Course: Big Data Management and Analysis

Group Members: Victor Fung, Shenghua You, Simon Ling

MTA Turnstile Analysis

Motivation and Objective

The MTA in New York has always been one of the primary modes of transportation for tourist and residents alike, either for residents of New York to get to work, home or travel around New York as tourists do.

The goal of the project is to analyze the data gathered by the MTA for their turnstiles, which would allow us to analyze any trends that occur throughout different seasons of the year, such as an influx at certain places during specific seasons of the year. We can also analyze trends such as how weather and price hikes can affect the riders of the MTA subway system.

Data Source

Data on MTA turnstile from May 5th, 2010 to April 1st, 2017

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

Weather Data for each day from May 5th, 2010 to April 1st, 2017

Landmark data for New York City

Methodology

To solve most popular places:

1. We will get the station that have most exits and sorted by seasons, then convert this station to coordinate.
2. Use the coordinate to search the popular places around this area

Big Data Challenges

As the turnstile data is not updated on a daily basis, but rather a weekly basis, we will need to find and match other data sources, such as weather to the exact day MTA updated the turnstile data. Afterwards, we will create a script to gather the data from the urls on the MTA website then

clean the data. The next problem would be for specific questions, such as finding local landmarks around specific stations, as we will need to map within a radius of each station, the possible landmarks around it.

Deliverables

A report will be provided which will contain the analysis of the data created to answer our project questions, along with any visual representations needed to help show the correlation between different elements in our proposed questions to help answer the project questions.