**About Dataset**

Data contains:

 Trip Duration (seconds)

 Start Time and Date

 Stop Time and Date

 Start Station Name

 End Station Name

 Station Latitude/Longitude

 Bike ID

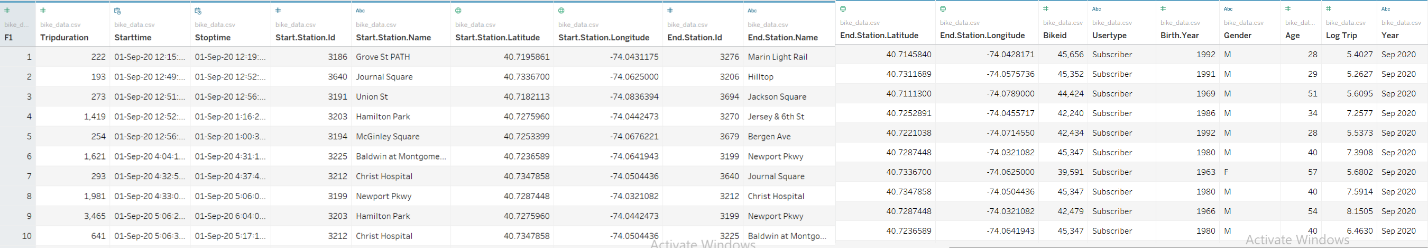
 User Type (Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member)

 Gender (F = Female; M = Male; NA)

 Year of Birth

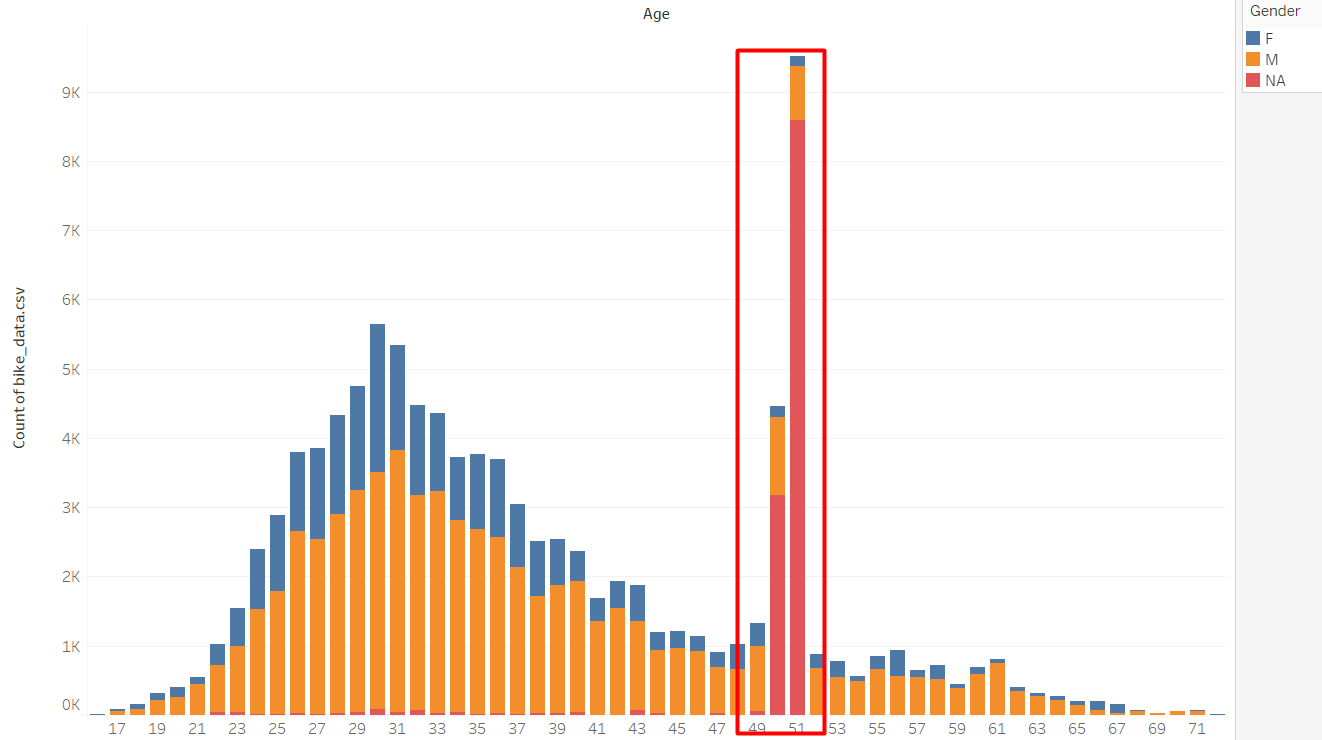
 F1 (row number from 1 to 103049)

