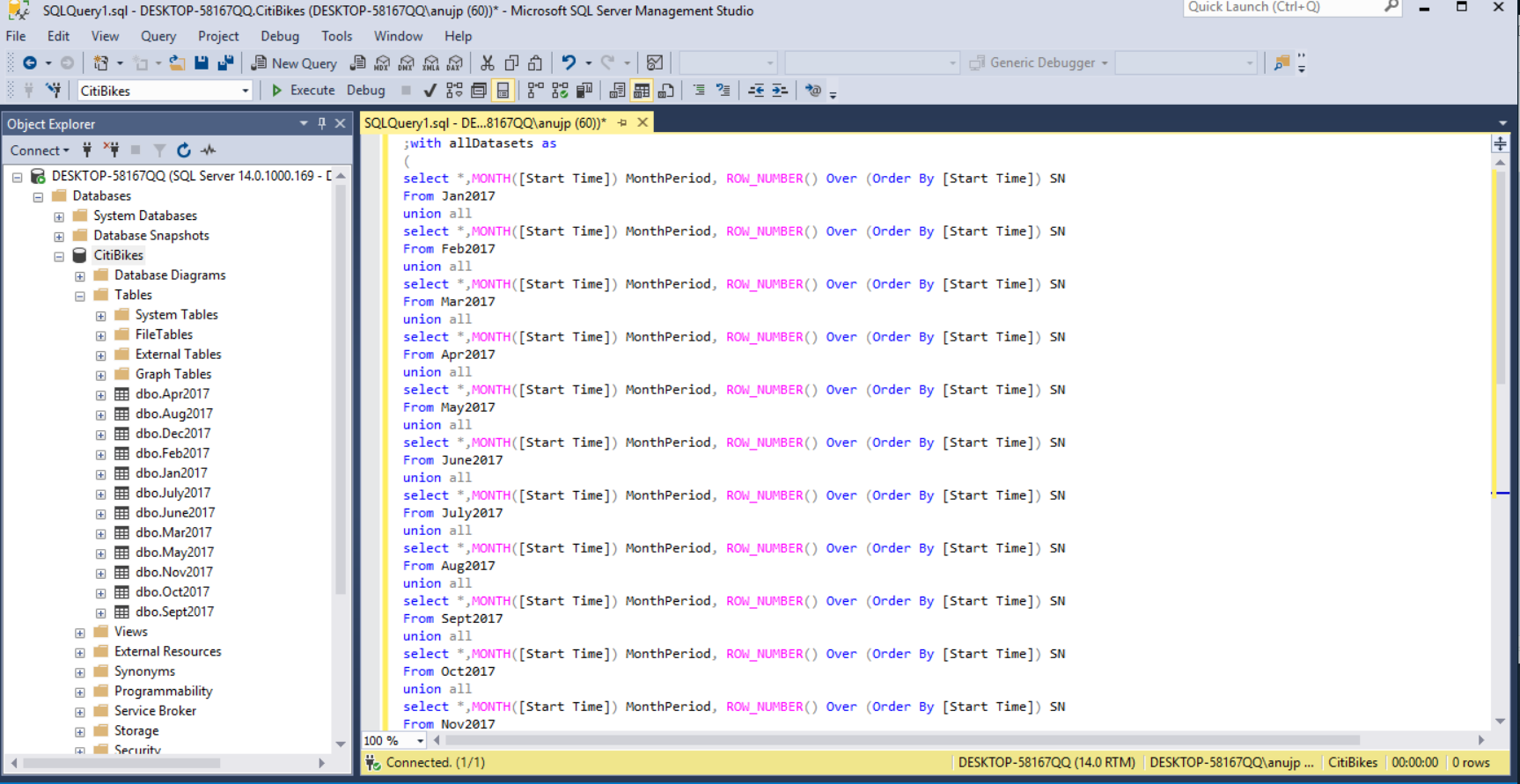
**TABLEAU ASSIGNMENT – CITIBIKE**

**Project Description:**

The data for this project is for the year 2017, from the Citi bike web site <https://www.citibikenyc.com/system-data>. The raw data has been aggregated outside of Tableau in a SQL server database (please see screen shots of queries that performed aggregation). This was done to minimize the overhead on Tableau so that visualizations are rendered at a faster pace. As, I have used Tableau public, the project is saved on a Tableau Public server and can be viewed at <https://public.tableau.com/profile/anuj.pandya#!/vizhome/AnujTableauHW/GenderBreakDown?publish=yes>. In case of difficulty opening the web page , please reach out to me for the login credentials. Additionally, below is a screen by screen breakout of the project.

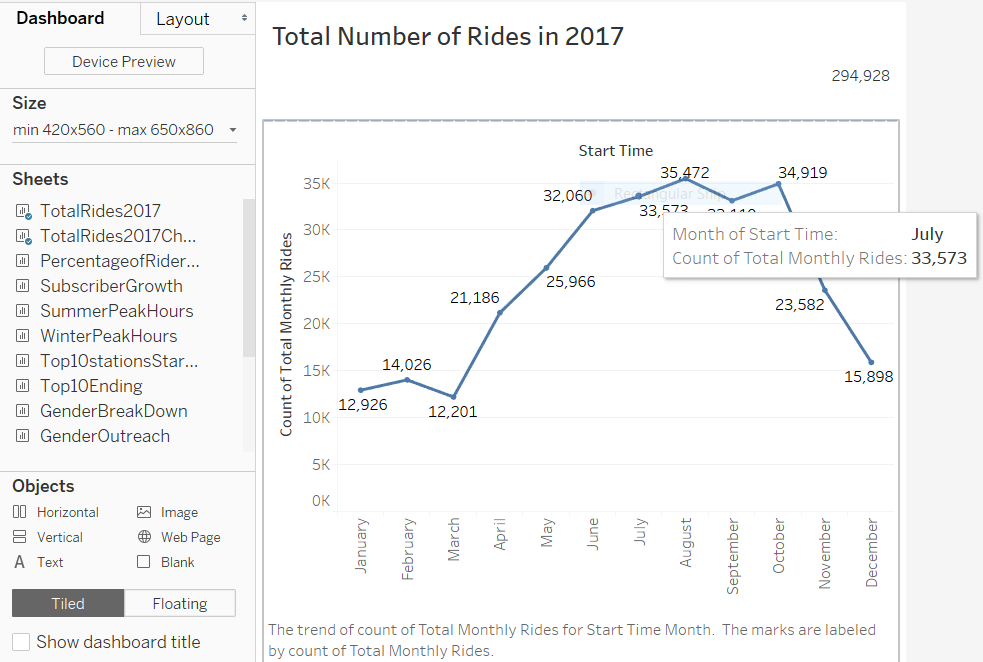
The aggregated dataset is on the GitHub along with the rest of assignment.

