Module 18 Challenge

Congratulations on your new job! As the new lead analyst for the New York Citi Bike program, you are now responsible for overseeing the largest bike-sharing program in the United States. In your new role, you will be expected to generate regular reports for city officials looking to publicize and improve the city program.

Since 2013, the Citi Bike program has implemented a robust infrastructure for collecting data on the program's utilization. Each month, bike data is collected, organized, and made public on the Citi Bike Data webpage.

However, while the data has been regularly updated, the team has yet to implement a dashboard or sophisticated reporting process. City officials have questions about the program, so your first task on the job is to build a set of data reports to provide the answers.

The questions I am most interested in answering are

1. How are riders using these bikes, for recreation or for commuting?
2. Are there significant differences between the Member riders and the Casual riders? For instance, does one group take longer trips in terms of duration and distance?
3. Are there significant differences between riders who use Classic bikes and riders who use Electric bikes?
4. What are the most popular stations for starting a ride?
5. What are the most popular stations for finishing a ride?
6. Are there differences in trends between summer and winter rides?

Methods:

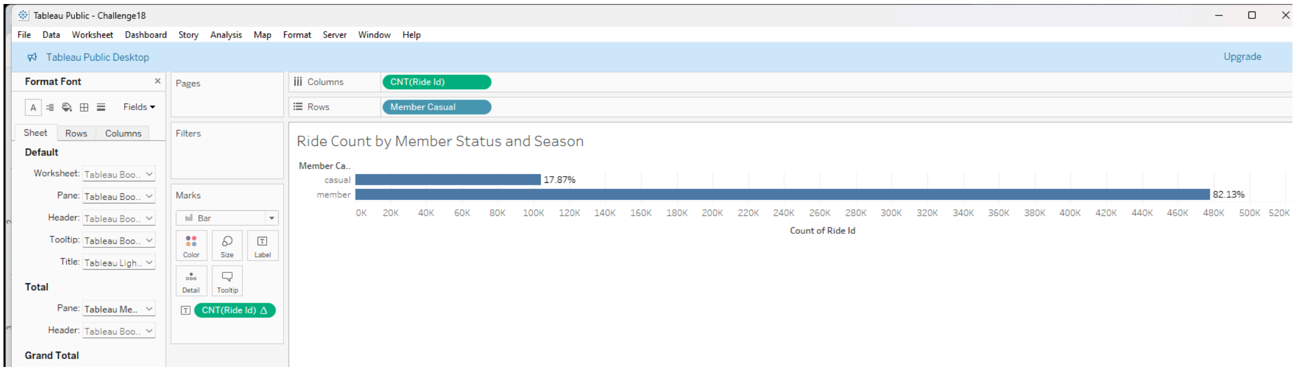
1. Because the amount of data available was so large, I selected June 2023 to represent summer rider behavior and December 2023 data to represent winter rider behavior.
2. I imported 7 csv files between June and December into VS Code as dataframes, and merged all data frames.
3. To reduce the volume of data to work with from over 5,000,000 records, I next took a sample of 10% of the total file. This brought the rows of the data frame down to 581,940 records.
4. To filter out bikes not in use, I considered only rideable_type values of classic_bike and electric_bike.
5. The ‘started_at’ and ‘ended_at’ columns were object types, so I converted them to datetime objects. This allowed me to further separate date and time information into ‘start_date’, ‘start_time’, ‘end_date’, and ‘end_time.’ It also allowed me to calculate the duration of each ride, in seconds, by subtracting ‘end_time’ from ‘start_time.’
6. I also created columns for ride day (day of the week) and ride month (June or December).

7. I used the haversine_distance function on the ‘start_lat’ (starting latitude), ‘start_lng (starting longitude) and ‘end_lat’ (ending latitude) and ‘end_lng’ (ending longitude) to calculate the distance in kilometers between ride starting points and ending points.
8. I used t-tests to determine whether the average duration and average distance of rides between ‘member’ and ‘casual’ riders and between ‘classic_bike’ and ‘electric_bike’ riders were significantly different.

Findings:

Citi Bike service users

- Users of the Citi Bike service are one of two types: Casual riders and Member riders
- While we can’t tell from the June and December 2023 data how many of each type of rider used the Citi Bike service, we can tell what proportion of the 10% data sample were made by Casual versus Member riders.
- Member-made rides made up the majority (82.13%) of all rides in that time period.

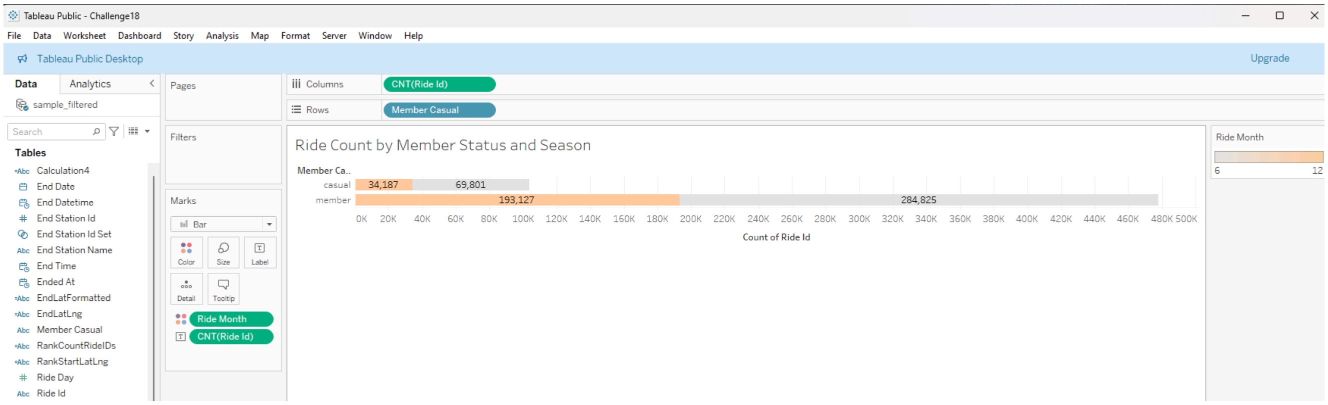


Columns: CNT(Ride Id)

Rows: Member Casual

Ride Count by Member Status and Season
Member Ca..
casual
member
17.87%
82.13%
0K 20K 40K 60K 80K 100K 120K 140K 160K 180K 200K 220K 240K 260K 280K 300K 320K 340K 360K 380K 400K 420K 440K 460K 480K 500K 520K
Count of Ride Id

- When we break those numbers down further into rides by season (with June data serving as a proxy for summer, and December data serving as a proxy for winter), we see that both Member and Casual riders make fewer rides in winter than in summer.
- This finding could have implications for bike maintenance frequency. Bikes are likely experiencing more wear and tear in the summer season than they are in winter.

