REPORT

REDEFINING UBER'S CUSTOMER JOURNEY WITH DATA-DRIVEN STRATEGIES

AUTHORS

GAUTHAM SATHYANARAYNAN & MAHROOSA IBTESAAM

Abstract

In this company business case, our primary objective is to optimize Uber's customer service process, thereby improving overall efficiency and enhancing user experience. Our focus is on three crucial areas: Enhancing Driver Matching, Accurately Estimating Arrival Times, and Improving the Efficiency of Customer Support. Additionally, we have employed the Analytical Hierarchy Process (AHP) to determine the most suitable city in Canada for launching a new feature in the Uber app.

To enhance driver matching, we conducted a comprehensive statistical analysis of data patterns related to demand. These patterns include variables such as the day of the week and ride durations. By leveraging K-means and DBSCAN clustering techniques, we identified temporal patterns and activity levels, providing valuable insights for predicting demand, optimizing supply, and ultimately improving the overall customer service on the Uber platform.

In order to improve arrival time estimates for Uber, we meticulously analysed historical data, taking into account variables such as traffic, distance, time of day, and day of the week. After comparing several regression models, we determined that the KNeighborsRegressor model yields the most accurate results, as measured by Mean Square Error (MSE) and Mean Absolute Error (MAE).

To enhance the efficiency of the customer support process, we implemented ML-based feedback categorization and automated responses. This streamlined support approach enhances resolution efficiency and ensures a smoother customer support experience for Uber users.

After following a rigorous AHP process, we have determined that Toronto is the optimal city in Canada to deploy any new feature in the Uber app.

Enhance Driver Matching

Enhanced driver matching is vital in ride-hailing apps like Uber, involving the pairing of passengers with the nearest available drivers and deploying more drivers in busy areas. This approach brings numerous advantages for passengers and drivers alike. Reduced waiting times enhance customer satisfaction, while efficient distribution of demand and supply leads to smoother operations and optimized service.

Drivers benefit from higher utilization rates, maximizing their earnings and minimizing idle periods. The proximity of matched drivers also enhances safety and reliability during rides. Strategically, employing more drivers during peak times allows for better dynamic pricing management and supports local economies by creating additional income opportunities.

Overall, enhanced driver matching gives ride-hailing companies a competitive edge, attracting users and drivers, while simultaneously delivering an improved transportation experience for all parties involved.

Our analysis aimed to enhance driver matching by exploring temporal patterns related to demand using two clustering methods: KMeans and DBSCAN. The dataset consisted of various variables, including trip start and end times, trip category, start and stop locations, trip distance, and purpose.

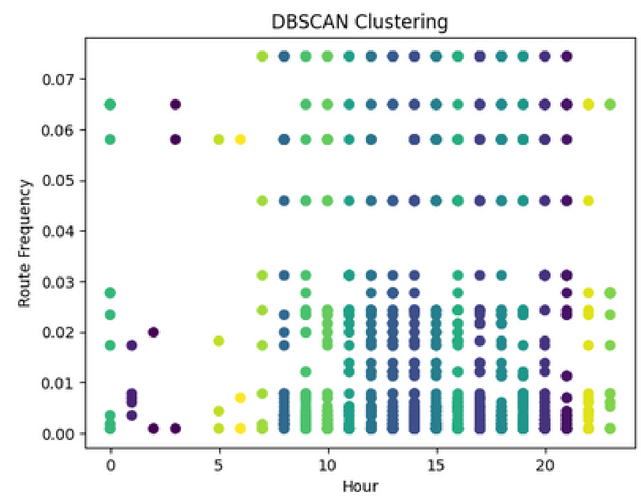
Clustering 1: KMeans Clustering

Using KMeans resulted in four distinct clusters, each representing specific time periods with varying activity levels. Cluster 1, corresponding to morning hours, exhibited the highest average count, indicating peak activity during this period. Cluster 2, representing early morning and late-night hours, had the lowest average count, suggesting minimal activity. Cluster 0 and Cluster 3 represented evening and afternoon hours, respectively, showing moderate levels of activity.

