
CYCLISTIC
Google Data Analytics Certificate Case Study

Business Task

Guiding questions for marketing program:

1. *How do annual members and casual riders use Cyclistic bikes differently?*
2. Why would casual riders buy annual memberships?
3. How can Cyclistic use digital media to influence casual riders to become members

I was tasked with answering the first question – to analyze the data to find insights into how the two types of riders use the bikes differently.

Description of Data Sources Used

- The data used was 12 monthly excel spreadsheets of all rides that month for the period April 2020 – April 2021, provided by the organization. Each sheet was a different month.
- Each row represented a ride with a unique ride id.
- The columns, apart from the ride id, were:

Column Name	Description
rideable_type	type of bicycle
started_at	date and time ride was started
ended_at	date and time ride was ended
start_station_name	name of the station where the ride began. This column contained nulls.
start_station_id	the id number, containing letters and numbers in some cases, of the start station. This column contained nulls
end_station_name	name of the station where the ride ended. This column contained nulls.
end_station_id	the id number, containing letters and numbers in some cases, of the end station. This column contained nulls.
start_lat	latitude of bike at start of ride
start_lng	longitude of bike at start of ride
end_lat	latitude of bike at end of ride
end_lng	longitude of the bike at end of ride
member_casual	whether the rider was a member or a casual rider
day_of_week	column added to represent the day of the week the ride started at numerically, where 1 = Sunday, 2 = Monday, etc

Documentation of cleaning and manipulation

- After inspecting the data types and some column values, I added a couple columns to the tables for ride length and a numeric value for the day of week the ride started at.
- The the 12 tables of monthly rides were loaded into SQL for the cleaning and transformation process.
- Data types were double check and inconsistencies and errors across tables were corrected.
- The 12 tables were combined into one table and then the data was validated and checked for errors.
- Errors and missing values were investigated for causes and remedies, then the effects of said values was assessed to determine how to treat them.
- Data was explored first by filtering and sorting.
- Data was manipulated to derive insights, as broken down in the Analysis Summary.

For the full deliverable, including the queries used, please see this GitHub link to the Documentation.

Analysis Summary

After data was loaded, compiled, cleaned, and validated, it was ready for analysis.

Relational tables were created for breaking down the data.

Data was aggregated across variables and segmented by casual and member rides to look for trends in the data.

The composite of rides by rider type was identified.

