GREEN CABS VS UBER IN NEWYORK CITY

M.Prabhakar [1] A.Sravanthi [2] Jakkula Sushma Reddy[3]

[1] Assistant Professor, Department of CSE, Narsimha Reddy Engineering College, Maisammaguda, Dhulapally, Secunderabad, Telangana, India

[2] Assistant Professor , Department of CSE,Narsimha Reddy Engineering College, Maisammaguda, Dhulapally, Secunderabad, Telangana, India

[3] M. Tech Student, Department of CSE, Narsimha Reddy Engineering College, Maisammaguda, Dhulapally, Secunderabad, Telangana, India.

.

Abstract - *The broad utilization of area based administrations has prompted an expanding accessibility of direction information from urban situations. These information convey rich data that are helpful for enhancing urban communities through activity administration and city arranging. However, it likewise contains data about people which can imperil their security. In this examination, we work with the New York City (NYC) taxi trips informational index freely discharged by the Taxi and Limousine Commission (TLC). This informational collection contains data about each taxus taxicab ride that occurred in NYC. An awful hashing of the emblem numbers (the ID comparing to a taxi) permitted the recuperation of all the emblem numbers and prompted a security break for the drivers, whose wage could be effortlessly removed. In this work, we start an investigation to assess whether "culminate" namelessness is conceivable and if such a personality divulgence can be maintained a strategic distance from given the accessibility of different arrangements of outer informational collections through which the shrouded data can be recouped. This is refined through a spatio-fleeting join based assault which coordinates the taxi information with an outer emblem information that can be effortlessly assembled by an enemy. Utilizing a reproduction of the emblem information, we demonstrate that our assault can re-distinguish more than 91% of the cabs that employ in NYC notwithstanding when utilizing an ideal pseudonymization of emblem numbers. We likewise investigate the viability of direction anonymization procedures and show that our assault can at present recognize a critical division of the taxicabs in NYC. Given the limitations in distributing the taxi information by TLC, our outcomes show that unless the utility of the informational collection is altogether traded off, it won't be conceivable to keep up the security of taxi emblem proprietors and drivers.*

Keywords-Big social data, Social set analysis, Social business, Visual analytics, geo-spatial, GIS, Taxi, Green cabs, Uber..

I INTRODUCTION

The NYC Taxi and Limousine Commission (NYCTLC) is a legislative office made in 1971, and is in charge of the authorizing and managing of New York City's yellow taxis, for-procure vehicles, para-travel, passenger vans and other extravagance limousine administrations. The NYCTLC licenses and controls around 50,000 vehicles and tallies 100,000 drivers. The paper exhibited here will center around the Green taxis, that were presented by the Five-Boro Taxi Plan, a NYCTLC activity that expects to take care of the demand. In August 2013, the NYCTLC introduced a fleet of Green cabs to the city of New York. These Green taxicabs were presented with the objective of giving the occupants of Brooklyn, Queens, the Bronx, and Upper Manhattan more access to metered taxis. Considering that the Yellow taxis like to work in the regions of NYC that are most thick in pick-ups (Manhattan and the air terminals), the accessibility of Yellow taxicabs has a tendency to be low in the external districts of NYC. Thus, Green taxis are not permitted to get road hails from the biggest piece of Manhattan (beneath 110th St. on the West Side, and beneath 96th St. on the East Side), or both of JFK or LaGuardia air terminals.

II RELATED WORK

The NYC Taxi and Limousine Commission (NYCTLC) is an administrative organization made in 1971, and is in charge of the authorizing and controlling of New York City's yellow cabs, for-procure vehicles, para-travel, passenger vans and other extravagance

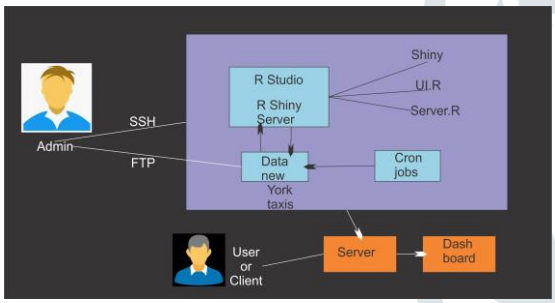
limousine administrations. The NYCTLC licenses and manages roughly 50,000 vehicles and checks 100,000 drivers. The paper exhibited here will center around the Green taxis, that were presented by the Five-Boro Taxi Plan, a NYCTLC activity that intends to take care of the demand surplus for taxi rides in the edges of New York City.

