**Abstract**

The goal of this project was to use and get familiar with Exploratory Data Analysis. This was achieved by using New York City Metropolitan Transportation Authority data along with two additional dataset which are Projected Population 2010-2040 – Summary and NYC Transit Data Turnstiles Station List. Most of the work that consume the time was cleaning the data especially having three datasets that all need to be compatible with each other's. After doing the cleaning, the analysis part begins where some statical have been calculated and some visualization was done. To sum up, three outcome can be use to conclude the project which were Most of the daily entries come from Manhattan, Brooklyn & Manhattan have the majority of stations and Future increase in population is Bronx & Brooklyn.

**Design**

This project originates can be seen as first import the data needed then start cleaning the data to make it easy to manipulate and visualize.

**Data**

The dataset contains more than 2 million rows for New York City Metropolitan Transportation Authority data, around 400 for NYC Transit Data Turnstiles Station List and 50 for Projected Population 2010-2040 – Summary.

**Algorithms**

For this project no specific algorithms was used, some build in libaraly and methods.

**Tools**

* Numpy and Pandas for data manipulation
* Scikit-learn for modeling
* Matplotlib and Seaborn for plotting

**Communication** slides and two-word files