International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 02, 2024. ISSN: 2348-6600
Reference ID: IJCS-497 PAGE NO: 3433-3441

PREDICTION OF UBER RIDES DATA ANALYSIS USING PYTHON

Mr.Gurbhej Singh
Department of Computer Science Engineering,
Chandigarh University,
Kharar, Punjab, India.
Email ID: gurbheje7751@cumail.in

Ms.Urvashi Bhatia
Department of Computer Science Engineering,
Chandigarh University,
Kharar, Punjab, India.
Email ID: urvashi.e8824@cumail.in

Abstract

Python is the best language for understanding and delving into real-world issues. Guido van Rossum introduced Python, a potent high-level object-oriented programming language. I have first introduced the features and characteristics of Python programming in this work. The paper also explains why Python is said to be the programming language with the quickest rate of growth. This paper's main goal is to calculate and examine the hourly and time zone-based supply and demand gap for Uber taxi services. Following study, we made an effort to determine the most troublesome time zone and, more precisely, the most problematic hour during which the supply and demand disparity is at its widest. We have suggested various ways to close this gap in order to raise customer happiness and boost the business success of the organization, based on the analysis conducted for this article. Python has been utilized as a programming and analysis tool to simplify this analysis.

Keywords: UBER, Sales, Languages.

Introduction

We are going to introduce the characteristics of python in this project. Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. This language is widely used in the recent projects. This language helps us to increase code readability and its syntax is easy to understand and easy to express the code in fewer lines. The language constructs enables the user to write program on both small and large scale. The primary characteristic of

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 02, 2024. ISSN: 2348-6600
Reference ID: IJCS-497 PAGE NO: 3433-3441

Python is its ability to support several programming paradigms, such as procedural, imperative, and functional programming, in addition to object-oriented programming. After a project is planned, resources are needed for the planning to be carried out (demand), and you anticipate having those resources accessible to satisfy those requests (supply).The study of the differences between these two is referred to as a supply and demand gap. We can take measures to close those gaps after analysis. In this paper, we have taken a case study of UBER cabs. We are going to identify the main cause of the problem (i.e. cancellation and non-availability of cars) and then analyse the data and observation to give the solution for these problems. After doing our study, we are able to explain to the customer the problem(s), their potential causes, and suggested solutions.

I. About Python Programming

Python is a high-level, interpreted, dynamic programming language for broad purposes that is built on object orientation. It doesn't repeat itself and is primarily focused on developing applications quickly. It is designed to focus on the readability of code and in this programmer can express the code in fewer lines. Python is used for both small and large scale development. For a project it is very important to choose the write programming language. We have to consider all the things (like cost, security, user-friendly, platform independence and many more) in a project. To avoid all the above problems we can consider an established a platform like Python. With several commercial advantages, all such restrictions may be disregarded while utilizing Python. Furthermore, because Python interacts with other languages so simply, development in Python spares us from future upgrades. According to W3Techs, Python is used by 1.4% of all the websites whose server-side programming language we know in December, 2023. Some of the top tech companies using Python include Google, Pinterest, Instagram, Spotify, and more. Python developers love the language because of its readability and simple syntax (www.botreetechnologies.com).

According to SlashData, there are now 8.2 million developers in the world who code using Python.

Fig. 1 A Sample bar graph using colors, which contrast well most used languages in project year wise

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 02, 2024. ISSN: 2348-6600
Reference ID: IJCS-497 PAGE NO: 3433-3441

A. HOW PYTHON ROCKS FOR RESEARCH AND ANALYSIS

Python's elegant and simple syntax makes it a great starting programming language for non-technical individuals who are not familiar with programming. The feature of this language is tractability with the help of this programmer can increase their skill set. Originally, only specialized data scientists would use this language, but more recently, the business analyst community has been very interested in the interactive Python computing environment for analysis and decision-making. Instant feedback and visualization options provided by python have made the work easy for many analysts to become skilled Python programmers while doing valuable work.

Python's fine-grained gain capabilities creates new opportunities. While many business analysts find considerable value in interactive analytic methodologies, the ability to construct fully functional, independent programs is as valuable. Sometimes, this capability enables business analysts to tackle "Big Data" topics that would seem to be the exclusive domain of highly skilled data scientists. Perhaps even more significant than this enhanced degree of independence is the ability for analysts to interact with these stakeholders more successfully because to their improved proficiency in data analysis and management. Business analysts may become even more effective mediators by gaining a deeper understanding of these specialist fields via acquiring a programming language that enables them to start making independent advances into such areas.