This is how the top 10 rows looks like

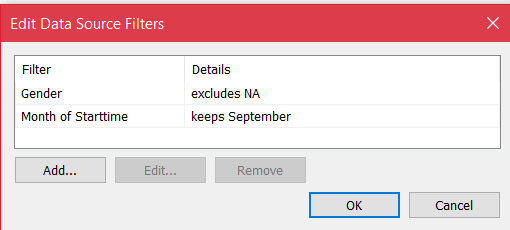


**Data Analysis**

Before beginning with data analysis, Let’s clean up the Gender and date column.

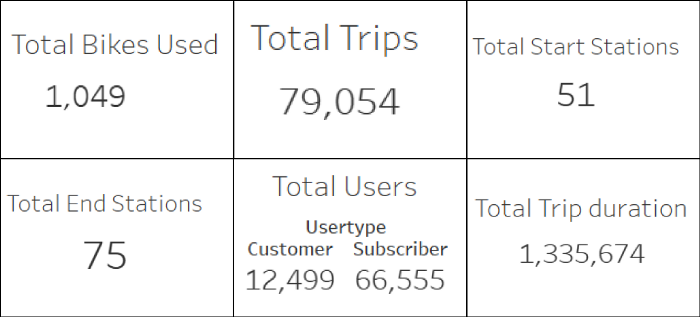


As we can see from this visualization that data contains falls value for the Gender “NA” . So we will first filter out the data. And also we will filter the date to include only September data.

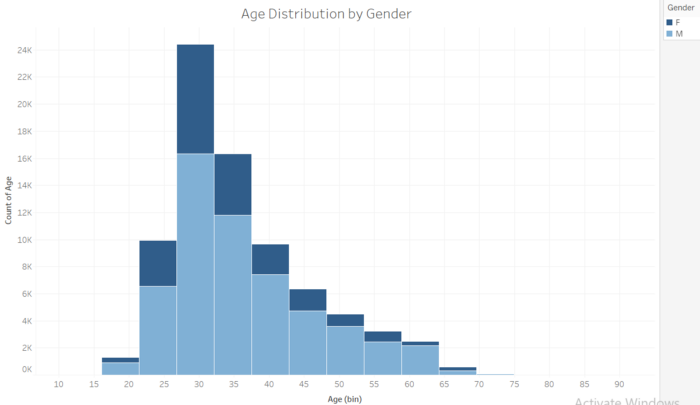


Now we are good to start with data analysis.

First let’s find out Total Bikes used, Total Trips taken, Total starting and ending Stations, Total Users and Total Trip duration(minutes)

