# Project : US Bikeshare

## Question/need:

* For the purpose of advertisement, we need to know the average age for bike’s riders and what is their gender.
* We need to know which trips are the longest trips and from which start stations to provide more bikes into those stations by taking in consideration that the bike whetherit is returned into the same station or returned on other station to minimize the lack of bikes in those stations.
* We need to know when the most common month, day and hour for all trips are,

to take the opportunity in the free traffic time for performing maintenance on the bikes.

* There are two types of users: Subscribers and customer. We need to know what is the most popular station where customer rent bikes from? To increase advertisement on those station to attract the customers for the different subscription plans.

## Data Description:

#### The data used in this project represent a bike share system provider for many major cities in the United States, to uncover bike share usage patterns. The data provided is only for New York city. The data used in this project in intended for Citibike company and advertisement company. The goal of this project is to help these company to decide for their work process based on the project result. For example, an advertisement company can determine their target customer based on the average age and gender of the bike’s rider.

#### Finally, We have multiple features in this date as the following:

|  |  |
| --- | --- |
| Feature | Example |
| **Start Time** | **(e.g., 2017-01-01 00:07:57)** |
| **End Time** | **(e.g., 2017-01-01 00:20:53)** |
| **Trip Duration** | **(in seconds - e.g., 776)** |
| **Start Station** | **(e.g., Broadway & Barry Ave)** |
| **End Station** | **(e.g., Sedgwick St & North Ave)** |
| **User Type** | **(Subscriber or Customer)** |

## Tools:

#### I will use python to tackle then questions mentioned above with the following libraries:

#### Panda: For data Processing and data cleaning. Matplotlib: For visualization, the relation between features.

**Q1:** **For the purpose of advertisement, we need to know the average age for bike’s riders and what is their gender.**

Instead of dealing with the age individually, Age has been classified into four main categories:

|  |  |
| --- | --- |
| Age | Category |
| Young Adults | 18-30 |
| Mid age Adult | 31-45 |
| old Adult | 46-70 |
| Senior Adult | More than 71 |

**Figure 2** shows that most of the riders are in the Mid-age adult category.

Chart, pie chart

Description automatically generated

**Figure 1: Age categories Percentages.**

However, **Figure 1** is not enough since it does not give us an indication of which station has a greater number of specific category of riders than other categories. Hence, we need to examine each start substation to get more knowledge about its most riders. For the purpose of saving time let’s take the first ten popular stations then we would have an idea about the other stations. After going into an intensive analysis as it can be seen from **Figure 2** for the ten most popular stations, we can notice here most of the riders fall in the categories of Mid-age Adult and old Adult. Hence, the result obtained from the above figure can be applied for all stations. However, we need to know the gender for the riders to determine the type of advertisement to be used.

Examining **Figure 3** that shows the gender percentage in each station for the top ten station. We can conclude that all the top ten popular station are most visited by male riders .As a result of that, advertisement should be focused on the riders on male with the categories of Mid-age Adult and old Adult.