SQL aggregation queries:



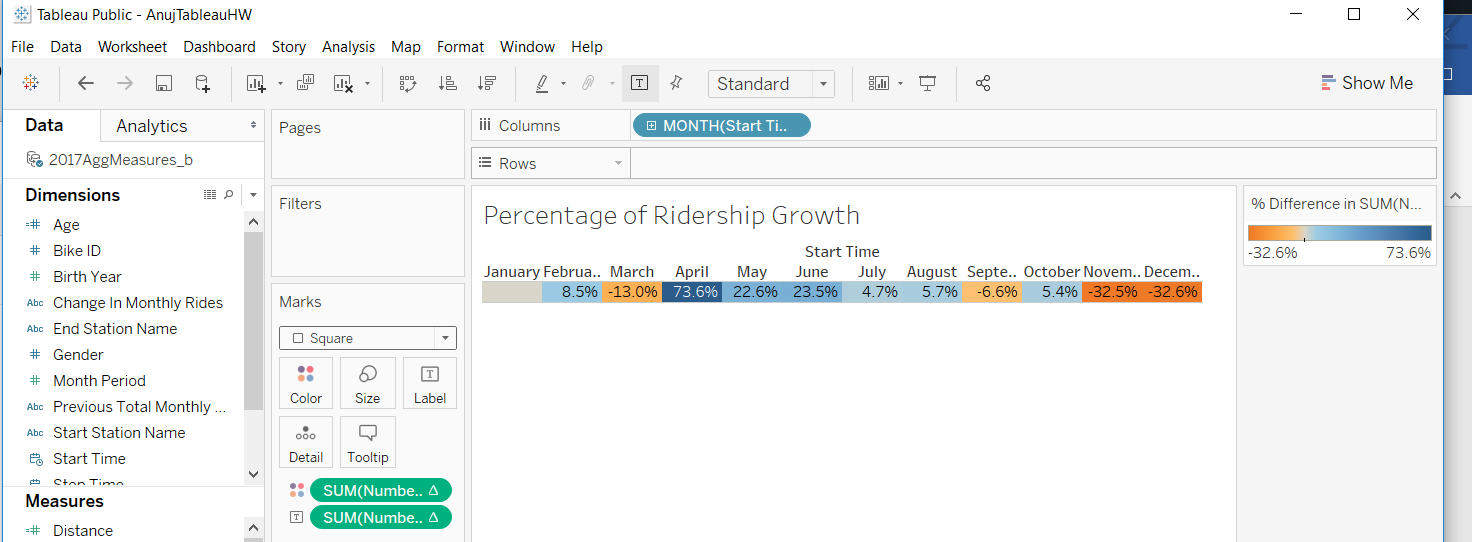
**Q1. How many Trips have been recorded during the chosen period?**

Period Chosen is the Year 2017. Plotted the Month of the Start Time as columns and the Count of the Number of Records as Rows.



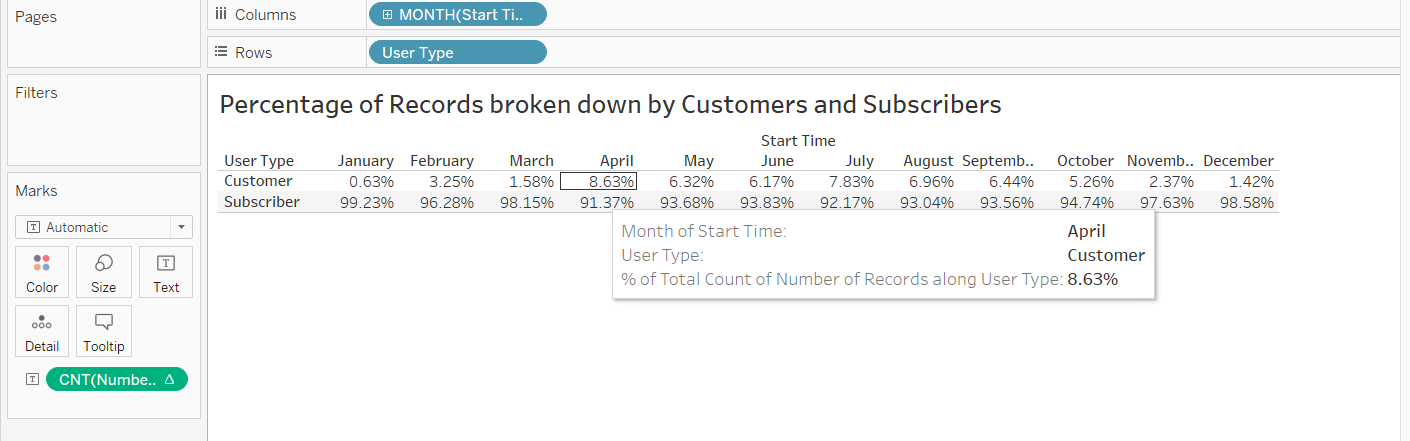
**Q2. By what percentage has total ridership grown?**

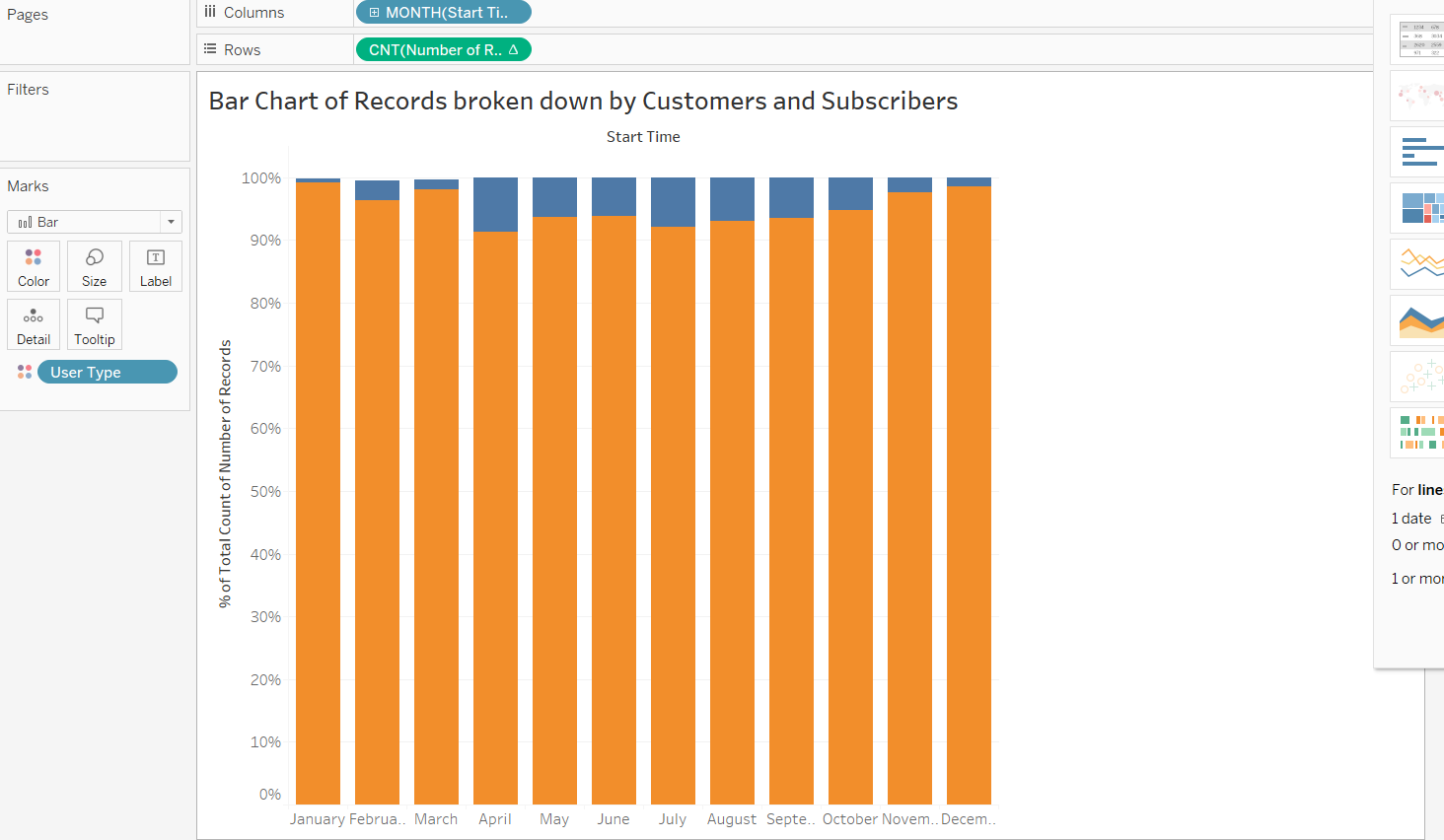
Ans. Plot the month as Columns and Number of Records as Rows. Change the Number of Rows to a COUNT and add a Table Calculation. Use “Percentage difference from” in Table calculation to get the percentage difference of the rides from the previous month. Use highlight table to show the change in colors.