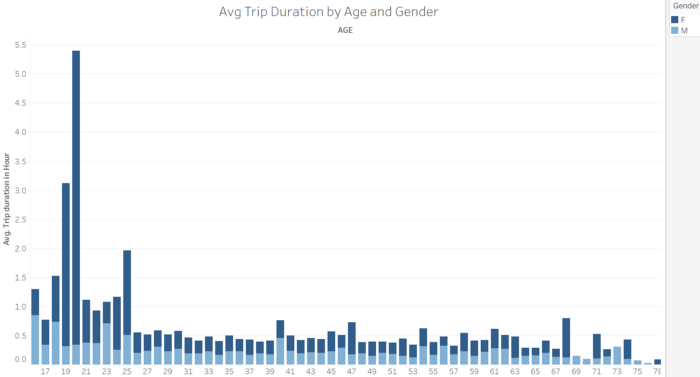


**Age Distribution**



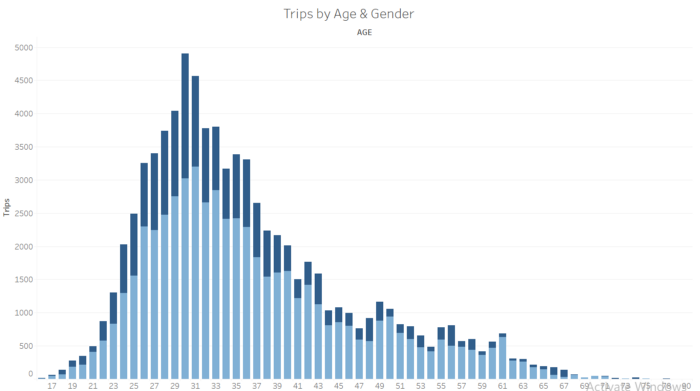
We can see most of the age lies between 27 to 36 and most of them are Male.

**Avg. Trip duration by Age & Gender**



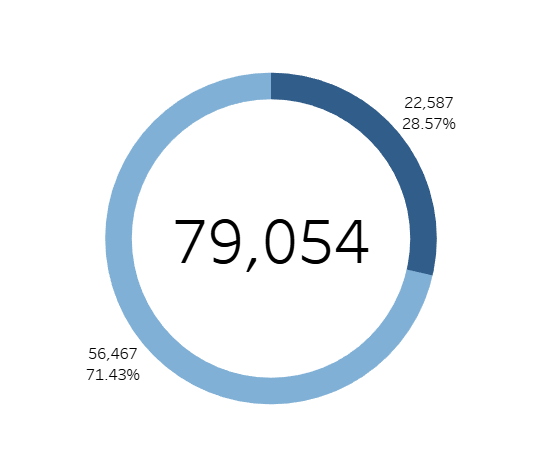
We can see that average trip duration of females of age 19 & 20 is highest.

**Total trips by Age & Gender**



We can see a slightly right skewed distribution. Most of the Trips are taken by males of age 25 to 38.

**Total trips by Gender**

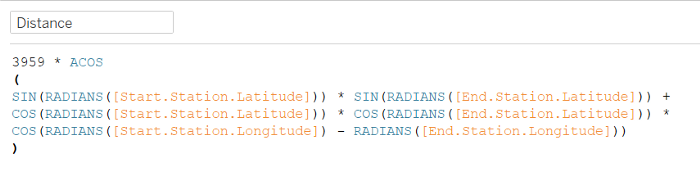


71.43% of Total trips are taken by Men and 28.57% of Total trips are taken by Female.

**Total Distance by Age**

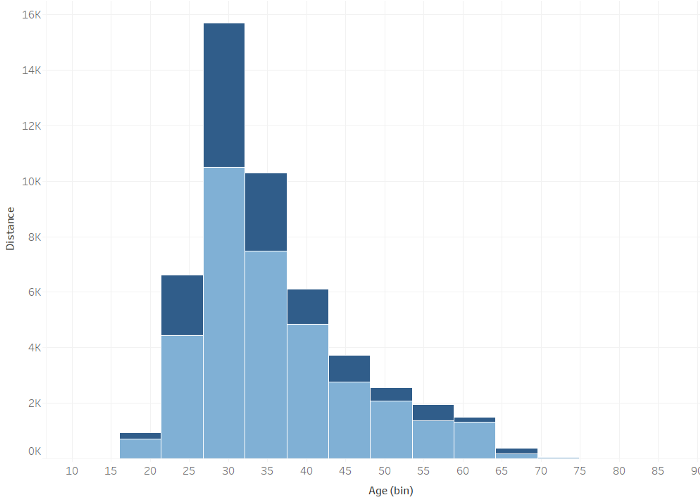
Since we don’t have a distance column , we will create one using latitude and longitude.

We will create a calculated field using the formula given below.