One specific strength of Python code is how easily high-level programming and low-level optimization can be balanced. According to what I've heard, when implementing comparable algorithms, you should write six lines of C/C++ code for every line of Python. Python code is supposed to be as high level as reasonable. But much like with other high-level languages, programming speed frequently comes at the expense of code speed.

B. SUPPLY AND DEMAND GAP

The relationship between supply and demand. There are two terms involved in the gap: supply and demand. While demand indicates the amount of a specific item or product that the consumer has requested, supply indicates how the item or product is available to the consumer. The idea behind gap is that the amount that is available determines how much of a certain item is in demand. As an illustration of the "thing," let's examine the flight tickets. The supply is the quantity of seats available on a flight. The demand is determined by the quantity of customers that desired to purchase a plane ticket. The amount that a customer is willing to pay for a flight ticket is determined by the market price of that ticket. Thus, when there is a demand for a flight

The demand for airline tickets soars during festivals and other special occasions, such as New Year's, that the supply is unable

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 02, 2024. ISSN: 2348-6600
Reference ID: IJCS-497 PAGE NO: 3433-3441

to keep up. Price, then, is the first indicator of that unequal connection.

Supply and demand is a model used in data science to determine market prices. It was suggested that, if the market remained intact, the cost per unit of a given good would fluctuate until it reaches a balance, at which time the amount supplied at the current price and the quantity sought at the current price would be equal. Economic equilibrium is the state of being in balance. Six attributes from the Uber dataset (masked dataset) were used in our research. These six characteristics are linked to every request a client makes and are:(i) Request ID: A special code for the request, (ii) Time of request: The moment the customer requested the travel,(iii) Cutoff time: The time and date of drop-off, should the travel be finished, The point at which the request was made is known as the (iv) pick-up point, and (v) driver id The driver's unique identifying number and (vi) The request's status: The trip's ultimate state, which could be either finished, canceled by the driver, or there weren't any cars available.

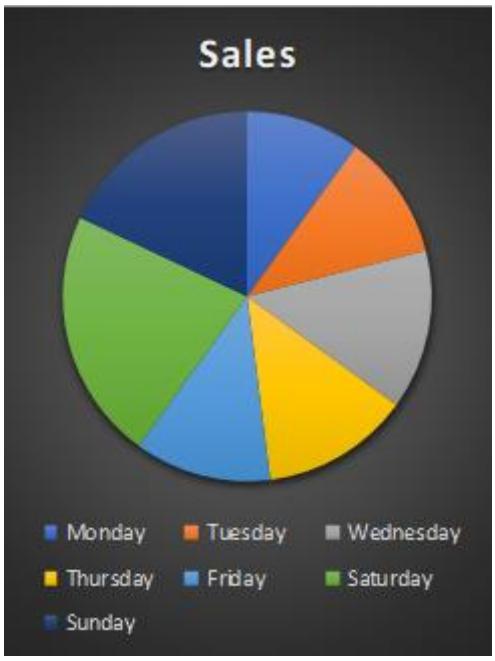

Fig. 2 A Sample pie chart using colors which contrast well supply and demand day wise.

C. DEMAND AND SUPPLY ANALYSIS OF UBER DATA

Uber, initially an IT company founded in 2009, expanded its services into cab transportation, offering convenient solutions for travelers, particularly those commuting to and from Airport located outside city centers. Many individuals have relied on Uber or similar cab services for Airport travel, a journey often fraught with challenges such as driver cancellations or car unavailability. These issues, commonplace among users, underscore the significance of user surplus in economic evaluations, influencing policies ranging from antitrust analysis to assessing the value of innovation. In 2014, David Sacks, CEO of Yammer and a notable figure in the tech industry, highlighted Uber's network

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 02, 2024. ISSN: 2348-6600
Reference ID: IJCS-497 PAGE NO: 3433-3441

effect, illustrating how the value of a product or service increases with its user base. Uber's surge pricing mechanism, reflecting the network effect, exemplifies its thriving ecosystem. Yet, user-facing challenges like driver cancellations and car shortages directly impact Uber's revenue streams. Such operational hurdles prompt a deeper examination of the underlying issues, especially focusing on driver cancellations, car availability, and the resultant surge pricing dynamics.