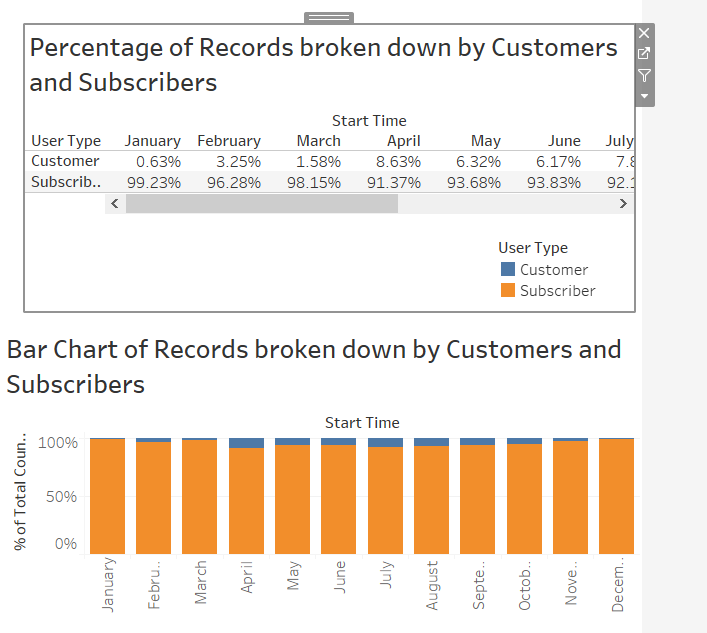


**Q3. How has the proportion of the Short-Term Customers and Annual Subscribers changed?**

Ans. Month is plotted on the Column. A calculation type of Percent of Total was performed. User Type (Customer or Subscriber) is plotted on the Rows. Numerical values and a bar chart are shown for analysis purposes. It is apparent that the subscribers are the main users and the customers tend to grow in the summer months.



