International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

A Machine Learning Framework for Profitability Profiling and Dynamic Price Prediction for the New York City Taxi Trips

Shylaja S, Kannika Nirai Vaani M

Abstract: ***The New York City Taxi & Limousine Commission's (NYC TLC) Yellow cabs are facing increased competition from app-based car services such as Ola, Uber, Didi, Lyft and Grab which is rapidly eating away its revenue and market share.***

Research work: In response to this, the study proposes to do profitability profiling of the taxi trips to focus on various key aspects that generate more revenue in future, visualization to assess the departure and arrival counts of the trips in various locations based on time of the day to maintain demand and supply equilibrium and also build a dynamic price prediction model to balance both margins as well as conversion rates.

Methodology/Techniques used: The NYC TLC yellow taxi trip data is analysed through a cross-industry standard process for data mining (CRISP-DM) methodology. Firstly, the taxi trips are grouped into two profitability segments according to the fare amount, trip duration and trip distance by applying K means clustering. Secondly, spatiotemporal data analysis is carried to assess the demand for taxi trips at various locations at various times of the day. Thirdly, multiple linear regression, decision tree, and random forest models are adopted for dynamic price prediction. The findings of the study are as follows, high profitable segments are characterized by airport pickup and drop trips, Count of trip arrivals to airports are more compared to departures from airports at any time of the day, and further analysis revealed that drivers making only a few numbers of airport trips can earn more revenue compared to making more number of trips in local destinations. Compared to multiple linear regression and decision tree, the random forest regression model is considered to be most reliable for dynamic pricing prediction with an accuracy of 91%.

Application of research work: The practical implication of the study is the deployment of a dynamic pricing model that can increase the revenue of the NYC TLC cabs along with balancing margin and conversion rates.

Keywords : Clustering, profitability profiling, machine learning, dynamic pricing, predictive modeling.

Revised Manuscript Received on March 30, 2020.
***** Correspondence Author
Shylaja S*, Business Analytics Specialization, Institute of Management, Christ (Deemed to be University), Bangalore, India.
Email: shylaja.s@mba.christuniversity.in
Kannika Nirai Vaani M, Assistant Professor, School of Business and Management, Christ (Deemed to be University). Bangalore, India.
Email: kannika.niraivaani@christuniversity.in

I. INTRODUCTION

An alternative to public transportation is private commercial transportation system. The Taxi Industry is one of the private commercial transportation systems. Taxi services based on GPS such as Lyft, Uber, and Ola, etc. have been into existence for the past few years. Such taxi services that are enabled by GPS have the potential to collect the location data related to every trip. This is considered a rich source of information to dig into meaningful insights related to demand from passengers and the mobility patterns of the passengers. The recorded GPS data has numerous applications in the Industry of private commercial transportation such as identification of popular pickup points, estimation of demand as per the day, week or month and monitoring of the traffic.

In general, some of the important questions to be addressed to increase the profitability of private commercial transportation businesses are (a) which are the popular destinations most passengers travel to? (b) Which are the locations where transport can find maximum departures (c) which time of the day or which day of the week or which month of the year the demand for public transport is high? (d) On what basis can the price of public transport be varied to maximize the revenue and so on? The answers to these questions help transport businesses in the planning of their operations. Therefore, the purpose of this study is to get insights into the above questions for yellow taxis of The New York City Taxi and Limousine Commission (NYCTLC).

In The United States being a large city, a taxi plays a very important role as it is the best substitute for public transportation. New York City taxi has green and yellow medallion taxi cabs that are licensed by NYCTLC and this study focuses on analyzing the records related to yellow medallion taxi cabs.

II. PROBLEM STATEMENT

NYC Taxi & Limousine Commission's (NYCTLC) Yellow cabs are facing increased competition from app-based car services such as Ola, Uber, Didi, Lyft and grab which is rapidly eating away its revenue and market share. In response to this, the study proposes to do profitability profiling of taxi trips to focus on various key aspects that generate more revenue in the future and also build a dynamic price prediction model.

III. LITERATURE REVIEW

A large part of the literature focuses primarily on spatial-temporal analysis and demand & price forecasting of the taxi trips. Popular destinations in New York City were uncovered by detecting popular drop-off locations using density-based clustering (Garcia, Avendano, & Vaca, 2018).