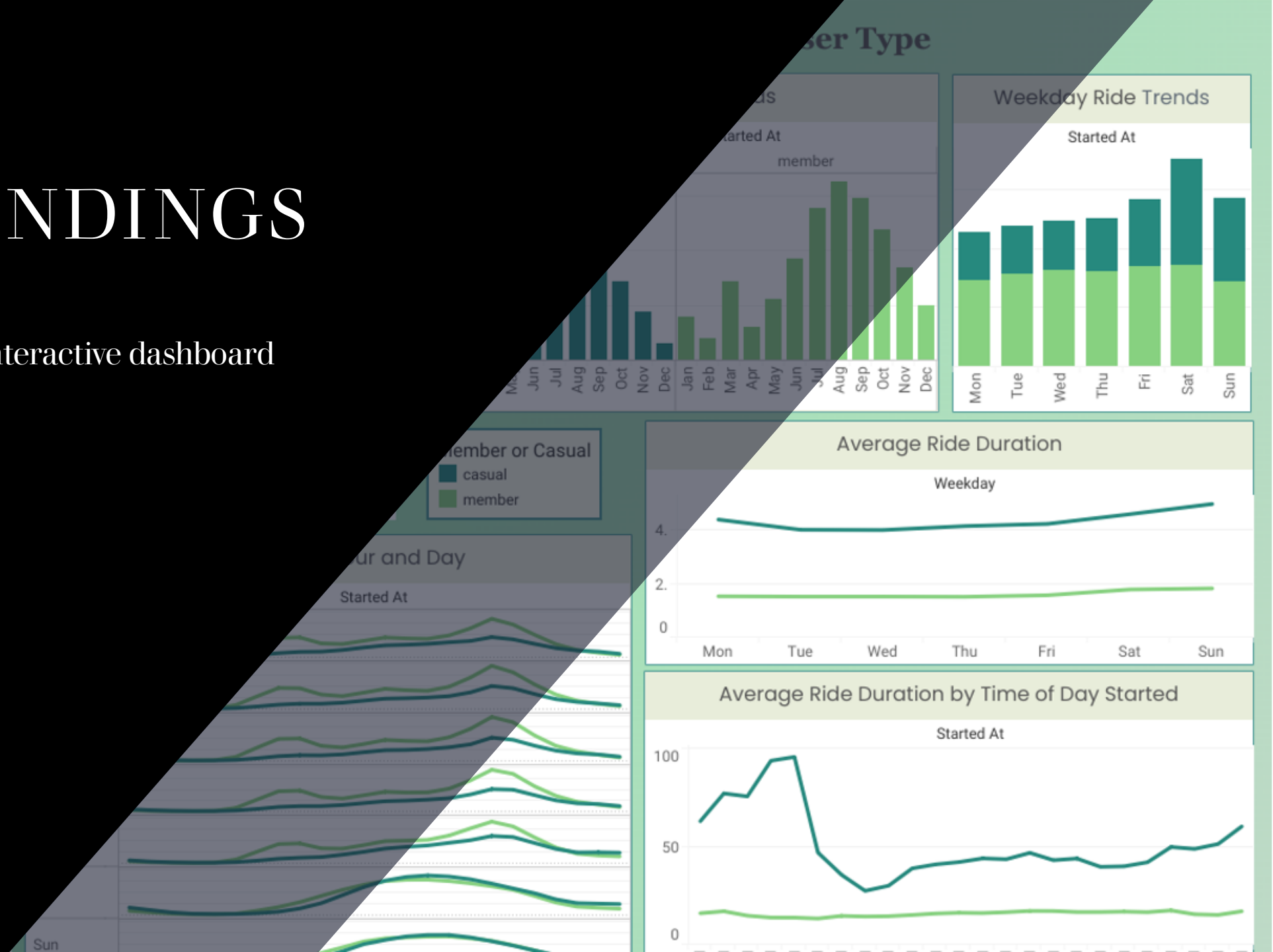
Trends and differences between rider types were identified for rides by weekday, time of day, and changes in ride duration over the week and over the day.

The aggregated and transformed data was loaded into Tableau for visualization.

5 visualizations and an interactive dashboard were created from analysis results.