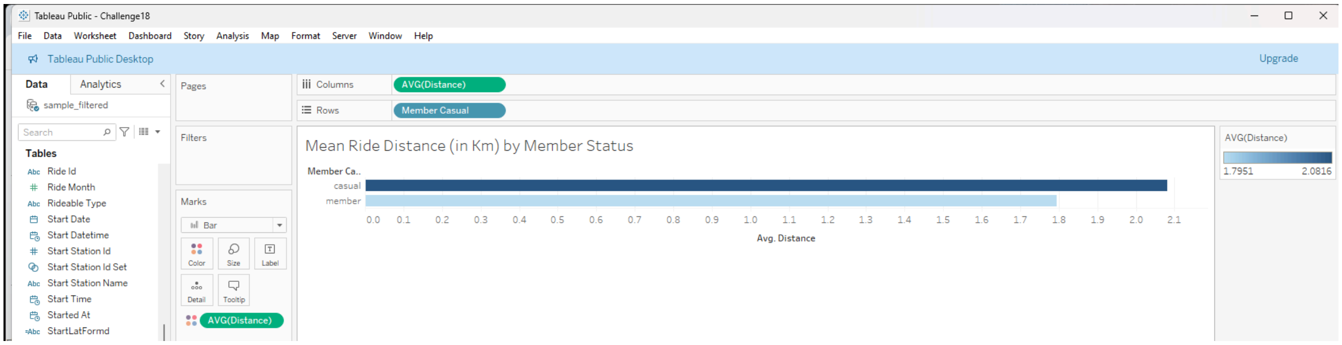


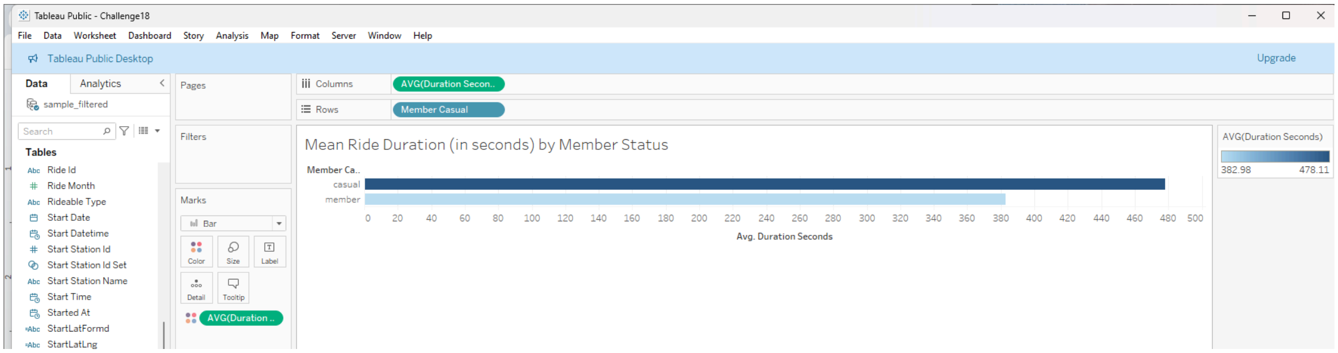
Columns: CNT(Ride Id)

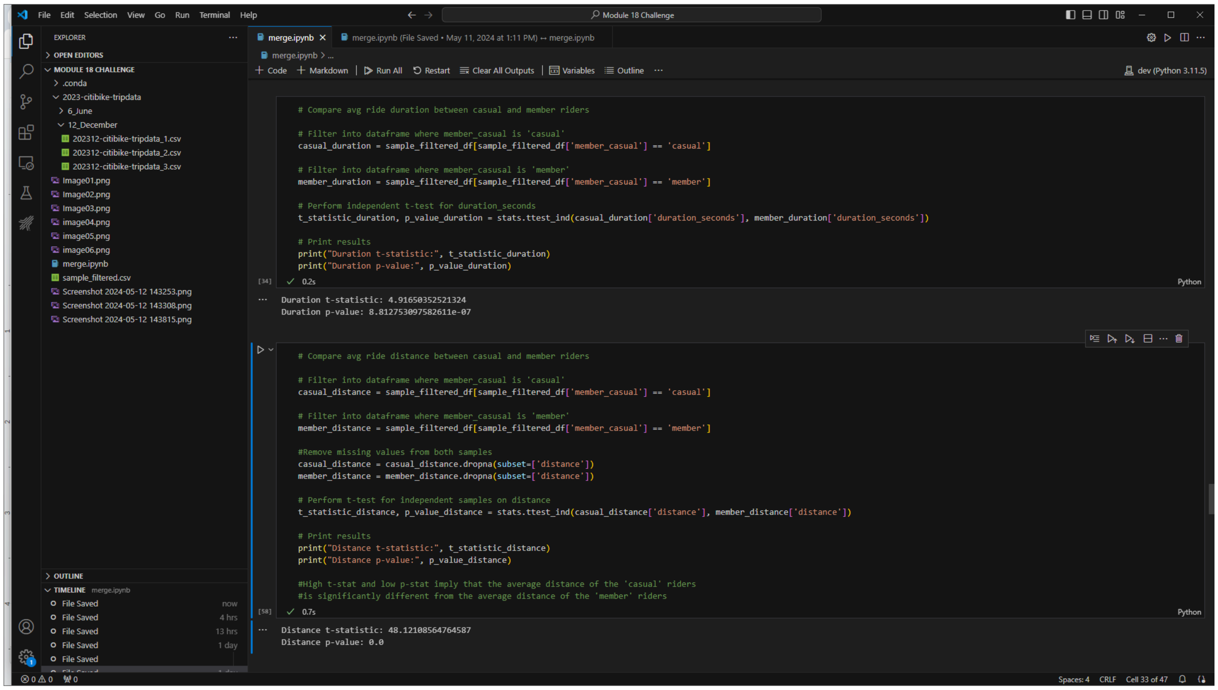
Rows: Member Casual

Ride Count by Member Status and Season
Member Ca..
casual
member
34,187
69,801
193,127
284,825
0K 20K 40K 60K 80K 100K 120K 140K 160K 180K 200K 220K 240K 260K 280K 300K 320K 340K 360K 380K 400K 420K 440K 460K 480K 500K
Count of Ride Id
Ride Month
6
12