Negative binomial regression model and Spatial association were implemented to analyze the travel patterns in urban areas for effective investment decisions in rapid transit infrastructure and service (Hochmair, 2016) and kernal density analysis was used to detect demand fluctuations (Markou, Rodrigues, & Pereira, 2017). The travel time estimation problem was addressed with a big data-driven approach using a simple baseline for Travel Time Estimation (Wang, Kuo, Kifer, & Li, 2014). With the aid of the Semi-Markov process and approximate dynamic programming (ADP) approach, time of the pricing was determined with supply and demand fluctuations of taxis and resulted in increasing the market revenue by 10% (Qian & Ukkusuri, 2017).

Hierarchical density-based spatial clustering (HDBSCAN), Random Swap clustering and sequential pattern mining were implemented to build a system that delivers traffic insights and recommendations to help taxi drivers with useful guidelines (Ibrahim & Shafiq, 2019). A taxi searching algorithm with an accuracy of 97.59% accuracy was built using distributed coordination and clustering to minimize the time of taxi reaching the passengers (Agrawal, Raychoudhury, Saxena, & Kshemkalyani, 2018). The problem of predicting the number of taxis required in zones at various times was solved using STVec (smooth transaction vector error correction model) and multi-outputs support vector regression (MSVR) model (Zhou, Wu, Wu, Chen, & Li, 2015). Visual Exploration is a challenge for big Spatio-Temporal urban taxi data, this was overcome by building a model that uses an adaptive level-of-detail rendering strategy to create a visualization that is clutter-free for results that are large (Ferreira, Poco, Vo, Freire, & Silva, 2013). The Markov predictor model built for prediction of taxi demand was 89% accurate and better by 11% compared to neural network predictors (Zhao, Khryashchev, Freire, Silva, & Vo, 2016). A platform for distributed spatiotemporal analytics was built to deal with the heterogeneous spatiotemporal dataset (Deva, Raschke, Garzon, & Kupper, 2017) and a simulation combined with effective indexing scheme and parallelization was built to get a scalable approach (Ota, Vo, Silva, & Freire, 2015). Various deep learning models such as Long short-term memory (LSTM) neural networks, adaptive boosting, decision tree regression model was built to detect pattern variation in taxi trips and the LSTM model was proven to be the reliable model (Najafabadi & Allahviranloo, 2018). The implications of social factors on driving behavior of drivers were proved to be dependent on each other based on correlation and social influence theory (Xu, et al., 2017).

The yellow taxi demand forecasting was achieved through spatiotemporal autoregressive (STAR) model which outperformed vector autoregressive (VAR) model (Safikhani, Kamga, Mudigonda, Faghih, & Moghimi, 2018). An ensemble of a tree-based model and a Long Short-Term Memory (LSTM) were implemented to predict taxi demand at LaGuardia airport and Long Short-Term Memory (LSTM) Recurrent Neural Network achieved an MAE of 48.1 and tree ensemble model achieved MAE of 56.9 (Coviensky, Katiyal, Agrawal, & Geary, 2017). For optimal assignment and incentive design in the taxi group ride, a Heuristic algorithm was implemented by which 47% of taxi trip mileage was saved (Qian, Zhang, Ukkusuri, & Yang, 2017). To explore the dynamics of human mobility patterns, interactive visual views are were designed through tensor decomposition (Shi, Lv, Seng, Xing, & Chen, 2019).

The big data was analyzed through Hadoop for the selection of areas that are commercially viable to increase the income of taxi drivers (Devabhakthuni, Munukurthi, & Rodda, 2018). Through data exploration, it was found that tourists tip the drivers around 0.5% to 0.6% more than localities (Neto, Nowak, & Ross, 2017) and it was also found that in the area of Manhattan. Uber is growing faster than yellow cabs (Poulsen, Dekkers, Wagenaar, Snijders, & Lewinsky, 2016). With the aid of GPS, the taxi travel time was predicted by implementing Linear Regression, Randomized K-Nearest Neighbor Regression and Support Vector Regression models and Support Vector Regression provided the best accuracy (Laha, Putatunda, & Sayan, 2017). Passenger- and speed-weighted efficiency (PSWE) proved that the taxi trips within Manhattan, LGA and JFK airports improve the route efficiency (Zhai, Bai, Peng, & Gu, 2019).