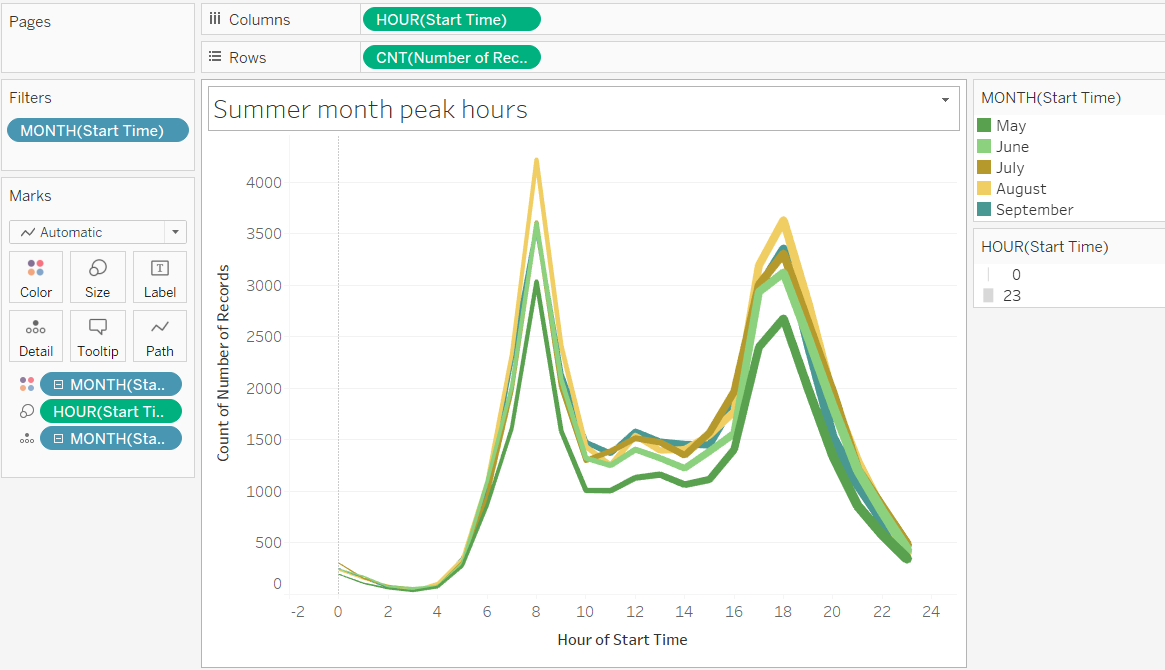




**Q5. What are the peak hours in which bikes are used during summer months?**

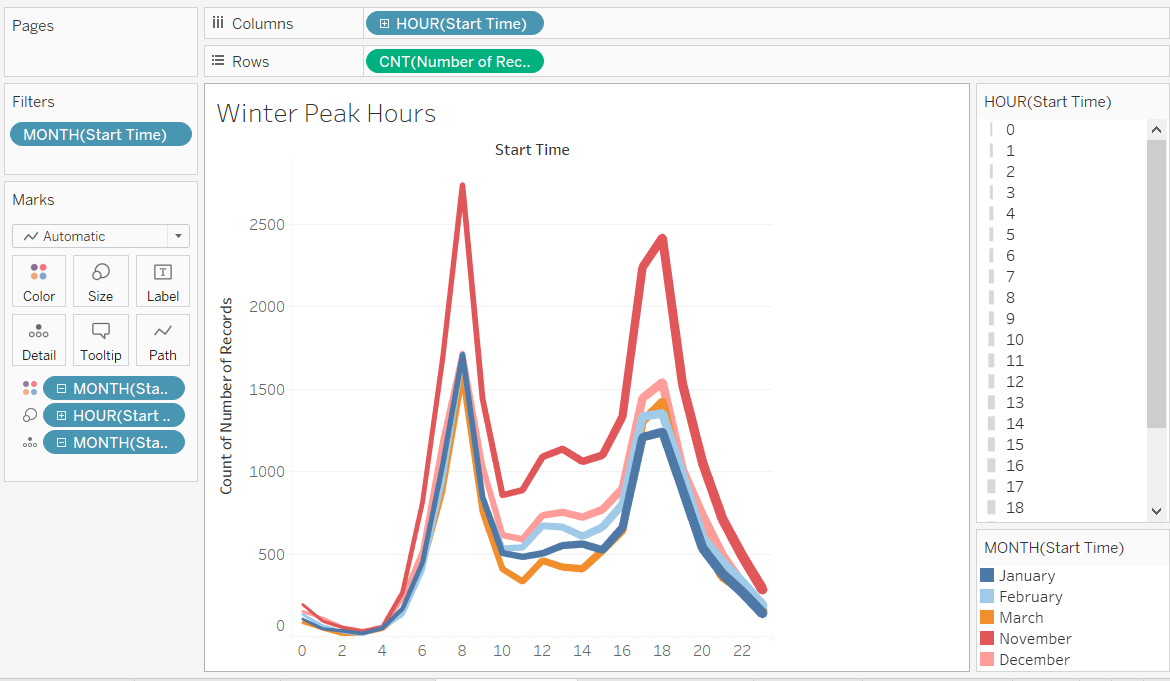
Ans. Used Hours from Start Time as columns, COUNT of the number of records as Rows. Filtered month for May thru September. Added Hour to size and month to color and detail.

It is apparent that the peak hours are 8AM and 6PM. This makes a lot of sense as this is peak rush hour time when people use bikes to get to their destinations.



**Q6. What are the peak hours in which bikes are used during winter months?**

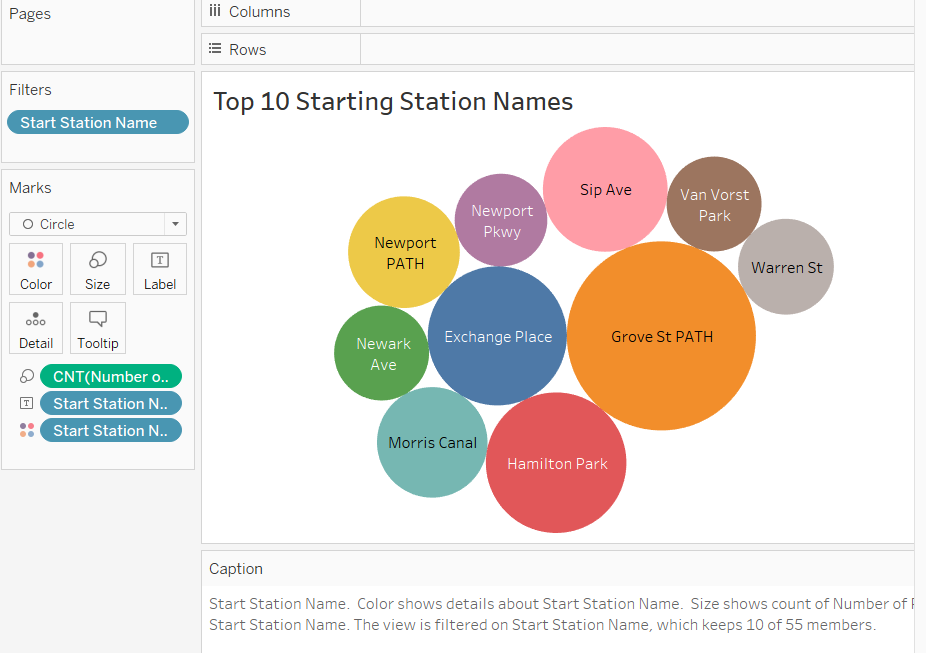
Ans. Used Hours from Start Time as columns, COUNT of the number of records as Rows. Filtered month for May thru September. Added Hour to size and month to color and detail.



**Q7. Top 10 Start Station Names?**

Ans. The Dimension Start Station Name was added to the columns and the Number of records added to the rows. A filter on the station name was added to show the top 10 by field.

Analysis: The locations in the figure are major hubs for public transportation for trains and buses. That’s the reason they are in the top 10 as bike riders could be biking to these hubs to travel further.



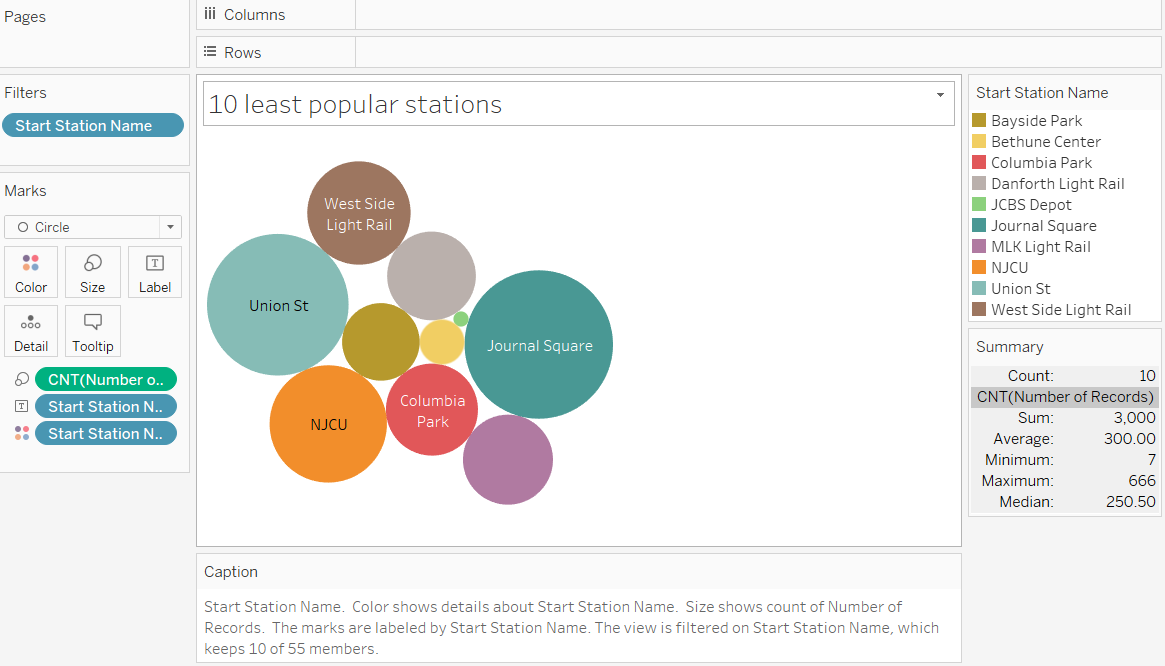
**Q7. Top 10 Ending Station Names?**

Ans. The Dimension End Station Name was added to the columns and the Number of records added to the rows. A filter on the station name was added to show the top 10 by field.