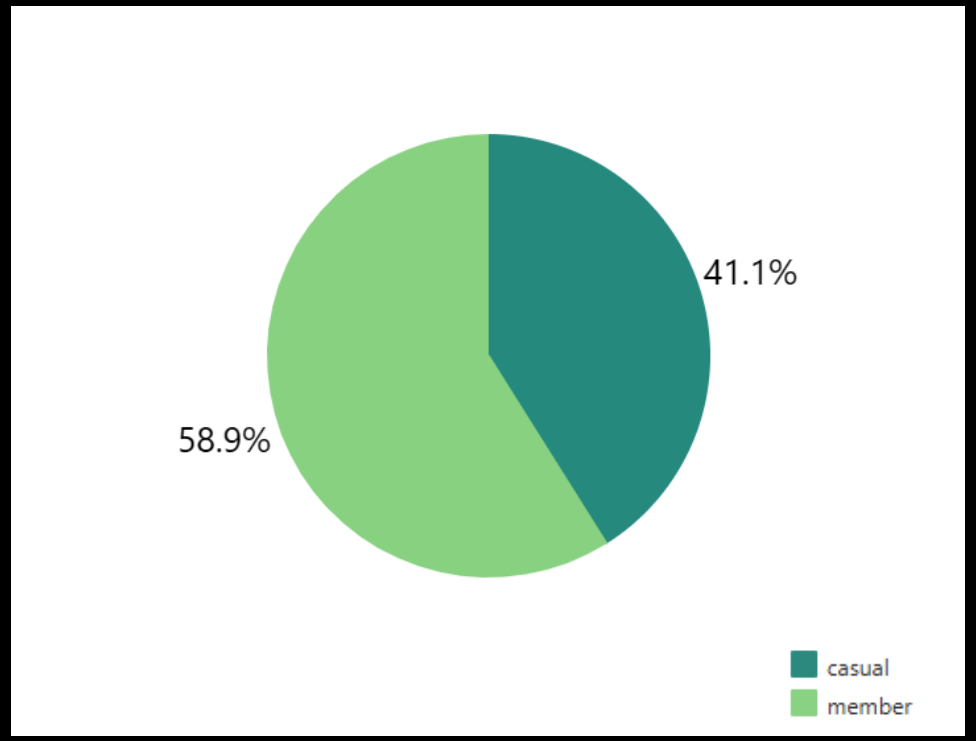
KEY FINDINGS

See here for the interactive dashboard of these findings



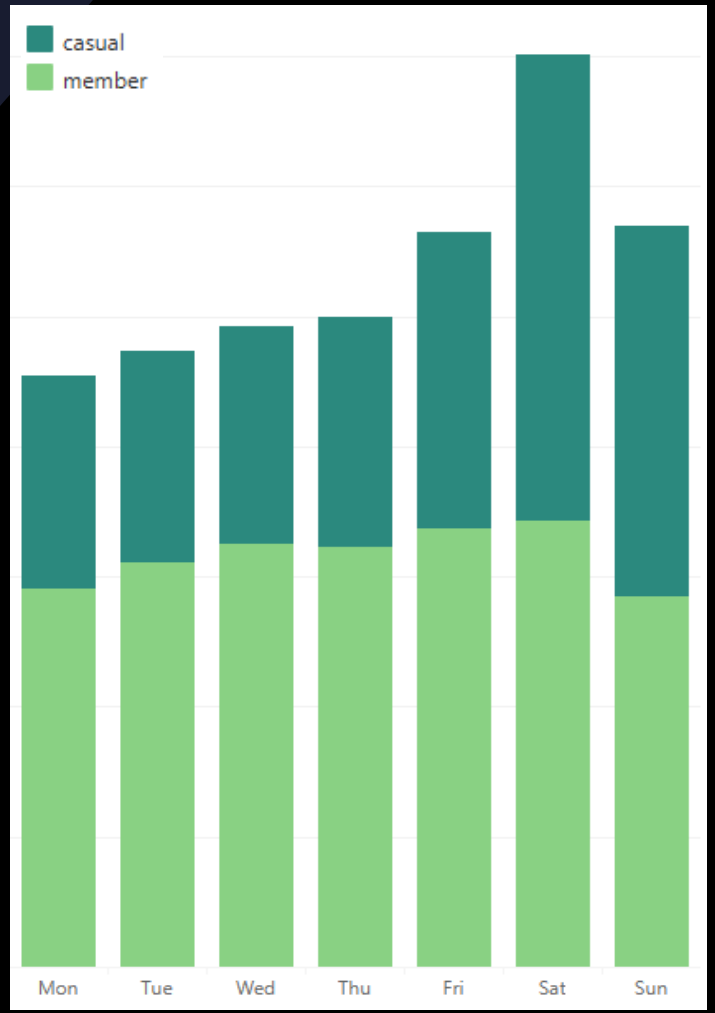
Total Rides by Rider Type

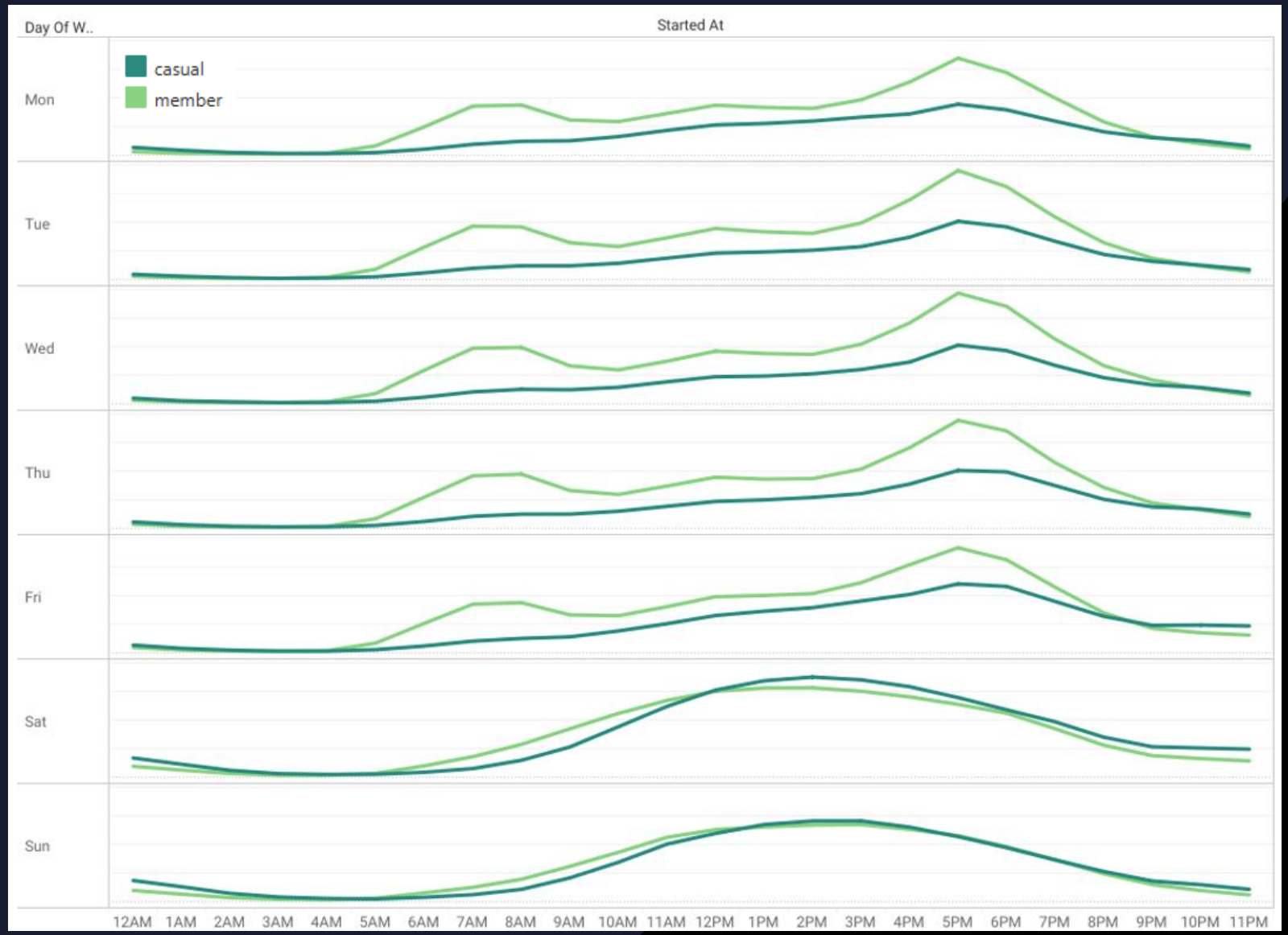
Nearly 60% of all rides from April 2020 through April 2021 were by members, and the rest by casual riders.



Rides by Weekday

- Members rode consistently throughout the week, with slight increases on Saturdays
- Casual riders rode rather consistently in smaller numbers Monday – Thursday, with large increases on the weekend, riding most on Saturdays and Sundays.



