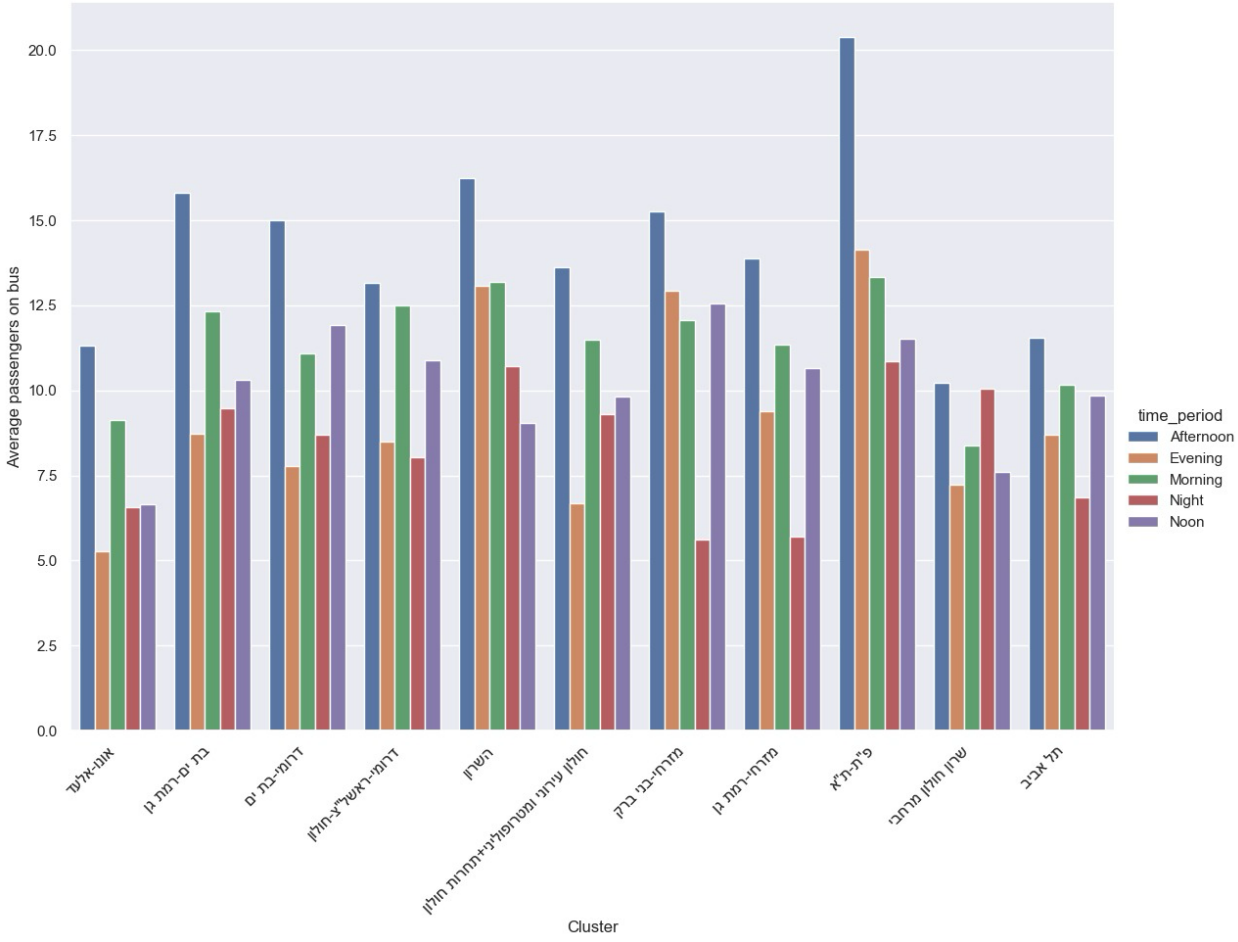
IML Hackathon – Conclusions and Suggestions