The game theory model was implemented through the pure strategy of Nash equilibrium (PSNE) to discover that beyond west 110th and 96th street, the pickups of yellow taxi dropped (Zhang, 2018). A Dynamic and Predictive Technique for pricing (ADAPT-Pricing) was adopted to reduce the taxi prices by 5% and also to increase the revenue by 15% simultaneously (Asghari & Shahabi, 2018). Convolutional LSTM network outperformed the LSTM network in predicting the demand of the taxis (Li, Sun, & Pang, 2018). A statistical experiment of two-way clusters proved that there was a positive relationship between the level of sunlight and tipping percentage in taxi rides with an increase of tipping percentage from 0.5 to 0.7 from dark sky day to a full sunshine day (Devaraj & Patel, 2017). The GPS data streams of the taxi were used for prediction of pick-up locations and spherical regression model was the best compared to the multivariate regression model in terms of accuracy-time trade-off criterion (Laha & Putatunda, 2018). This paper utilizes the concepts of the above mentioned methodologies to detect the key aspects that generate more revenue for the industry.window for it.

IV. OBJECTIVES

The study aims to analyze various key aspects of taxi trips that generate more revenue by achieving the following objectives:

1. To classify the trips into various segments based on profitability.
2. To carry out spatiotemporal analysis to assess the net migration of taxi from one location to another at various times of the day to maintain demand and supply equilibrium of cabs.
3. To build a dynamic price prediction model to balance margin and conversion rates.

V. EXPERIMENTAL METHODOLOGY

The Cross Industry Process for Data Mining (CRISP-DM) methodology is followed for model building focusing on four modules involving data collection, data preparation, exploratory data analysis and model building, and diagnostics.

A. Data Collection

The models have been built on a publically available dataset. The NYC yellow taxi trip data include variables that capture the pick-up and drop-off locations with its latitude and longitude, pick up and drop off time stamp with dates, trip distance, trip fares, types of rates, types of payment and the count of passengers reported as per the driver. The data contains 583359 observations and 19 features. A second dataset is merged with the first dataset is imported and consists of 263 rows and 6 attributes. This second dataset consists of information on pickup and drop off locations. The final merged dataset has 583359 observations with 28 attributes.

B. Data Preparation

The data comprise of no missing values, however, there is the presence of junk values. The junk values are imputed through mean and mode techniques. The categorical variables are transformed into the format required for model building through the 'hot encoding' technique in python. K means clustering is applied to data to divide the trips into high and low profitable segments. Profitability profiling is given in Table 1. The data is divided into train and test data, where 70% of the data is used for training the model and the remaining 30% for testing.

Table I: Taxi Trip profiling

Cluster and frequency	Average fare amount	Average trip distance	Taxi trip Profiling
Cluster 1 - 397228	$13.7	2.4 kms	Low profitable trips
Cluster 2 - 170240	$69	17.2 kms	High profitable trips

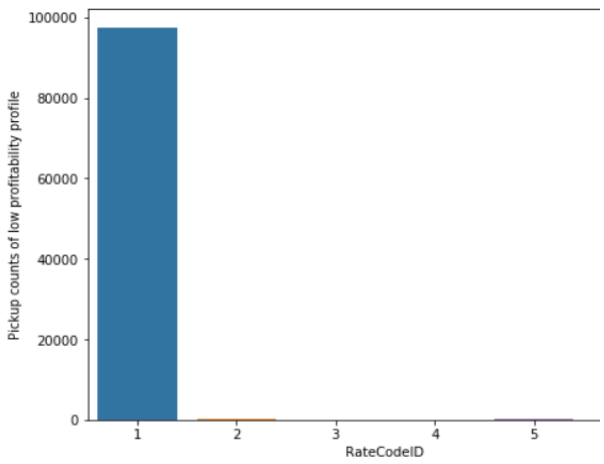
C. Exploratory Data Analysis

Figure 1 of EDA represents the plot of segment wise trip pickup counts vs. rate code ID. The low profitable trips focus only on the local pick up and drop off locations, hence they have only standard rates. Whereas the high profitable trips have most picks ups from JFK airport (Rate code ID2) followed by pickups from Newark airport (Rate code ID 3) which are from Newark and JFK airports and minimal pickups in local based on the standard rate and from Nassau or Westchester.