This paper employs a masked dataset akin to those managed by Uber's data analysts to delve into these challenges. The primary goal is to unravel the root causes behind driver cancellations and car shortages, offering strategic recommendations for improvement. Through meticulous analysis, the study aims to delineate the supply and demand gap across various time slots, categorized as early morning, morning, afternoon, evening, night, and late night. Furthermore, the analysis extends to discerning the supply and demand gap concerning pickup points, specifically delineated as Airport-to-city and city-to-Airport trips. By scrutinizing these parameters, the study seeks to provide actionable insights to mitigate the prevalent issues hampering Uber's operations and revenue potential.

Understanding the supply and demand dynamics at different times of the day and across distinct pickup points is pivotal. For instance, the early morning rush to catch flights might strain available resources, leading to increased cancellations and car unavailability. Conversely, late-night pickups might witness similar challenges due to reduced driver availability.

Moreover, the disparity in supply and demand between Airport-to-city and city-to-Airport trips warrants attention. Factors such as traffic patterns, Airport regulations, and commuter behavior influence these dynamics, necessitating tailored strategies to address the specific needs of each route.

In essence, this paper endeavors to offer a comprehensive analysis of Uber's operational challenges, emphasizing the critical interplay between supply and demand dynamics. By identifying key pain points and proposing strategic interventions, the study aims to enhance Uber's service reliability, customer satisfaction, and revenue sustainability in the competitive ride-sharing landscape.

II. Problem Solving and Analysis Methodology:

STEPS

1. Data Loading and Column Derivation:

Begin by loading the dataset and creating new columns based on specific requirements and existing data columns. For instance, utilize the 'Request timestamp' to derive columns such as Request time, request hour, and Request time_Zone.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 02, 2024. ISSN: 2348-6600
Reference ID: IJCS-497 PAGE NO: 3433-3441

2. Time Zone and Hourly Analysis:

Segment the requests (demand) and supply (completed / cancelled / no cars available) based on time zone and hourly intervals for better visualization and understanding of the problem.

3. Consideration of Pickup Point:

Incorporate the pickup point (Airport/City) as an additional factor during the analysis of the demand and supply gap. This helps in providing a more comprehensive view of the situation.

4. Data Analysis and Solution Proposals:

Analyze the derived data to identify the pickup points and time zones where the gap between supply and demand is most significant. Based on the analysis, propose solutions aimed at reducing this gap, thereby improving the efficiency and reliability of the service.

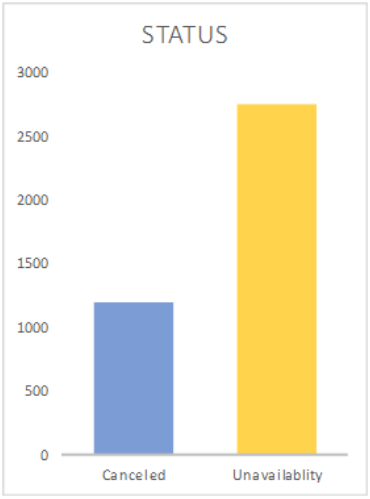

Fig. 3 A Sample bar show no of cancelled and unavailability of Cab

D. KEY OBSERVATION:

Mainly, the trends observed suggest that people predominantly travel from the city to the Airport until the afternoon, while from the Airport to the city, the traffic increases from evening to late night.

1) City to Airport:

In this direction, cab requests face challenges mainly due to the unavailability of cars, which is comparable to cancellations. Both cases show the 25th percentile (represented by the blue color).

The presence of outliers, indicated by dotted lines, signifies extreme values.

