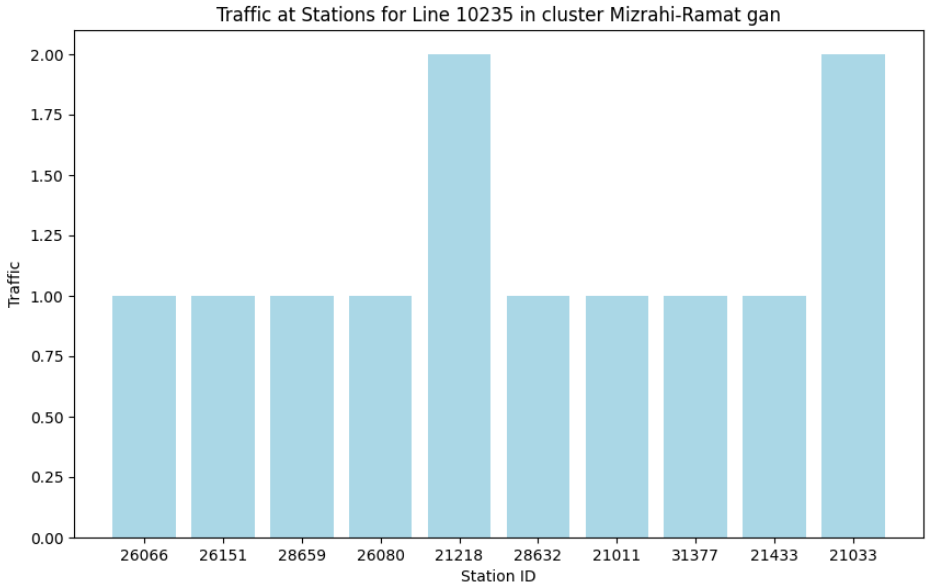
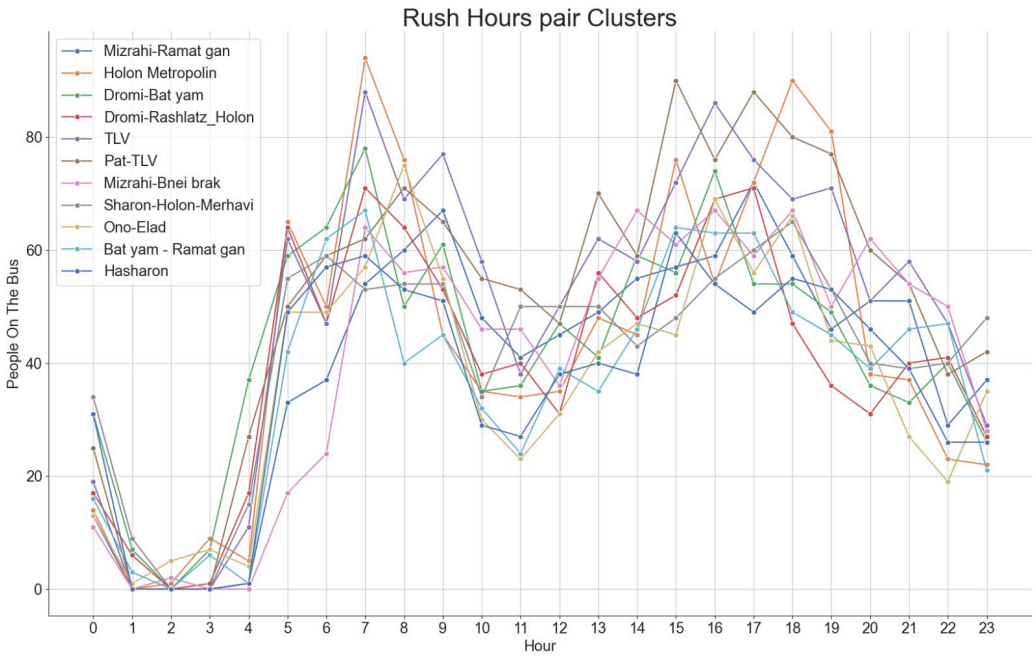
Part C:

