Name: Faisal Mirza faisal.a.mirza11@gmail.com GitHub: faisalmirza11

Project Proposal Template

Background:

My friend Yasser is a movie writer that works at a film production company called Universal Pictures, he told me that he came up with a horror movie idea that fully takes place in the New York City subway, but in order to film a movie in the subway, the production company needs to know where and when are the best places and times that they can film a movie at that had the least amount of people. I told my friend that the subway transit traffic data is provided online by the MTA, and that me being a data scientist who also happens to like horror movies, I got excited and I offered that I could help them in finding the best stations and times that had the least amount of people, so that they could film a movie there. So, my friend Yasser told the production company about me, and they agreed for me to help them.

Question/need:

Framing question: What is the best time and location for having the least amount of people in the subway for filming a movie?

Who benefits: film production companies seeking to film in the subway. Example: Universal Pictures.

Data Description:

The dataset is provided by the New York Metropolitan Transportation Authority and it contains information collected from the turnstile devices in the stations, some of this data include:

- Control Area, Unit, SCP: which all represent an individual turnstile.
- Station: represent the station name where the turnstile is located at.
- Entree, Exit values: which shows the number of entries, exits of the station cumulative.
- Date, Time: which represents the date and time of a snapshot of the turnstile info.

Scope:

- The data used will cover a three-month period where it is preferably not a summer break to avoid late outgoing tourists, residents.
- The dataset used will roughly have 700k rows, and it will have 12 columns.

Name: Faisal Mirza faisal.a.mirza11@gmail.com GitHub: faisalmirza11

Tools:

Technologies: SQL, sqlite3, Python, Jupiter notebook.

Libraries: Numpy, Pandas, Matplotlib, Seaborn.