Uber Data Analysis Using Python

Ragini Pal[1], Ankita Mondal[2], Pratik Parab[3], Prachi Gurav[4]
[1, 2] Dept of Computer Engineering
[1, 2] St. John College of Engineering and Management, Palghar

Abstract- *Data Analytics has helped organizations enhance and develop their exhibition for quite a long time. Information examination and perception has supported us with a few advantages, not many of them being recognizing arising patterns, concentrating on connections and examples in information, investigation top to bottom and clincher are the experiences we draw from these examples. This task is tied in with understanding one such informational collection of uber from New York City and is very part to comprehend the utilization of information investigation and perception. It is created with the assistance of ' machine inclining calculations' and 'Python' programming language utilizing libraries like pandas, numpy, matplotlib. pyplot. Through projects like this, we can acquire information on different complex tasks acted in information representation. It will empower us to perceive the examples in information of this immense association and gives basic bits of knowledge of undiscovered data. Additionally guide us in understanding the activities of matplotlib. pyplot.*

Keywords- Uber, Data analytics, Data visualization, Python programming, Machine Learning Algorithms, pandas, numpy, matplotlib.pyplot.

I. INTRODUCTION

Uber has arisen as driving organization in the arrangement of new transportation choices inside the contemporary world. Uber, then, is fundamentally occupied with systems administration, and all the organization's arising activities can be conceptualized as far as basically giving a medium through which the pertinent stockpile can get together with the significant interest. Investigation is an enormously developing specialty that individuals apply in their organizations to give it a lift. This is all the more an information representation project that will upgrade our insight towards involving the python library for getting the information and for fostering an instinct for understanding the clients who benefit the excursions. In this task, we are investigating Uber's Data to foresee the admission/cost of specific taxi. At the same time examining other potential results like the reason for Uber, normal distance gone by a Uber, and so on.

II. LITERATURESURVEY

A writing review in a task report addresses the review done to aid the fruition of an undertaking. A writing study likewise depicts an overview of the past existing material on a subject of the report. A writing survey is an insightful paper, which incorporates the current information including meaningful discoveries along with hypothetical and strategic commitments to a specific point.

1. Uber Data Analysis using GGPLOT
Authors: Mrunal Patil, Vidya Kumari, Adarsh Patil, Laxshmikant Ahire and Umakant Mandawkar.

How data is used for visualization and how time and place affect customer trip is explained. Use of ggplot2 package lead to better data visualization of Uber data but parameter such as rate was not predicted.[1].

2. Uber Related Data Analysis using Machine Learning
Authors: Rishi Srinivas, B.Ankayarkanni and R.Sathya Bama Krishna.

This research shows basic outline of trips travelled with respect to longitude and latitude. It suggests that datasets used focus on outskirt of place which provides less information related local area travels and their analysis.[2].

3.RealTimeUberDataAnalysisofPopularUberLocationinKubernatesEnvironment
Authors: T.M Gunswardena, k. P.NJayasena.

Real time data analysis was done with and without kubernates environment and performance handling of CPU was also shown. This research provide information on real time uber data with less precision and CPU performance is shwn which is useful for effecient an fast processing.[3].

4.Preliminary Exploration of Uber Data as an Indicator of Urban Liveability.
Authors: Aguinaldo Bezerra Gisliany Alves, Ivanovitch Silva, Pierangelo Rosati, Patricia Takako Endo ans Theo Lynn.

This paper represented a study using Uber ETA data as simple, low cost and international urban liveability indicator. Solution that aims to overcome the issue of waiting time is shown but apparently it fails to give information such as purpose of Uber, travelling time, predicting rates, etc.[4].

III. PROPOSED METHOD

This study aims to analyze Uber datasets to predict the price of an Uber, at the same time analyzing other possible outcomes such as purpose of Uber, average distance travelled,etc. and calculating CPU speed and efficiency for the modules and algorithm used.