Key conclusions:

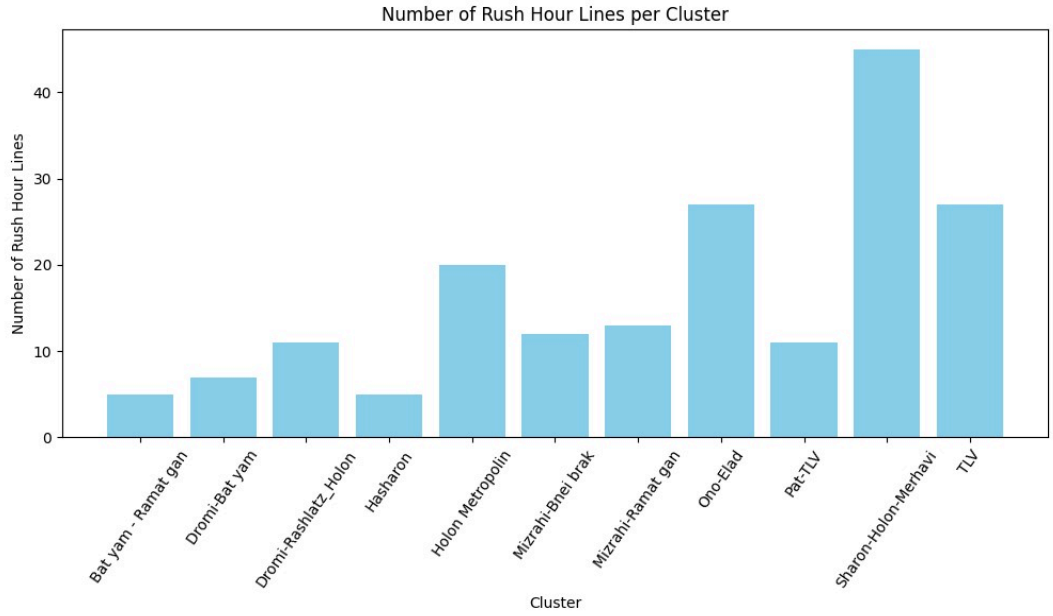
In our data exploration, we discovered a few interesting patterns.
First, for each station we calculated its traffic. This computation was done in order to analyze the busiest stations each line passes through. These stations could be a major factor in the bus' total travel time. A busy station can cause delays, eventually leading the bus to be late to subsequent stations. This calculation was done for every line in each cluster, in order to understand the relation between busy stations and delayed bus lines. The computation for station n took into consideration the number of passengers which boarded from station n-1, the number of passengers which boarded on station n, and the number of passengers which continued to station n+1. An example of this calculation for line 10235 in cluster Mizrahi-Ramat gan is shown in the graph below:



An improvement we concluded from this analysis is to add more stops before/after high traffic stations if a bus line has too many busy stations. This way, each bus will spend less time waiting for passengers to get on or off the bus, reducing delays.
Secondly, we wanted to better understand how different hours affect the bus capacity. Based on the amount of people who were on the bus, we concluded whether there was a 'rush'. If more than a certain number of passengers were on the bus at a given hour, it was labeled as a rush hour. That number served as a lower bound, which helped us classify which hours were busier than others. In addition, we plotted a graph for each cluster indicating the amount of people on the bus at any given hour of the day. As seen in the picture, the rush hours are seen as 'peaks' in the graph for each colored line, representing a cluster.



An improvement we concluded from this analysis is to increase the number of lines at rush hours in each cluster. The addition of bus lines can increase the availability and reliability of the lines, reducing the time people wait to board stations and the crowds on the bus.
Lastly, we sampled each one of the clusters in order to evaluate the varied usage between different regions of israel. Based on the rush hours we classified as explained above, we further analyzed the number of busy lines in a cluster, which supplied us with insights about the usage of public transport in a given cluster.



Our main recommendation from this analysis is to allocate more resources to the busiest clusters. This is due to the reason that many citizens in a busy cluster are using public transportation.