**Data Analysis for Marketing**

**Abstract**

The project focuses on analyzing the MTA data to figure out the pattern of the traffic in the metropolitan area. The main purpose is to know the busiest month of the year and the amount of people moving in and out of stations. The location was chosen based the stations’ traffic.

**Design**

The project started after Fast Food Co asked for help to open their first branch. The client wants to maximize their sale by ensuring that they will cover the operation cost of their grand opening. Furthermore, the clint would like to create an offers timetable based on the amount of traffic. The less people in front of the store the more offers will be given. The rent of the location and the opening of the branch is going to take place at the beginning of the busiest month. There are some necessary things that we need to take care of for the sake of efficiency and understanding:

* Dividing the data into 12 months. This will speed up the process.
  + It is required to use the latest data.
* The busyness is calculated by the sum of ENTRIES and EXITS.

**Data**

For this project, I have used a full year of data files. To ensure efficiency, I have used A divide-and-conquer methodology. This was done by separating the data based on the months.

**Algorithms**

In order to reduce the amount of working with 12 months I have used functions efficiently. For instance:

* A function will take the numbers of the week to pull the data from the directory of this project. The function returns a concatenated DataFram of several txt files.
* A function that fixes the columns spacing.
* Multiple functions that add columns of the hourly entries, hourly exits, day of the week, the week number and so on. This will help in analyzing data easily.
* The MTA pairing had many examples in cleaning and organizing data. I have used all of them in my project.

**Tools**

* Numpy and Pandas for data manipulation
* Matplotlib and Seaborn for plotting
* Date to time
* SqlAlchemy