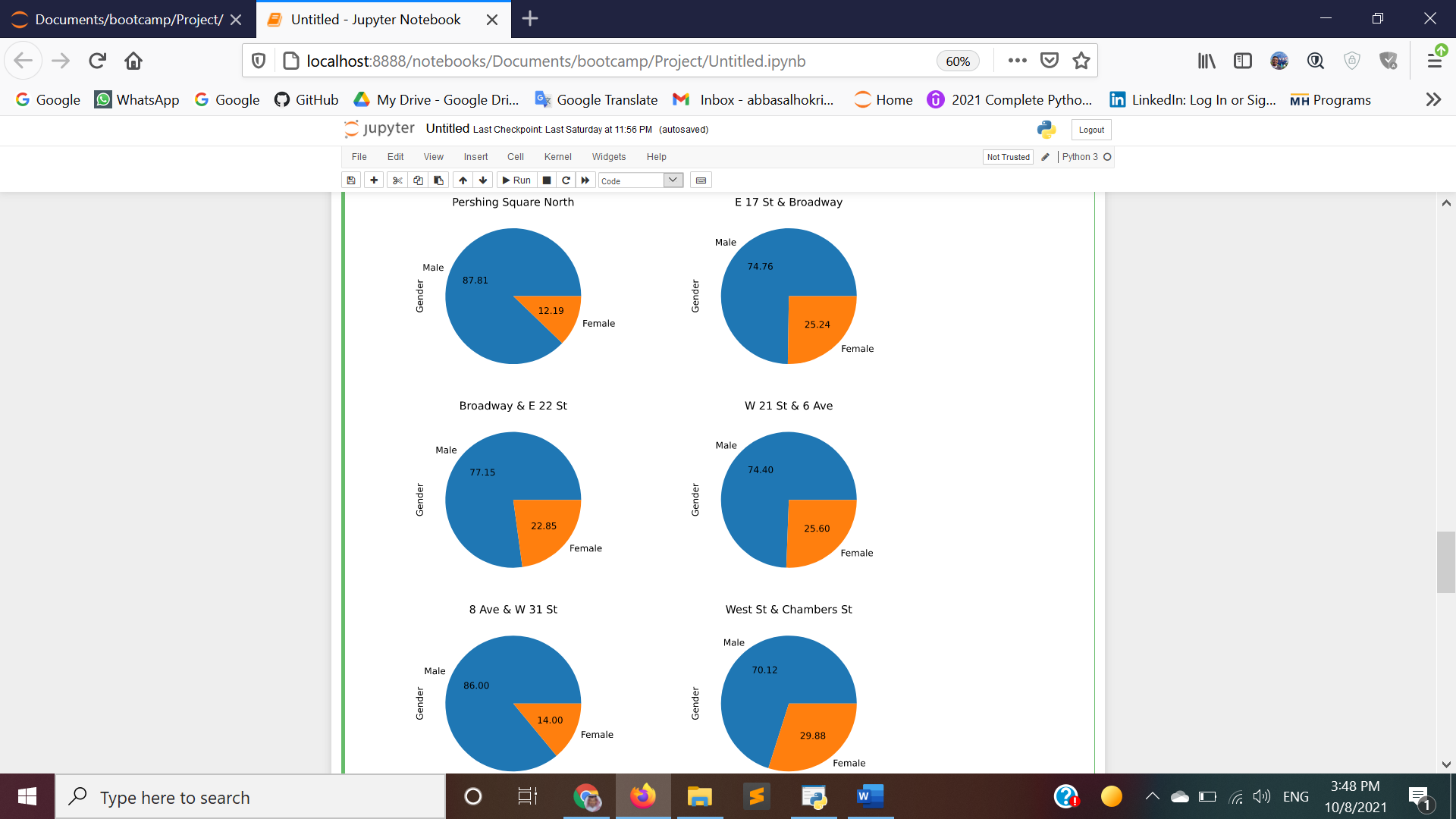
Chart, pie chart

Description automatically generatedChart, pie chart

Description automatically generatedChart, pie chart

Description automatically generated

**Figure 2: Age categories percentage in each station for the top ten station.**



A picture containing graphical user interface

Description automatically generated

**Figure 3:Age categories percentage in each station for the top ten station.**

**Q2: We need to know which trips are the longest trips and from which start stations to provide more bikes into those stations by taking in consideration that the bike whether it is returned into the same station or returned on other station to minimize the lack of bikes in those**

Having a close look at **Figure 4,** we can notice that the most poupler trips in terms of start and end stations. Hence, the bikeshare company might consider making an offer for the these stations, if the riders return the bike to the same station he/she rented from, he/she will have a special discount. By making this offer, we will make sure that there would be a redduancy to supply the most pouplar station with required number of bikes. Instead of transorming the bikes from stations by the company expense.

Graphical user interface, text, application

Description automatically generated

**Figure 4:The most popular trips in terms of start and end stations.**

**Q3: We need to know when the most common month, day and hour for all trips are, to take the opportunity in the free traffic time for performing maintenance on the bikes.**

To solve this problem, we should start by shown the most month had most trips over the others. As it can be seen from **Figure 5,** the most common month is June, and the least month is January From **Figure 5** , we can see when the temperature is low, bike rental is decrease and this happens on January , February ,and March. From this data we can decide that the overall maintenance can be conducted during January. However, we need to do maintenance in each month hence, we need to know what the least days in terms of bikes rental are.

Chart, pie chart

Description automatically generated

Figure 5: Bike rental in each month in compared to the total months. with base month = 2700

After going into an investigation, we can notice from **Figure 6**, day 31 is the least bike rental for all the six months in the data given. Day 31 has 3730 bike rentals for all stations.

Moreover, we need to know when is the least time for bike rental for all station in all days to do daily maintenance. Also, from the same data we can find out when is the most crowded time of the day to have more bikes ready for rental.

From **Figure 7** we can notice that most bike rentals happened in the afternoon and the least happened at night.

Chart, pie chart

Description automatically generated

Figure 6: Bike rental in each day in compared to the total days with base day = 2703

Chart, pie chart

Description automatically generated

Figure 7: Bike rental in each day in compared to the total days with base time = 2700