Figure 2 represents plot of segment wise pickup count of the taxi for an hour of the day and day of the week. The graph shows the comparison between low profitable trips (upper plot) and high profitable trips' (lower plot) pickup count of taxi for an hour of the day and day of the week. The low profitable trips have more demand on Thursday, Friday and Saturday. But has a drop in demand for taxis on Sunday. However, the high profitable trips have uniform demand throughout the week except on Friday, the demand is high.

Further, an interactive visualization through a New York City map is developed in this section of the study. The study in this section tries to explore the net migration of taxis from one location to another and if it depends on the time of the day. Therefore arrival and departure counts of taxis for each location for every hour are computed. The analysis is represented on an interactive New York City map and markers are added and they are assigned with two colors. If there are more number of departures than arrival, they are represented by orange circles and if there are more number of arrivals than departures, they are represented by blue circles.

Figure 3 represents a spatiotemporal graph at noon. The plot of net departures at noon of the day at LaGuardia airport with 411 departures and 735 arrivals. Figure 4 represents spatiotemporal graph at 6 PM. spatiotemporal visualization gives the plot of net departures at 6 p.m. of the day at Midtown Manhattan with 1165 departures and 2738 arrivals. Figure 5 represents spatiotemporal graph at 10 PM. spatiotemporal visualization gives the plot of net departures at 10 p.m. of the day at Times Square with 751 departures and 2168 arrivals.



Rate code ID 1= Standard rate 2=JFK 3=Newark 4=Nassau or Westchester 5=Negotiated

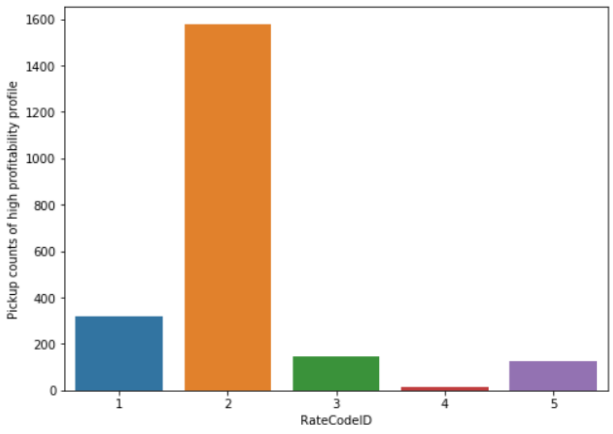


Fig1.The Plot of segment wise trip pickup counts vs. rate code ID.

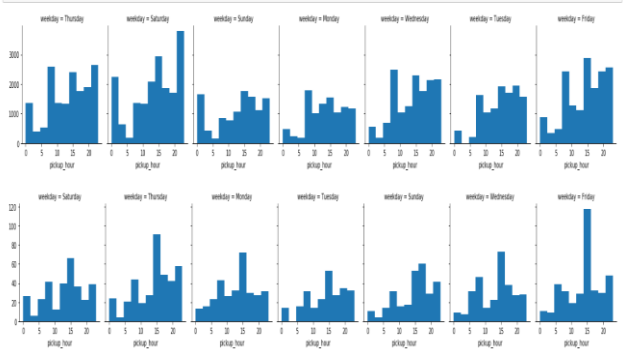


Fig.2 Plot of segment wise pickup count of the taxi for hour of the day and day of the week.

```
# plot net departures at 12PM
trip_counts = get_trip_counts_by_hour(12)
plot_station_counts(trip_counts)
```

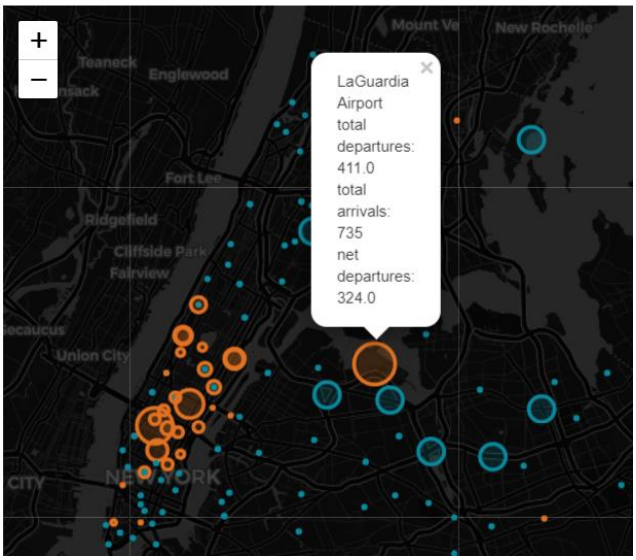


Fig.3 Spatiotemporal graph at 12 PM

```
# plot net departures at 6PM
trip_counts = get_trip_counts_by_hour(18)
plot_station_counts(trip_counts)
```