The "NYC Taxi Data Set," a verifiable store of 750 million rides of taxi emblems over a time of four years (2010-2013). This informational collection gives rich (bunch) data on the developments in a urban system as its residents approach their day by day life. We display an unearthly investigation of taxi development in view of the diagram Fourier change, which requires the ghostly disintegration of a vast coordinated, scanty lattice. Critical contemplations toward dealing with this network are talked about. Preparatory outcomes demonstrate that our technique enables us to pinpoint areas of co-conduct for movement in the Manhattan street organize.

Today, there are around 13,000 taxicabs being used in New York City consistently—however by outline they generally get and drop off a solitary traveler or gathering. Some famous transportation new businesses, for example, Uber and Lyft, offer ride-sharing alternatives, yet vehicles commonly have space for just two travelers at most.

Research distributed in Proceedings of the National Academy of Sciences in 2014 found that 80 percent of Manhattan taxi outings could be shared by two riders, yet the work didn't consider new riders joining after an excursion has just started. Likewise, the 2014 work and different investigations of ride sharing either confine the quantity of riders or they don't think about the impacts of giving clients a chance to pick diverse get and drop-off areas from each other, Alonso-Mora says. So the genuine advantages for huge limit vehicles haven't been resolved previously.

III DESIGN OF THE WORKFLOW:



Here first collect real dataset from DATA.GOV. Now divide real data into different chunks. To do this task we apply fixed size chunking algorithm. In fixed chunking algorithm initialize the number of chunks and size of chunks is to be generated for example size of 64 MB. It shows record is separated into different lumps of size 64MB.

Subset is the procedure of example will get which transitional keys and qualities. Every mapper must decide for the greater part of its yield (key, esteem) sets which reducer will get them.

IV EXPERIMENT RESULTS

Using trip data records, how does NYCTLC's share of rides per zip code compare to Uber's in the outer neighbourhoods of New York?.

1) Meaningful Fact #1: Green cabs are just as popular as Uber on the weekend. The distribution of rides according to weekends versus weekdays comparison is very similar in regards to Green cabs and Uber as shown in Fig. 4. Also, the distribution is close to equal in both cases with approximately 40% of the rides occurring during the weekends. It should be noted, though, that the distribution is not really equal in terms of days as the weekends constitute 2.5 day and the weekdays 4.5 days. This means that even though the visualization deceives the interpreter to think of the distribution as a close to 50/50, one should realise that there are more rides taken place during one weekend day than one week day.

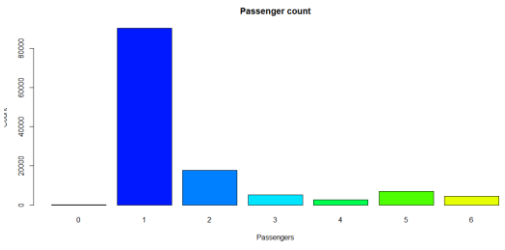
2) Meaningful Fact #2: Weekdays versus weekend rides per hours. To make up for this difference in days with 4.5 weekday days and 2.5 weekend days, we took the total of number of rides occurring during weekdays and divided them by the total number of hours in

4.5 weekdays. Similarly we took the total number of rides occurring in weekends and divided that by the total number of hours in those 2.5 weekend days. The bar charts in Fig. 5 show the difference between the average rides per hour in weekends and weekdays for Green cabs and Uber respectively in total numbers, on the left, and in percentage increase, on the right. The difference is clearer in the right bar chart, as it shows a 3% higher increase of Uber rides per hour in the weekends i.e. compared on average hours, Uber increases during weekends by 48% while Green cabs increase by 45%.