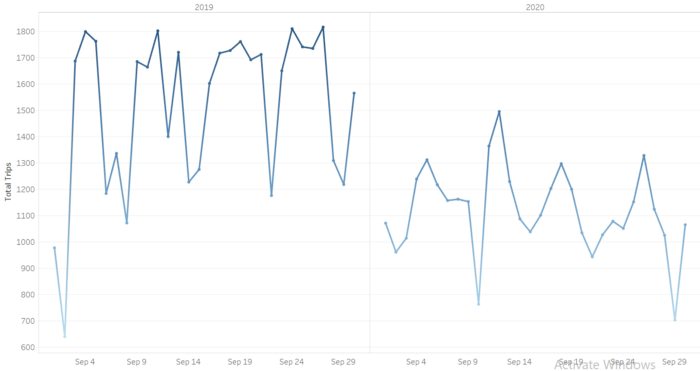
This calculates distance in miles. If we want it in Kilometers replace 3959 with 6371.

Note: The above calculation uses the Great Circle distance formula to balance complexity with accuracy, and uses the average radius of the Earth.



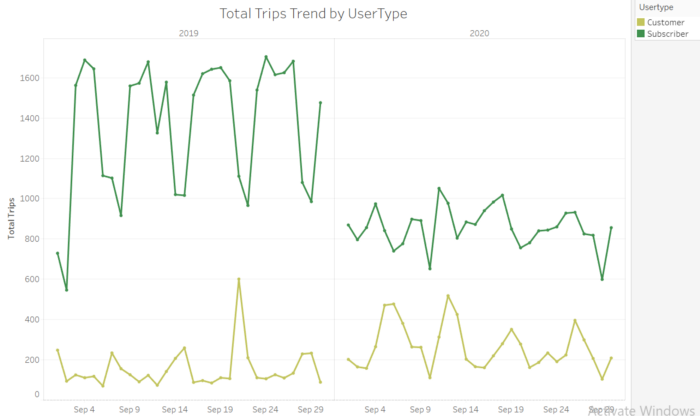
This visual shows the total distance by age and gender. As we can see men of age 27–38 have travelled larger distance.

**Total Trips Trend**



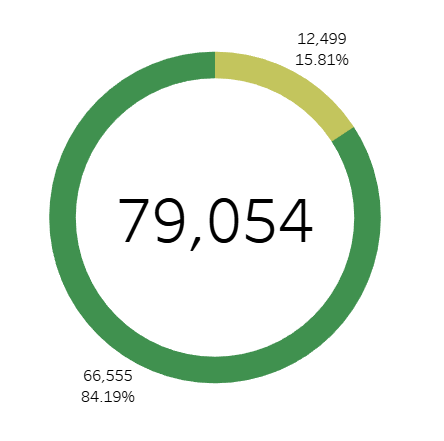
This visual shows the clear trend of Total trips. We can see that trips are decreased during 2020. May be these happened because of covid, since most people were at home and couldn’t go to their offices.

**Total Trips Trend by User Type**



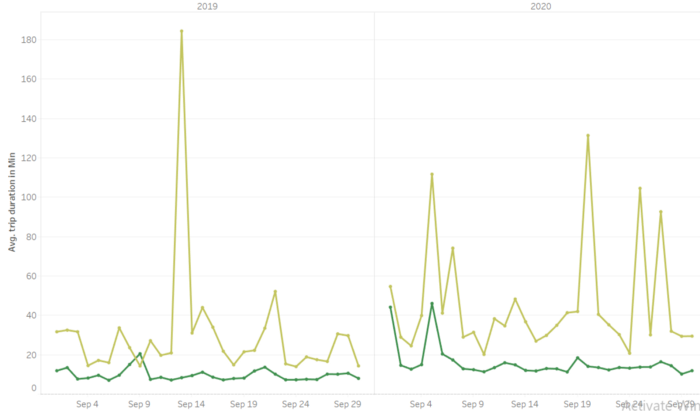
This visual shows us the total trip trend by user type. We can see that trend of customers were almost constant even in 2020 but trip taken by subscribers have been reduced. May be because most of the subscribers were either students or working professions and since offices and colleges were closed hence trips are decreased in 2020.

