

# MTA Turnstile Data Analysis

T5 Data Science Bootcamp

Written by: Asma Althakafi

23 September 2021

## Contents

<b>Introduction .....</b>	<b>3</b>
<b>Data Description .....</b>	<b>3</b>
<b>Tools .....</b>	<b>4</b>
<b>MVP Goal.....</b>	<b>4</b>
<b>References .....</b>	<b>5</b>

## Introduction

The Metropolitan Transportation Authority (MTA) is North America's largest transportation network. The MTA network comprises the nation's largest bus fleet and more subway and commuter rail cars than all other U.S. transit systems combined [1].

Café chain wants to open new branches in New York. I will analyze the MTA Turnstile Data to see which stations have a lot of people. This analysis will help the Café chain to determine the places where it will open new branches.

## Data Description

The data I will use is MTA Turnstile Data will obtain from <http://web.mta.info/developers/turnstile.html>.

The data was collected from May 2010 to September 2021. I will take a sample of the data, which is the latest three months: August, July, June.

The features of the data are [2]:

- C/A: Control Area.
- UNIT: Remote Unit for a station.
- SCP: Subunit Channel Position represents a specific address for a device.
- STATION: Represents the station name the device is located at.
- LINENAME: Represents all train lines that can be boarded at this station.
- DIVISION: Represents the Line originally the station belonged to.
- DATE: Represents the date in (MM-DD-YY) format.
- TIME: Represents the time for a scheduled audit event in (hh:mm:ss) format.
- DESc: Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours).
- ENTRIES: The cumulative entry register value for a device.
- EXIST: The cumulative exit register value for a device.

I will add more features which are:

- Day: the name of the day.
- Entries\_num: number of people entered.
- Exist\_num: number of people exited.

## Tools

The tools I will use are SQLite to store data in a database. Jupyter Notebook to use Python language to clean, analyze and visualize the data.

## MVP Goal

The MVP will consist of three things:

- Database that contains the MTA turnstile data.
- Report that contains analysis and visualization of the MTA turnstile data.
- Presentation to present the analysis results.

## References

[1] [Online]. Available: <https://new.mta.info/about-us>.

[2] [Online]. Available:  
[http://web.mta.info/developers/resources/nyct/turnstile/ts\\_Field\\_Description.txt](http://web.mta.info/developers/resources/nyct/turnstile/ts_Field_Description.txt).