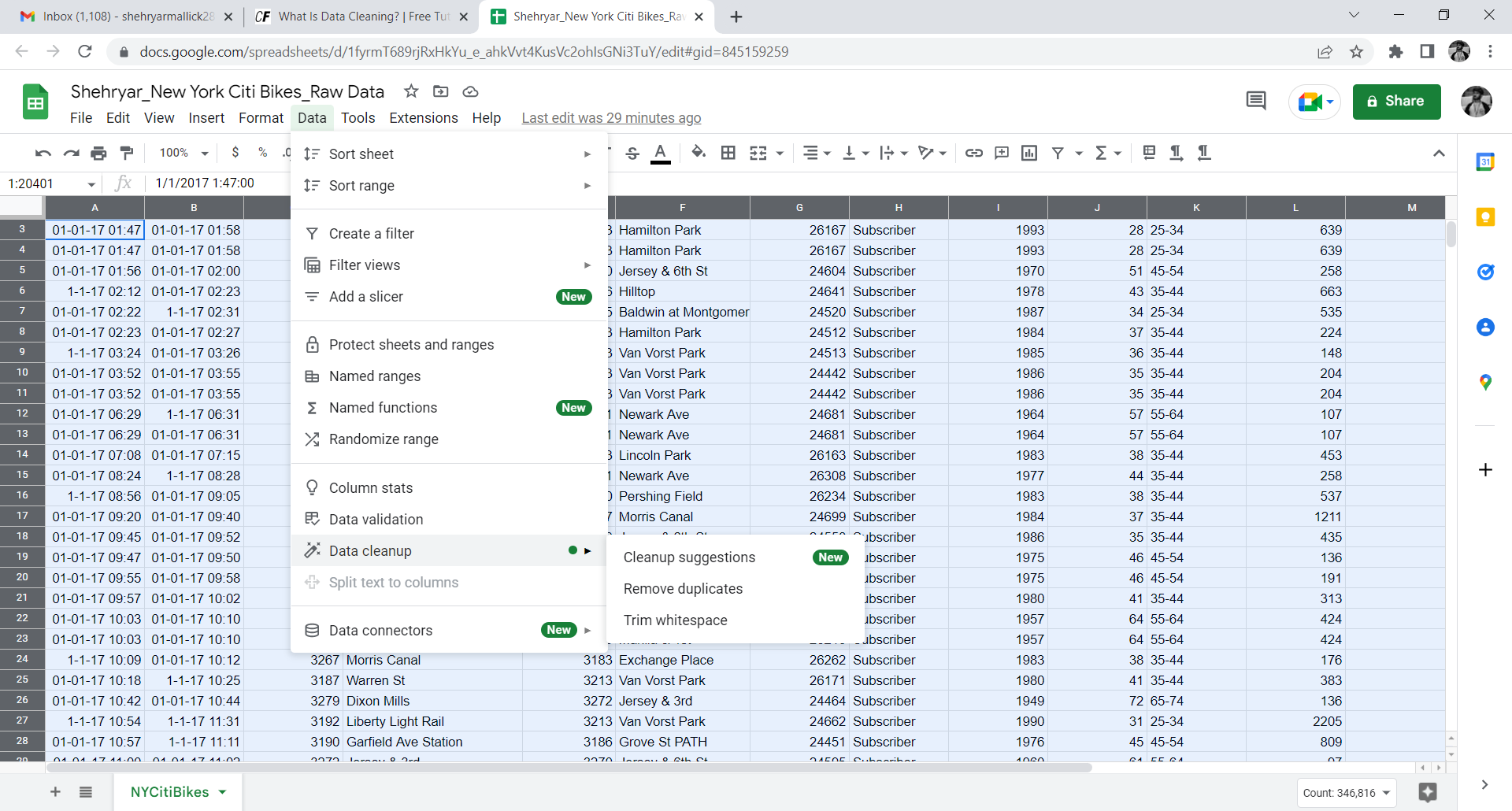
I’ve worked on the New York Citi Bikes Data to answer the following questions:

