Problem Statement:

According to NYPD recent report the crime rate in the subway has increased by 93% compared to last year. Public safety is a concern to transit agencies and the NYPD and making an action to solve this issue is required at this critical time. Subway crimes were never this high and there are specific times and stations that have a higher crime rate than others. In order to find a solution, more investigation about those time and places is a must. Also, finding out the demographic for those who are more targeted in crimes help in developing more specific solution. Moreover, knowing the background of the most wanted subway criminals could help in tracking the root of the issue.

Data Description:

- The data that will be used are MTA database and NYPD monthly report.
- Getting data such as the total number of crimes per station and information about victims and criminals all will be collected from the monthly crimes report
- Information about the busy hours and average numbers of ridership on each of those times will be obtained from the MTA database.
- From analyzing both data sets a better idea will be formulated about when is the most crimes occur and who are the most targeted group or the most violent ones.
- The sample will be all the people who use NYC subway in a regular basis.

Tools:

- SQLITE to combine the databases and Python and its libraries including NumPy, Panda, Matplotlib to analyze and visualize the data.