- When we compare the mean distance and mean duration (in seconds) of rides between Members and Casual riders, we see that - on average - Members are taking shorter rides in terms of both time and distance. The differences in mean distance and duration are statistically significant,as t-test p-values are less than 0.05 in both cases.
- The difference in mean distance and mean duration between Member and Casual riders suggests that Casuals are doing some sightseeing or leisure cycling, while Members are using CitiBikes to get from point A to point B under time constraints.







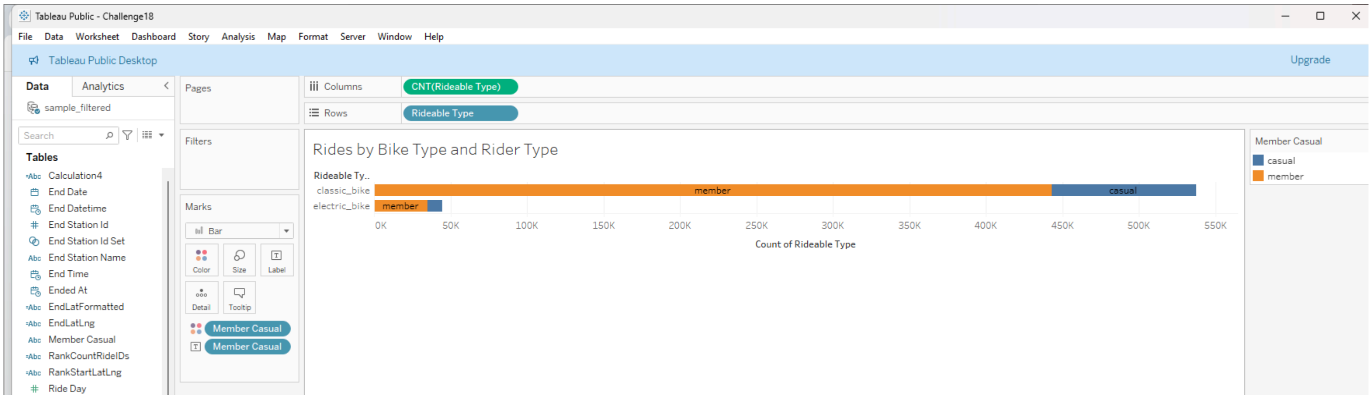
```
# Compare avg ride duration between casual and member riders

# Filter into dataframe where member_casual is 'casual'
casual_duration = sample_filtered_df[sample_filtered_df['member_casual'] == 'casual']

# Filter into dataframe where member_casusal is 'member'
member_duration = sample_filtered_df[sample_filtered_df['member_casual'] == 'member']

# Perform independent t-test for duration_seconds
t_statistic_duration, p_value_duration = stats.ttest_ind(casual_duration['duration_seconds'], member_duration['duration_seconds'])

# Print results
print("Duration t-statistic:", t_statistic_duration)
print("Duration p-value:", p_value_duration)
```

```
Duration t-statistic: 4.91650352521324
Duration p-value: 8.812753097582611e-07
```

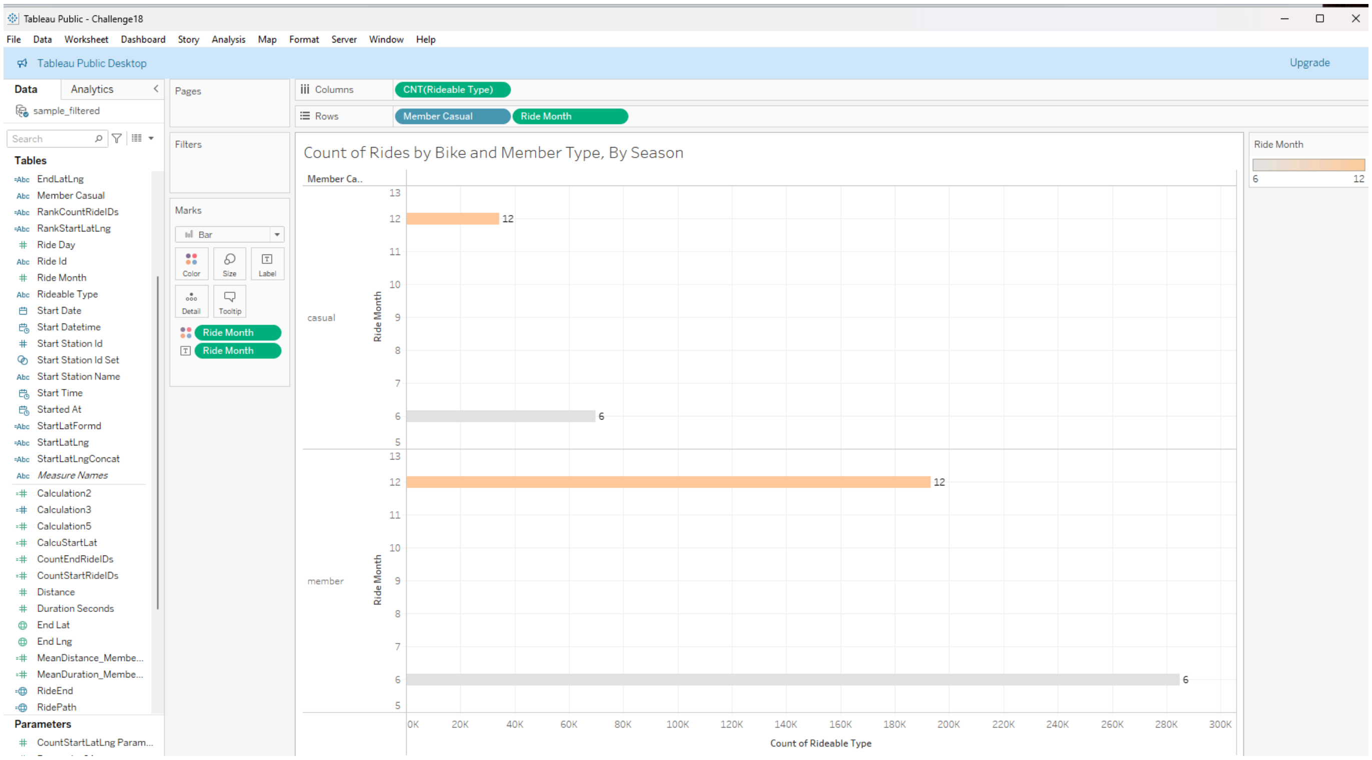
```
# Compare avg ride distance between casual and member riders

# Filter into dataframe where member_casual is 'casual'
casual_distance = sample_filtered_df[sample_filtered_df['member_casual'] == 'casual']

# Filter into dataframe where member_casusal is 'member'
member_distance = sample_filtered_df[sample_filtered_df['member_casual'] == 'member']

#Remove missing values from both samples
casual_distance = casual_distance.dropna(subset=['distance'])
member_distance = member_distance.dropna(subset=['distance'])

# Perform t-test for independent samples on distance
t_statistic_distance, p_value_distance = stats.ttest_ind(casual_distance['distance'], member_distance['distance'])

# Print results
print("Distance t-statistic:", t_statistic_distance)
print("Distance p-value:", p_value_distance)

#High t-stat and low p-stat imply that the average distance of the 'casual' riders
#is significantly different from the average distance of the 'member' riders
```

```
Distance t-statistic: 48.12108564764587
Distance p-value: 0.0
```