Analysis: The locations in the figure are major hubs for public transportation for trains and buses. That’s the reason they are in the top 10 as bike riders could be biking to these hubs to travel further.



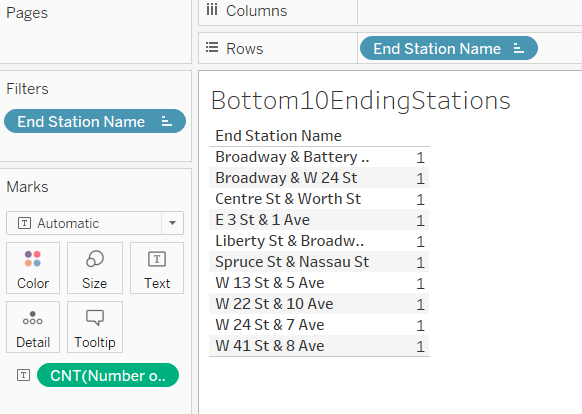
**Q8. Bottom 10 Start Station Names?**



**Q9. Bottom 10 Start End Names?**

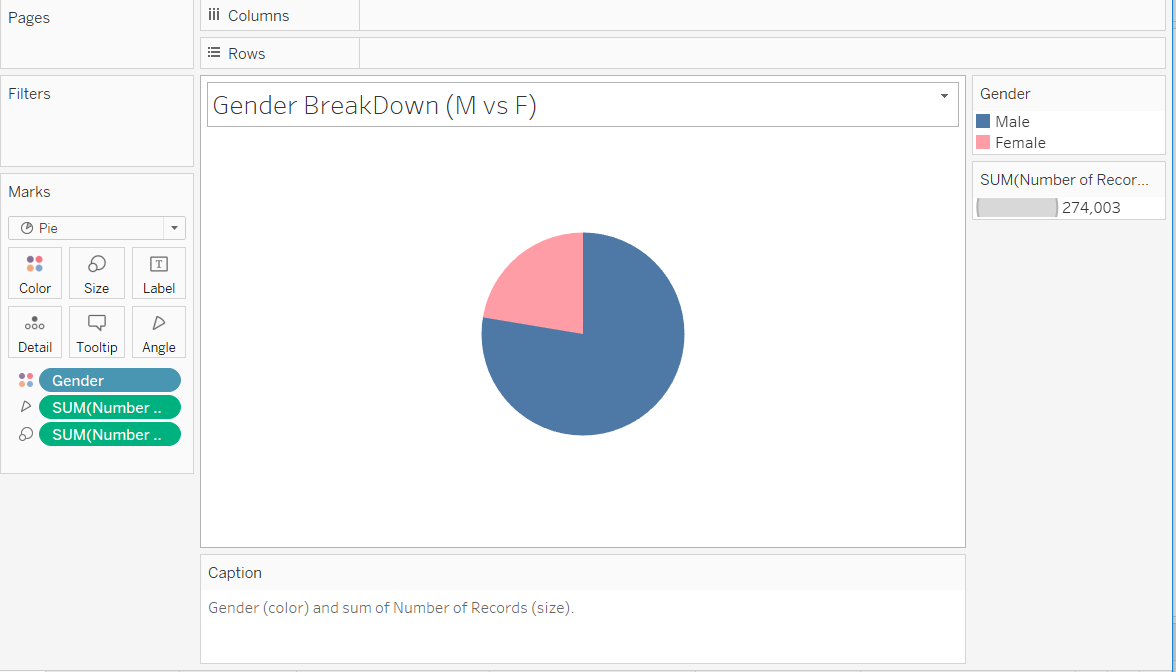
Ans. The 10 least popular stations all have count of 1, hence all bubbles are the same size. Please see text table below showing the number count.





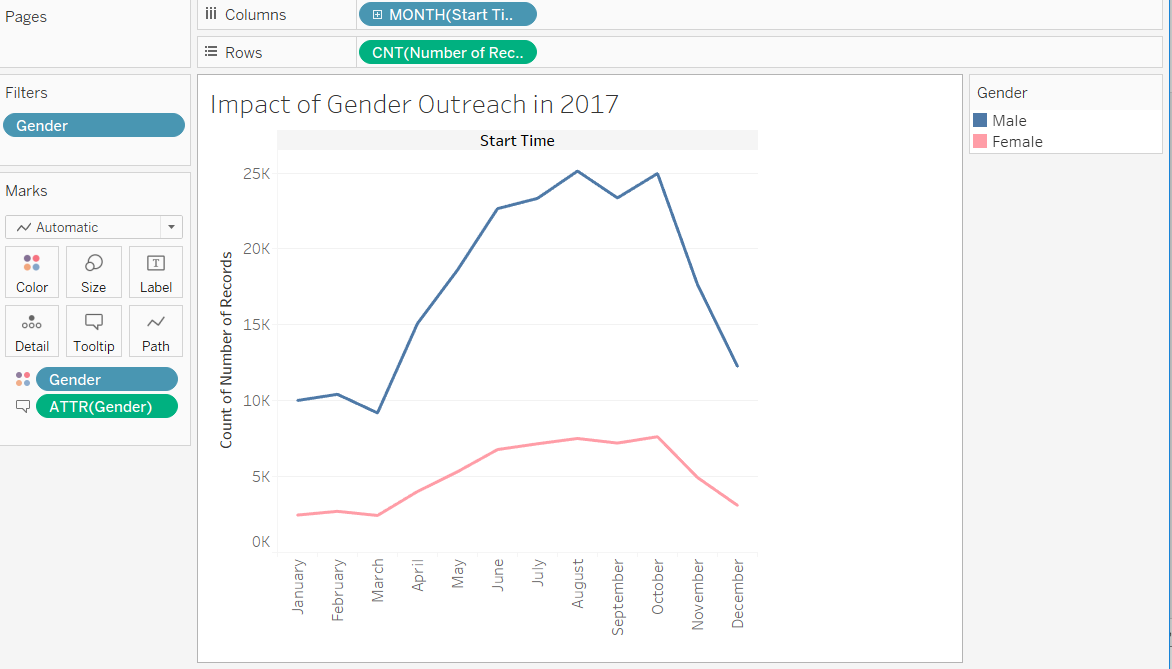
**Q10. Gender breakdown of active participants?**

Ans. Sum the number of records in columns and gender on rows. Edit the alias on gender 1 – Male , 2 – Female. Hide the Unknown gender. Ensure the gender is a dimension. Display results in a pie chart.