**Total Trips by User type**



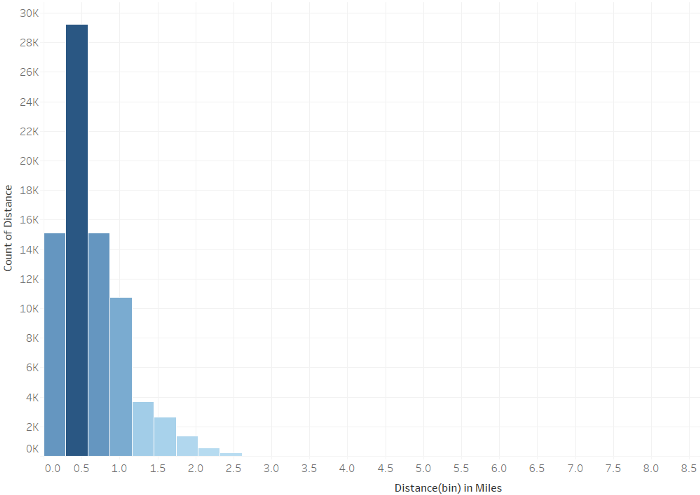
84.19% of Total trips are taken by Subscribers and 15.81% of Total trips are taken by Customers.

**Avg. Trip Duration(min) Trend by User Type**



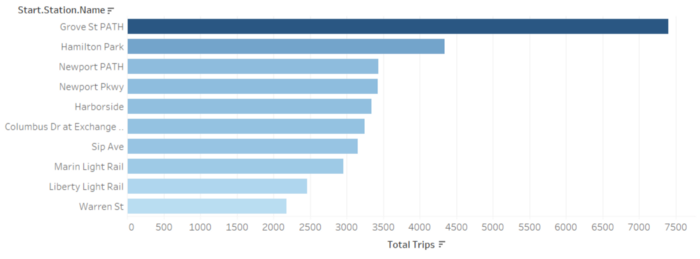
Even though subscribers have taken more trips but the Avg. Trip duration are higher for customers than subscribers. May be because the subscriber’s offices are nearby.

**Distance Distribution**



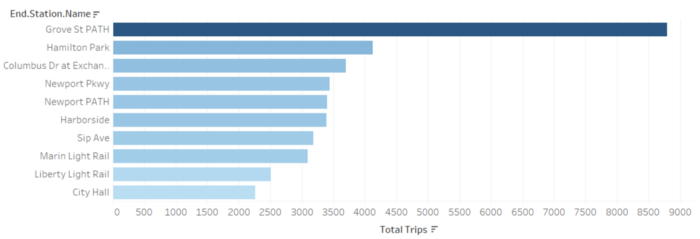
We can see most of the travel distances are below 1.2 miles.

**Top 10 Start Stations**



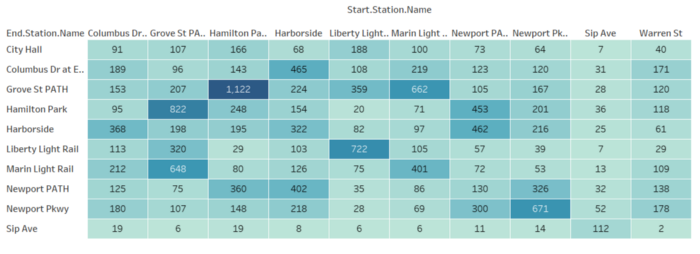
Here are the Top 10 start stations. We can see Grove St PATH is most active among other stations. May be because most of the users live near this station,

**Top 10 End Stations**



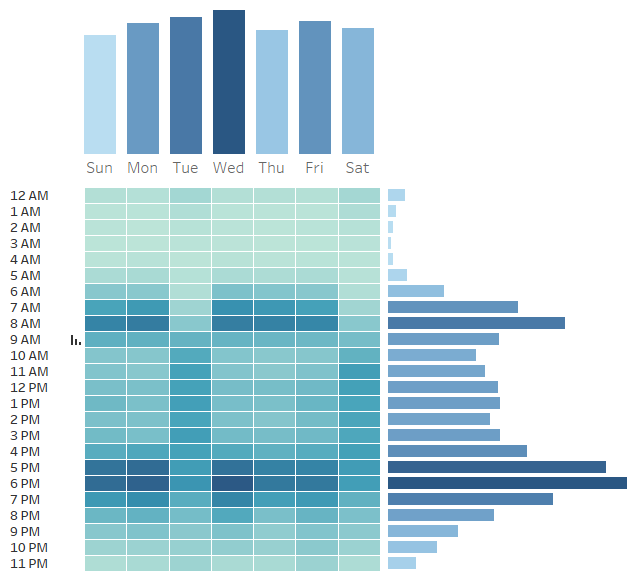
And here are the Top 10 end stations , again we can see Grove St PATH toping the list.