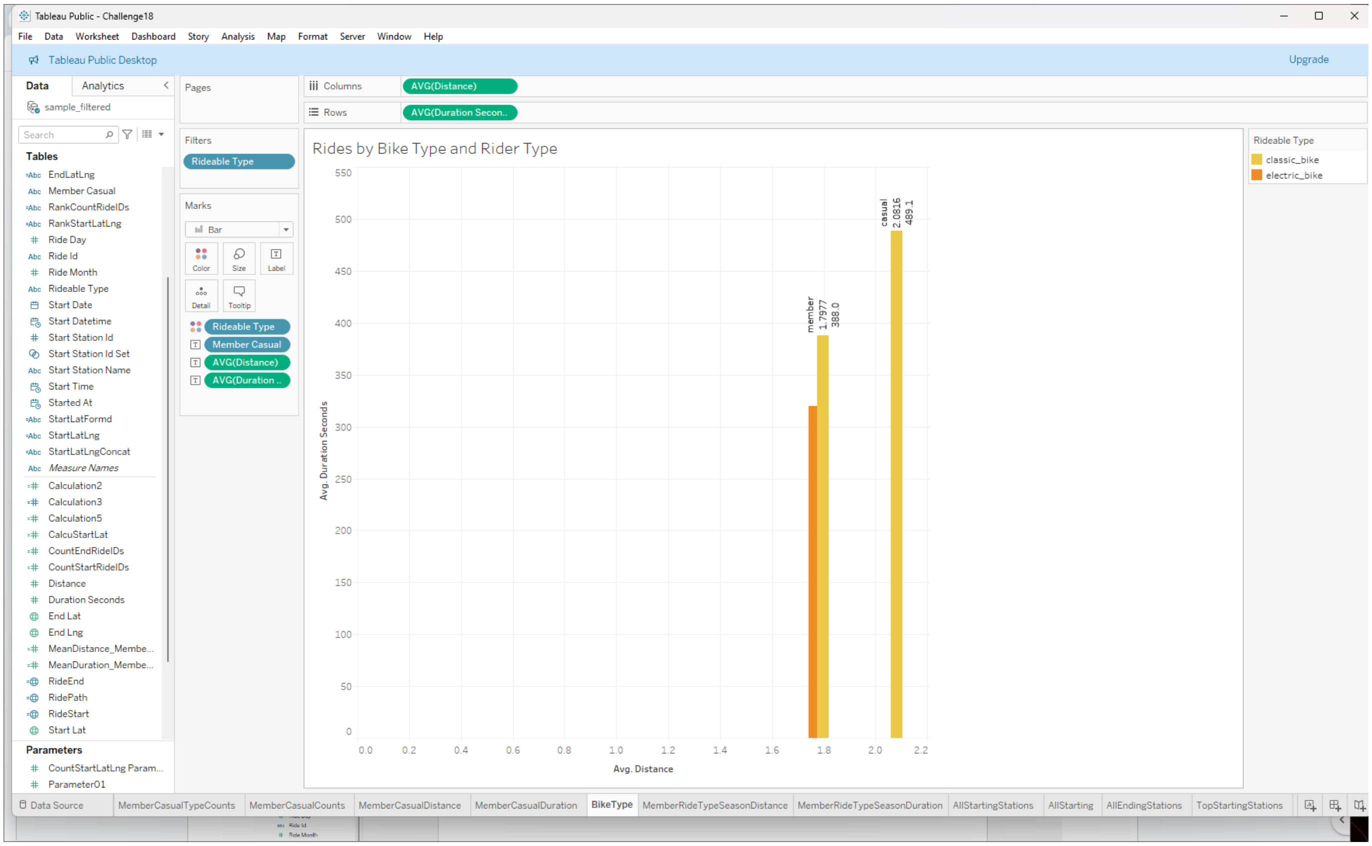
- There is a third dimension to the bike rides to consider, Ride Type. Bikes are either Classic or Electric. As we can see in the graph immediately below, far fewer rides are made on Electric bikes. This is likely because there are fewer Electric bikes than Classic.



- When we add seasonality to the graph, we can see that both Member and Casual riders are using electric bikes, but Members are using them much more extensively than Casuals, in both seasons. This may indicate that Electric bikes aren't as available as Classic bikes are when Casual riders are interested in riding. It could also suggest that Casual riders are not as familiar with Electric bikes and that they prefer Classic bikes.
- If the City decides to eventually replace Classic bikes with Electrical bikes, educational signage and marketing may be needed to transition Casual riders to Electric bikes and avoid a decrease in CitiBike usage.