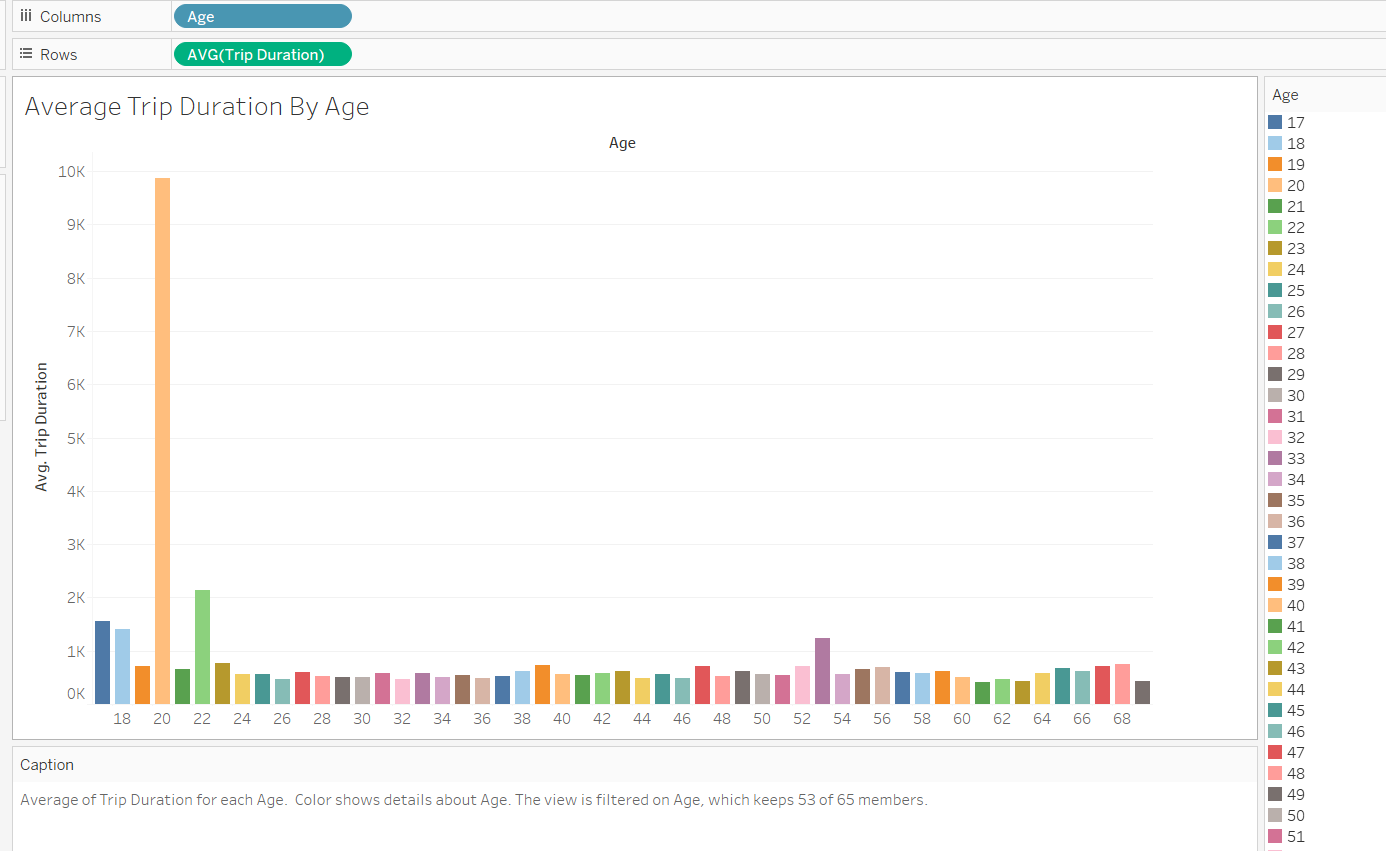
**Q11. Gender Outreach.**

Ans. Display a Time – Series (month of the start date), Gender as a discrete dimension and the count of the Number of records. Color code the Gender.



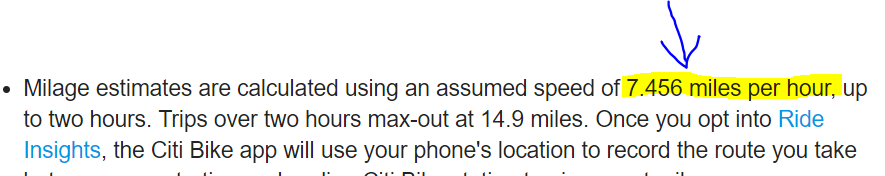
**Q12. How does the average Trip duration change by age?**

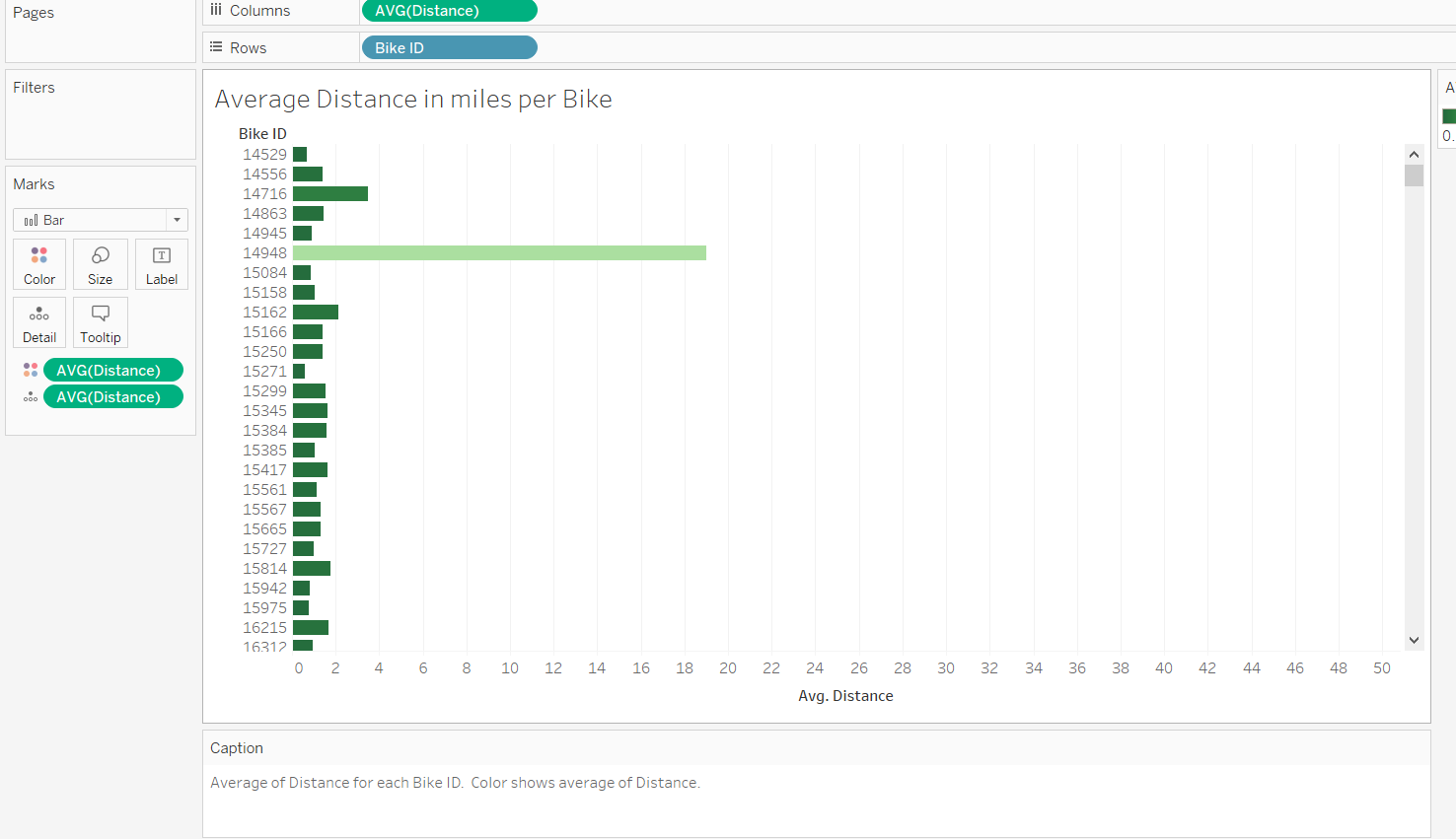
Ans. Create a calculated field “Age” by using formula: YEAR(TODAY()) - [Birth Year]. Add Age as a dimension and the Average of the measure “Trip Duration” as a column.