**Station Heatmap**



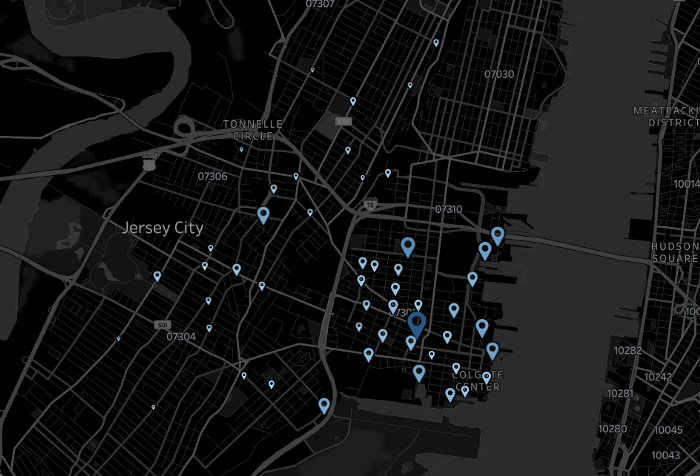
Highest no. of trips are taken from Hamilton Park to Grove St PATH.

**Heatmap of Hours by Weekday**



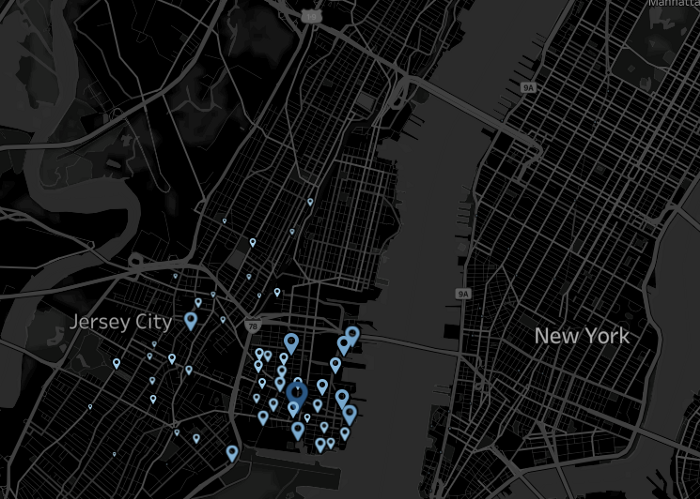
We can see that most of the trips are taken at 7,8 AM and 5,6 PM. And Trips taken on Sundays are less. This is may be because users go to office or college at 7,8 AM and return home at 5,6 PM. And since Sunday is holiday we can see the no. trips are less.

**Start Station Map**



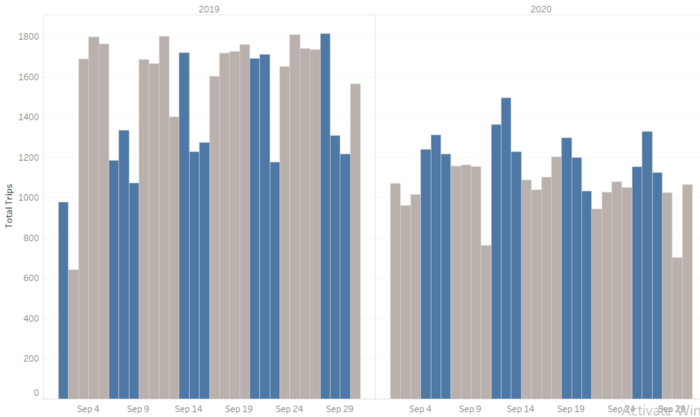
Size of the pointer tells about the no. trips taken from particular station. We can see most of the stations are at east side.