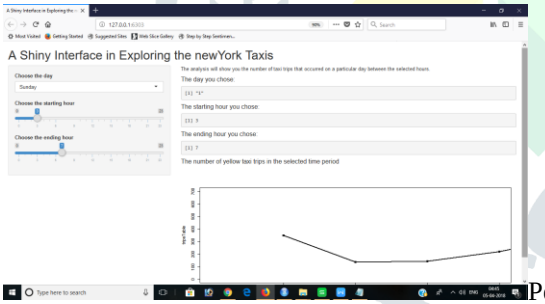
3) Meaningful Fact #3: There is no clear correlation between the negative and/or positive growth of Uber and Green cabs. With the explosive growth of Uber, one could imagine that when looking at both negative and positive growth, Green cabs would see a negative growth where Uber is experiencing a positive growth, i.e. a 'takeover' growth by Uber. However, as seen in the growth visualization below this is not the case in all areas

	Var1	Freq
1	1	61043
2	2	66367

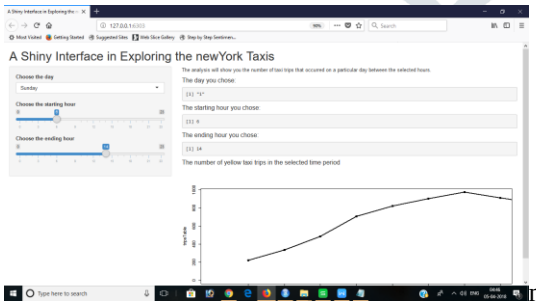
Vendors Green taxis and Yellow Taxis



maximum passenger count



Peak hour trips made with in the time lap



maximum rides @ time intervals

V CONCLUSION

In this paper, we conducted classification model for analyzing the data of taxis which found to be more effective than the statistical models, creating the subsets for the required data based on the probability models. The results obtained are pretty much helpful for the organization in arranging the cabs in peak hours at different location which in return provides enough profits for the company and the analysis helps in identifying the better locations for the cabs to be maintained

.

VI REFERENCES

The analysis is done for only the data available online. The same can be implemented to the data of national and local taxis like OLA Cabs, Radio Cabs etc.. This will results in economic growth of the company and also improve the financial value of the cabs. The application seems to works with almost 200000 lines of the data, The same can be implemented with hadoop and R. This can be applied using the Rstatistics which works well in the hadoop and R framework. This produces a huge amount of data analytical platform with the best use of all available resources. Finally implementing the analysis with shiny makes ease of anlysing the data and improved the economic value of the products

VI REFERENCES

[1] Yukun Zhou, Dan Feng, Wen Xia, Min Fu, Fangting Huang, Yucheng Zhang and Chunguang Li, "Information Deduplication Techniques", 2010 International Conference on Future arrangement Technology and Management Engineering, IEEE 2010, pp. 430-433.

[2] Wen Xia, Hong Jiang, Dan Feng and Lei Tian, "MUCH: Multithreaded Content-Based File Chunking". IEEE Transactions on Computers, IEEE 2015, pp. 1-6.

[3] Won, Lim and Min, "Set out: A Deduplication-Aware Resemblance Detection and Elimination Scheme for Data Reduction with Low Overheads", IEEE Transactions on Computers, IEEE 2015, pp.1-14.

[4] Zhi Tang and Youjip Won, "SecDep: A User-Aware Efficient Fine-Grained Secure Deduplication Scheme with Multi-Level Key Management", IEEE 2015, pp. 1-4.

[5] Qinlu He, Zhanhuai Li and Xiao Zhang, "Multithread Content Based File Chunking System in CPU-GPGPU Heterogeneous Architecture", Communications and Processing, IEEE 2011, pp. 58-64.

[6] Bin Lin, Shanshan Li, Xiangke Liao and Jing Zhang, "A Study on Data Deduplication Techniques for Storage", 2014 Sixth International Conference on Advanced Computing(lCoAC), IEEE 2014, pp.161-166.

[7] E. Manogar and S. Abirami, "ReDedup: Data Reallocation for Read Performance in Deduplication System", 2013 International Conference on Advanced Cloud and Big Data, IEEE, pp.117-124.

[8] Apache-Hadoop, http://www.hadoop.apache.org/wiki/apachehive

[9] http://hadoop.apache.org