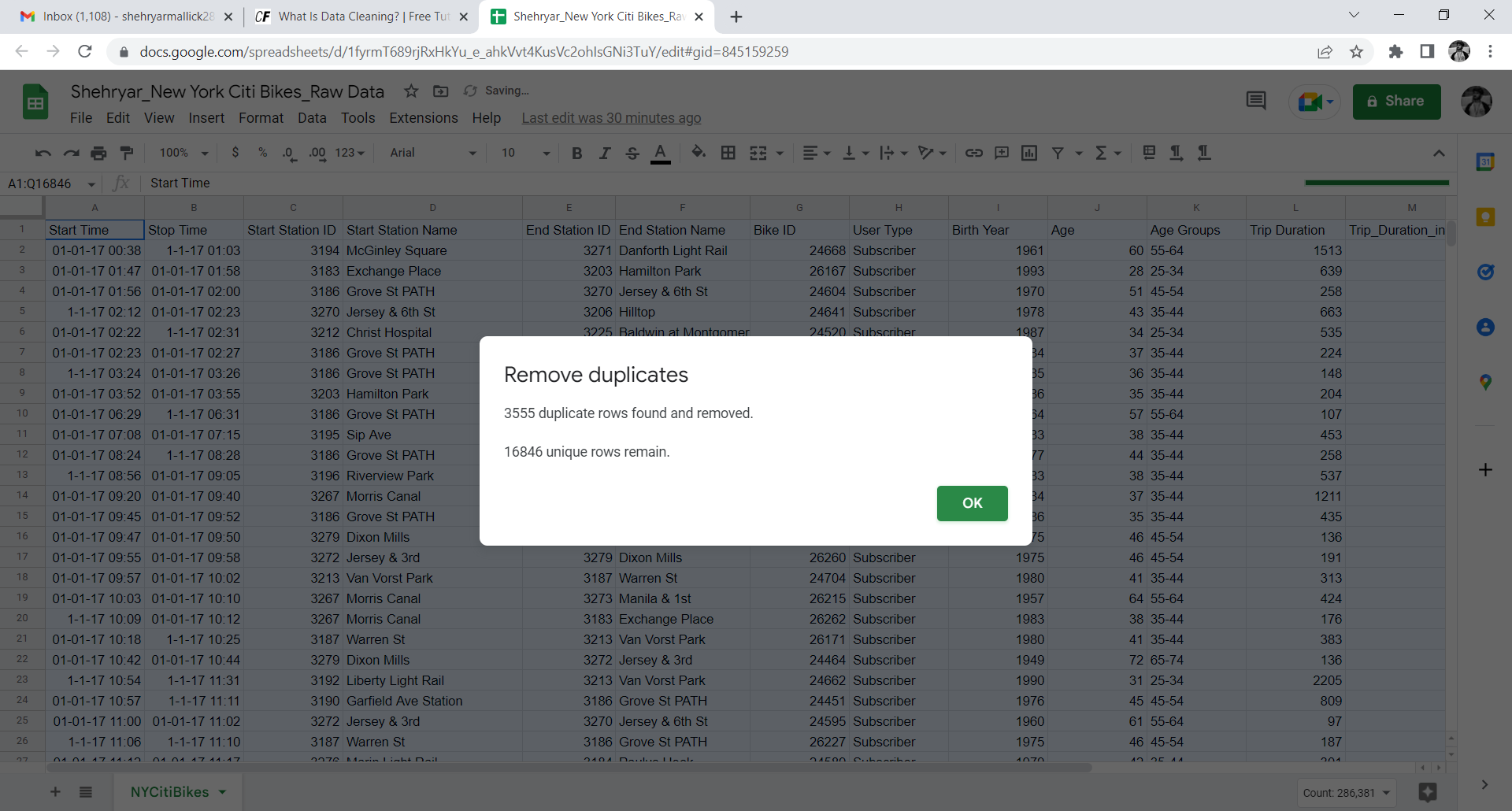
1. What are the most popular locations across the city for NY Citi Bike rental?
2. How does the average trip duration vary across different age groups, and over time?
3. Which age group rents the most bikes?
4. How does bike rental vary across the two user groups (one-time users vs long-term subscribers) on different days of the week?

