Tia Rashed Alzanan
Project Proposal
24 Sep 2021

Metropolitan Transportation Authority Commercials Pricing

Overview:

As many people in New York City use The Metropolitan Transportation subways as main transportation. In this project I plan to analyze MTA data to locate the busiest stations to help us determine commercials prices

Goal /Questions :

Determining the prices of advertisements in the stations according to the crowding.

What are the stations that have the most traffic?

What is the average traffic for each station during weekends, and weekdays?

Datasets:

All datasets are from MTA website (http://web.mta.info). Below is the deception of data:

C/A: control area and we have 750 control areas, the datatype is string.

UNIT: Remote unit for a station and we have 470 units, the data type will be string.

SCP: Subunit Channel Position represents an specific address for a device,the data type is string.

STATION: Represents the station name the device is located at, the number of stations is 379.

LINENAME: Represents all train lines that can be boarded at this station, the number of lines is 114. And the data type will be string.

DIVISION: Represents the Line originally the station belonged to

DATE: Represents the date (MM-DD-YY), the data type will be data_time.

TIME: Represents the time (hh:mm:ss) for a scheduled audit event, the data type will be data_time.

DESC: Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours), the data type will be string.

ENTRIES: The cumulative entry register value for a device, and the data type is Integer.

EXIST: The cumulative exit register value for a device, and the data type is Integer.

Features:

New Features will be added as following:

Traffic: Sum up the number of entries and exits to check how crowded the place is the data type will be an integer.

Exact_entries: Calculate the exact entries by subtracting the cumulative number from the previous same device. The data type will be an integer.

Exact_exits: Calculate the exact exits by subtracting the cumulative number from the previous same device. The data type will be an integer.

Tools:

To analyze the data, I will use some tools, which are as following :

- Python Language
- Panda Module
- Numpy Module
- Matplotlib
- DB browser