International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 03 | March - 2023 Impact Factor: 7.185 ISSN: 2582-3930

Cancabs - Efficient, Data Driven Commute Analytics

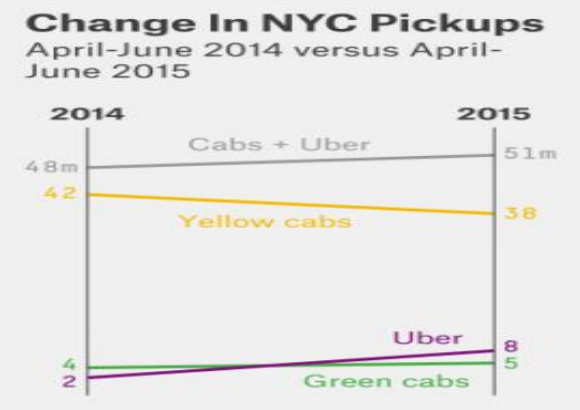
Harsh Gupta[1], Aksh Sood[2] and R. Lakshminarayanan[3]

[1] Department of Networking and Communications, SRM Institute of Science and Technology, Kattankulathur, 603 203, Chennai, India

Introduction

CabCabs is A daily commute cab solution for millions traveling to develop a more efficient and stress-free travel using data models and efficiency charts. Car lending for cab services based on user location estimation. It operates on scale. We thrive to achieve the best user satisfaction and experience using clustering and segmentation analysis. In this paper, we look forward to solving the problem of scale, and how it affects the different aspects of the market. In our case, the market of travel. There has been a shift in resources in this market.

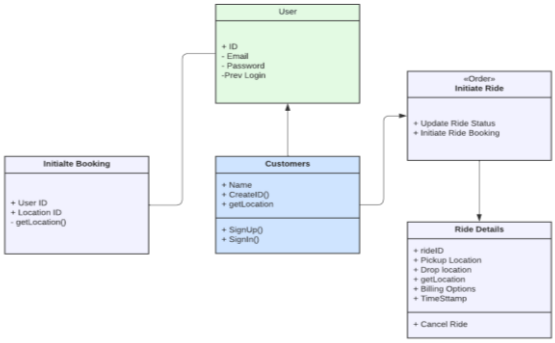
In this Study, we are implementing A daily commute cab solution for millions traveling to develop a more efficient and stress-free travel using data models and efficiency charts. Car lending for cab services based on user location estimation. It operates on scale. We thrive to achieve the best user satisfaction and experience using clustering and segmentation analysis.



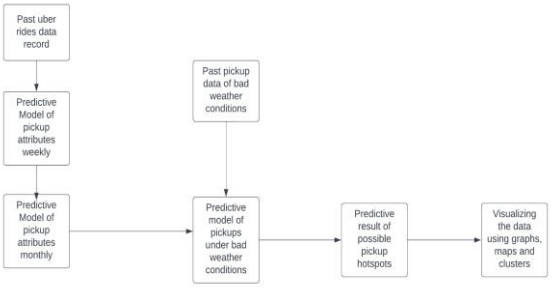
In this paper, we look forward to solving the problem of scale, and how it affects the different aspects of the market. In our case, the market of travel. There has been a shift in resources in this market.

Following our previous research, the following UML diagram is derived. Since we built this AI model based off of a real world

application like Uber, the UML diagram remains largely user specific.



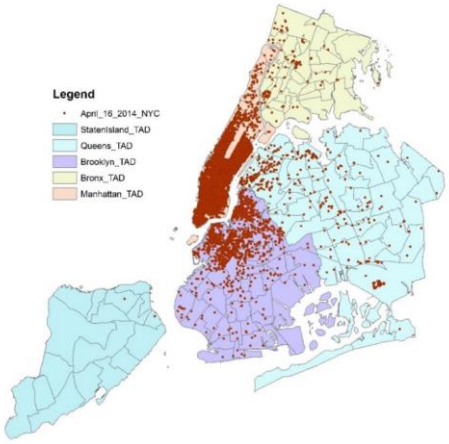
In this paper, we look forward to solving the problem of scale, and how it affects the different aspects of the market.



With reference to the above Block diagram we can understand the intent of the project and the description it portrays.

Related Work

Based on our research, Similar attempts have been made to replicate the efficiency of our model. For example, The paper titled "Predicting Short-Term Uber Demand Using Spatio-Temporal Modeling: A New York City Case Study" written by Sabiheh Faghih, Abolfazl Safikhani, Bahman Moghimi and Camille N. Kamga have collectively worked towards a similar goal. Quoting their abstract, "The demand for e-hailing services is growing rapidly, especially in large cities. Uber is the first and a well-known e-hailing service in New York City and the United States.In New York City, the demand for Uber has grown as compared to that for yellow cabs between 2014 and 2015. However, it's possible that this demand is not evenly dispersed throughout time or space. We can more clearly understand and estimate the demand for e-hailing services by using spatiotemporal time series model scan. In this study, two spatio-temporal models (Spatial-Temporal Autoregressive (STAR); least absolute shrinkage and selection operator) and one temporal model (Vector Autoregressive (VAR)) are examined for their ability to make predictions applied on STAR (LASSO-STAR)) and for different scenarios (based on the number of time and space lags), and applied to both rush hours and non-rush hours periods. The results show the need of considering spatial models for taxi demand."