- Note that when Member riders use Electric bikes, they tend to ride a very similar average distance in Kilometers as when they ride Classic bikes. This reinforces the idea that Members are using bikes to get from point A to point B. That their average ride duration in seconds is noticeably shorter on electric bikes indicates that electric bikes get Members to their destinations more quickly than Classic bikes.



Columns: AVG(Distance)

Rows: AVG(Duration Secon..

Filters: Rideable Type

Marks: Bar

Rideable Type
Member Casual
AVG(Distance)
AVG(Duration ..

<u>Citi Bike Stations</u>

- In the sample analyzed, every ride had a starting latitude, a starting longitude, an ending latitude, and an ending longitude. Because any station can be both a starting station and an ending station, two separate maps indicate where the sample rides started and where they ended.
- The clusters of bike stations are very similar between the starting and ending maps. The most notable differences are that the ending locations are spread wider than the starting locations in these directions: southeast, northeast and west of Manhattan. This suggests that commuters are riding home from lower and midtown Manhattan, but not as frequently riding into the city by bike.
- Additionally, the most popular ending locations are clustered in both lower and midtown Manhattan and in the outer boroughs and Jersey City.
- The ten most popular starting and ending stations are in midtown and lower Manhattan. In fact, most of these stations ranked in the top 10 for starting and ending rides.

