SDAIA Academy T5C04 Bootcamps: Data Science

EDA Module Project Proposal

MTA Turnstile data: an exploratory analysis

Date: September 23, 2021

Prepared for: Dmitry Denisov

Prepared by: Nada Rambu

Background

WomenTechWomenYes (WTWY) is a fictional organization that aims to draw attention to their upcoming summer gala. Taking advantage of New York MTA turnstile data, the stations with the highest traffic can be recognized in order to determine the optimal placement for their street teams, allowing them to reach out to a wider segment of people who are interested.

Data Description

The Metropolitan Transportation Authority (MTA) is providing a weekly data about their turnstile usage across New York city (Find the data on the following link: http://web.mta.info/developers/turnstile.html). Each file contains more than 190000 row. where each row of the dataset represents a single entry or exist through a turnstile. This data will be used to build a regression model that predicts the stations with highest ridership. Though perhaps finding additional hidden correlations that could lead to valuable conclusions. Hypothetically, the data collected throughout the previous summer seasons (i.e., June, July, and August) will be explored for the purpose uncovering any underlying ridership patterns during summer.

Tools

The following tools will be used to carry on the project:

1. The data will be stored in a PostgreSQL database, then handled via Python environment using SQLAlchemy toolkit.
2. Pandas library will be used to create data frames for easier data manipulation.
3. Matplotlib, Plotly and Seaborn libraries will be used to visualize and discuss the results of the analysis.