MTA Turnstile Data Analysis

SEPTEMBER 2

SDAIA ACADMY (T5)
Written by: Sara AlAbdulsalam

أكاديمية سدايا
SDAIA Academy

Table of Contents

Introduction

Metro stations is one of the public transportations that can help with decreasing the street traffic of big cities such as New York which has the largest rapid transit system in the world by number of stations, with 472 stations in operation [1], a one train can take more than 600 cars off the road [2], it is also a better choice for the individuals since it is less expensive comparing to the expenses of cars gas, and less stressful. However, with crowdedness of metro station that can overwhelm people making them tend to use cars instead, we can reconstruct and be able to help MTA better design subway entrances, exists, and locate the dysfunctioning turnstiles to manage and distribute congestion areas using insights from public MTA Turnstile Data[3].

Data description

The data that will help us taking insights about the nature of metro stations congestions days, hours, and stations is a from MTA website that provides a series of data files containing numbers of cumulative entries and exits by stations, turnstile, with their dates and time specified. The metro data records are weekly produced and mostly collected every 4 hours.

In this project and in order to carry out the insights I will use the first half of the current year 2021 data.

Features provided in the dataset are [4]:

- C/A = Control Area (e.g., A002) which is a string
- UNIT = Remote Unit for a station (e.g., R051) which is a string
- SCP = Subunit Channel Position represents an specific address for a device (e.g., 02-00-00) which is a string
- STATION = Represents the station name the device is located at which is a string
- LINENAME = Represents all train lines that can be boarded at this station
  Normally lines are represented by one character. LINENAME 456NQR represents train server for 4, 5, 6, N, Q, and R trains. which is a string
- DIVISION = Represents the Line originally the station belonged to BMT, IRT, or IND which is a string
- DATE = Represents the date (MM-DD-YY) which is a data type
- TIME = Represents the time (hh:mm:ss) for a scheduled audit event which is a time type
- DESc = Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours)
  Audits may occur more than 4 hours due to planning, or troubleshooting activities.

Additionally, there may be a "RECOVR AUD" entry: This refers to a missed audit that was recovered. Which is a string

- ENTRIES = The cumulative entry register value for a device which is integer
- EXIST = The cumulative exit register value for a device which is integer

I will add more features which are the following:

- Turnstile_location = which is a combination of C/A + unit + SCP can be used to locate the near by places around the turnstile on google map
- entries_num = which is the number of entries for the station timestamp observed by taking the difference of the cumulative entries and the previous one.
- exits_num = which is the number of entries for the station timestamp observed by taking the difference of the cumulative exits and the previous one.
- Weekday = the weekday name to distinguish between weekdays and weekends.
- Congestion = which is the number of entries and exists added up to know how busy the station is.

Tools description

To carry out the project and explore the data, I will be using Jupyter lab to use python language. In addition to Python library which are:

Matplotlib, and Seaborn for data visualization.

Numby, and Panda for data read and write operations.

Conclusion

MTA Turnstile data analysis and exploring will give us insights about the congestion areas which can help in reconstructing the design of the MTA, this can be carried by using python data visualization and manipulation libraries. The expected results is to have crowded stations on the weekends and morning work hours therefor more managements can be utilized, also stations near to work and entertainment areas can have more turnstile because its expected to have more entrances than others.

References

[1] https://en.wikipedia.org/wiki/New_York_City_Subway

[2] https://translink.com.au/travel-with-us/benefits-of-public-transport

[3] http://web.mta.info/developers/turnstile.html

[4] http://web.mta.info/developers/resources/nyct/turnstile/ts_Field_Description.txt