Predicting crime rates using taxi rides in NYC

Carlos Petricioli   Varsha Muralidharan   Valerie Angulo

New York University
{cpa253,vm1370,vaa238}@nyu.edu

12/15/2017

Big Data Analytics Symposium - Fall 2017

Analytics Project

Predicting crime rates using taxi rides in NYC

Team

- Carlos Petricioli (cpa253)
- Varsha Muralidharan (vm1370)
- Valerie Angulo (vaa238)

Abstract

- We study crime rates and its relationship on how people use of taxis in New York City.
- Our hypothesis is that people are less likely to walk in areas subjectively deemed more dangerous and will instead opt to use more reliable and immediate transportation such as designated taxis.
- WE FOUND THAT _____.

Predicting crime rates using taxi rides in NYC

Motivation

Typical user of this application

The scientific community, the citizens and demographics themselves.

Insight that we'll get

The objective is to get a better sense of the way that crime relates to the use of taxis in NYC.

Goodness

cleaning data

we made sure that −

model testing

we made sure that −

Data Sources

Taxi rides data form TLC (*Link*)

- It covers years from 2009 to June 2017.
- The yellow taxi trip records include:
  - pick-up and drop-off dates/times,
  - pick-up and drop-off locations,
  - trip distance,
  - itemized fares,
  - rate types,
  - payment type,
  - passenger counts.

**NYPD Complaint Data (*Link 1, Link 2*)**

- This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) from 2006 to year to date data.

NOAA Weather stations data *(Link)*

The *Integrated Surface Database (ISD)* consists of global hourly ansynoptic observations compiled from numerous sources into a single common ASCII format and common data model.

- ISD's complete history of hour-by-hour readings for one user-specified weather stations
- We selected:
  - Central Park
  - JFK
  - Laguardia

Design Diagram
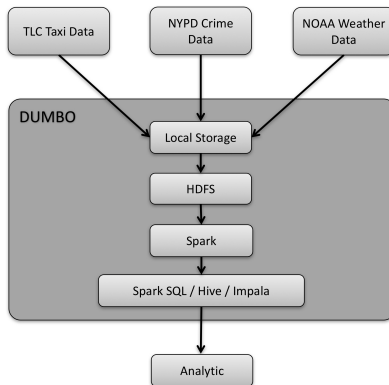
Platform:

- NYU HPC cluster (Dumbo)

Results

1. Result 1
2. Result 2
3. Result 3

Obstacles

1. Obstacle 1
2. Obstacle 2

Summary

Brief wrap-up!

Acknowledgements

References

📄 J. Bendler, T. Brandt, S. Wagner, and D. Neumann.

Investigating crime-to-twitter relationships in urban environments - facilitating a virtual neighborhood watch.

In M. Avital, J. M. Leimeister, and U. Schultze, editors, *ECIS*, 2014.

📄 Federal Bureau of Investigation, Uniform Crime Reporting.

Offenses Known to Law Enforcement by State by City.

2016.

📄 M. Traunmueller, G. Quattrone, and L. Capra.

Mining Mobile Phone Data to Investigate Urban Crime Theories at Scale, pages 396–411.

Springer International Publishing, Cham, 2014.

H. Wang, D. Kifer, C. Graif, and Z. Li.

Crime rate inference with big data.

In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 635–644, New York, NY, USA, 2016. ACM.

Thank you!