Abstract:

The goal of the project is to use specific dates range from Metropolitan Transportation Authority (MTA) turnstile data to determine the number of persons using the stations in summer season & winter season. I used January & February for winter season, July & August for summer season. I started by observing only one station, taking dates of winter & summer seasons in consideration, then extracting single day entries & calculating the median of 7 days entries, and ended by Inserting data into understandable plots.

Design:

This project is requested by a marketing agency called TPC, to launch a marketing campaign aims to increase public transportation popularity.

Data:

The dataset contains 3,088,413 rows & 11 columns, I used 12,845 rows & 4 columns.

The extraction of number of persons in one station is done by substitution of each entry & the entry be for it, and substitution of entries & exits.

Algorithm:

* making a mask to show specific columns & assigning it to variable.
* getting the true values of 1-day entries & exits
* returning the Median

Tools:

* NumPy and Pandas for data manipulation
* Seaborn for plotting
* PowerPoint for presentation visualizations