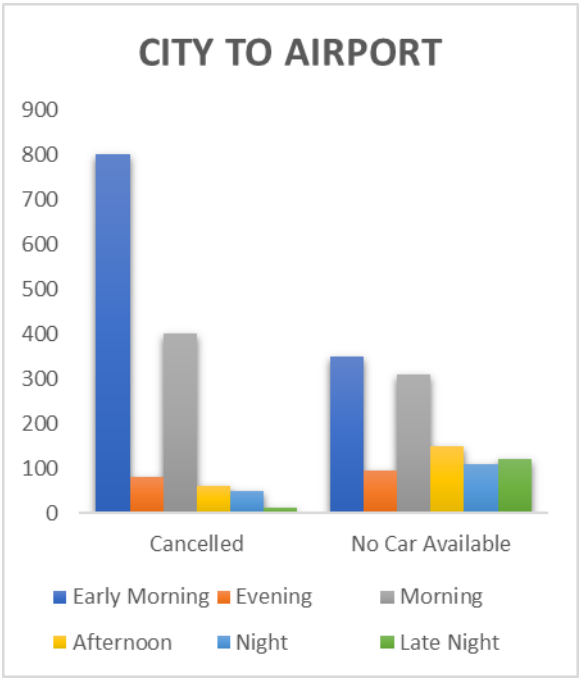

Fig. 4 A Sample bar show cancelled and no car available city to airport

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 02, 2024. ISSN: 2348-6600
Reference ID: IJCS-497 PAGE NO: 3433-3441

2) Airport to City:

Conversely, the cancellation of requests is more prevalent compared to unavailability, as indicated by the 75th percentile (represented by the orange color).

Similar to the other direction, outliers exist, pointing to exceptional cases.

In both scenarios, unavailability tends to surpass cancellation rates, particularly during the full daytime slot.

Further analysis delves into:

a) Supply-demand gap: Identifying disparities between the number of requests and available cars.
b) Time slot analysis: Determining which time slots exhibit the most pronounced gaps.
c) Hourly analysis: Identifying specific hours where the gap is most prominent.
d) Root cause analysis and solutions: Seeking to understand the reasons behind the observed trends and proposing solutions accordingly.

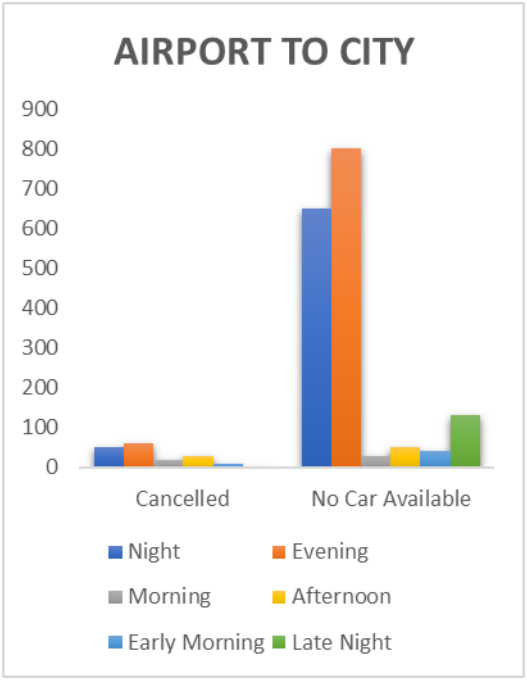

Fig. 5 A Sample bar show cancelled and no car available airport to city

E. OBSERVATIONS IN COUNT :

1. Number of Trips Cancelled: From Airport to city = approx. 200
2. From city to Airport= approx. 1000
3. Difference= approx. 800
4. Number of Cars Unavailable Available From Airport to city = approx. 1800
5. From city to Airport= approx. 900
6. Difference= approx. 900

Trip Cancelled is more from city to Airport Reason behind this might me waiting time for a driver is more because once driver will go to Airport he will get next passenger when next flight will arrive, so waiting time (which may be more) depends on Flight's

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 02, 2024. ISSN: 2348-6600
Reference ID: IJCS-497 PAGE NO: 3433-3441

arrival time. Another reason might be that number of cars available is more from Airport to city as campared to city to Airport. Reason behind this is because ones driver will get a passenger for city, happily they will come because next drive easily they will get from city/Airport. Overall if we see the figures, passengers are not getting cabs (either due to cancellation/ no cars available) from city to Airport. There must be some Problem when we are traveling from city to Airport. We have tried to list the possible problems and accordingly suggested some solutions to deal with those problems in a way to that we can reduce supply and demand gap, which was the objective of my paper. I have also tried to find the problematic time zone I.e hour during which number of requests are maximum. This is the only time when we can analyze the supply and demand gap correctly. Fig 4 shows the number of requests made for cab time zone basis and fig 5 shows the number of requests hourly basis. Each color represents one time zone. From Fig 6, it is clear that particularly at 18th hour (Evening 6 P.M) supply and demand gap is more. Reason behind this is may be because of more traffic in the evening, cabs are stuck in traffic. Morning and evening is office hours, so cabs will be busy in those trips so for/to Airport trips supply and demand gap will be more.