Clustering 1: KMeans Clustering

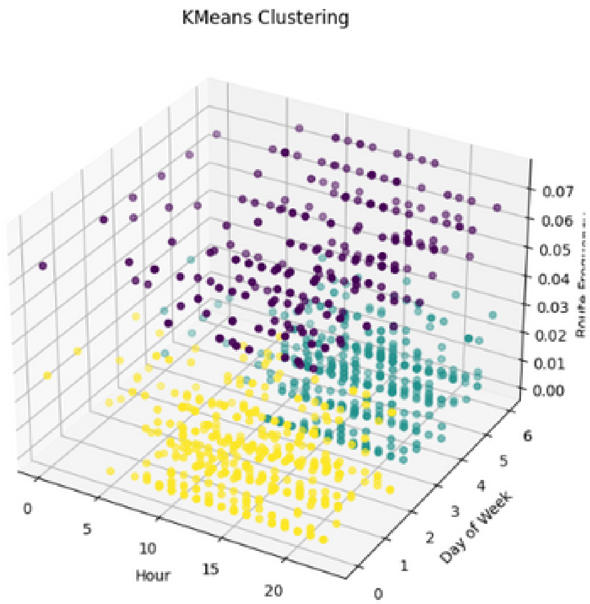
The implemented DBSCAN yielded 16 clusters, including one cluster labeled as noise or outliers. Each cluster represented a specific hour and its average count, providing a more granular view of temporal trends.

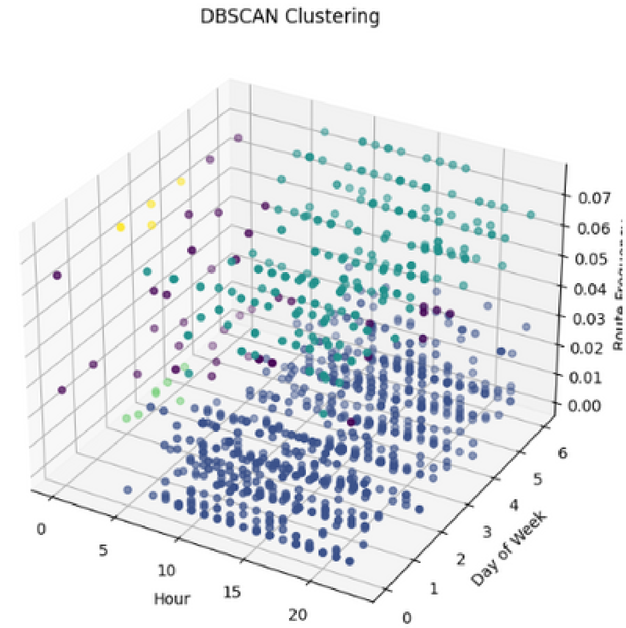
Certain clusters, such as Cluster 8 (hour 15) and Cluster 9 (hour 18), showed high average counts, indicating significant activity during those specific hours. Other clusters, such as Cluster 13 (hour 0) and Cluster 2 (hour 20), had relatively lower average counts, suggesting quieter periods.



Clustering 2: KMeans Clustering

Clustering 2 with KMeans divided the data into three clusters based on hours and days of the week. Cluster 0 encompassed a broad range of hours throughout the day and all days of the week, with a moderate average count. Cluster 1 included specific days and hours, exhibiting higher activity. Cluster 2 represented different days and hours, displaying the highest average count, indicating significant activity during those specific time frames.





Clustering 2: DBSCAN Clustering

With DBSCAN resulted in four clusters, including one noise cluster. Cluster 1 represented hours occurring throughout the day on all days of the week with a moderate average count. Cluster 0 encompassed hours from 5 to 23 on all days of the week, indicating significant activity during these periods. Cluster 1 exhibited similar hours but excluded hour 5, showing moderate activity. Cluster 2 included only hours 0, 1, and 2 on specific days, suggesting lower activity during those time frames.