**Q13. What is the average distance in miles that a bike is ridden?**

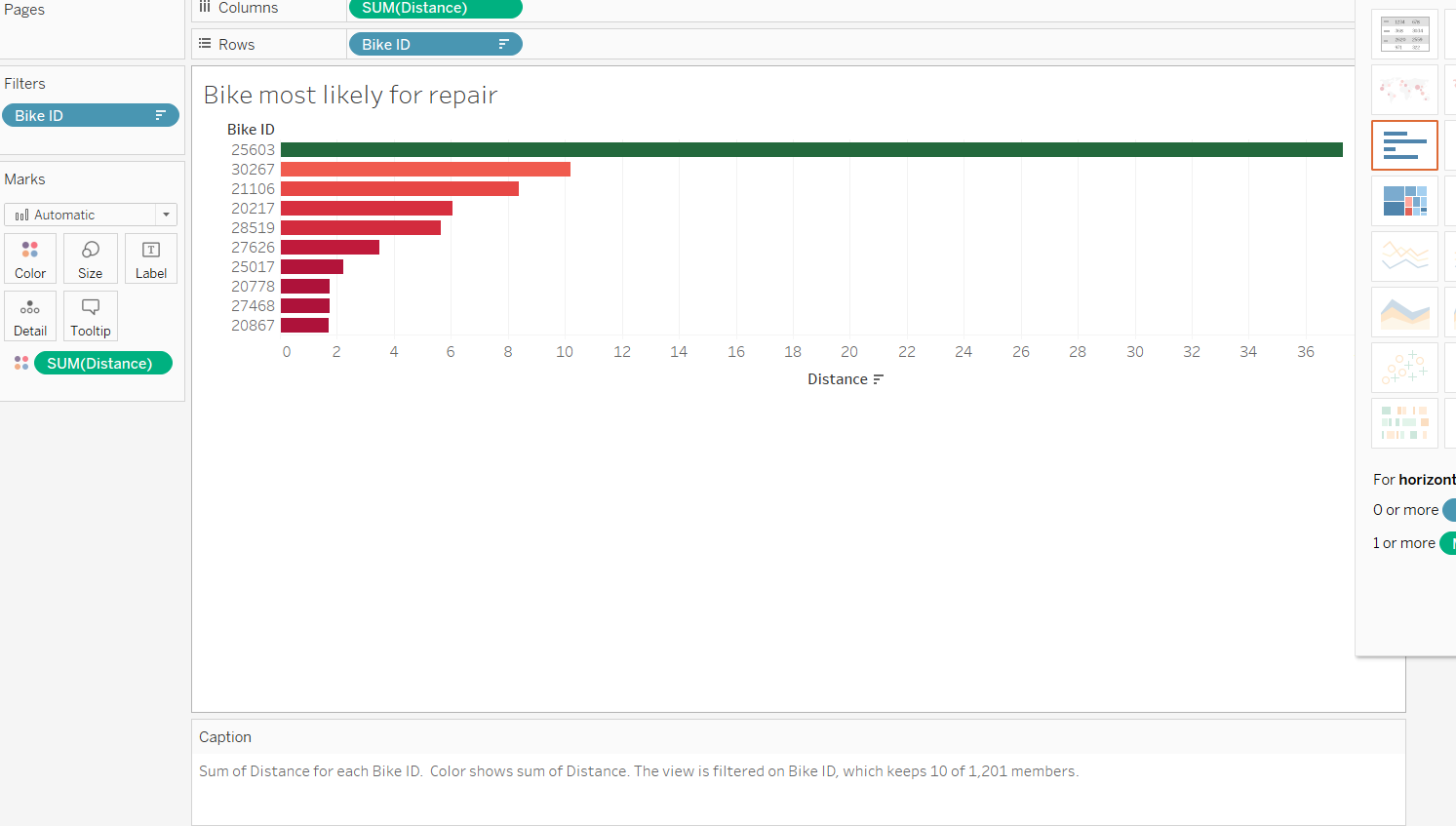
Ans. Create a new calculated field “Distance” by using formula : 7.456 \* [Trip Duration]/3600, as suggested on the Citi bike web site.





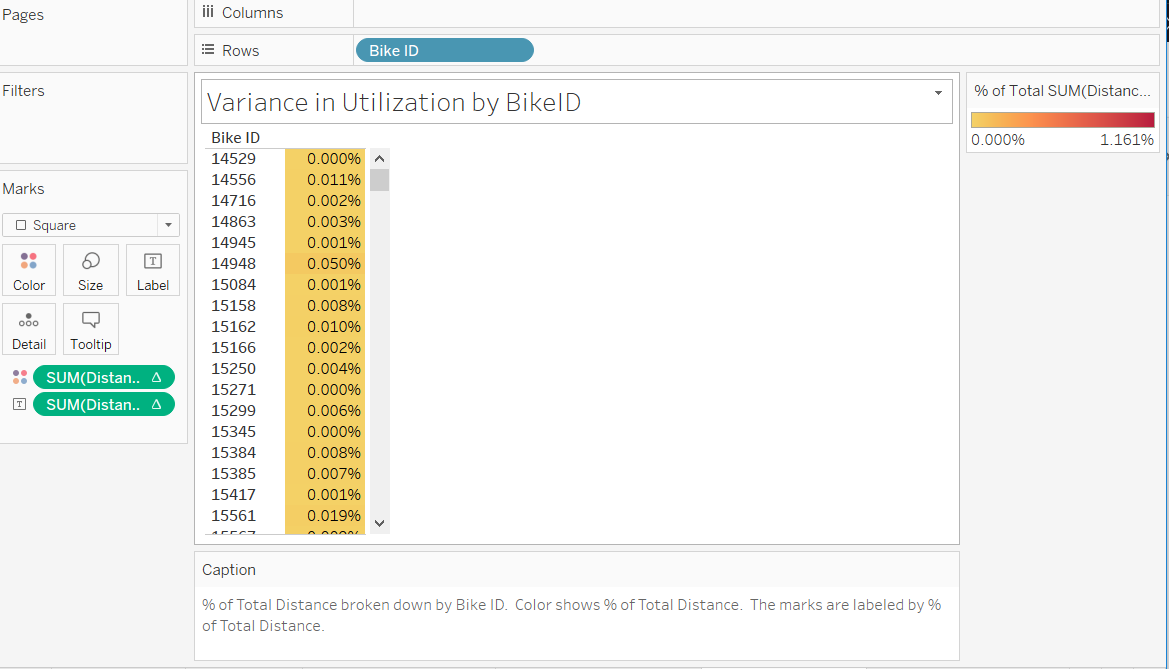
**Q14. Which bike ID is most likely for repair?**

Ans. The bike that has travelled the most distance is most likely for repair.