**End Station Map**



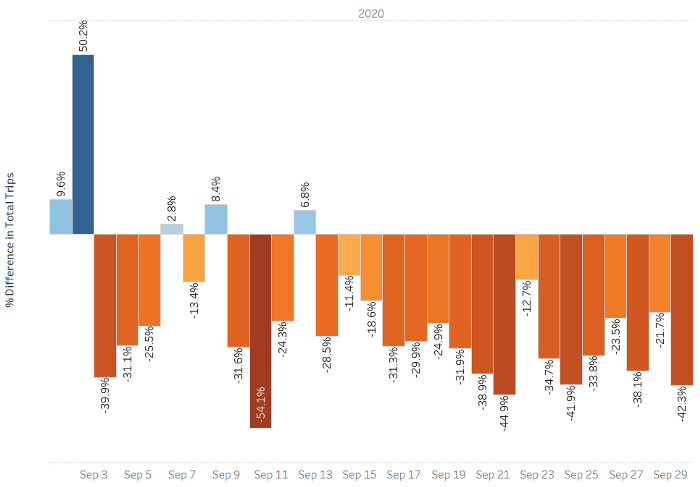
These are the locations of end station. Again we can see most of the end stations are nearby to the start stations. But there are few end stations near New York.

**Total Trips comparison between Weekends & Weekends**



The blue shaded columns shows the total trips during weekends(Fri-Sun). We can see that for 2019, most of the trips are taken during weekdays (grey shaded) where as for 2020, most of the trips are taken during weekends. May be again these trends are because of covid.

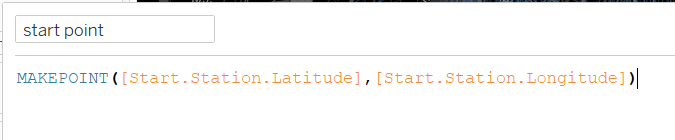
**Percentage of difference in total trips for 2020 & 2019**

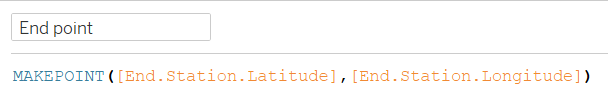


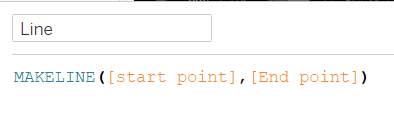
As we can see total trips are reduced during most of the days in 2020.

**Line Map**

To create a line map we need to create 3 calculated field as shown below.



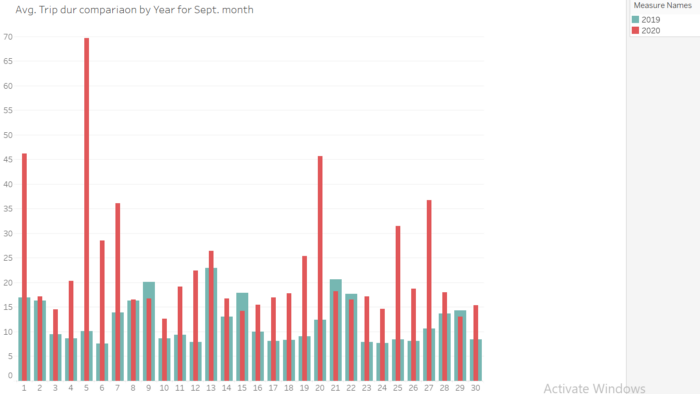






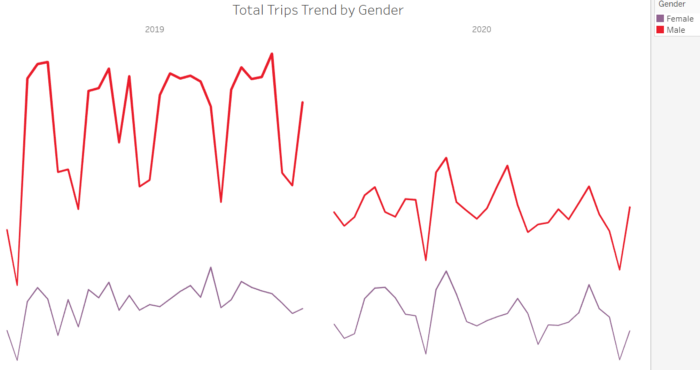
As we can see most of the lines are cluttered. This visual is not so helpful for our analysis.