Comparing the two sets of data, both KMeans and DBSCAN algorithms were employed. The second set, however, introduced the inclusion of days of the week, enabling deeper insights into temporal patterns. The number and composition of clusters differed between the two sets, reflecting variations in data grouping and providing valuable information for optimizing driver matching strategies based on demand patterns at different times and days.

Accurately Estimating Arrival Times

Accurate estimation of arrival times is vital for ride-hailing apps like Uber, ensuring a seamless experience for passengers and drivers alike."Accurate ETAs are important for positive user experience. But ETAs are most difficult to get right since it involves many factors like traffic, time of the day, weather, etc." (Singh, n.d.).

Precise arrival time estimates contribute to customer satisfaction by allowing passengers to plan efficiently and avoid unnecessary waiting. Reduced wait times enhance the convenience and appeal of the service, potentially attracting more users. Additionally, dependable arrival time predictions build trust and reliability, leading to fewer ride cancellations due to uncertainty.

For drivers, accurate estimates enable better time management, enhancing their efficiency and productivity. This can potentially lead to to better earning potential and job satisfaction for Uber's network of drivers

Ride-hailing companies can strategically leverage this data to optimize operations, analyze traffic patterns, and deploy drivers more effectively, gaining a competitive advantage. Ultimately, providing accurate arrival time estimates is integral to customer retention, safety, and the overall success of ride-hailing apps in the competitive transportation landscape.

To enhance the accuracy of arrival time estimation for Uber, a comprehensive analysis was conducted utilizing various regression models and historical data. The goal was to predict the time taken for rides starting from different locations, considering factors such as traffic, distance, time of day, and day of the week.

Six regression models were evaluated for their predictive performance using 5-fold cross-validation, and their Mean Squared Error (MSE) and Mean Absolute Error (MAE) were calculated to assess their accuracy.

Linear Regression Model:

The Linear Regression model exhibited poor predictive accuracy, with an MSE of 996399268312170836787 2 and with a significantly elevated MAE of 29794387657.08.

Decision Tree Regressor Model:

The Decision Tree Regressor model demonstrated improved predictive performance compared to LinearRegression, with a lower MSE of 496.75 and a MAE of 9.89.

Random Forest Regressor Model:

The RandomForestRegressor model further improved the predictive accuracy, achieving a lower MSE of 308.48 and a MAE of 8.94 compared to the DecisionTreeRegressor model.

K-Neighbors Regressor Model:

The K NeighborsRegressor model outperformed all previous models, displaying the lowest MSE of 285.78 and the smallest MAE of 8.50, indicating superior predictive accuracy for estimating arrival times.

SVR (Support Vector Regressor) Model:

The SVR model obtained an MSE of 306.11 and a MAE of 8.78, showing slightly higher average absolute difference compared to the KNeighborsRegressor model.

.

XGB Regressor Model:

The XGB Regressor model exhibited slightly improved performance compared to the SVR model, achieving an MSE of 298.41 and a MAE of 9.40.

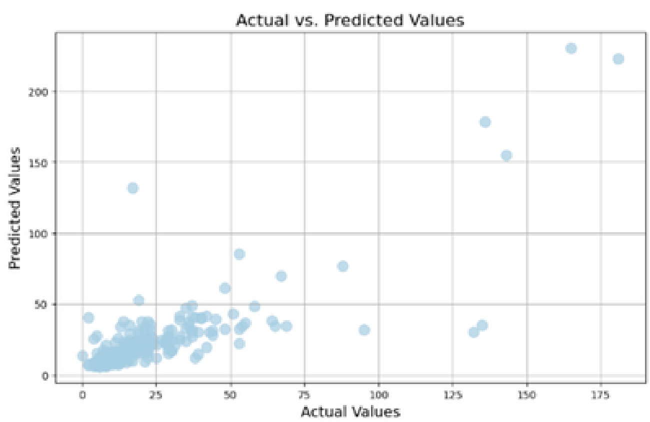
In conclusion, the K-Neighbors Regressor model outperformed other regression models in predicting accurate arrival times. Its lower MSE and MAE values testify to its superior predictive capability, making it the prime choice for Uber's ETA estimation.

