

Introduction

Dunkin donuts started a breakfast menu camping, they are looking to expand more branches in New York Metropolitan to help them reach crowded location.

Question

To make Dunkin donuts camping succeed the company raised questions to be answered by the data scientist of the company

Who is my target audience?

What is the preferred time for my target audience?

What is the preferred stations for my target audience?

Data description

The data set will be used for the analysis is MTA Turnstile to determine what is the most used stations during the breakfast period the next table shows the feature of the data

| Name | Type |
|-----------|--------|
| C/A | Text |
| UNIT | Text |
| SCP | Adress |
| Station | Text |
| Line name | Text |
| Divisions | Text |
| TIME | Time |
| DATE | Date |
| DESC | Text |
| EXSIT | Number |
| ENTRIES | Number |

Data have been collected over 3 months with over 2 million observations and the feature that are expected to be analyzed are number of users that enter and exit the station and the time

Tools:

The tools that will be used are SQL to maintain needed data and SQLAlchemy to move data to python and pandas for data manipulation and matplotlib for visualization

MVP :

The MVP for the project will contain the most crowded station during the breakfast time