A. System Architecture

The system architecture for the given module is as follows:

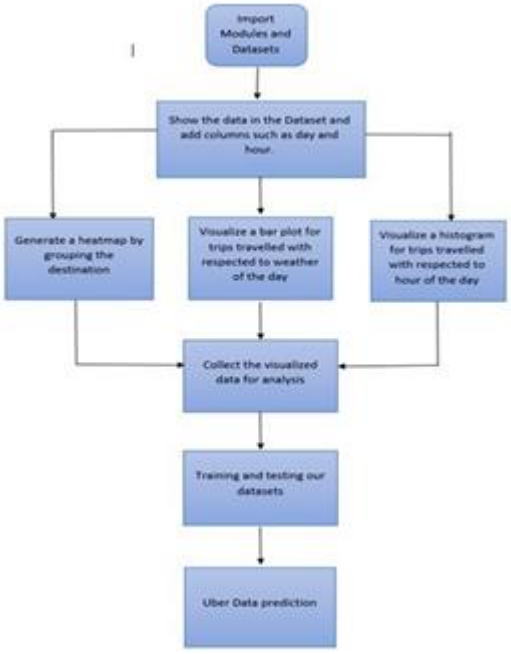
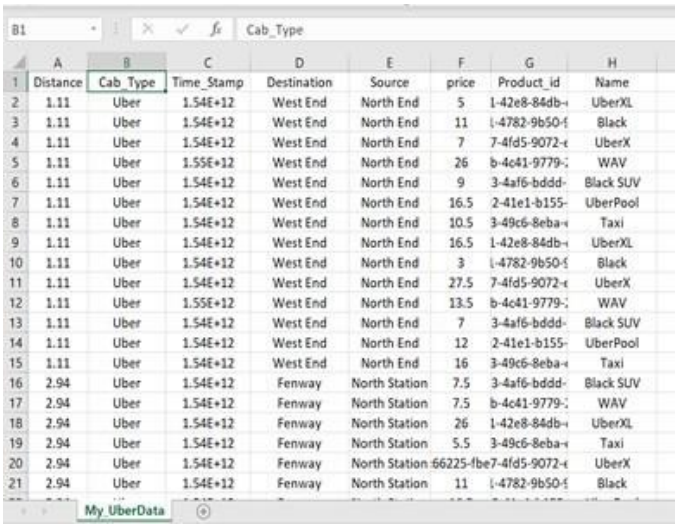


Fig 1. System Architecture

B. Raw Data(Dataset)

Crude information is the information that is gathered from a source, however in its underlying state. It has not yet been handled - or cleaned, coordinated, and outwardly introduced. Crude information can be physically recorded or composed, recorded, or consequently input by a machine. You can track down crude information in an assortment of spots, including data sets, records, accounting sheets, and, surprisingly, on source gadgets, like a camera. Crude information is only one kind of information with possible energy.



	A	B	C	D	E	F	G	H
1	Distance	Cab_Type	Time_Stamp	Destination	Source	price	Product_id	Name
2	1.11	Uber	1.54E+12	West End	North End	5	1-42e8-84db-	UberXL
3	1.11	Uber	1.54E+12	West End	North End	11	1-4782-9b50-1	Black
4	1.11	Uber	1.54E+12	West End	North End	7	7-4fd5-9072-e	UberX
5	1.11	Uber	1.55E+12	West End	North End	26	b-4c41-9779-3	WAV
6	1.11	Uber	1.54E+12	West End	North End	9	3-4af6-bddd-	Black SUV
7	1.11	Uber	1.54E+12	West End	North End	16.5	2-41e1-b155-	UberPool
8	1.11	Uber	1.54E+12	West End	North End	10.5	3-49c6-8eba-e	Taxi
9	1.11	Uber	1.54E+12	West End	North End	16.5	1-42e8-84db-e	UberXL
10	1.11	Uber	1.54E+12	West End	North End	3	1-4782-9b50-1	Black
11	1.11	Uber	1.54E+12	West End	North End	27.5	7-4fd5-9072-e	UberX
12	1.11	Uber	1.55E+12	West End	North End	13.5	b-4c41-9779-3	WAV
13	1.11	Uber	1.54E+12	West End	North End	7	3-4af6-bddd-	Black SUV
14	1.11	Uber	1.54E+12	West End	North End	12	2-41e1-b155-	UberPool
15	1.11	Uber	1.54E+12	West End	North End	16	3-49c6-8eba-e	Taxi
16	2.94	Uber	1.54E+12	Fenway	North Station	7.5	3-4af6-bddd-	Black SUV
17	2.94	Uber	1.54E+12	Fenway	North Station	7.5	b-4c41-9779-3	WAV
18	2.94	Uber	1.54E+12	Fenway	North Station	26	1-42e8-84db-e	UberXL
19	2.94	Uber	1.54E+12	Fenway	North Station	5.5	3-49c6-8eba-e	Taxi
20	2.94	Uber	1.54E+12	Fenway	North Station	66225-fbe7	4fd5-9072-e	UberX
21	2.94	Uber	1.54E+12	Fenway	North Station	11	1-4782-9b50-1	Black

Fig 2. Raw Dataset

C. Data Importing

A huge amount of trip data will be collected from uber for training and testing data.