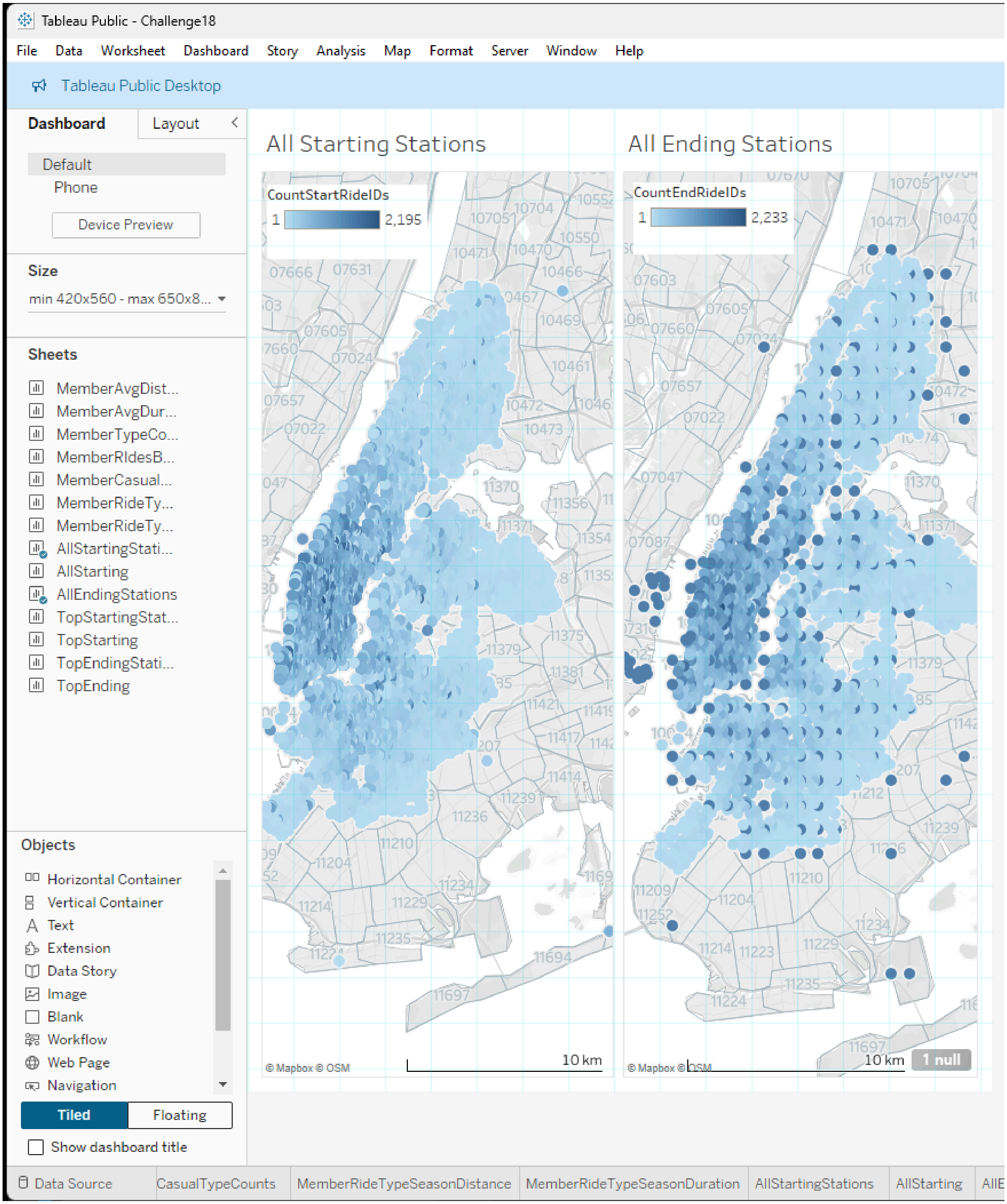


Tableau Public - Challenge18

All Starting Stations

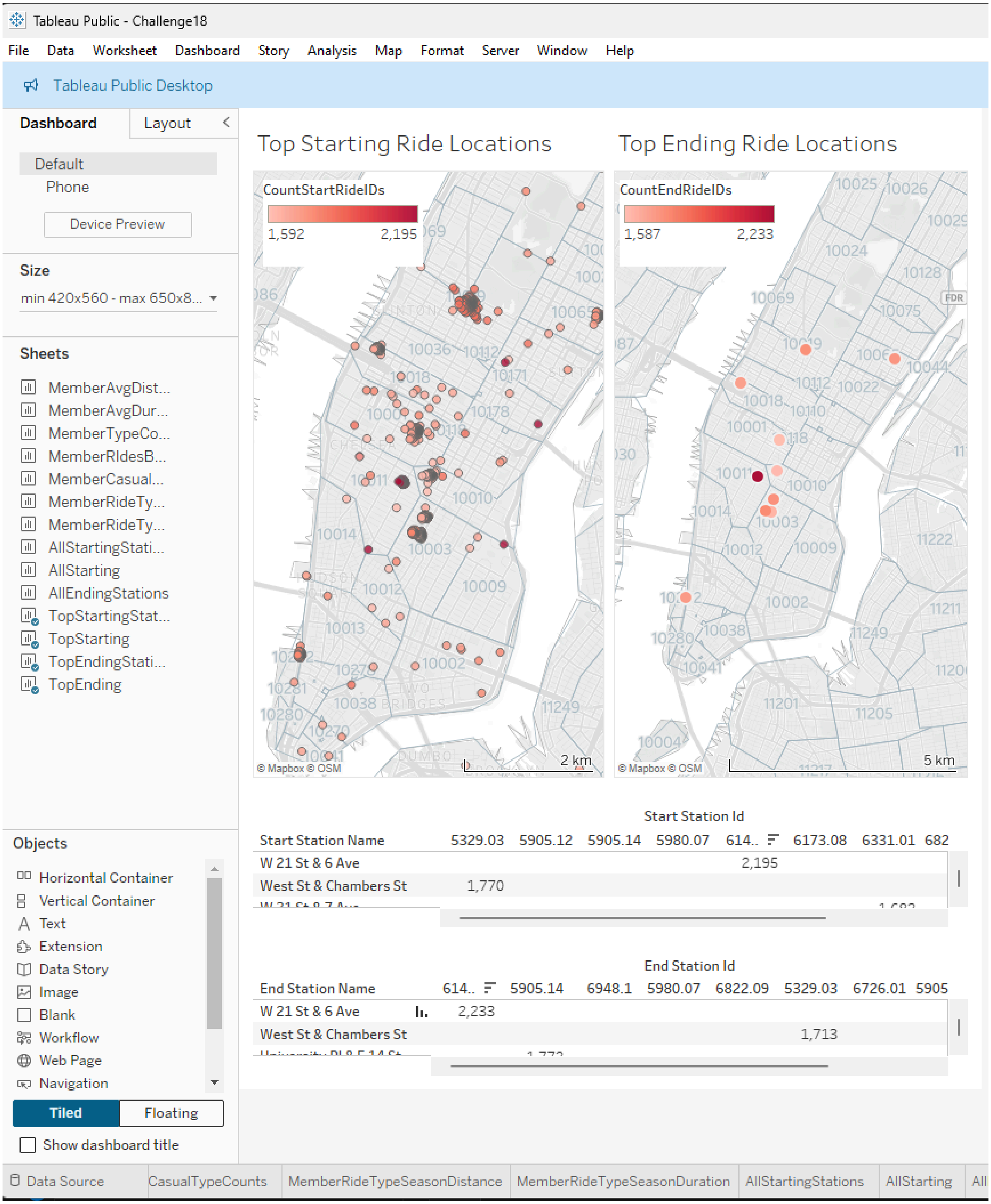
CountStartRideIDs 1 – 2,195

© Mapbox © OSM 10 km

All Ending Stations

CountEndRideIDs 1 – 2,233

© Mapbox © OSM 10 km 1 null



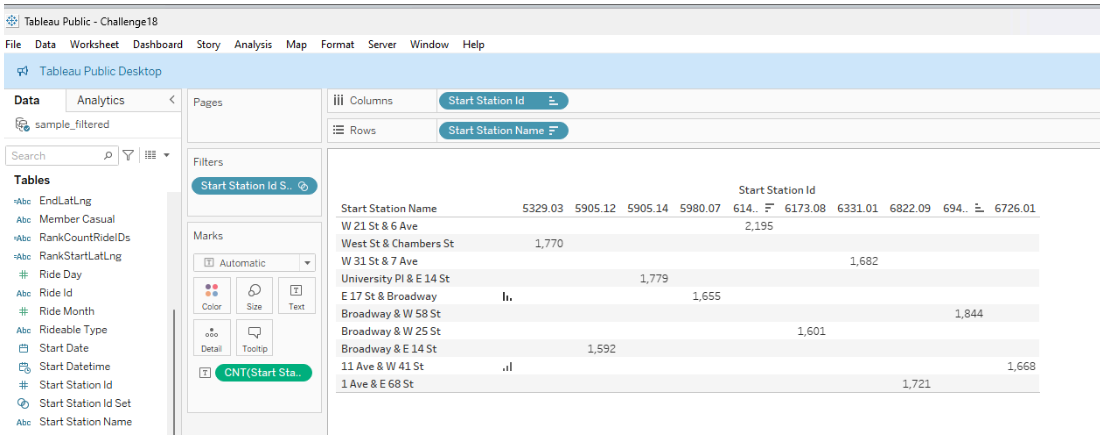
Start Station Id

Start Station Name	5329.03	5905.12	5905.14	5980.07	614..	6173.08	6331.01	682
W 21 St & 6 Ave					2,195			
West St & Chambers St	1,770							
W 21 St & 7 Ave							1,693	

End Station Id

End Station Name	614..	5905.14	6948.1	5980.07	6822.09	5329.03	6726.01	5905
W 21 St & 6 Ave	2,233							
West St & Chambers St						1,713		
University Pl & E 14 St		1,772						

