

## **Part 3.3:**

### **Data exploration and suggestions to improve public transportation:**

The plot in figure 1 represents the total number of passengers which boarding some bus at each hour of the day. From this plot we can conclude that the rush hours are during the morning, around

7:00-9:00 and also during 14:00-16:00.

We would like to check whether the distribution of the number of buses which operating at each hour is suitable for those rush hours. To do so, we create the plot in figure 2.

The plot in figure 2 represents the total number of active buses, and the total number of passengers who boarding some bus for each hour of the day.

As we can see, there is a pretty good match between the number of people who get on the buses and the number of buses at each hour.

In other words, during hours when there is a larger amount of people who need to use the public transportation, there are more active buses. Thus, in this aspect the public transportation is managed well.

Next, we create the plot in figure 3.

The plot in figure 3 describes the following:

Each point represent coordinate of (longitude, latitude), which corresponds to a station's location in Israel's map. The color of each point represents the range of the number of passengers that boarded some bus from this station.

Green color represents passengers\_up between 0-5.

Blue color represents passengers\_up between 6-50.

Pink color represents passengers\_up between 51-200.

Purple color represents passengers\_up between 201-300.

Orange color represents passengers\_up between 301-500.

Red color represents passengers\_up>500.

We can see that there are a lot of green points, which indicate that there are a lot of stations which are used by only few passengers (at most 5).

As a result, one of our suggestions for improving the public transportation is to cancel these stations that are used by very small number of citizens in order to invest the resources invested in their maintenance and the maintenance of the buses that stopped there at other stations and other buses which operates in more busier areas or maybe for adding new bus lines and new stations in those busier areas.

Moreover, we can conclude from the plot in figure 3 that most of the red, orange, and purple points are located in the area which blocked in the following range of the longitude and latitude:

$32.0 \leq \text{latitude} \leq 32.1$  and  $34.75 \leq \text{longitude} \leq 34.85$

As we mentioned before, those points colors represent the largest number of passengers which boarding buses from these stations. Thus, new bus lines could be introduced in this area, in order to moderate the load between the different bus lines in the stations over there.

In addition, we create new table which described for each line Id the average number of passengers for trip. Base on this table we create the plot in figure 4 which described the 5 lines id with the highest average number of passengers for trip, which are colored in the plot in red, and described the 5 lines id with the lowest average number of passengers for trip, which are colored in the plot in green.

As a result, on one hand, we can conclude from this plot that the buses with line's Id: 35002 , 81001 , 24092 , 12161 , 29082 contains more than 70 passengers in average for trip, which indicates that these bus lines should have **increased** frequency.

On the other hand, we can conclude from this plot that the buses with line's Id: 10235 , 10576 , 36079 , 32075 , 10616 contains at most 10 passengers in average for trip, which indicates that these bus lines should have **reduced** frequency.

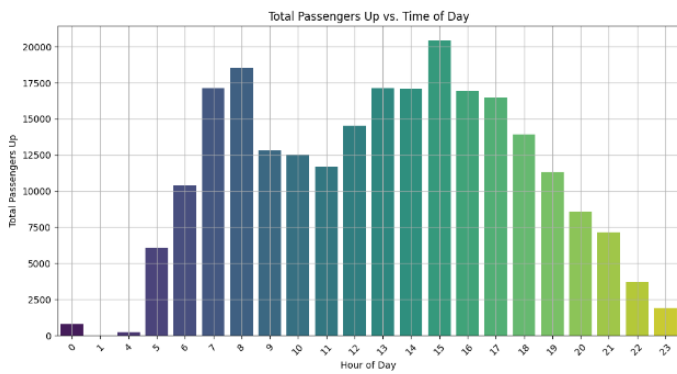


Figure 1

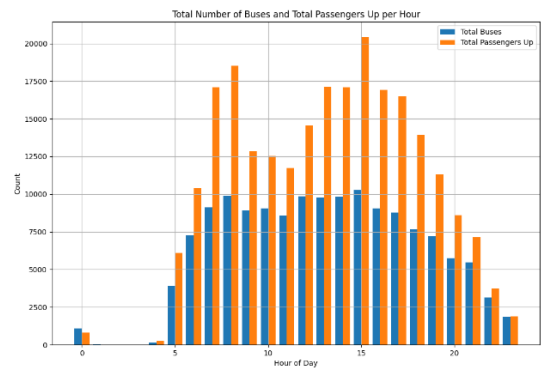


Figure 2

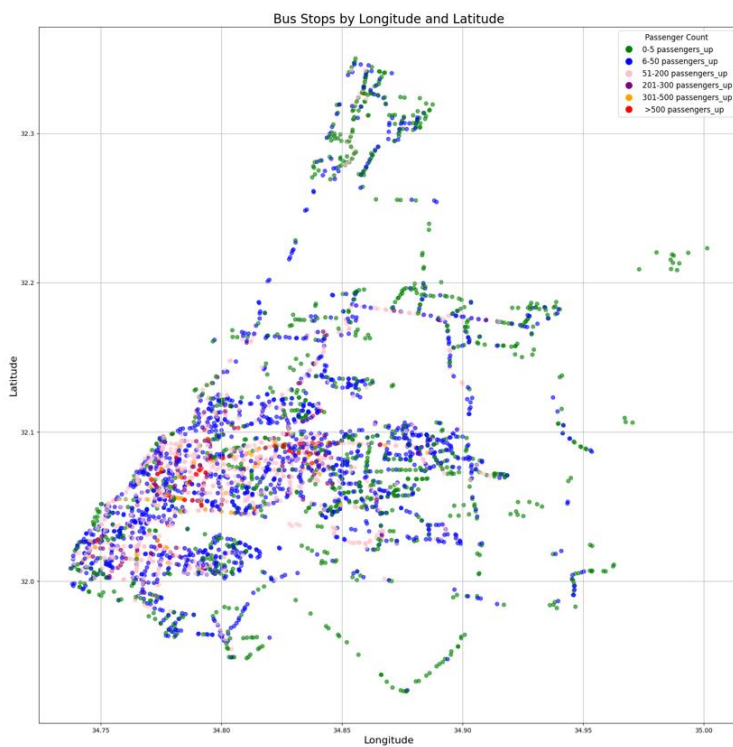


Figure 3

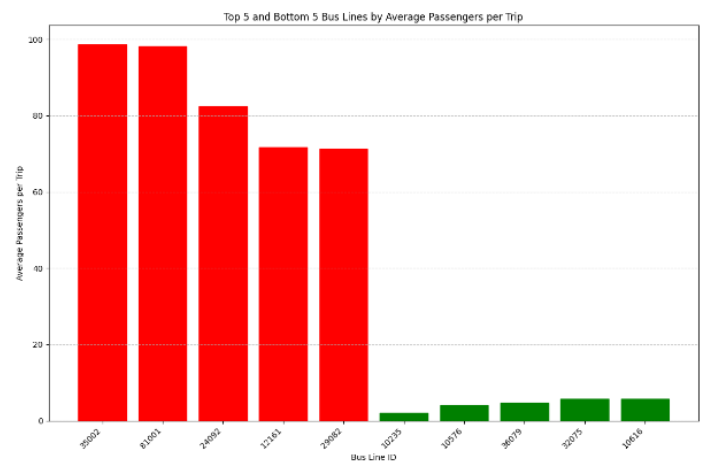


Figure 4