

Project proposal

Background

Biscuit company , My goal of this project is promoting a new product for biscuit company that produced new biscuit and must to be real tested from people so we can from MTA dataset know the daily traffic will accrue in which time in the day. and which day of the week, and my strategy to distribute free biscuit when they enter the station so people can know more about my product .my target is the most busiest stations Such as ,top 5 crowded stations in (NYK),united states .

Question:

- How long Does advertising take to work ?
- what are the five most active stations?
- what are the most active days of the week?
- what is the best way to promote a product with less time ?

Data Description:

MTA dataset

From this dataset of subway of New York we can now number of entries and exits by station

The dataset from 3 months ago has a lot of rows and 11 columns

Columns are

C/A: control area (A002)

UNIT: Remote Unit for a station (R051)

SCP: Subunit Channel Position represents a specific address for a device
(002-00-00)

Station: Represents the station name the device is located at

Line Name: Represents the line originally the station belonged to BMT, IRT or IND

Date: Represents the Date(MM-DD-YY)

Time: Represents the time (hh:mm:sss) for a scheduled audit event>

DESC: represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours)

Entries : the cumulative entry register value for a device

Exits: The cumulative exit register value for a device

The dataset is available on the following website :

<http://web.mta.info/developers/turnstile.html>

Tools:

Data analysis with the following:

- Python
- Pandas
- NumPy
- SQLAlchemy
- Matplotlib
- seaborn