But before I could dive into the data I needed to clean the data set as preprocessing is quite an important step and can affect the analysis.

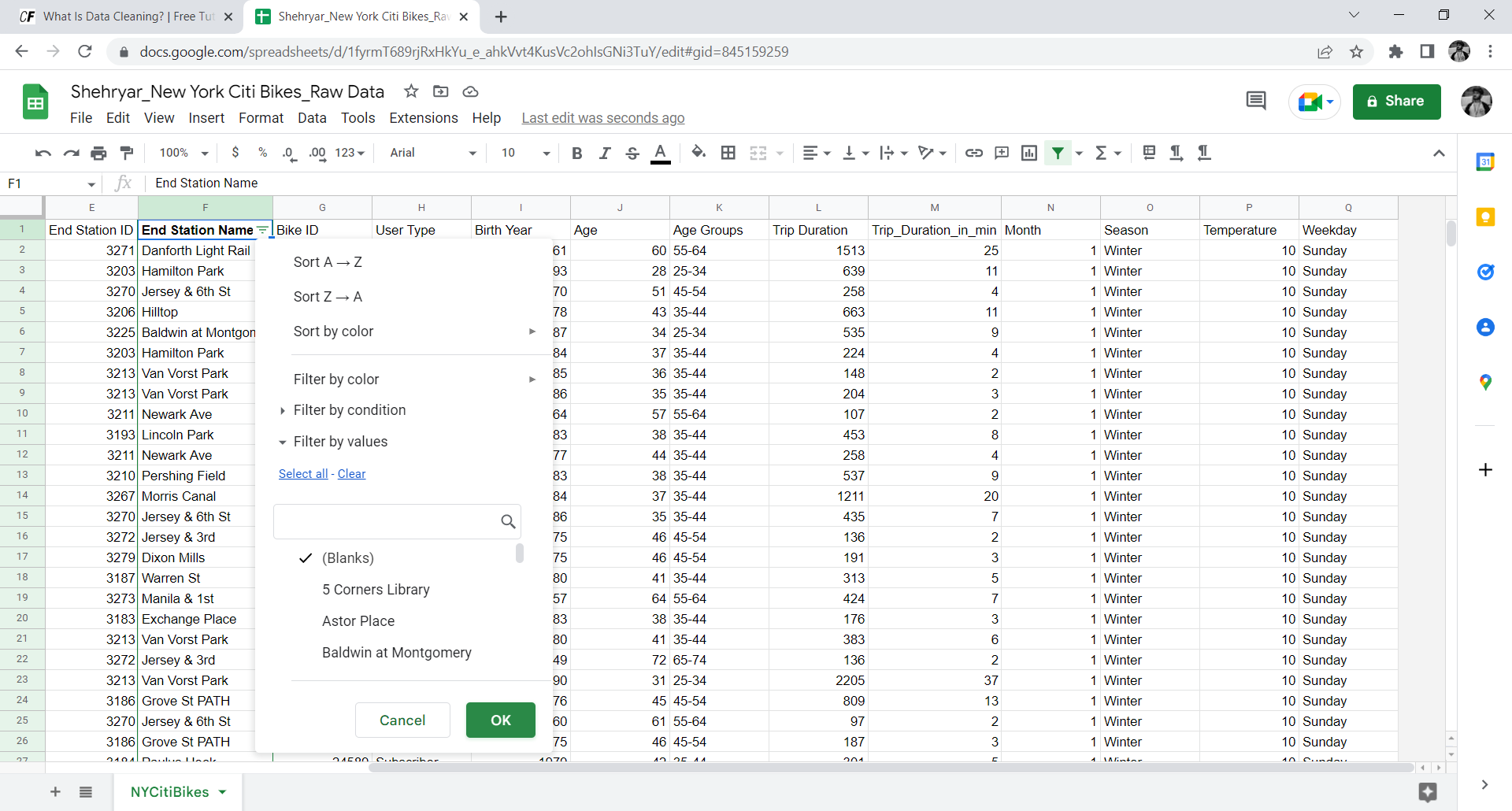
I carried out the following steps:

