# Project Proposal for Turnstile Data

صورة تحتوي على نص

تم إنشاء الوصف تلقائياً

### **Introduction:**

With the use of The New York subway MTA turnstile that show of entries, linename, unit and exits by station, turnstile, date and time. Data files are produced weekly. Data records collected every 4 hours per day. We will be able to select the most crowded stations and fill them with vending machines for games nearby schools, books machines nearby business areas, snacks and drinks in the most station has long waiting times.

**Question/need:**

Our goal of analysis this data is to find the most crowded stations to fill them with vending machines for games, books, flowers, snacks and drinks depending on where the station is located near to schools, colleges or business areas, customizing machines to fill their time with something useful and fun . According to that we will find the stations that had the highest and lowest traffic in weekdays and weekends, wither the traffic in Peak Time high or low?, how many hours people spend in each station to provide many number of machine in that station. Economic companies will benefits form exploring the data.

**Data Description:**

### We will use data about the patterns of transit traffic in New York City: [MTA turnstile data](http://web.mta.info/developers/turnstile.html), collect and clean our data set and then we will perform EDA to better understand our data:

|  |  |
| --- | --- |
| Station | Represents the station name the device is located at |
| date | Represents the date |
| time | Represents the time for a scheduled audit event |
| entries | The cumulative entry register value for a device |
| exits | The cumulative exit register value for a device |
| station\_code | C/A + unit, locating a station |

The unit of analysis is the group. After modeling and analysis data economic companies will be able to benefit from crowded in each station and helping people to spend their waiting time in a fun and useful day.

**Tools:**

**The tools that we will use is exploratory data analysis in pandas ,** **visualization libraries**(such as matplotlib and seaborn) and Numpy. And till now we are not planning to use any additional tools

Done by:

-Alnirah AL Qahtani. -Ghadah AL Harbi