



Relation Between Entertainment and Work with MTA Services

Overview:

In this project, I plan to find the relation between the crowd of people who use the MTA service and their work and the entertainment. Then check who gets car rent or use public transport, the check for the three famous locations and which have the most traffic. I will study data for one year.

Goal:

Study the relation between the entertainment location and the work area with the crowd of MTA service.

Datasets:

All the Source data coming from the MTA website (<http://web.mta.info>) and I will get some information from Google Maps to explore the locations, but before that, I must clean the data, then add the column that I need while I study the data. For example, and I will add some information like traffic and day names and recalculate the entries and exits. Below is the deception of data:

Control_Area: this control area and we have 750 control areas, and the data type is string.

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

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

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

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

DIVISION: Represents the Line originally the station belonged to

BMT : Brooklyn–Manhattan Transit Corporation.

IND: Independent.

IRT: Interborough Rapid Transit.

PTH: PATH Port Authority Trans-Hudson.

RIT: Rochester Institute of Technology.

SRT: Scarborough Subway.



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:

I will add four more columns for the traffic and the exact number for entries and exits and the day names as following:

Traffic: Add the entries and the exits to check how the location was the crowd. and 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.

Day_name: Based on the date will return the name of date, and the data type is a string.

Tools:

To analyze the data, I use different tools like SQL Browser, Excel, Notepad ++, Jupyter, python language. And I use a different library from python for Example requests, urllib, DateTime, numby, panda, matplotlib, SciPy, statistics.

Conclusion:

What I expected after study the data was that the work hours for the workday are more traffic than the off-work hours. And during the weekend afternoon hours are busier than morning hours. And we will find the difference between who rent a car and who use public transportation. Also, figure out which entertainment area has more traffic.