Exploratory Data Analysis (EDA) Proposal for Outfront on the MTA turnstiles

Leena AlQasem

Leenabdulh@gmail.com

Banan Alhethlool

banan.alhethlool@gmail.com

- **Introduction:**

The first project for Data science Bootcamp T5 is called the Exploratory Data Analysis (EDA) for the MTA dataset turnstile. Below we will shed light on the company that we cooperate to assist them with their issue, dataset description, and finally, the tools used in this project.

- **Background:**

- **Company information:** Outfront is one of the largest media and outdoor advertising companies. It helps people, places, and businesses grow stronger by providing an integrated and target platform. With that, it also owns and operates outdoor advertising displays.

- **Problem statement:** Outdoor commercial advertising is necessary for brands and international companies. It helps grab customers' attention through displaying ads. Therefore, knowing rush hours, idle times, and whether Covid-19 has an effect or not is essential to Outfront.

- **Value for the company:** Due to the nature of the subway audience, ads are seen by millions of riders multiple times a day. Therefore, it is a cost-effective option for both Outfront and businesses planning to work with them.

- **Dataset:**

The Metropolitan Transportation Authority is North America's largest transportation network, serving 15.3 million people across a 5,000-square-mile travel area surrounding New York City through Long Island, southeastern New York State, and Connecticut. MTA dataset is published publicly with weekly updated data.

- **Scope:** We will compare the dataset from 23 Feb to 23 Mar in 2019, from 22 Feb to 21 Mar in 2020 and from 27 Feb to 27 Mar 2021.

- **Tools:**

- **Technologies:** Jupyter Notebook, Python, SQL and SQLlite.

- **Libraries:** Pandas, NumPy, Matplotlib and Seaborn.