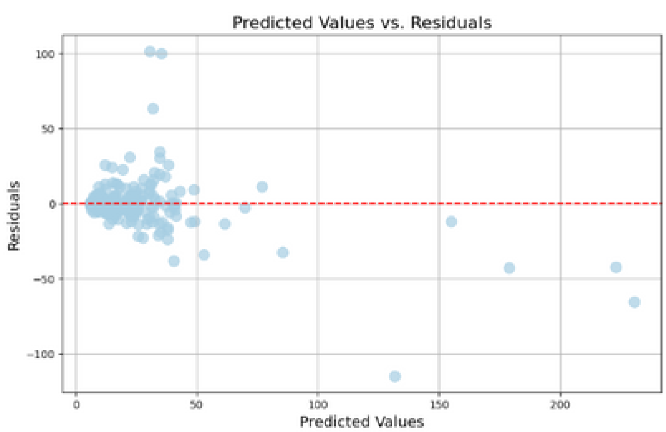
Thus the implementation of the model would enhance the reliability of ETA predictions, boosting user satisfaction and overall service efficiency of Uber.

K-Neighbors Regressor Analysis and Hyperparameters

After choosing the K-Neighbors Regressor as our model, we found that it performs optimally with 'metric' set to 'euclidean', 'n_neighbors' set to 9, and 'weights' set to 'distance'.

Additionally, the performance for the model was visualized and verified by plotting the following graphs:





After plotting predicited values vs actual values, and residual values vs predicited values we observed that the result is that the predicited values strongly aligns with the actual values.

Improving the Efficiency of Customer Support Process

An efficient customer service process holds paramount importance in the context of ride-hailing apps like Uber, directly impacting user satisfaction and overall platform success. A streamlined and responsive customer service mechanism ensures that users' queries, concerns, and issues are swiftly addressed, enhancing their overall experience.

According to McKinsey, "AI technologies could potentially deliver up to $1 trillion of additional value each year, of which revamped customer service accounts for a significant portion" (n.d., para. 1).

Timely support during rides, after trips, or while making bookings is crucial in this fast-paced transportation service. Moreover, the ability to identify and promptly resolve problems related to rides, payments, or account matters fosters trust and reliability among users.

Positive user experiences, in turn, contribute to customer retention and brand reputation. Automating certain aspects of customer service, such as using chatbots for initial responses, not only facilitates cost savings for Uber but also ensures scalability as the platform continues to grow.

The data collected through customer service interactions provides valuable insights that can drive data-driven improvements in Uber's services and user experience.

Offering round-the-clock availability and deploying automation provide a competitive advantage, differentiating Uber from its ride-hailing competitors. Ultimately, an efficient customer service process, bolstered by automation, is instrumental in ensuring exceptional user experiences, solidifying customer loyalty, and sustaining Uber's position in the dynamic ride-hailing industry.

To achieve this, we have automated the customer service process for Uber by leveraging BERT-based text classification and automated response generation. The goal is to efficiently categorize customer complaints and generate relevant automated responses tailored to each complaint category. By automating this process, Uber can handle a large volume of customer feedback more effectively, provide timely responses, and enhance overall user experience.

To achieve this, we first preprocess the customer complaints, truncating overly long reviews and converting the category labels into numerical format using label encoding. The complaints are then one-hot encoded to create binary vectors representing each category. The dataset is split into a training set and a test set to train and evaluate the BERT-based model.

.