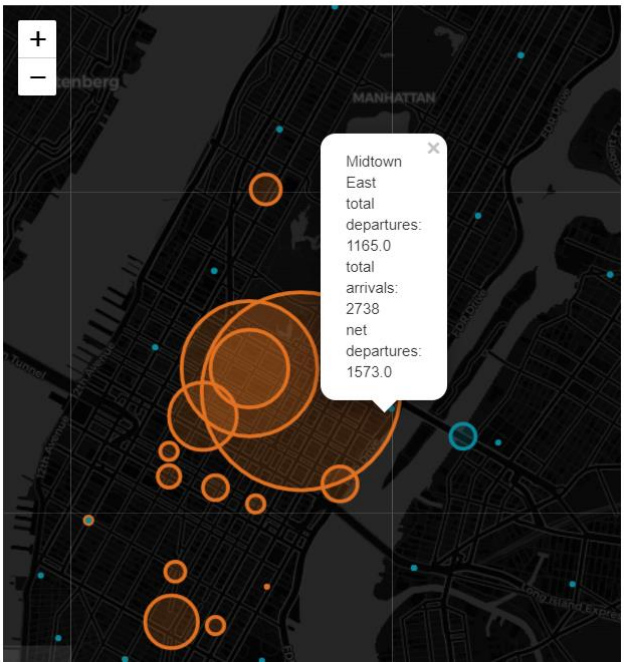


Fig4. Spatiotemporal graph at 6 PM

```
# plot net departures at 10PM
trip_counts = get_trip_counts_by_hour(22)
plot_station_counts(trip_counts)
```

Fig5. Spatiotemporal graph at 10 PM

D. Model building and diagnostics

Dynamic price prediction models based on multiple linear regression, decision tree, and random forest are proposed. The features for the model building were extracted through principal component analysis. The features are pickup hour, travel distance, pick up latitude and longitude, and drop off latitude and longitude, distance traveled and day of the week. Further, the linear regression model, decision tree and the random forest are built on the training data. The set of rules or problem solving operations for each of these algorithms related to this study are given in below sections.

VI. MULTIPLE LINEAR REGRESSION

Certain assumptions need to be satisfied before building the multiple linear regression model, each of these assumptions and their results related to the data is as follows: (a) The dependent variable must be scalar: This assumption is satisfied as the dependent variable is 'fare amount' which is scalar. (b) No Auto-correlation: This condition is measured by the Durbin Watson coefficient. As the coefficient value is 1.935 and lies between 1.9 and 2.08, the dataset considered satisfies the condition of no Autocorrelation. (c) No multicollinearity: As the features in figure 6 show no correlation and hence the condition of no multicollinearity is satisfied. The multiple linear regression model is implemented as the major assumptions are satisfied.

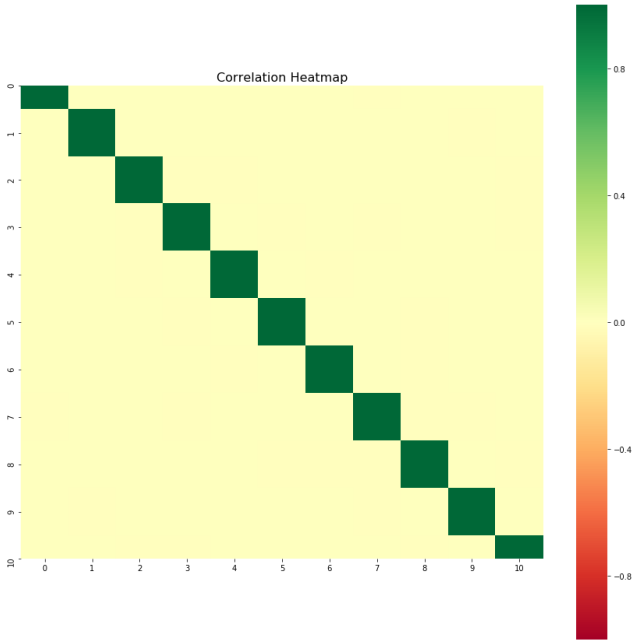


Fig6. Correlation heat map

VII. DECISION TREE

A regression model can be built in the form of a tree structure with the aid of decision tree. The dataset is broken into smaller subsets with leaf and decision nodes. Greedy search method is employed by the decision tree and the ID3 algorithm is used in decision tree regression model and the results are predicted based on standard deviation reduction, the same set of rules is adopted in the model for taxi price prediction.

