CORE
Provided by UB ScholarWorks
Metadata, citation and similar papers at core.ac.uk

NYC Taxi Trip and Fare Data Analytics using BigData

Umang Patel, Anil Chandan

Department of Computer Science and Engineering, University of Bridgeport, Bridgeport, CT 06604, USA

Abstract

Traditionally the data captured from the NYC Taxi & Limousine commission was physically analysed by various analyst to find the superlative practice to follow and derives the output from it which would eventually aids the people who commute via taxis. Later during early 2000 the taxi services where exponentially developed and the data capture by NYC was in GB's, which was very difficult to analyse manually. To overcome these hitches BigData was under the limelight to analyse such a colossal dataset. There were around 180 million taxi ride in city of New York in 2014. BigData can effortlessly analyse the thousands of GB within a fractions seconds and expedite the process. This data can be analysed for several purposes like avoiding traffics, lower rate where services are not functioning more frequency than a cab on crown location and many more. This information can be used by numerous authorities and industries for their own purpose. Government official can use this data to deliver supplementary public transport service. The company like Uber can use this data for their own taxi service

Introduction

In present day transportation dataset contains large quantity of information than the preceding data. For specimen, from the year 2010 TLC using the Global Positioning System(GPS) data type for every taxi trip, including the time and location (latitude and longitude) of the pickup and drop-off. In this a complete traffic data which contains nearly 180 million rows of data in the year 2014. Due to huge amount of data, this data is example of "BigData." Using BigData it's easy to develop procedures to clean and process the data so it used for analyse the raw data into useful way in transportation service.

The core objective of this is to analyse the factors for demand for taxis, to find the most pickups, drop-offs of public based on their location, time of most traffic and how to overcome the needs of the public. In Fig 1. will provide you information about the most pickup location in New York City.