We then initialized the BERT tokenizer and loads a pre-trained BERT model fine-tuned for sequence classification tasks. The model is then trained using the training data, optimizing the parameters to minimize categorical cross-entropy loss. During training, the model learns to recognize patterns and relationships in the text data to accurately predict the category labels for unseen complaints.

After training, we defined two functions: "predict_category" and "automated_response. predict_category" uses the trained BERT model to predict the category label of a given complaint with an accuracy of **0.9574**. The numerical label is then converted back into the original category label using the label encoder. automated_response takes a complaint as input, determines the category label using predict_category, and generates an automated response tailored to the predicted category. These automated responses can provide appropriate and consistent feedback to customers, addressing their concerns and inquiries promptly.

According to McKinsey, "AI technologies could potentially deliver up to $1 trillion of additional value each year, of which revamped customer service accounts for a significant portion" (n.d., para. 1).

Timely support during rides, after trips, or while making bookings is crucial in this fast-paced transportation service. Moreover, the ability to identify and promptly resolve problems related to rides, payments, or account matters fosters trust and reliability among users.

Positive user experiences, in turn, contribute to customer retention and brand reputation. Automating certain aspects of customer service, such as using chatbots for initial responses, not only facilitates cost savings for Uber but also ensures scalability as the platform continues to grow.

AHP to determine the most suitable city in Canada for launching a new feature in the Uber app

The Analytical Hierarchy Process (AHP) plays a pivotal role in the decision-making process of selecting the most suitable city in Canada for launching a new feature in the Uber app.

With numerous factors to consider, such as population density, traffic patterns, local regulations, and user demographics, AHP provides a structured and systematic approach to evaluating these complex criteria. Its ability to accommodate both subjective and objective factors empowers decision-makers to strike a balance between data-driven insights and expert judgment.

AHP's consistency checks validate the judgments made during the evaluation, ensuring the reliability and robustness of the decision-making process. Additionally, the transparency offered by AHP allows stakeholders to comprehend how different criteria are weighted, leading to greater buy-in and support for the chosen city among team members.

The sensitivity analysis capability of AHP provides valuable insights into the decision's robustness, offering a deeper understanding of critical factors influencing the outcome. By assessing how changes in criteria weights impact the final decision, Uber can adapt its strategies and be well-prepared for various scenarios.

Furthermore, the optimization of resources, such as marketing efforts, driver on boarding, and operational costs, becomes more effective when selecting the most suitable city. AHP's consideration of risk factors and uncertainties aids Uber in assessing potential challenges associated with each city and devising appropriate risk mitigation strategies.

As Uber continues to expand its services and features, AHP's scalability proves invaluable, allowing the evaluation of multiple cities simultaneously. In conclusion, the Analytical Hierarchy Process (AHP) enables Uber to make well-informed decisions, optimize resources, and enhance the success of their new feature rollout in the chosen city by providing a comprehensive and data-driven decision-making framework.

In this analysis, we utilized the Analytic Hierarchy Process (AHP) to evaluate and select the most suitable city for launching a new feature in Uber app among Toronto, Vancouver, Montreal, Ottawa, and Calgary. The main criteria considered for assessment were Population, Income, and Transportation, each encompassing several sub-criteria to ensure a comprehensive evaluation. The main criteria pairwise comparison matrix provided me with the relative importance of these factors, and we derived their corresponding weights using the AHP function.

After conducting the evaluation, we identified Toronto as the most favourable city for expansion, achieving the highest overall score of 1,902,766.74.

