For this assignment, the 201902-fordgobike-tripdata.csv dataset is used. First of all, the data is loaded and inspected. Some wrangling and cleaning on the data is performed.

The start and station id’s are converted from a float to an int data type. The station id’s together with the longitude and latitude tells use where the station is in the area.   
These locations are later displayed in a map of the area.

The year of birth is converted to a datetime object. Several rows have null values in this field. Those rows are deleted from the dataset.

The start and end time are also converted to datetime objects. Further is the start time used to extract two new fields. Day, containing the day (Monday thru Sunday) and hour, containing the starting hour (0 thru 23)

The dataset contains a gender field. In the exploration we found three types. Male, female and Other. Also, a number of rows have null values in this field. The rows with other and null values are removed from the dataset. The gender field is converted to a categorical field type.

The data contains a user type field indicating if the user is a subscriber or a customer. This field is converted to a categorical field.

There is a social program to make it possible renting bikes for low incomes. The rides within this program are indicated in the bike share field which is also converted to a categorical field.

After all the wrangling and cleaning a test is done the check if the start station id’s and the end station id’s point to the same station. The check is performed by comparing it visually on the name. and programmatic by comparing the latitudes.

All stations, 329 in total, are put in one dataframe stations. All rental information is put into the dataframe bikes. 183215 rows.

During the wrangling several visuals are used to gather insights in the data. After the wrangling the fields are explored using several visuals.

The goal of this assignment is to find if there are differences between generations. A generation such as Millenials or Baby boomers is a group of people grouped by year of birth. We explore the duration of a ride, the ratio male to female, the ratio subscriber to customers, the bike share program. We also explore if there are differences in which stations are more often used by a certain generation.

We enhance the bike dataframe with a generation field containing the four generations found in the dataset (X,Y,Z and baby boomers). In this process we remove a number of rows to limit the set to these four generations. The actual exploration in the bikes dataframe is on 160565 rows.

In the first exploration one single field basis we find the ratio’s between males (76%) and females (24%). The ratio between subscribers (90%) and customers (10%). The ratio of the four generations is found with the generation Y being the largest one.

In this univariant exploration we find that the duration of a ride is short. 510 seconds or less for 50% of the rides. There are several longer rides up to 80.000 seconds.

The weekdays are the days most rides happen and during daytime with two peaks, one in the morning and one in the evening.

The duration of the rides does not really depend on the starting or ending station. Most of them have approximately the same durations with a few exceptions.

The difference in size of the four generations is considerable. There are almost 70% of the rides conducted by generation Y people. This against 4,5 % by baby boomers.

To compare the numbers we take the relative portion of the generation. In this way we observe that generations X,Y and baby boomers have less females while Generation Z has more females then the total ratio.

Generation Z has more customers than the overall ratio while the baby boomers have the smallest number of customers. In the same generation Z we find almost the double number of participant in the bike share program with baby boomers coming second. Interesting to see is that the baby boomers travel longer than the generation Z which have in general shorter rides.

We also investigated where the rides start and end. From this follows that The X, Y and baby boomers follow a similar pattern while the generation Z has a different pattern meaning that they start in general more from other stations then the other three generations.

We did some further exploration and can point on a map per station which generation starts or end mostly in that particular station. It is not that large part of the city are dedicated to a specific generation, but one can see that some area’s in the city have more from a certain generation. For example, in the neighborhood of the Berkley university we observe that a lot of the stations are the starting and end point for Generation X.

The X, Y and baby boomers have a similar pattern if we explore the start and end time of the rides. They all have peak in the morning and the evening. Generation Z has only a small peak in the evening. They als rides bikes all day of the week. The other three generation use the bikes more often during the week instead of the weekend.

Some other observations are that females have longer rides. Longer rides we find also in the weekend and obviously during day time the rides are longer then the other hours of the day.

Finaly we did some explorations with more fields, but always in relation to the generation. We then observe the following females have always longer rides then males in all generations. Also customers have longer rides than the subscribers, with females again longer then males.

With respect to the starting and end points we already noted that the Generation Z has a different set of stations preferred to start from. When looking to customers/subscribers from generation Z we observe this difference also however it is smaller for the customers.

We also discussed the usage of the bikes during weekdays en weekend. The same pattern holds when dividing it in males/females.

A final observations is that the different generations use all stations to rent a bike. The duration of the bike depends a little bit on which station it is rented from, which gender and which generation is renting. However one cannot conclude that some particular stations are over favorited above mentioned features.