Results

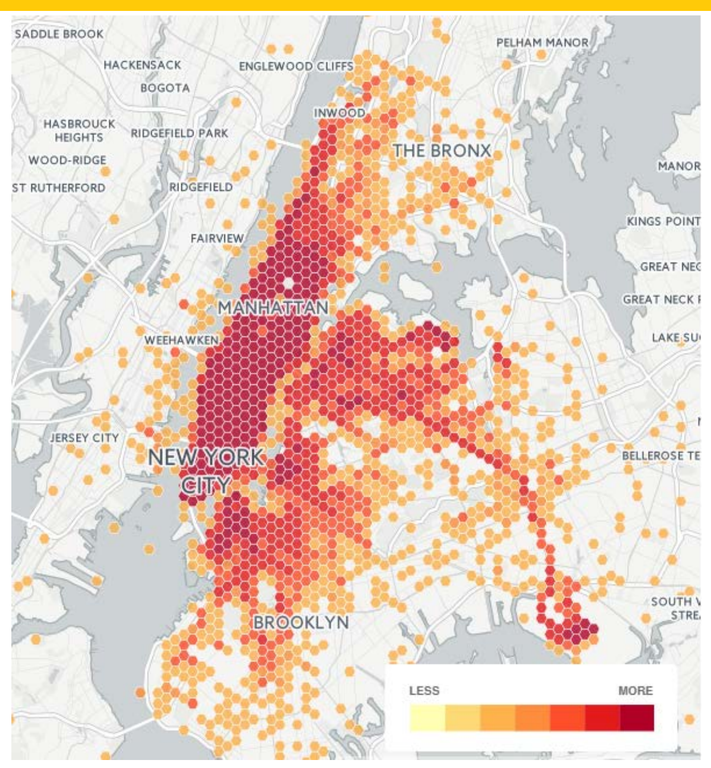

Fig. 1 Most Pickup Location in New York

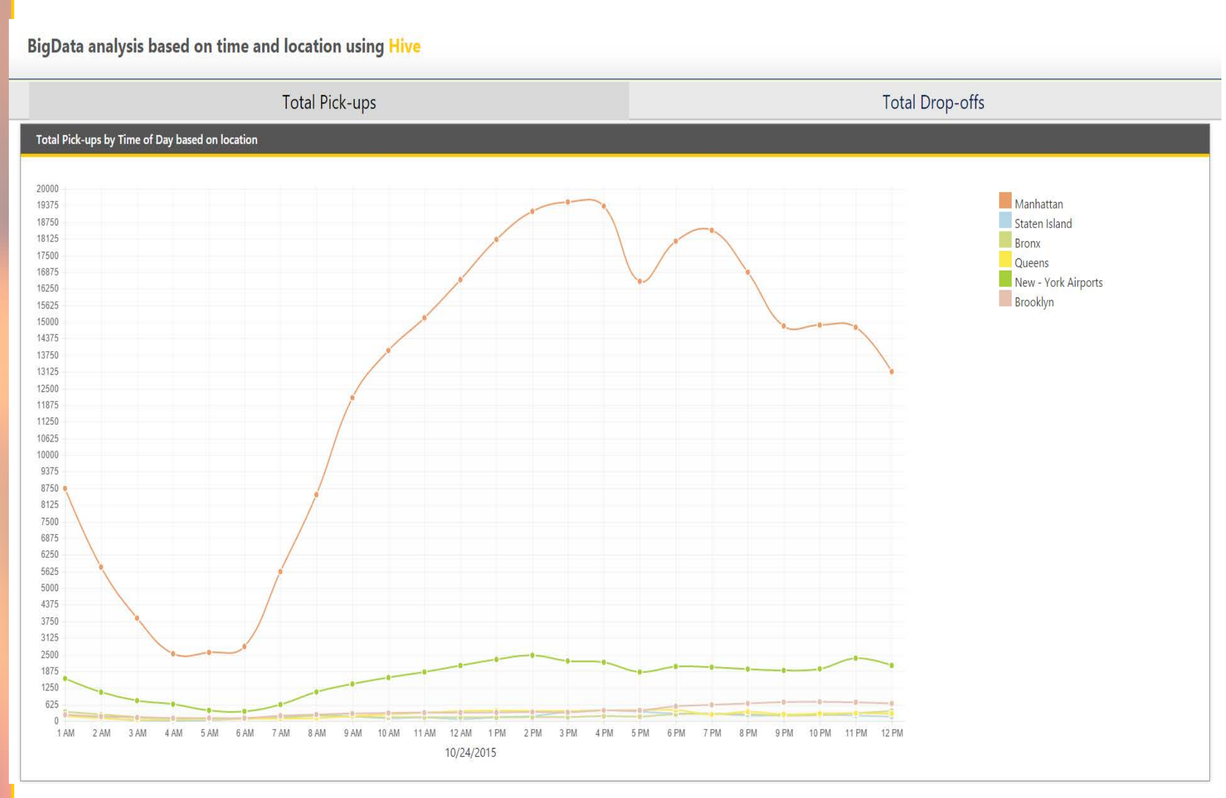

Fig. 2. Average Pick-ups by Time of Day based on location

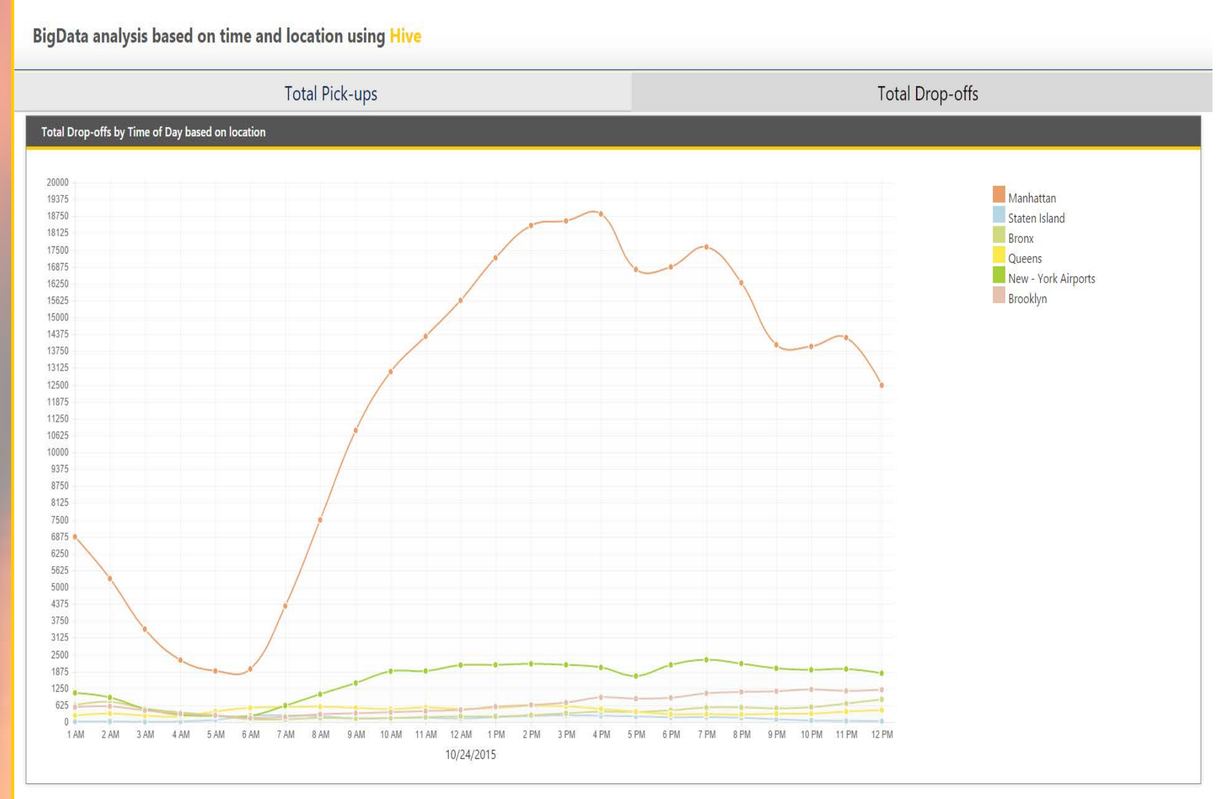

Fig. 3. Average Total Drop-offs by Time of Day based on location

Conclusion

We conclude the usefulness of these project of trip provides planners, engineers, and decision makers with information about how people use the transportation system. In this case, by identifying the factors that drive taxi demand, forecasts can be made about how this demand can be expected to grow and change as neighbourhoods evolve. As decisions are made regarding the regulation of the taxi industry, the provision of transit service, and urban development, these models are useful for forming a complete and holistic vision of how travel patterns and use of modes can be expected to respond.

Problem Definition

- Analysis on Individual
  - Driver with most distance travelled.
  - Driver with most fare collected.
  - Driver with most time travelled.
  - Driver with most efficiency based on distance and time
- Analysis on Region
  - Most pick up location.
  - Most drop off location.
- Analysis based on time and location
  - Average Total Pick-ups and Drop-offs by Time of Day based on location
  - This section will through some light on how we determine some complex analysis by using Hive In this analysis we will determine average of total pickup and drop-offs by time in a day based on location
  - Fig 2 and Fig 3 shows the output of the total pick-ups and drop-offs by every hours of a day based on location
- Analysis based on Fare
  - Average Driver Fare Revenue per hour (Gross and Net).

Performance Evaluation

Problem Definition	Time In Seconds	
	MapReduce	Hive
Driver with most distance travelled	20	15
Driver with most fare collected	15	12
Driver with most time travelled	22	19
Driver with most efficiency based on distance and time	45	35
Most pick up location	10	8
Most drop off location	9	8

Reference: [1] NYC Taxi & Limousine Commission. http://www.nyc.gov/html/tlc/html/about/about.shtml,
[2] Taxicab fact book http://www.nyc.gov/html/tlc/downloads/pdf/2014_taxicab_fact_book.pdf