UBER SUPPLY AND DEMAND GAP ANALYSIS USING PYTHON

Dr. Divya Kashyap

Assistant Professor

Indian Academy Degree College Autonomous, Bangalore, India

Abstract: Python is a most suitable language for both learning and analyzing real world problem. Python is a powerful high level, object-oriented programming language introduced by Guido van Rossum. In this paper I have first introduce the python programming characteristics and features. Paper also provides information about the reasons behind python being credited as the fastest growing programming language. The general objective of this paper is to estimate and analyze the supply and demand gap of Uber cabs services hourly and time zone basis. After analysis we have tried to get the most problematic time zone and in that specifically the most problematic hour where supply and demand gap is maximum. By using the analysis done in this paper, we have proposed some solutions to minimize this gap to improve the customer's satisfaction and company's business performance. To make this analysis easy we have used python as a programming and analysis tool.

I. INTRODUCTION:

In this paper, we are going to introduce the characteristics of Python. Python is a general-purpose, high-level programming language which is widely used in the recent times (TIOBE Software Index ,2011) (Radar.oreilly.com,2006)(Tecosystems: A report,2011). Its design philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than would be possible in languages such as C (Summerfield, Mark. Rapid GUI Programming with Python and Qt.). The language constructs enable the user to write clear programs on both a small and large scale (Kuhlman, Dec 15,2013). The most important feature in Python being it supports multiple programming paradigms, including object-oriented, imperative and functional programming or procedural styles. Python supports a dynamic type system and automatic memory management and has a large and comprehensive standard library.Once planning for a project is done then resources for that planning is required (demand) to get executed required (demand) and the resources that you expect to be available to meet those demands (supply). Analysis of difference between these two is known as supply and demand gap analysis. After analysis we can take action to minimize those gaps.

In this Paper we have taken a case study of UBER cabs. I will identify the root cause of the problem (i.e. cancellation and non- availability of cars) and then analyze the data and observations to give the solution for the same. As a result of our analysis, we should be able to present to the client the root cause(s) and possible hypotheses of the problem(s) and recommend ways to improve them.

II. ABOUT PYTHON PROGRAMMING

Python is a high-level Object oriented based, interpreted and general purpose dynamic programming language which doesn't repeat itself and mainly focuses on rapid application development. Its design philosophy focuses on the readability of the code and programmers can express concepts in fewer lines of code that would be possible. The language provides constructs designed to allow clear programs both small and large scale development.

Suppose our company is a startup and planning for product development, then selecting the right programming language is very important and difficult. With a tight fixed budget, we need to consider about numerous things like security, platform independence, user-friendliness, HR cost and many more. To avoid all these we can consider an established platform like Python. While using python all such constraints can be ignored with number of business benefits. Moreover, Development in python gives us the freedom from upgradations in future as python can interact easily with other languages.

According to W3Techs, Python is used by 0.2% of all the websites whose server-side programming language we know. Ubuntu.com, Stackoverflow.com, Digg.com, Wish.com etc. are some of the popular sites which
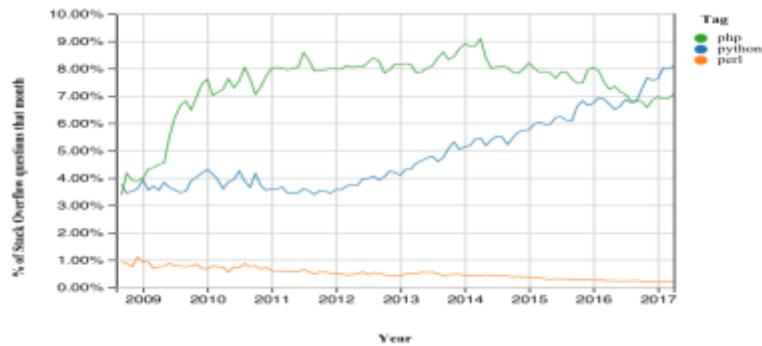

Fig 1: Yearly % of stack overflow question that month

use Python(https://w3techs.com/technologies/details/pl-python/all/all). In fact, hug and overloaded websites like Facebook and Instagram have some major dependencies on Python.

As per BuiltWith(https://trends.builtwith.com/framework/Python), approximately 974,112 websites are using Python currently and it is considered to be one of the top 10 programming languages even after 29 years of its inception.

