

Big Data Analytics

Uber Taxi Trip and Fare
Data Analytics using BigData

Submitted by:

Rohit J (4nm17cs148)

Sahil Bhanastarkar (4nm17cs156)

Table of contents

1. Abstract
2. Introduction
3. Tools used
4. Expected results
5. References

ABSTRACT

As there is an amassed evolution in the metropolitan zones, urban data are apprehended and have become certainly manageable for first-hand prospects for data-driven analysis which can be recycled for improvement of people who live in urban areas. This particular project highlights, the prevailing focus on the dataset of Uber taxi trips and fare. Later during early 2000 the taxi services where exponentially developed and the data captured by Uber was in GB's, which was very difficult to analyze manually. To overcome these hitches BigData was under the limelight to analyze such a colossal dataset. There was around 180 million taxi ride in the city of New York in 2014. BigData can effortlessly analyze the thousands of GB within a fractions seconds. This data can be analyzed for several purposes like avoiding traffics, lower rate where services are not functioning more frequently and many more. This information can be used by numerous authorities and industries for their own purpose. Government officials can use this data to deliver supplementary public transport services. The company like Uber can use this data for their own taxi service.

Introduction

Transportation has been proved as the most vital service in large cities. Diverse modes of transportation are accessible. In large cities in the United States and cities around the world, taxi mode of conveyance plays a foremost role and used as the best substitute for the general public use of transportation to get their necessities. For instance, by today in New York, there are nearly 50,000 vehicles and 100,000 drivers.

In order to provide a very good taxi service and plan for effective integration in the city transportation system, it's very important to analyze the demand for taxi transportation. The dataset provides relating information such as where taxis are used, when taxis are used and factors which tend the public to use a taxi as divergent to other modes of transportation.

In this complete traffic data which contains nearly 180 million rows of data in the year 2014. Due to the huge amount of data, this data is an example of "BigData." Using BigData it's easy to develop procedures to clean and process the data so it used in a useful way in transportation service.

TOOLS USED

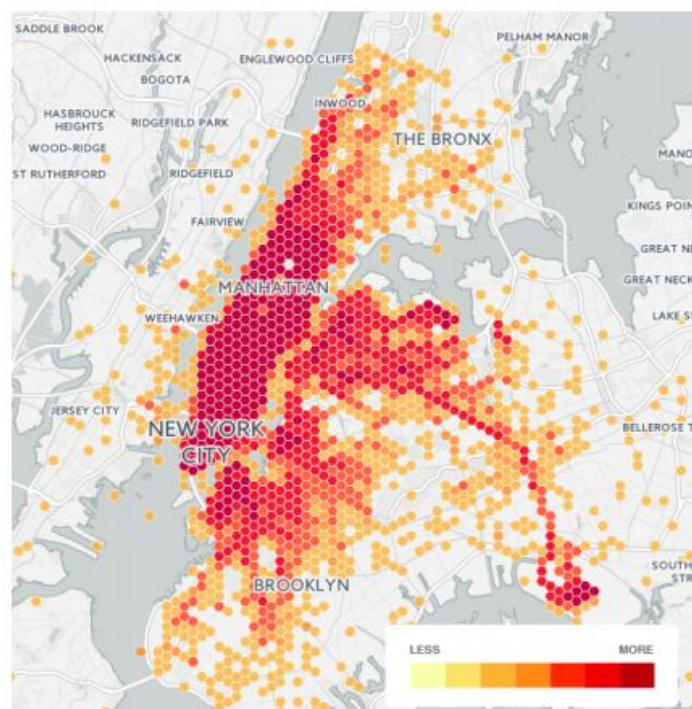
❖ Hadoop

❖ MapReduce

❖ Hive

EXPECTED RESULTS

The first analysis i.e. Analysis on Individual can help for determining the ability of the individual driver. We can determine the ability like efficient, quick, accuracy, etc. which can be helpful in evaluating the individual and reward the individual who is doing great work or train the individual who is struggling to do the good work.



In the second analysis, we determine which region has the highest pickup and drop-off location, it helps the vendor to provide more taxis where there is more pickup and lessen the number of taxis where there is more drop-off

References

- ❖ <https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city>