Their model looks similar to the heat map though it doesn't account for the weather changes in a city like New York City.

A data file is taken and extracted of which the features and information is extracted

The date, time, day of the week, month day , hour of the day are extracted and saved in respective variables

The following features are processed and converted to pivot table

The pivot table is used to display various graphs journey by week, month ,hour day of different months at different times

The given data is then used to display the heatmap of the given location showing the various density of customers in a particular cluster

Another feature is extracted from the data to provide details of the weather conditions of the region

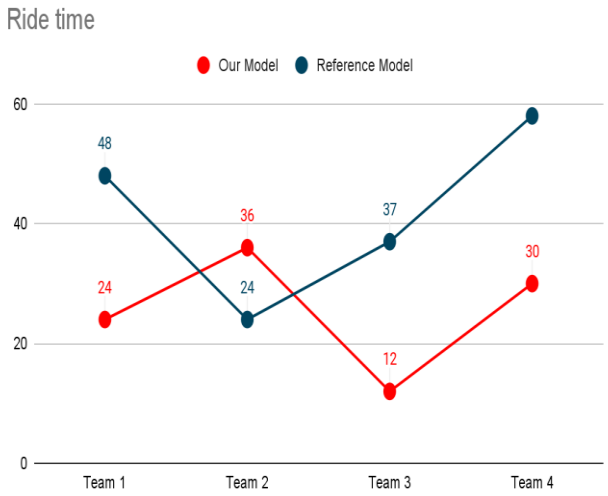
The weather condtions and the date time data are clubbed together .

Clustering is performed on the following data and then displayed the the map using hexagonal dataframes

Feature Set

As mentioned before, our model relies on scale and other information to calculate the best journey path. Completion of this project was done in two phases. Each phase adding a level of complexity to better predict ETA's (Estimated Time of Arrival) and quicker journeys. Keeping safety, reliability and availability in mind, the first phase involves cab density in a given area. We have taken an example of the New York City map. Based on the cab density data research done earlier, our model calculates a real time path for the closest cab. Not just Proximity, but the driver / Passenger preference is also kept in the loop. The implementation of this feature is attained using several useful python libraries. Some of these libraries are Pandas, Seaborn, MLP Toolkits, MatplotLib and basemap.

Coming to the second phase of the project. It deals with the weather information gathered from various sources. This not only helps with traffic prediction, but also gives the user crucial information on their destination's weather conditions. The type of transport offered to the user may also change based on inter-changeable conditions of the weather. Weather based transport systems can take input from multiple sources of weather and supply a more accurate measurement of the weather on a regular basis. The user may also help with these predictions by suggesting changes in the atmosphere voluntarily. The implementation of this feature is attained using several useful python libraries. Some of these libraries are h3, JSON, Folium, geoJSON and Branca Colormap.



Conclusion

The demand for Uber in Manhattan, New York City, is captured using a novel modeling approach that is presented in this study. The Manhattan TAD level and 15-min time intervals are used to aggregate the Uber pick-up data. With the help of this aggregate, a fresh spatiotemporal modeling strategy can be used to comprehend demand both spatially and temporally. LASSO-STAR and STAR, two spatio-temporal models, were created

using data from Uber pick-ups over the course of a normal day, and the performance of the models was assessed using MSPE. The MSPE findings showed that it is strongly advised to employ the LASSO-STAR model as opposed to the STAR model. While this is going on, the established spatio-temporal time series models' forecast accuracy can be increased by being aware of nearby demand data. It has also been discovered that the kind of weight matrix employed in spatio-temporal modeling can influence how well the models function. The effect of Uber on normal cabs will be analyzed using a change-point detection approach as part of this research. Additionally, future research will incorporate exogenous variables into the time series models to account for additional transport demand-related data, such as ridership on the subway and buses, demand for bicycles, weather, etc.

Data Availability

The information and data required to replicate or check this study are available with the respective authors and may be requested to be presented.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] https://www.researchgate.net/publication/317421456_How_cities_use_regulation_for_innovation_the_case_of_Uber_Lyft_and_Sidecar_in_San_Francisco

[2] https://www.researchgate.net/publication/331486131_Predicting_Short-Term_Uber_Demand_in_New_York_City_Using_Spatiotemporal_Modeling

[3] https://www.researchgate.net/publication/321604577_Predicting_Short-Term_Uber_Demand_Using_Spatio-Temporal_Modeling_A_New_York_City_Case_Study

[4] https://www.researchgate.net/publication/322971828_Putting_Data_in_the_Driver's_Seat_Optimizing_Earnings_for_On-Demand_Ride-Hailing

[5] Sturm, Bob L. "The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use." arXiv preprint arXiv:1306.1461 (2013).

[6] https://www.researchgate.net/publication/311766052_A_genetic_algorithm_approach_for_topic_clustering_A_centroid-based_encoding_scheme

[7] https://www.researchgate.net/publication/305214471_Clustering_Data_Streams

[8] https://www.researchgate.net/publication/351912026_Uber_Related_Data_Analysis_using_Machine_Learning