

Analysis of Parking Violations in New York City Metro Area

Team Members

Justin Nichols: jnich56@lsu.edu

Bryce Lee: blee57@lsu.edu

Dataset

The dataset contains parking violation data around the NYC metro area. We plan to use information regarding the issue date, registration state, violation code, plate type, vehicle make, and vehicle body type.

Link to data: <https://www.kaggle.com/new-york-city/nyc-parking-tickets>

The dataset contains 4 separate files.

1. Parking_Violations_Issued_-_Fiscal_Year_2014__August_2013__June_2014_.csv (1.87 GB)
2. Parking_Violations_Issued_-_Fiscal_Year_2015.csv (2.86 GB)
3. Parking_Violations_Issued_-_Fiscal_Year_2016.csv (2.15 GB)
4. Parking_Violations_Issued_-_Fiscal_Year_2017.csv (2.09 GB)

Big Data Frameworks

1. Apache Hadoop MapReduce
2. Apache Hue
3. Cloudera VM pseudo-distributed mode

Application Description

The goal of the big data application is to find correlations between which types of vehicles (plate type, vehicle make, and vehicle body type) received the costliest violations depending on the season of the year (spring, summer, fall, winter). We want to analyze if, for example, sports cars or convertibles receive more frequent and/or costlier violations in summer compared to traditional sedans or SUVs. Conversely, in the winter, would more traditional cars be ticketed at a higher rate than sports cars or convertibles since people are less likely to drive cars more suitable in warmer months than in colder ones. The real-world value associated with our analysis is knowing which vehicle types receive the costliest violations and are more likely to be ticketed when driving in and around a city. This could be useful for people traveling to a city as well as for local city governments to know the types of cars that more often violate parking laws.