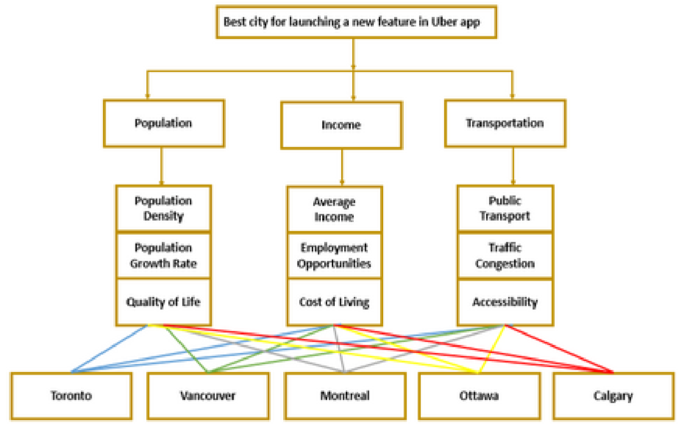
This result was a direct outcome of the AHP-based methodology, enabling an objective and transparent decision-making process.

Overall, the AHP-based evaluation approach offers a structured and strategic framework for selecting an optimal city for expansion.

After obtaining the main criteria weights, we proceeded to determine the sub-criteria weights within each main criterion category. By analysing sub-criteria pairwise comparison matrices, we calculated the average weights for each sub-criterion, ensuring their contributions to the final decision were accounted for.

To arrive at a comprehensive evaluation, we combined both the main criteria and sub-criteria into a unified set of all criteria. Utilizing this consolidated set, we computed the final weights for each criterion by considering their relevance to the overall expansion decision. This approach allowed for a systematic and unbiased comparison of all candidate cities.

The city data for Toronto, Vancouver, Montreal, Ottawa, and Calgary provided essential values for each criterion and sub-criterion, aiding in the computation of individual city scores. These scores were calculated by multiplying each criterion value with its corresponding weight and then summing the products to determine each city's overall performance based on the established criteria.



The computed city scores are as follows:
Toronto: 1,902,766.74
Vancouver: 452,098.04
Montreal: 1,149,482.39
Ottawa: 665,062.10
Calgary: 884,708.35

After conducting the evaluation, we identified Toronto as the most favorable city for expansion, achieving the highest overall score of 1,902,766.74. This result was a direct outcome of the AHP-based methodology, enabling an objective and transparent decision-making process.

Overall, the AHP-based evaluation approach offers a structured and strategic framework for selecting an optimal city for expansion.

This systematic methodology empowers decision-makers like us to make informed choices, considering multiple criteria and sub-criteria to enhance the likelihood of successful and sustainable expansion ventures.

Refrences

- McKinsey & Company. (n.d.). The next frontier of customer engagement: AI-enabled customer service. Retrieved from https://www.mckinsey.com/capabilities/operations/our-insights/the-next-frontier-of-customer-engagement-ai-enabled-customer-service

- Tomlin, C. J., Lygeros, J., & Sastry, S. (2006). A game theoretic approach to controller design for hybrid systems. In Proceedings of the 2006 American Control Conference (pp. 551-556). Retrieved from https://people.eecs.berkeley.edu/~tomlin/papers/conferences/rlt06_gnc.pdf

- ScienceDirect. (n.d.). The next frontier of customer engagement: AI-enabled customer service. Retrieved from https://www.sciencedirect.com/science/article/pii/S2772424722000257

- Uber. (n.d.). The next frontier of customer engagement: AI-enabled customer service. Retrieved from https://www.uber.com/en-CA/blog/deepeta-how-uber-predicts-arrival-times/

- PLOS ONE. (n.d.). The next frontier of customer engagement: AI-enabled customer service. Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0275030

- Qualtrics. (n.d.). Cluster analysis: Uncover patterns in your data. Retrieved from https://www.qualtrics.com/experience-management/research/cluster-analysis/

- National Center for Biotechnology Information (NCBI). (n.d.). A Study of Clustered Data and Approaches to Its Analysis. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6634702/

- Sage Journals. (n.d.). Data-Driven Method for the Prediction of Estimated Time of Arrival. Retrieved from https://journals.sagepub.com/doi/10.1177/03611981211033295?icid=int.sj-abstract.similar-articles.6