Projcrt proposel:
MTA Data Analysis

samah alrefaei

MTA Data Analysis

introduction: As data science student at SDAIA My first idea project exploring the MTA'S Turnstile data my goal in this project is to encourage riders to use New York subways in peak times to use this data calculate peak times for several stations , then make an advertisement for a specific coffee shop Use MTA to calculate peak times from 5 top stations and make Advertisement for a specific coffee shop.

Data description:

First: I will study the MTA data and the following is the data 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-
STATION	Represents the station name the device is located at
LINENAME station	Represents all train lines that can be boarded at this station Normally lines are represented by one character. LINENAME 456NQR repersents train server for 4, 5, 6, N, Q, and R trains.
DIVISION	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:ss) for a scheduled audit event
DESC	Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours) 1. Audits may occur more that 4 hours due to planning, or troubleshooting activities. 2. Additionally, there may be a "RECOVR AUD" entry: This refers to a missed audit that was recovered.
ENTRIES	The comulative entry register value for a device

solution:

Using MTA Turnstile and New York City census data, and analyse to find best placement for street teams.
Finding most busiest subway stations, most crowded day and best time slot .

Tools:

In this analysis data I use more tools like: python , pandas, matplotlib,numby ,jupyter.

concolusion:

In this project I need use MTA data for boost ridership; and increase revenue. These changes will help ensure that the MTA remains the strongest.