Top 10 Starting Stations



Columns: Start Station Id

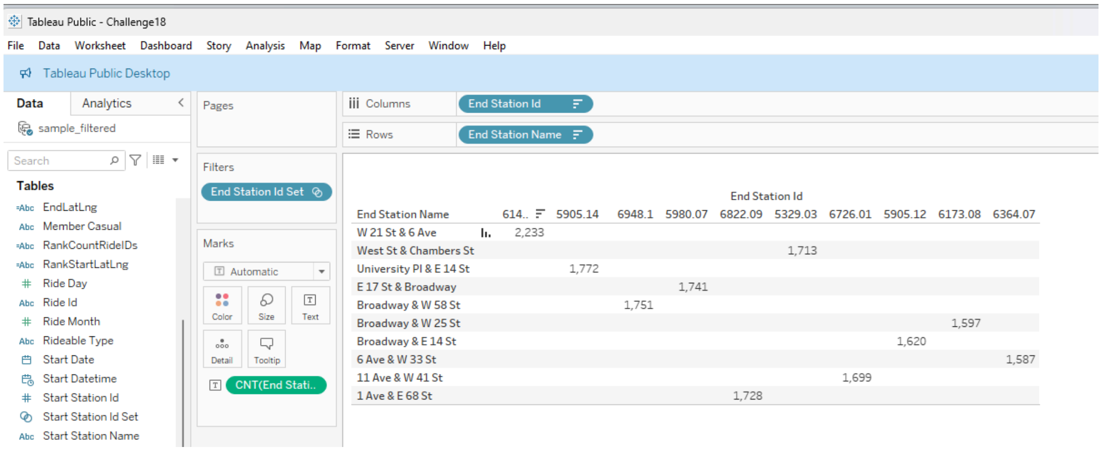
Rows: Start Station Name

Filters: Start Station Id S..

Marks: CNT(Start Sta..

	Start Station Id									
Start Station Name	5329.03	5905.12	5905.14	5980.07	614..	6173.08	6331.01	6822.09	694..	6726.01
W 21 St & 6 Ave					2,195					
West St & Chambers St	1,770									
W 31 St & 7 Ave							1,682			
University Pl & E 14 St			1,779							
E 17 St & Broadway				1,655						
Broadway & W 58 St									1,844	
Broadway & W 25 St						1,601				
Broadway & E 14 St		1,592								
11 Ave & W 41 St										1,668
1 Ave & E 68 St								1,721		

Top 10 Ending Stations



Columns: End Station Id

Rows: End Station Name

Filters: End Station Id Set

Marks: CNT(End Stati..

	End Station Id									
End Station Name	614..	5905.14	6948.1	5980.07	6822.09	5329.03	6726.01	5905.12	6173.08	6364.07
W 21 St & 6 Ave	2,233									
West St & Chambers St						1,713				
University Pl & E 14 St		1,772								
E 17 St & Broadway				1,741						
Broadway & W 58 St			1,751							
Broadway & W 25 St									1,597	
Broadway & E 14 St								1,620		
6 Ave & W 33 St										1,587
11 Ave & W 41 St							1,699			
1 Ave & E 68 St					1,728					

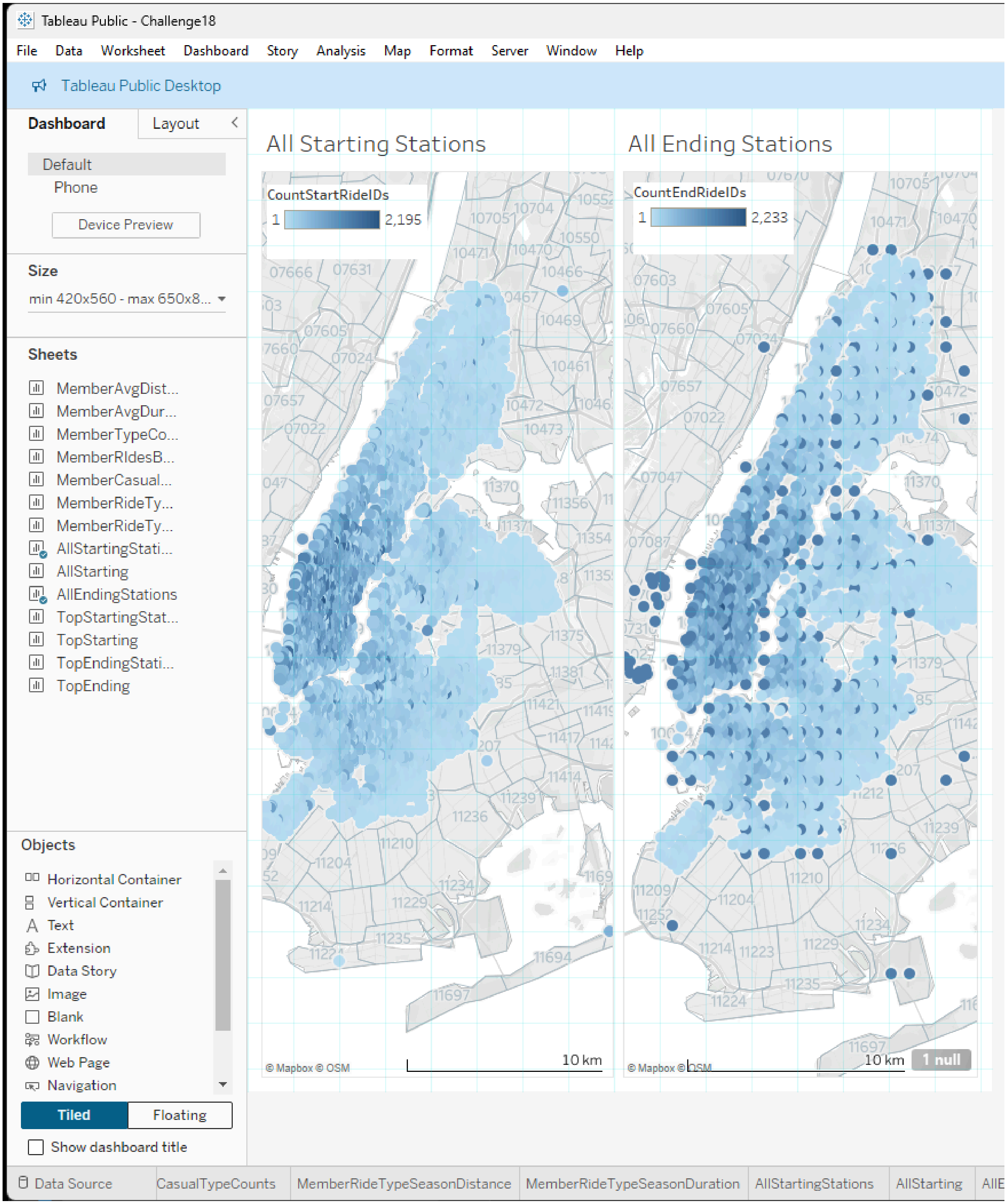
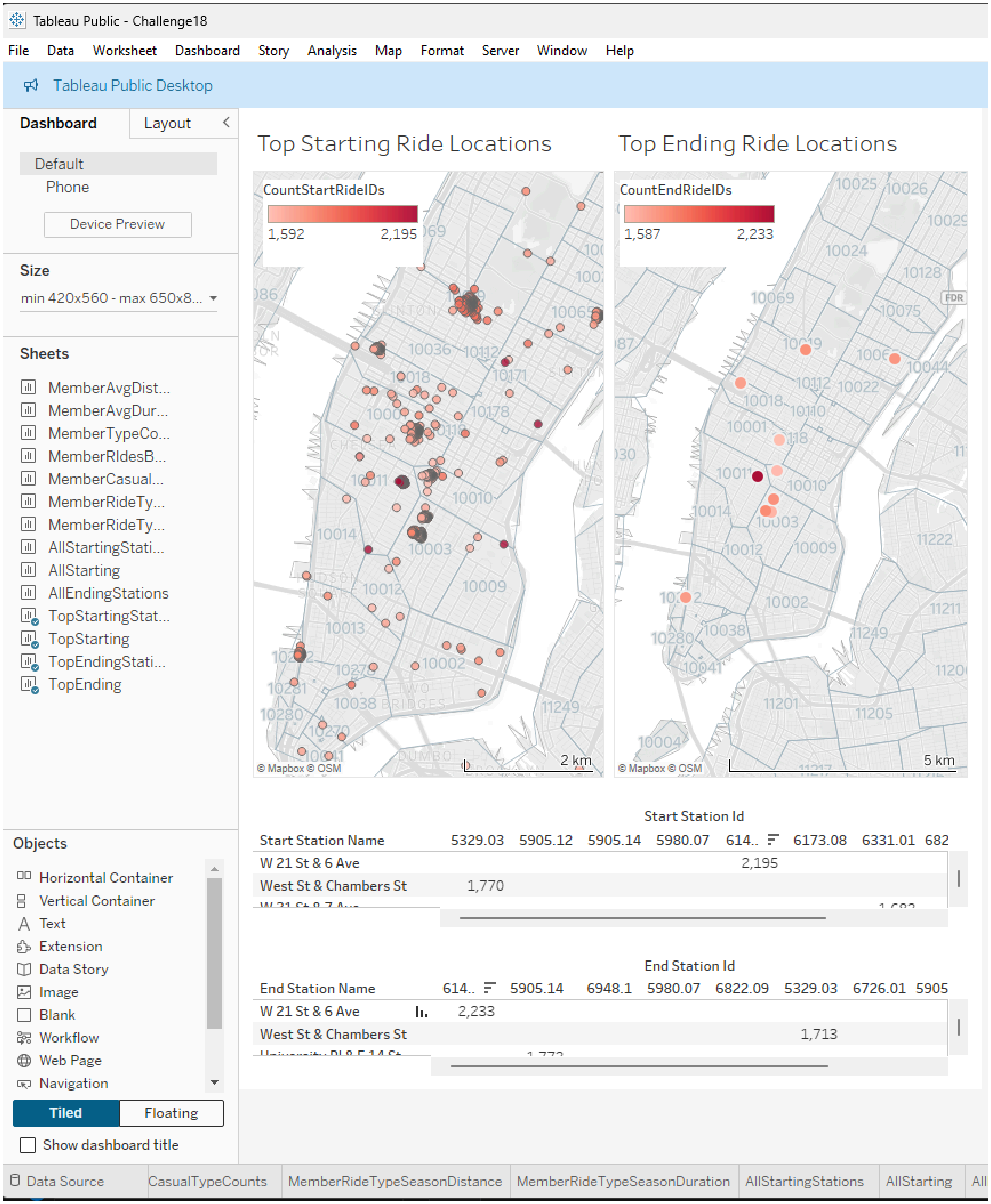