III. WHY PYTHON ROCKS FOR RESEARCH AND ANALYSIS?

Simplicity and elegance of syntax of python, makes this programming language as an excellent first programming language for non-technical people those who don't know programming. Tractability is another feature of this language, with which programmer can grow their skill set. Before this language was common among specialized data scientists only, but later the interactive iPython computing environment has also attracted great attention within the business analyst's community 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.

Fine Gaining capability with of python opens up further possibilities. While interactive analytic techniques are of great use to many business analysts, being able to create fully functioning, independent programs is of similar value. This ability can sometime allow business analysts to address the so-called "Big Data" questions that can otherwise seem the sole province of specialized data scientists. More important than this higher level of independence, perhaps, is the fact that this increased facility with data analysis and handling allows analysts to communicate more effectively with such stakeholders. Through learning a programming language which allows them to begin making independent inroads into such areas, business analysts gain a better perspective on these specialized domains, and this allows them to function as even more effective intermediaries.

The ease of balancing high-level programming with low-level optimization is a particular strong point of Python code. Python code is meant to be as high level as reasonable — I've heard that in writing similar algorithms, on average you would write six lines of C/C++ code for every line of Python. However, as with most high-level languages, you often sacrifice code speed for programming speed.

IV. SUPPLY AND DEMAND GAP

Supply and demand Gap involves two terms: one is supply and another is demand. Supply implies how the item or product is available to consumer whereas demand reflects the quantity of that particular item or product requested by the consumer. Gap is the concept that the demand of something depends largely on how much of it is available. Let's look at the Flight tickets as an example of the "thing". The number of seats in a flight makes up the supply. The number of consumers who wanted to book the flight ticket makes up the demand. The market price of a flight ticket compromise of how much a consumer will pay to have that ticket. Therefore, when demand of a flight ticket skyrockets—says, during Festival or any special occasion like New Year etc — supply can't keep up. Accordingly, price is the first thing to reflect that unequal relationship.

In Data science, supply and demand is a model to calculate the price in market. It proposed that, consider everything is unchanged in market, the price of a unit for a particular product will vary until is balances at a point, where quantity requested at the current price and quantity provided at the same current price will be equal. This state is of equilibrium is known as economic equilibrium. During our analysis we have taken six attributes from the Uber dataset (masked dataset). These six attributes associated with each request made by a customer are:(i) Request id: A unique identifier of the request,(ii)Time of request: The date and time at which the customer made the trip request,(iii)Drop-off time: The drop-off date and time, in case the trip was completed, (iv)Pick-up point: The point from which the request was made and (v)Driver id: The unique identification number of the driver and (vi)Status of the request: The final status of the trip, that can be either completed, cancelled by the driver or no cars available.

V. DEMAND AND SUPPLY GAP ANALYSIS OF UBER DATA

Initially Uber , founded in 2009, was a IT company , later they jumped into cab services also.You may have some experience of traveling to and from the airport (considering usually airports are located far from the city). Have you ever used Uber or any other cab service for this travel? Did you at any time face the problem of cancellation by the driver or non-availability of cars? From many years analyst and economist are focusing on the importance of user surplus while making welfare calculations. User surplus is a critical input to many economic policies such as antitrust analysis, non-market goods's analysis and calculating the value of innovation (e.g. Williamson 1968,Willing 1976, Bresnahan 1986).Back in 2014, David Sacks, CEO and founder of Yammer, former CEO of Zenefits and current member of the PayPal Mafia, tweeted his napkin sketch of Uber's network effect(Jasmine Lee, 2017):A network effect is the phenomenon in which a product or service gains more value the more people use it. A network effect is only as good as the levels of participations from both sides of the supply chain: the product or service and the people who use that product or service. And if Uber's surge pricing system is any indication, the network effect is alive and thriving.