VIII. RANDOM FOREST

A supervised machine learning technique which use ensemble method of learning to predict the outcome, in this study for price prediction. A technique of bagging that works by aggregation of various decision trees with the following modifications:

1. The splitting of variables at each node is based hyperparameter. This enables the ensemble technique not to rely much on an individual variable, therefore using all the variables to its fullest potential.
2. Further while building the splits, random sample of the data is considered for building each tree resulting the avoiding the problem of over fitting. As a result of above two steps the built trees are not highly correlated leading to a reliable results from the model.

Table 2 indicates the feature importance score of the random forest model implying trip distance, trip duration and pickup hour has the highest influence of price prediction.

Table II: Feature Importance score in Random Forest

Feature	Feature Ranking
Trip distance	0.772963
Trip duration	0.172500
Drop off Longitude	0.018166
Pickup hour	0.013449
Pickup Longitude	0.008346
Drop off Latitude	0.005636
Pickup Latitude	0.004831
Thursday	0.001652
Tuesday	0.001219
Sunday	0.000743
Saturday	0.000495
Friday	0.000391
Monday	0.000255
Wednesday	0.000050

IX. COMPARISON OF MODELS:

The built models are tested on 30% of previously unseen data. The models are compared based on their root mean square errors (RMSE) and their variance score computed on test data.

Table III: RMSE and Variance scores

Model	RMSE value	Variance score
Multiple linear regression	3.6083	0.87
Decision Tree	3.6083	0.85
Random Forest	2.9634	0.91

The model with the least error and highest variance is considered as the best model. Therefore, the Random Forest model is finalized for deployment as it has the least RMSE values compared to other models and highest variance score as shown in table 3 inferring there is 91% of variability in the dependent variable ‘fare amount’ caused by the independent variables.

X. MODEL DEPLOYMENT:

The finalized random forest model was deployed by building a web application using the FLASK module in python as shown in figure 7. The business impact of this deployed model is illustrated in figure 8 and figure 9 with a comparison of taxi fare prediction on a Monday at 8 am from Eastchester to Times Square and the same trip on the same day but at evening 7 pm during peak hour. The predicted price for the former trip is $217.86 and the latter trip is $256.82. Therefore, the increase in revenue earned by the yellow cab driver for the same trip during peak hours is $38.96. Hence, a conclusion can be drawn that the dynamic pricing model has a significant business impact.

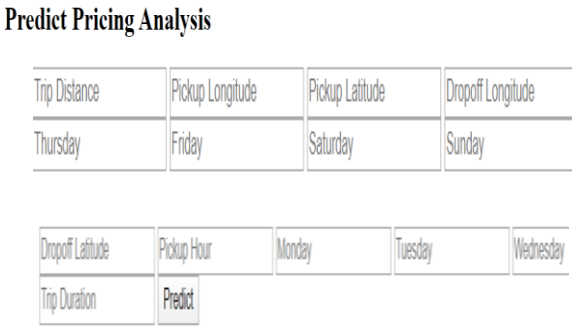


Fig7. Deployed random forest model

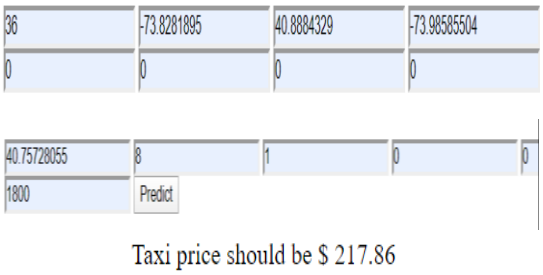


Fig8. Predicted taxi fare from Eastchester to Times Square on a Monday at 8 am.