Tableau Public - Challenge18
File
Data
Worksheet
Dashboard
Story
Analysis
Map
Format
Server
Window
Help
Tableau Public Desktop
Dashboard
Layout
Default
Phone
Device Preview
Size
min 420x560 - max 650x8...
Sheets
MemberAvgDist...
MemberAvgDur...
MemberTypeCo...
MemberRIdesB...
MemberCasual...
MemberRideTy...
MemberRideTy...
AllStartingStati...
AllStarting
AllEndingStations
TopStartingStat...
TopStarting
TopEndingStati...
TopEnding
Objects
Horizontal Container
Vertical Container
Text
Extension
Data Story
Image
Blank
Workflow
Web Page
Navigation
Tiled
Floating
Show dashboard title
All Starting Stations
CountStartRideIDs
1
2,195
© Mapbox © OSM
10 km
All Ending Stations
CountEndRideIDs
1
2,233
© Mapbox © OSM
10 km
1 null
Data Source
CasualTypeCounts
MemberRideTypeSeasonDistance
MemberRideTypeSeasonDuration
AllStartingStations
AllStarting
AllE



Top Starting Ride Locations

CountStartRideIDs

1,592 — 2,195

© Mapbox © OSM

2 km

Top Ending Ride Locations

CountEndRideIDs

1,587 — 2,233

© Mapbox © OSM

5 km

	Start Station Id							
Start Station Name	5329.03	5905.12	5905.14	5980.07	614..	6173.08	6331.01	682
W 21 St & 6 Ave					2,195			
West St & Chambers St	1,770							

	End Station Id							
End Station Name	614..	5905.14	6948.1	5980.07	6822.09	5329.03	6726.01	5905
W 21 St & 6 Ave	2,233							
West St & Chambers St						1,713		

Ideas for Further Analysis

- Are the outbound rides ending at some of the most popular ending stations occurring during similar times of day and on similar days of the week? Do these trips coincide with breaks in train availability?
- How does travel and popularity of stations vary between weekdays and weekends?
- If the outer stations that are among the most popular ending stations are not used as much as beginning stations, what does that imply for relocating bikes back to the more popular starting stations? Does CitiBike need to transport bikes back into Manhattan if riders don't take them into the city?
- How many Classic and Electric bikes were in circulation in June 2023? How many were in circulation in December 2023? Did Electric bikes increase in number between summer and winter?

All maps and visualizations can be found at Profile - christine.jauregui | Tableau Public under Challenge18.