Noam Rosenmann, Shani Sova, Shahar Cohen, Tal Hollander and Tal Neumann

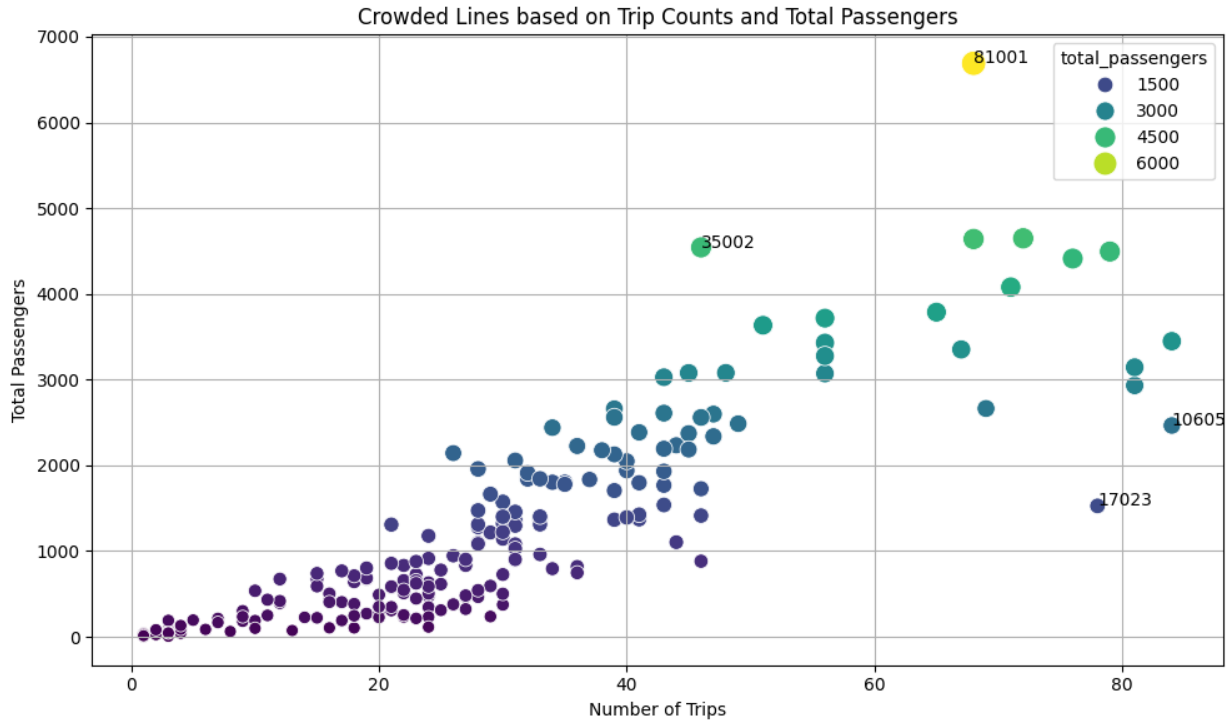
Conclusion/suggestion 1 – redirecting resources



we see that overall, there are more passengers on the bus on average in all the clusters in the afternoon. we see that some clusters have less passengers in average in all the parts of the day (אונו-אלעד, שרון חולון מרחבי). On the other hand, in בני ברק מזרחי and ,ת"א פ"ת the average amount of people is higher.

We would suggest that more buses should be added to the afternoon rides, and more in the areas with a higher average amount of passengers. For example, in lines that have on average 5 people on the bus in the evening, reduce the frequency of the line to use the buses in line that have more than 10 people on them in the same time period.

To determine what lines need higher frequency or the opposite, we calculated how crowded are the lines. We used polynomial fitting to determine what lines are outliers from the "rational" crowd. To do so, we calculated residuals (vertical distances from the line) and set the **hyper-parameter**, which is the distance from the line we allow, to be 5 times the median distance and marked the lines.



From the graph above, we suggest increasing the frequency of the 81001 and 35002 lines. Although the 81001 line's frequency is relatively high, it serves the most people in the data we were given, and therefore is quite crowded. and on the other hand – reduce the frequency of the 17023 and 10605 lines, which serve less passengers.