**Avg. Trip duration comparison by Year for Sept. month**



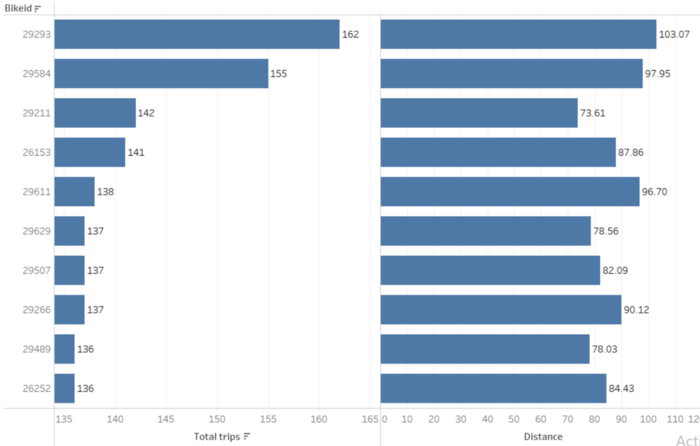
We can see that average trip duration for 2020 is more than for 2019.

**Total Trips trend by Gender**



We can see total trips taken by female is remained constant even in 2020. But male users are reduced.

**Top 10 Bike id’s**

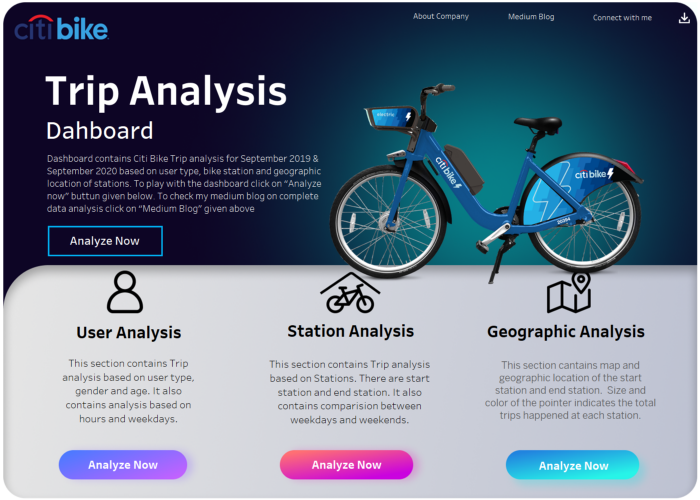


These are the top 10 bike ids by total trips. It’s total trips and distances are shown.

**Dashboard**

I created totally 4 pages

**Home**



This contains the overview of the dashboard and about 3 sections, that is User analysis , station analysis and geographic analysis.

**User analysis**



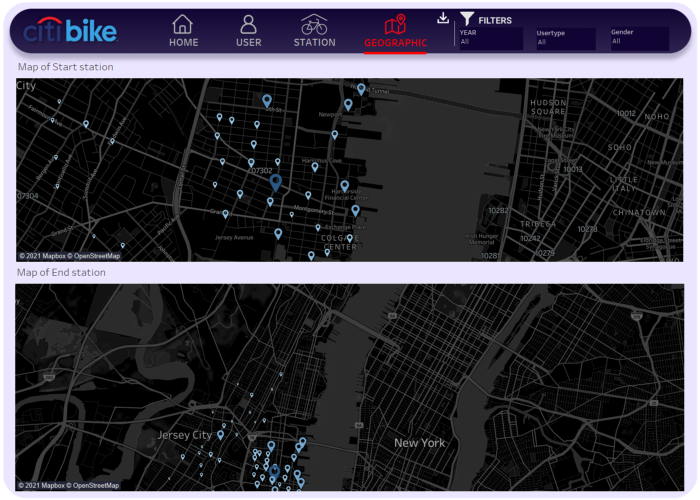
This section contains Trip analysis based on user type, gender and age. It also contains analysis based on hours and weekdays.

**Station analysis**



This section contains Trip analysis based on Stations. It also contains comparison between weekdays and weekends.

**Geographic analysis**



This section contains map and geographic location of the start station and end station. Size and color of the pointer indicates the total trips happened at each station.