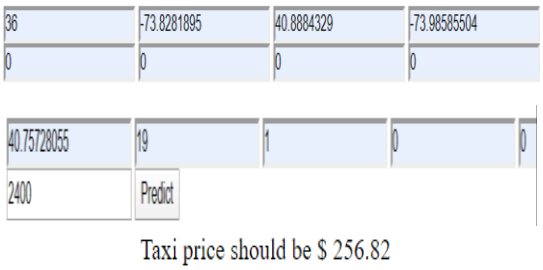


Fig9. Predicted taxi fare from Eastchester to Times Square on a Monday at 7 pm.

XI. FINDINGS

The findings for profitability profiling, spatiotemporal analysis and dynamic price prediction of taxi trips are as follows, though there are only 170240 trips that belong to high profitable segment compared to 397228 trips that belong to low profitable segment, the average revenue earned by high profitable segment trips is $69 and average revenue earned by low profitable segment trips is $13. This is because the average trip distance and trip duration traveled by high profitable segment trips are high compared to low profitable segment trips. The low profitable segment trips are the ones that are restricted to local destinations and most of the high profitable segment trips are the ones that travel from local locations to JFK, Newark, and Nassau or Westchester destinations.

The demand for low profitability characteristic trips increases gradually from 6 a.m. onwards and peaks at 7 p.m. and decreases gradually. The demand for high profitability characteristic trips increases 4 a.m. onwards and dips after 7 a.m. onwards. Later, the demand increases from 11 a.m. onwards and peaks at 2 p.m. and gradually decreases. The demand for low profitability characteristic trips (local destination trips) is high on Fridays and Saturdays whereas dips on Sundays. The demand for high profitability characteristic trips (airport trips) is uniform across days of the week except on Fridays, the demand is high compared to other days. JFK airport is the popular destination for taxi trips at all times of the day, LaGuardia airport has maximum arrivals between 10 a.m. and 1 p.m. and between 6 p.m. and 8 p.m., Times Square is the most popular destination at any time of the day for taxi trip, however, the arrival counts to this place is maximum between 5 p.m. to 10 p.m. compared to other destinations. After 5 p.m. the arrivals at Midtown Manhattan increases drastically.

Principal component analysis for feature extraction extracted the important features contributing to price prediction of taxi trips. The important features are trip distance, trip duration, pickup latitude, pickup longitude, drop off latitude, drop off longitude and weekday. Random forest is the most reliable model for dynamic price prediction with an accuracy of 91%.

XII. DISCUSSION

In summary, the study segregates the trips into highly profitable and low profitable segments. The analysis revealed that drivers making only a few number of airport trips can generate more revenue compared to making more trips in local destinations. Popular destinations and their demand trends were uncovered through spatiotemporal analysis to maintain the demand and supply equilibrium. The data is analyzed through CRISP-DM methodology with the aim to build a dynamic price prediction model that predicts optimal prices that can balance both margins as well as conversion rates. As there are millions of taxi trips recorded every year, implementing this project in a big data platform and integrating with customer records is the scope of future study. Also, some customers cancel the bookings, therefore the study can further be extended in predicting the taxi booking confirmation or rejection from a customer after generating the dynamic price for the taxi trips.

XIII. CONCLUSION

The study is carried out in a structured approach with the aid of CRISP-DM methodology. The understanding of variables under investigation is the key aspect of building the solution for the identified issue and to fulfill the given objectives Numpy, pandas, scikit learn and matplotlib libraries in python plays an important in executing the project. K-means clustering, principal component analysis, multiple linear regression, decision tree and random FOREST and the techniques adopted in the study to achieve the goals and objectives. This study discovers the key aspects of the trips that generate more revenue through profitability profiling. K means clustering method segregated the trips into highly profitable and low profitable segments based on fare amount, trip distance and trip duration. Spatiotemporal analysis reveals the popular destinations with maximum arrival counts at various times of the day. The random forest regression model is a reliable model for dynamic price prediction with an accuracy of 91%. The study results indicate that airport destination trips are highly profitable. Also, only fewer airport destinations or arrival trips earn more revenue to drivers compared to many local destination trips. Therefore, through the analysis, the study has provided the findings that are key aspects to take strategic business decisions to increase the revenue to NYCTLC. Finally, though the study has its own limitations, they can be considered as future scope of study and can be explored for new insights in future for the given problem statement and objectives of the study.

REFERENCES

1. Agrawal, A., Raychoudhury, V., Saxena, D., & Kshemkalyani, A. D. (2018). Efficient Taxi and Passenger Searching in Smart city using Distributed Coordination.
2. Hochmair, H. H. (2016). SPATIO-TEMPORAL PATTERN ANALYSIS OF TAXI TRIPS IN NEW YORK CITY.