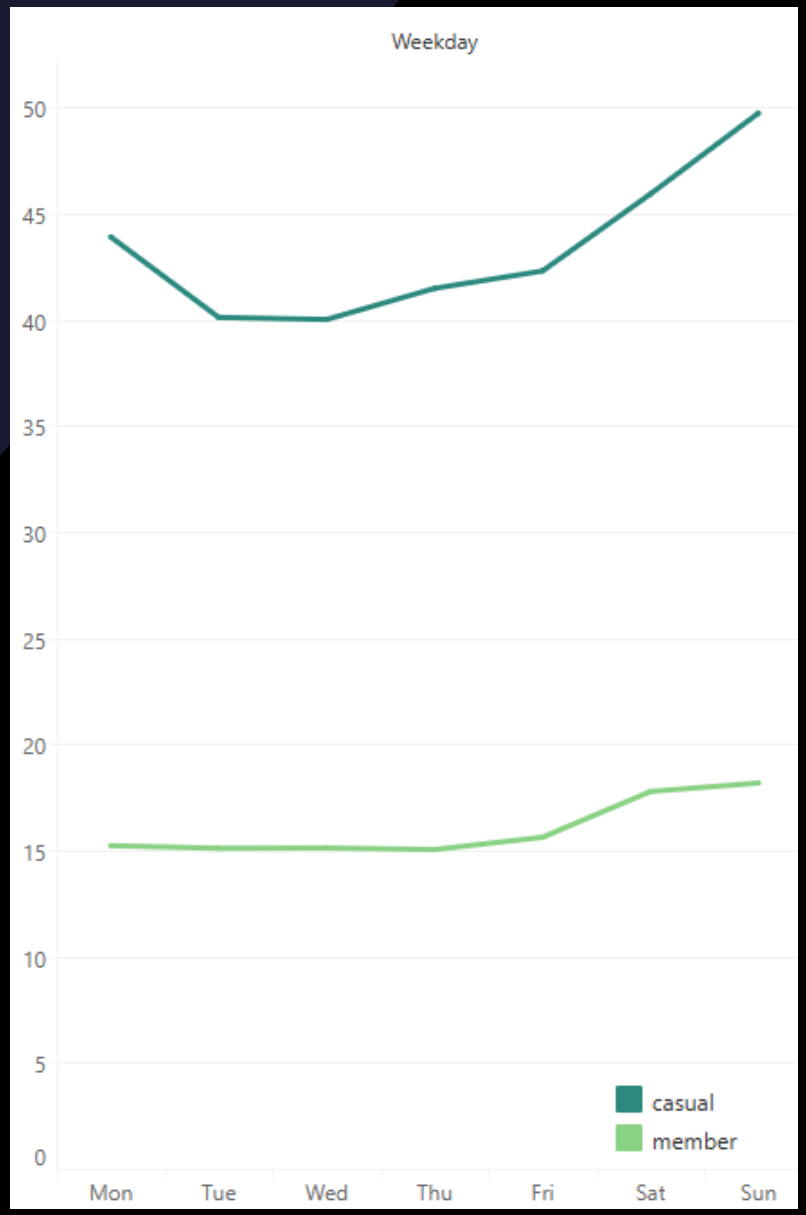


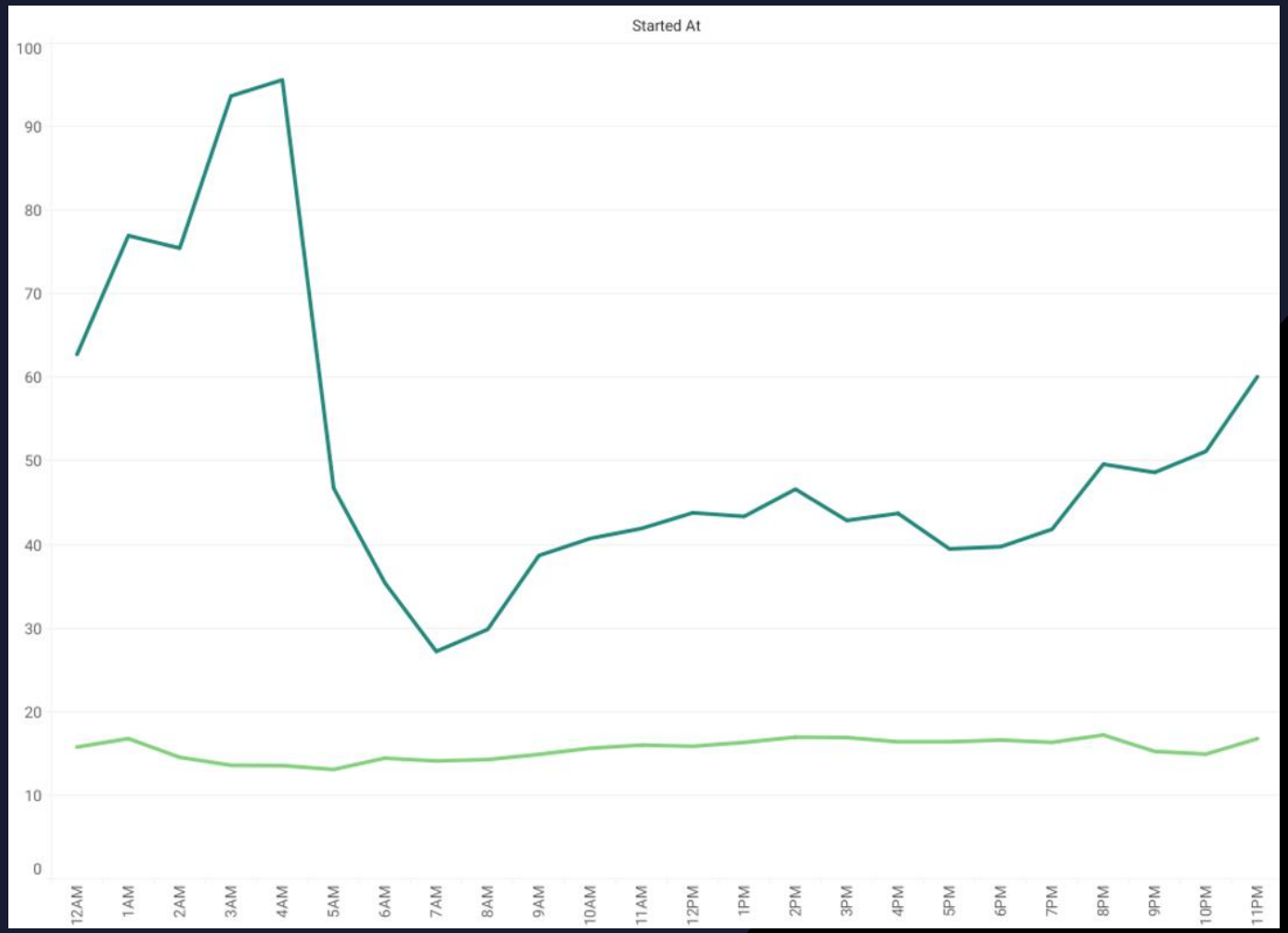
Rides by Hour & Day of Week

- The number of rides is rather consistent across weekdays for both groups, with similar behavior between groups on weekends.
- Members ride more frequently earlier in the day than casual riders.
- Casual rides see small but consistent increases throughout the day, peaking at 5 – 6 pm, where member rides also peak.

Ride Duration by Weekday

- Both groups saw increases in ride durations at the end of the week, with a more significant increase in duration for casual riders.
- Casual riders take much longer rides that vary throughout the week, compared to member riders, who are likely taking the same route Monday through Friday.





Ride Duration by Hour of Day

- Member riders take roughly the same length of ride throughout the day.
- Casual riders are less consistent, with their longest rides happening in the early hours of the day, between 12 – 4 am.

Recommendations

One tactic is to offer a certain amount off rides at a particular time for select months to new members.

The areas of focus would be:

- in the early hours between 12 – 4 am, when casual riders are taking significantly longer rides.
- On Saturdays and Sundays, when casual riders are taking more and longer trips.
- In the mornings, where encouraging riding earlier in the day could push casual riders to change their method of commuting to bike.

For example: “Rides 50% off on weekends for new members for the first 3 months”