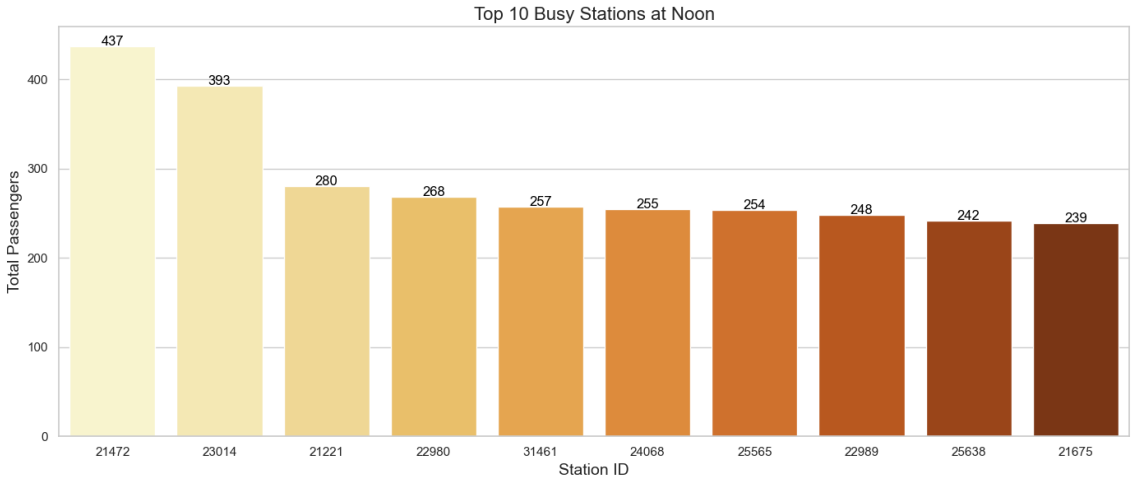
Conclusion/suggestion 2 – ensuring proper streetlights and shading

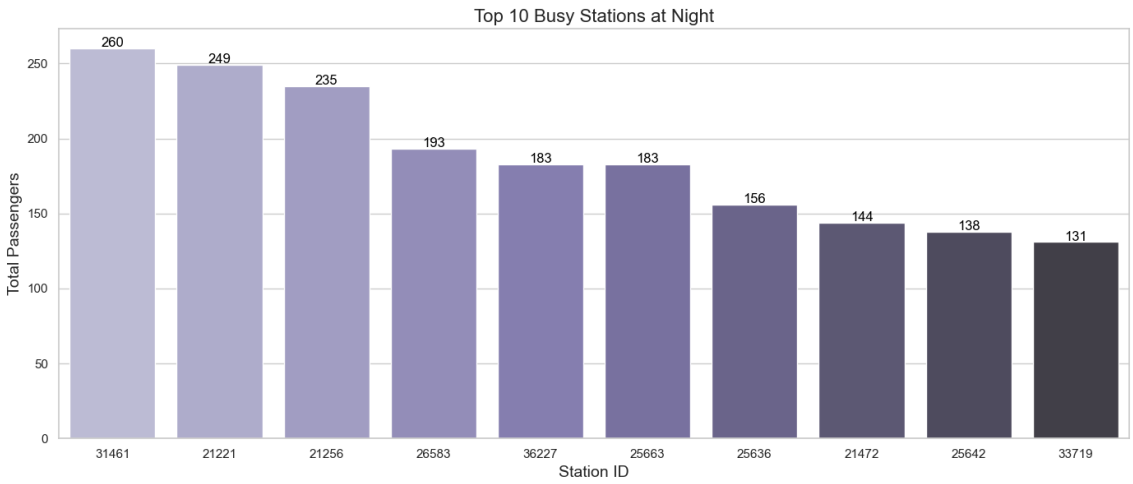
Considering the last few weeks' extreme temperatures, we were concerned about the public transport's user's well-being – as we are aware that many of them are elderly, and even young people shouldn't stay in the sun for long. Therefore, we found out what stations are usually crowded in the noon time (11:00-15:00), by identifying the amount of people boarding to a bus. At noon, when the UV radiation from the sun is the strongest, and we suggest adding additional shading to those stations.

In addition, we wanted to make sure that people using public transportation during the night will feel safe at the stations, and that the bus driver will se them clearly. In

the same way, we found out what stations are the busiest during the nighttime (23:00-7:00) and we suggest adding proper street lighting to those stations.

In the graph we plotted the top 10 of each type, but we suggest determining a threshold (e.g., x passengers) to choose stations.





Conclusion/suggestion 3 – canceling unused stations

We wanted to verify the necessity of all the stations in the database. To do so, we filtered out the stations in the database that don't have any passengers boarding or leaving the buses going through those stations.
We did it by using the passengers_up field to check if passengers board on a bus in those station and checking if the passengers_continue decreased compared to last station to check for passengers who left the bus.

We made sure to consider the sample size of the stations, and we set **hyper-parameter** of a minimum of 3 mentions of a station in the database to determine it as unnecessary.
Using this method, we found 10 stations that have no passengers boarding or leaving the bus. We suggest further inspections of those stations, and if the data provided matching the reality – removing those station and possibly save fuel and reduce carbon emissions in the areas of the stations.