3. Xu, T., Zhu, H., Zhao, X., Liu, Q., Zhong, H., Chen, E., & Xiong, H. (2017). Taxi Driving Behavior Analysis in Latent Vehicle-to-Vehicle Networks: A Social Influence Perspective. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 1285-1294).
4. Asghari, M., & Shahabi, C. (2018). ADAPT-Pricing: A Dynamic And Predictive Technique for Pricing to Maximize Revenue in Ridesharing Platforms. *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, (pp. 189-198).
5. Capretz, & M., M. A. (2019). Spatiotemporal Forecasting At Scale. *Electronic Thesis and Dissertation Repository. 6316.*
6. Coviensky, A., Katiyal, A., Agrawal, K., & Geary, W. (2017). Estimating Demand for Taxis at LaGuardia Airport.
7. Deva, B., Raschke, P., Garzon, S. R., & Kupper, A. (2017). STEAM: A Platform for Scalable Spatiotemporal Analytics. *A Platform for Scalable Spatiotemporal Analytics. Procedia Computer Science*, 731–736.
8. Devabhakthuni, K., Munukurthi ,., & Rodda, S. (2018). Selection of Commercially Viable Areas for Taxi Drivers Using Big Data. *Devabhakthuni, K., Munukurthi, B., & Rodda, S. (2018). Selection of Commercially Viable Areas for Taxi Drivers Using Big Data. Smart Innovation, Systems and Technologies.*
9. Devaraj, S., & Patel, P. C. (2017). Taxicab tipping and sunlight. *PLOS ONE*.
10. Ferreira, N., Poco, J., Vo, H. T., Freire, J., & Silva, C. T. (2013). Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips.
11. Garcia, J. C., Avendano, A., & Vaca, C. (2018). Where to go in Brooklyn: NYC Mobility Patterns from Taxi Rides. *WorldCIST*, 203–212.
12. Howard, A. J., Lee, T., Mahar, S., Intrevado, P., & Woodbridge, D. M.-k. (2018). Distributed Data Analytics Framework for Smart Transportation. *IEEE 4th Intl. Conference on Data Science and Systems.*
13. Ibrahim , R., & Shafiq, O. (2019). Detecting taxi movements using Random Swap clustering and sequential pattern mining. *Journal of Big Data volume*.
14. Laha, A. K., & Putatunda, S. (2018). Real time location prediction with taxi-GPS data streams. *Transportation Research Part C: Emerging Technologies*, Pages 298-322.
15. Laha, Putatunda, A. K., & Sayan. (2017). Travel time prediction for GPS taxi data streams. *Indian Institute of Management Ahmedabad*.
16. Li, P., Sun, M., & Pang, M. (2018). Prediction of Taxi Demand Based on ConvLSTM Neural Network. *International Conference on Neural Information Processing.*
17. Markou, I., Rodrigues, F., & Pereira, F. C. (2017). Use of Taxi-Trip Data in Analysis of Demand Patterns for Detection and Explanation of Anomalies. *Transportation Research Record*, 129-138.
18. Najafabadi, S., & Allahviranloo, M. (2018). Inference of Pattern Variation of Taxi Ridership Using Deep Learning Methods: A Case Study of New York City. *International Conference on Transportation and Development*

AUTHORS PROFILE

Shylaja S, received the bachelor's degree in Electronics and Communication Engineering from Global Academy of Technology affiliated to Visvesvaraya Technological University (VTU) in the year 2017. Currently, a final year MBA Student in Business Analytics Specialization at Institute of Management, Christ (Deemed to be University), Bangalore, India. She will be completing her post-graduation by March 2020 and is highly focused on research work and a competent academic career. She has presented research papers in national conferences. She is keen on exploring different areas in the field of Analytics. She has done notable interdisciplinary projects in the area of data mining and predictive analytics.

Dr. Kannika Nirai Vaani M holding MCA, M.Phil, Ph.D(Comp.Sci).She is currently working as Assistant Professor in School of Business and Management, Christ (Deemed to be University). She has 17 years of experience from IT Industry, Academics and research. Her expert areas are Datamining, Big Data Management, Data Modelling and Data Warehouse, Data and Business Analytics, Artificial Intelligence and Machine Learning.