bikes than other bike types
- Not likely to use docked bikes
- Less likely than Casual Riders to take longer trips in the late morning/afternoon
- AVG lat/long location about the same as Casual Riders (just east of Goose Island)

Customer Profiles (Casual Riders)



Casual Riders

- Likely tourists
- More likely to take trips to tourist destinations
- More likely to have longer trip durations
- More likely to travel on the weekends
- Most traveled day of the week: Saturday
- Top Start/End station: Streeter Dr & Grand Ave (id: 13022) (also top round trip station), a top tourist destination
- Slightly more likely to take trips later in the day
- Trip volumes peak in the evenings during the week and in the afternoons on the weekends
- More likely to take trips in the Summer months
- More likely to visit more different stations
- Less likely to travel more different routes (start/end station combos), except in the early mornings
- Trips per station likely to vary less widely
- Trips per route likely to vary more widely
- Twice as likely as Members to take round trips
- More likely to take round trips from more different stations
- More likely to use electric bikes than other bike types
- More likely than Members to use docked bikes; docked bike trips have longer duration
- More likely than Members to take longer trips in the late morning/afternoon
- AVG lat/long location about the same as Members (just east of Goose Island)



Recommendations

1. Determine how Annual Membership could be more attractive for tourists and make requisite additions/changes

- consider discounted membership rates/promotions for frequent trips to top tourist destinations, on weekends, for longer trips, and for trips during the Summer months

2. Direct the majority of marketing communications to tourists and feature top tourist destinations/activities (i.e. the Navy Pier and DuSable Harbor) on the weekends in the Summer months in marketing content

- it may be helpful to identify and target tourists who visit Chicago more than once per year

3. Implement partnerships with nearby tourist spot establishments for top tourist stations and advertise resulting value-added services/offers for customers who become Members



Additional Data to Improve Recommendations

- Customer demographic and psychographic data for each Member/Casual Rider
- Unique customer IDs for trips (to follow customers from one trip to the next over time)
 - can be used to personalize marketing content for individual customers as well as further distinguish between Members/Casual Riders
- Fill in the remaining missing data for start and end locations
- Figure out why especially electric bikes are potentially causing so many nulls and fix the issue



QUESTIONS/GUIDANCE