F. PROPOSED SOLUTION:

After looking at the data and what we've observed, I've got some ideas to make Uber better. If we follow these suggestions, we can make sure there are enough cabs for everyone when they need them. When the gap between the number of people wanting rides and the number of available cabs is smaller, Uber will do even better, and the company will make more money.

One way to do this is by encouraging people to share rides. We could offer discounts or gift vouchers to those who choose to share a cab. Also, during busy times, we could increase the price a bit. This might make some people wait to book until a less busy time or choose to share a ride.

To make things even better, we could give extra benefits, like discounts, to those who book two hours before or after the busy times. We could call this period "happy hours" and make it a special time with extra perks.

To deal with cancellations, Uber could make the penalty for canceling a ride higher. This might make people think twice before canceling.

For drivers going from the city to the Airport during busy times, we could give them more money. This could make drivers more interested in accepting rides during these times and less likely to cancel.

These are just some ideas, but if we try them out, we might see some positive changes in how Uber works!

III. Conclusion

Python is experiencing remarkable growth as a programming language, primarily driven by a significant increase in its application in big data, data science, and machine learning. A recent analysis conducted by Stack Overflow, the largest online

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 12, Issue 1, No 02, 2024. ISSN: 2348-6600
Reference ID: IJCS-497 PAGE NO: 3433-3441

developer community, has established this connection. Utilizing Python, I conducted an analysis on a masked Uber dataset. Upon examining graphs and observations, it is evident that there is a notable supply-demand gap, particularly during the evening around 6 P.M., from the city to the Airport. To enhance performance and address this gap, various strategies such as promoting shared rides and introducing incentives during "happy hours" can be implemented Fig. 1 shows an example of a low-resolution image, which would not be acceptable, whereas Fig. 2 shows an example of an image with adequate resolution. Check that the resolution is adequate to reveal the important detail in the figure.

Acknowledgment

1. "Programming Language Trends - O'Reilly Radar". Radar.oreilly.com. 2 August 2021.
2. https://w3techs.com/technologies/details/pl-python/all/all.
3. https://ieeexplore.ieee.org/document/9432347.
4. https://www.analyticsvidhya.com/blog/2021/10/end-to-end-predictive-analysis-on-ubers-data/
5. Summerfield, Mark. Rapid GUI Programming with Python and Qt.
6. Williamson, Oliver. 1968. Economies as Antitrust. The Welfare Trade-Offs. American Economic Review 58(1): 18-36.

References

1) L.K. Poulsen, D. Dekkers, N. Wagenaar, W. Snijders, B. Lewinsky, R.R. Mukkamala, et al., "Green Cabs vs. Uber in New York City", *2016 IEEEInternational Congress on Big Data (BigData Congress)*, pp. 222-229, June 2016.
2) S.S. Faghih, A. Safikhani, B. Moghimi and C. Kamga, *Predicting Short-Term UberDemand Using Spatio-Temporal Modeling: A New York City Case Study*, 2017, [online] Available: .
3) S. Guha and N. Mishra, "Clustering data streams" in Data stream management, Berlin, Heidelberg:Springer, pp. 169-187, 2016.
4) M. Ahmed, E.B. Johnson and B.C. Kim, *The Impact of Uber and Lyft on TaxiService Quality Evidence from New York City*, 2018.
5) S. Wallsten, The competitive effects of the sharing economy: how is Uber changingtaxis, Technology Policy Institute, vol. 22, pp. 1-21, 2015.
6) D.N. Sotiropoulos, D.E. Pournarakis and G.M. Giaglis, "A genetic algorithmapproach for topic clustering: A centroid-based encoding scheme", *2016 7th International Conference on Information Intelligence Systems Applications (IISA)*, pp. 1-8, July 2016.