Name: Haneen Alhomoud
Email: Haneen.77@hotmail.com

MTA Project Proposal

This is the first project for the T5 Data Science Boot Camp, where we perform Exploratory Data Analysis (EDA) on the MTA turnstiles data

1- Background:

- **Company info:** Blood donation center provides busses, booths and centers designated for blood donations.
- **Problem/opportunity:** The lack of precise geographical informatics to allocate the centers busses and booths properly in order to reach the highest number of donators.
- **Value for the company:** the ability to receive more donations, and reduce the operational fees.

2- MTA Turnstile Dataset:

2.1 About MTA:

The Metropolitan Transportation Authority is North America's largest transportation network, serving a population of 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.

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. The MTA's operating agencies are MTA New York City Transit, MTA Bus, Long Island Rail Road, Metro-North Railroad, and MTA Bridges and Tunnels.

2.3 Data Description:

Data obtained from http://web.mta.info/developers/turnstile.html.

2.4 Columns description:

Field Name	Description
C/A	Control Area (A002)
UNIT	Remote Unit for a station (R051)
SCP	Subunit Channel Position represents an specific address for a device (02-00-00)
DATE	Represents the date (MM-DD-YY)
TIME	Represents the time (hh:mm:ss) for a scheduled audit event
DESC	Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours)
ENTRIES	The comulative entry register value for a device
EXITS	The cumulative exit register value for a device

2.5 Scope:

- The sample size is set to be 3 months (*Saturday, September 01, 2018 - Saturday, September 01, 2018),* the sample is collected in normal occasions.

3- Tools:

- **Technologies:** SQL, SQLlite, Python, Jupyter Notebook
- **Libraries:** Numby, Pandas, Matplotlib, Seaborn