The first thing I did was select the whole data and removed the duplicates using the following steps.

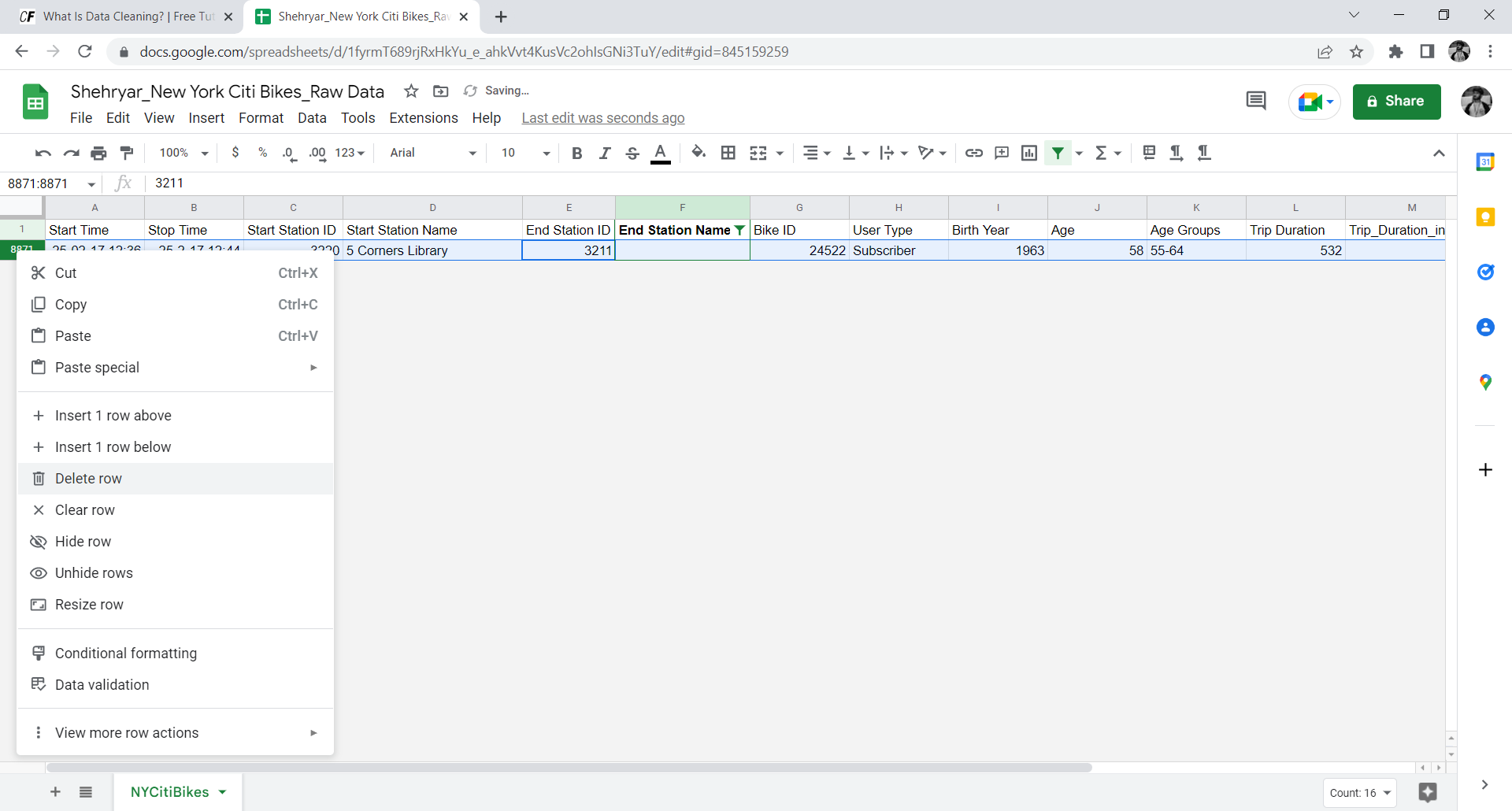




Upon removal of the duplicates the next step we carried out was I created a filter for the End station Name column and selected all the blank rows.

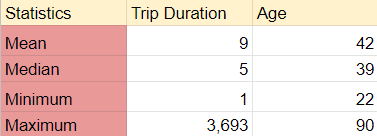