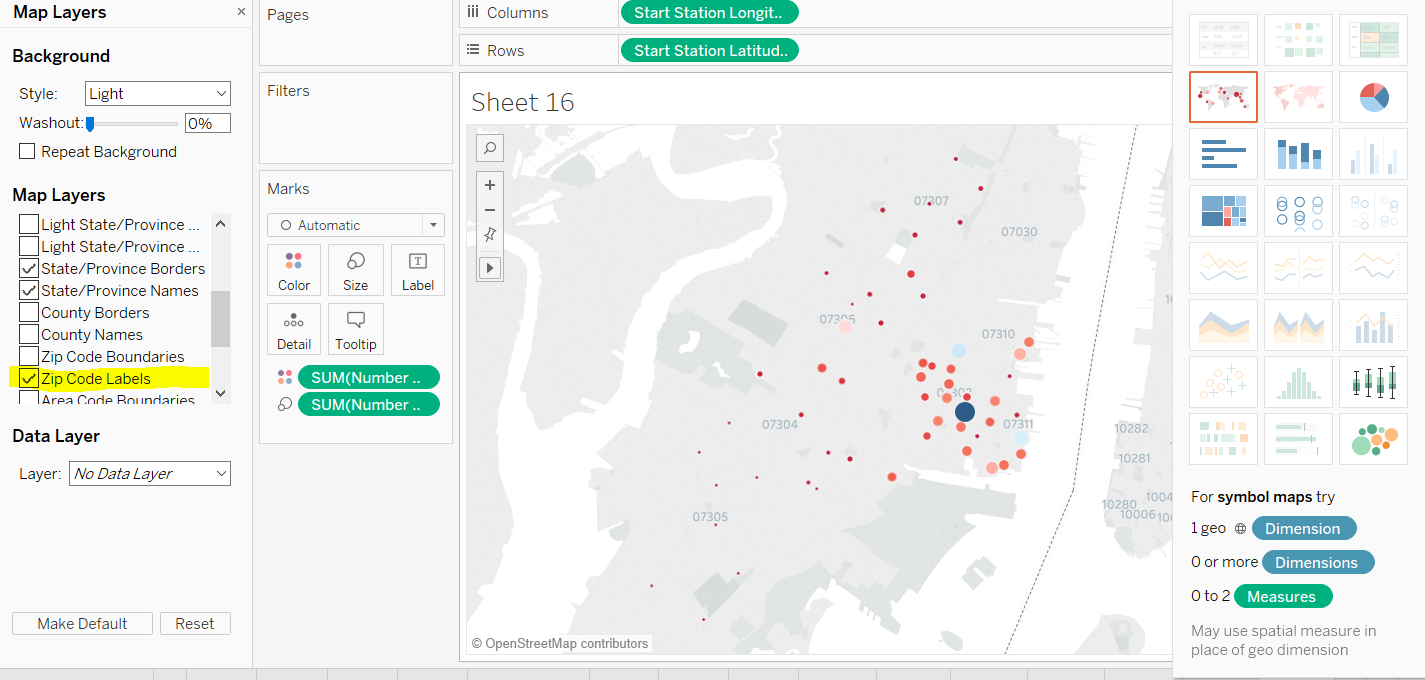
**Q15. How variable is the utilization by Bike ID?**

Ans. Add a Table calculation for the Sum of the distance. Add the bikeID as a dimension. The calculation should be the Percent of Total. The numbers depict the variance in utilization.



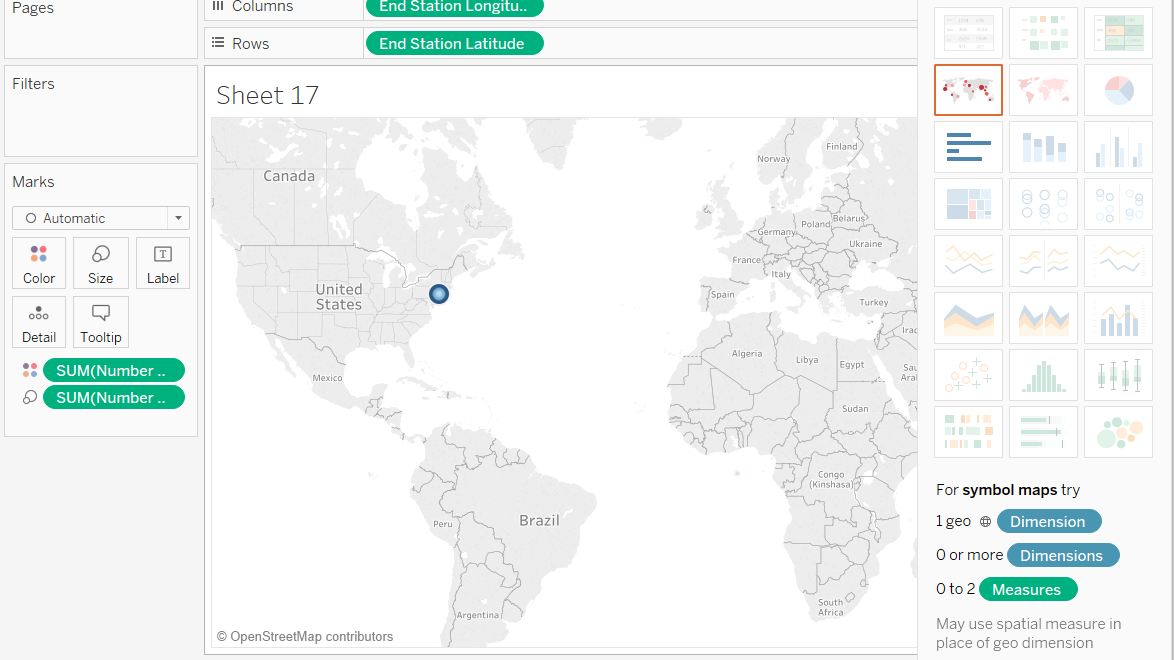
**Q16a. Static Map that plots all bike stations with a visual representation of the most popular locations to start a journey.**

Ans. Plot Start **Latitude on Rows** and Start **Longitude in Columns**. Show the number of records as size and color code. Zip Code overlay is achieved by adding a map layer and checking the Zip Code layer.



**Q16b. Static Map that plots all bike stations with a visual representation of the most popular locations to end a journey.**

Ans. Plot End **Latitude on Rows** and End **Longitude in Columns**. Show the number of records as size and color code. Zip Code overlay is achieved by adding a map layer and checking the Zip Code layer.



Q17. Dynamic map that shows how each stations popularity changes overtime?

Ans.

Q7. What are the top 10 stations for starting a journey?

Ans.

Q4. Today what are the peak hours in which bikes are used during summer months?

Ans.