Well, if these are the problems faced by customers, these issues also impact the business of Uber. If drivers cancel the request of riders or if cars are unavailable, Uber loses out on its revenue. Let's consider more about such problems that Uber faces during its operations. In this paper we have taken masked data set which is similar to what data analysts at Uber handle, to address the problem Uber is facing - driver cancellation and non-availability of cars leading to loss of potential revenue and surge pricing. In this paper we will try to identify the root cause of the problem (i.e. cancellation and non-availability of cars) and recommend ways to improve the situation. Identify and analysis of the Supply and Demand Gap time slots basis and later hourly basis. Time slots: Early morning, Morning, Afternoon, Evening, Night, Late Night. We will also try to identity the Supply and Demand Gap of cabs based on pickup points and time slots, where pickup points:

i.   Airport (to city)
ii.   City (to airport)

VI. PROBLEM SOLVING AND ANALYSIS METHODOLOGY: STEPS

1. Load the data and drive the new column based on requirement and existing data column.
 Ex: Based on 'Request timestamp', Request time, request hour and Request time_Zone are derived.
2. Divided the request (demand) and supply(completed/cancelled/ no cars available) time_zone basis and hourly basis to visualize the problem.
3. One more factor i.e. Pickup point (Airport/City) also considered during analysis of demand and supply gap.
4. Analyze the derived data to get the Pickup point and time zone where gap between supply and demand is maximum and suggested some solutions to reduce this gap.

V. KEY OBSERVATIONS:

Mostly people are going from city to airport till afternoon and from airport to city from evening to late night.
City to Airport: Cab requests are not satisfied mostly because of unavailability of cars as compared to cancellation. Both the cases (cancellation or unavailability) it is showing 25th percentile. (Blue color).
Airport to City: Cancellation of request is more as compared to unavailability. Both the cases (cancellation or unavailability) it is showing 75th percentile.(Orange color). In both the plots, some values are extreme high as we can see with dotted line showing outliers.
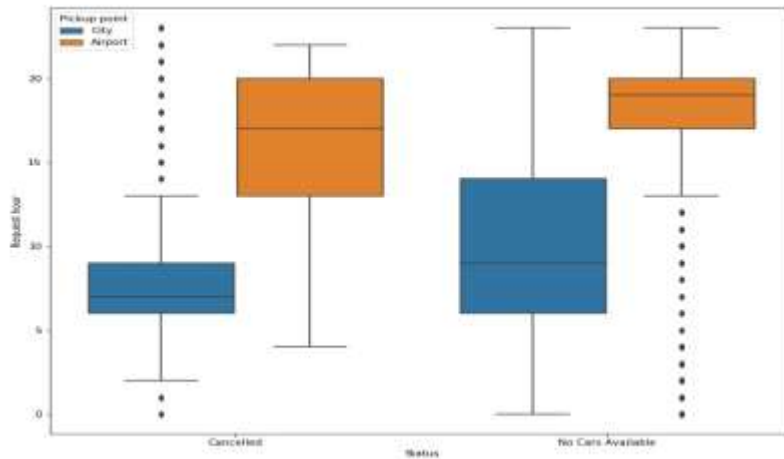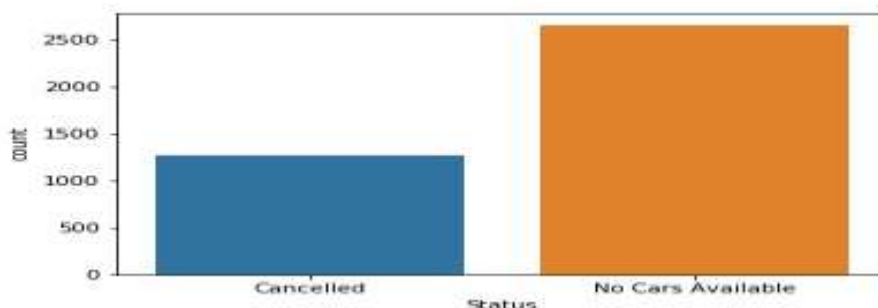

Fig 2: Cancellation or Unavailability of Cab hourly basis

We are considering slot as full day, for both pickup points (Airport/City) unavailability is more as compared to cancellation.