Once all the empty rows were selected I got rid of them.



Since most of the data set was cleaned, I now move onto the next step of my data analysis that is the descriptive statistics part.

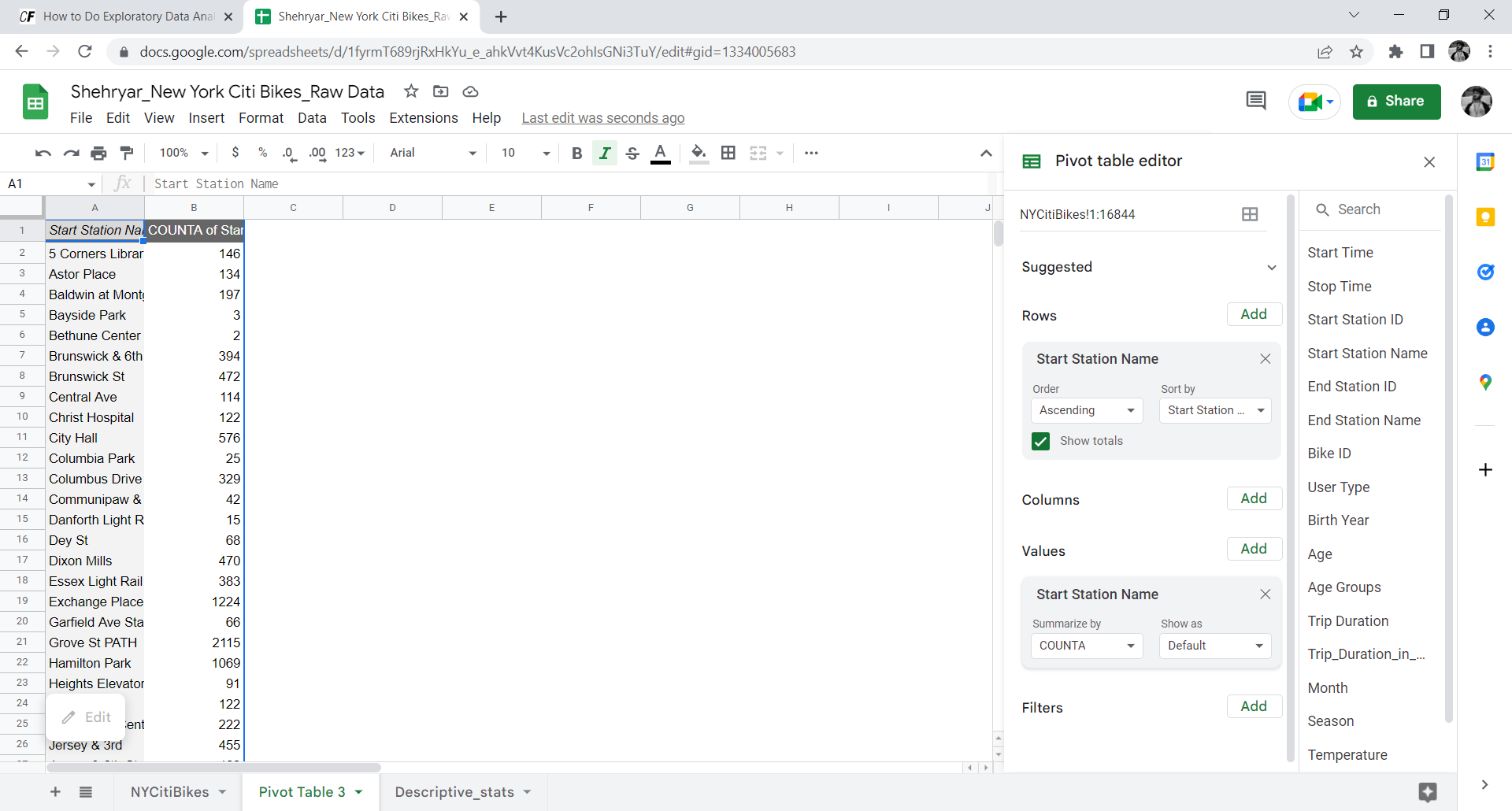
First things first let’s I got an idea of the data I was working with by creating a descriptive stats chart as shown:



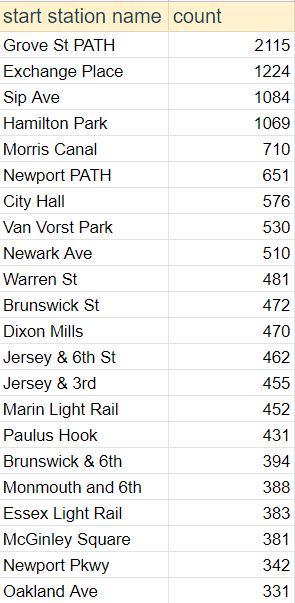
Once I got the general idea of the important columns of our data we now work on the individual questions.

The first question was: **What are the most popular locations across the city for NY Citi Bike rental?**

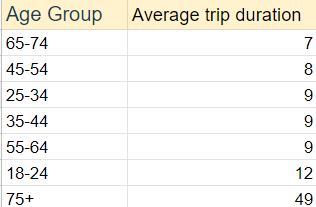
To answer this question I calculated the frequency distribution to identify the popular pickup points. I created pivot table for this purpose as shown:



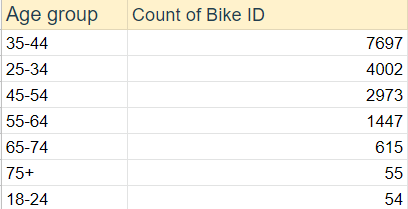
Once we create a pivot table we can basically extract the essential information we need to focus for a particular question. **(Pivot tables are created in the same manner for the other question as well).**



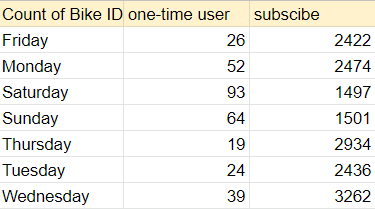
**How does the average trip duration vary across different age groups, and over time?**

****

**Which age group rents the most bikes?**

****

**How does bike rental vary across the two user groups (one-time users vs long-term subscribers) on different days of the week?**



Oof! That seemed like a lot of work but now for the fun part you must’ve heard *“A picture is worth a thousand words”* and that is quite true, even though you can pretty much evaluate how different KPIs are answered but the stakeholders aren’t analysts, they need visualizations to quickly comprehend what’s going on and this is where the visualization comes into picture, google sheets provide visualization of data which I used to make interactive visualizations of the above questions as shown.

