Abstract

The objective of this exploratory data analysis is to identify zip codes within New York City that have relatively low public Wi-Fi service with respect to commuter traffic as proxied by NYC subway commuter volumes. I leveraged publicly available data sets to identify traffic volumes by zip code, and cross-referenced high- and low-volume zip codes with existing hotspot densities to determine which zones have relatively lower available bandwidth, assuming all commuters would use Wi-Fi if available. My conclusion from this analysis identifies 18 zip codes of interest that could serve as a starting point for commercial testing of a subscription-based public Wi-Fi service.

Design

The background of this project assumes a hypothetical scenario in which a for-profit telecom company ('ABC Telecom') has conducted market research that determines there is market demand for enhanced Wi-Fi availability throughout New York. The conclusions of this analysis could serve as a starting point for market testing subscription-based Wi-Fi services that would open up new revenue streams for ABC Telecom.

Data

The raw (pre-cleaned) data set was constructed by joining 3 distinct data tables to generate 7,761,000 distinct data points of 13 data fields. The three standalone data tables were:

- NYC MTA turnstile data obtained for free from the Metropolitan Transportation Authority (link);
- NYC Wi-Fi hotspot locations obtained for free via NYC OpenData (link); and
- a proprietary list of NYC subway stations, their line names, and associated zip codes created manually by me.

Key data fields in the raw data set included turnstile name, station, station zip code, date and time of data entry, cumulative turnstile entries and exits, and number of public Wi-Fi hotspots associated with the respective zip code of the station/turnstile. The primary data drivers of conclusion for this exploratory analysis are total volume by zip code and the associated hotspot density (number of hotspots relative to commuter traffic) of the respective zip code.

Algorithms/tools

- SQLite for data aggregation
- SQLAlchemy for querying data into my Python environment
- Python Pandas
- Python Matplotlib
- Microsoft Excel for illustrative tables and minor visualizations

Communication

The findings of this exploratory analysis are principally communicated in the presentation associated with this document.