Next we have tried to analyze the data more deeply to the find:

a. Supply demand gap

b. Which time slot gap is more?

c. Particularly which hour it is most

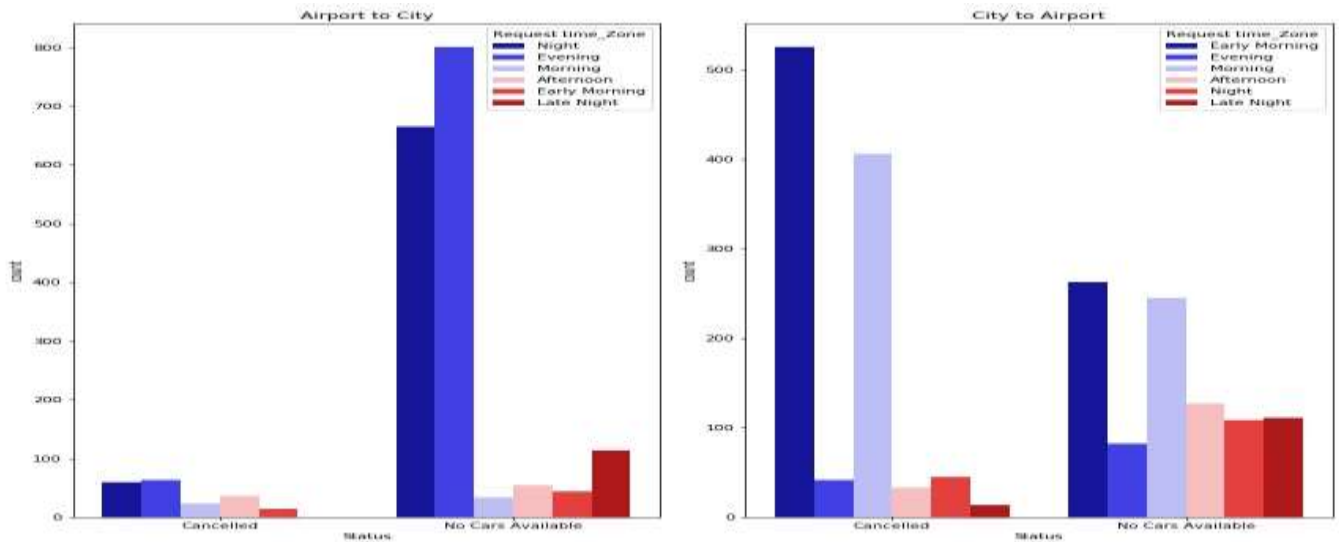d. Try to find the reason and then propose some solution for the same.

Fig 4:
Count of Cancellation and Unavailability based on Time Slots

OBSERVATIONS:
(I)FROM AIRPORT TO CITY

Overall Unavailability is more as compared to cancellation but if we see closely, during night and evening, difference between two is huge.

Reason:

i. May be number of flights are more during this duration.

ii. Usually people will not prefer to travel during night hour so the number of cabs reaching to airport would be less. Hence affecting availability.

Proposed Solution:

i. Provide shared trip option(if not available)

ii. (if available)Introduce extra offers for shared trips

NOTE: late night hours usually drivers want to finish their trip and want to go home so they will accept trip for any location without cancellation.

(II)FROM CITY TO AIRPORT

Overall Cancellation is more as compared to unavailability but if we see closely during morning and early morning, difference between two is huge.

Reason:

 i. The number of passenger available from airport to city would be less during this slot hence waiting time for the cab driver after reaching to airport would be more for airport to city trip. Hence they will not prefer to go airport during this duration and cancel the trip.

ii. For morning cancellation, city traffic(office hours) seems to be one more factor.

OBSERVATIONS IN COUNT

1. Number of Trips Cancelled:
    From airport to city = approx. 200
    From city to airport= approx. 1000
    Difference= approx. 800
2. Number of Cars Unavalible Available
    From airport to city = approx. 1800
    From city to airport= approx. 900
    Difference= approx. 900

a. 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 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 5 shows the number of requests made for cab time zone basis and fig 6 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.
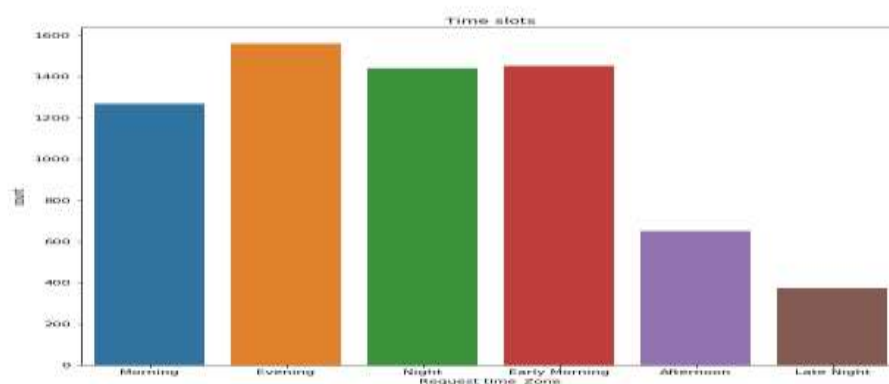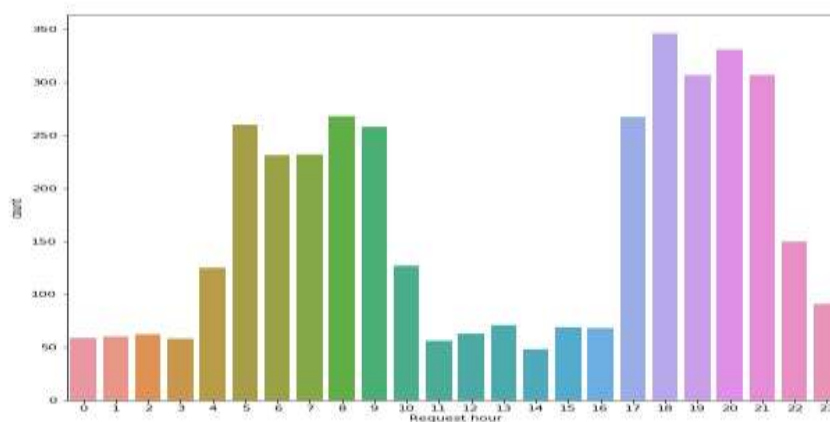

Fig 5: Number of requests time zone basis

Fig 6: Number of requests hourly basis

VI. PROPOSED SOLUTION:

After analyzing the Figures and observations, I am proposing some solutions for the observed problems. By considering these proposed solutions we can reduce the supply and demand graph of uber cabs. If supply and demand gap is reduced, it will surely have a direct positive impact on the performance of the company and company's profit. We can give some extra benefits are sharing cabs like some discount or gift vouchers etc. By increasing the price during these problematic time slots we can reduce the number of individual request so will prefer to book either in sharing or during other slots. By giving some extra benefits (concession) for booking 2 hours before and after these time slots also we can improve the performance. We can mention this 2 hours before and after duration as a happy hours with extra benefits.

To handle the cancellation condition Uber can increase the penalty amount for cancellation. By Increase the profit share of the driver for city to airport trip during this period, will increase the interest among drivers to not to cancel the request.

CONCLUSION

Python is the fastest-growing programming language is being powered by a sharp uptick in its use for big data, data science and Machine learning. The link has been established by a new analysis by Stack Overflow, the question and answer hub that is home to the world's largest online developer community. By using python, I have tried to analyze the Uber data set which is masked. After analyzing graphs and observations I am concluding that evening time around 6 P.M from city to airport, Supply-Demand Gap is more. We can reduce this gap and improve the performance in many ways like sharing cabs, happy hours etc.

REFERENCES:

1. Williamson, Oliver. 1968. Economies as Antitrust Defense: The Welfare Trade-Offs. American Economic Review 58(1): 18-36.

2. Jasmine Lee. April 28. 2017. How Uber leverages Supply and demand in their pricing model.

3. https://w3techs.com/technologies/details/pl-python/all/all

4. https://trends.builtwith.com/framework/Python

5. TIOBE Software Index . 2011. TIOBE Programming Community Index Python.

6. "Programming Language Trends - O'Reilly Radar". Radar.oreilly.com. 2 August 2006.

7. "The RedMonk Programming Language Rankings: January 2011 – tecosystems". Redmonk.com.

8. Summerfield, Mark. Rapid GUI Programming with Python and Qt.

9. Kuhlman, Dave. Dec 15, 2013. A Python Book: Beginning Python, Advanced Python, and Python Exercises.

10. Willig, Robert D. 1976. "Consumer's Surplus Without Apology." American Economic Review. 66(4): 589-597

11. Bresnahan, Timothy, 1986, "Measuring the Spillovers from Technical Advance: Mainframe Computers in Financial Services." American Economic Review 76(4): 742-755.

12. Jasmine Lee , April 28, 2017, "How Uber Leverages Supply and Demand in Their Pricing Model".