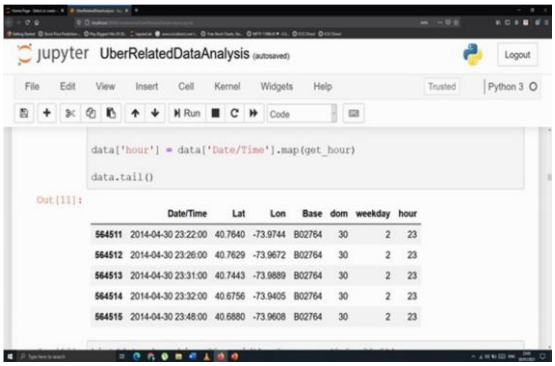


Fig 3. Processing the dataset

D. Data Visualization

Information perception is the portrayal of information through utilization of normal designs, like graphs, plots, infographics, and even activitys. These visual presentations of data impart complex information connections and information driven bits of knowledge in a manner that is straightforward.

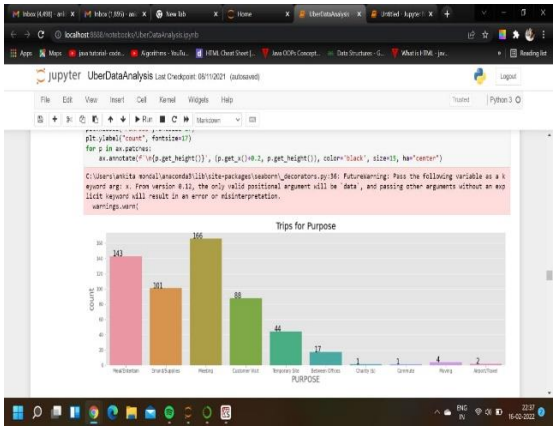


Fig 4. Purpose of Trip

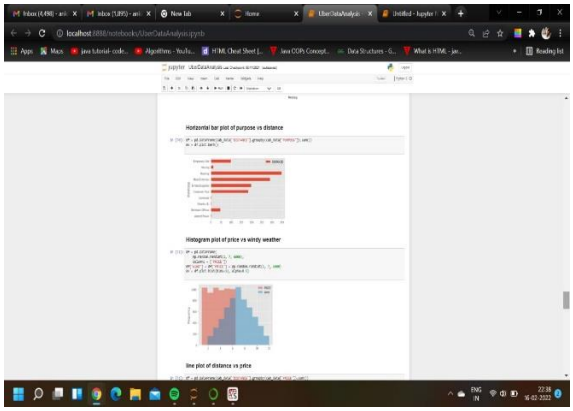


Fig 6. Histogram for Prices Vs Weather

E. Algorithm

Linear regression: It is a machine learning algorithm based on supervised learning. It is mostly used for finding out the relationship between variables and forecasting.

Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable.

*y = a_0 + a_1 * x → Linear Equation*

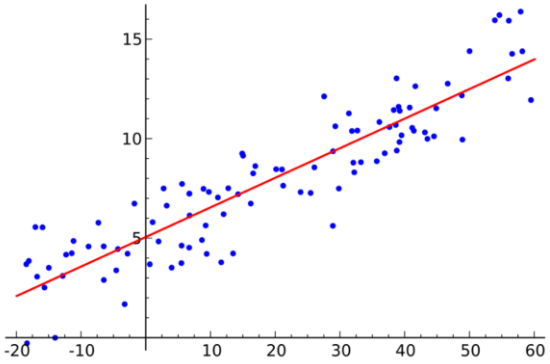


Fig 7. Linear Regression

F. Working

1. Importing libraries such as pandas, numpy, matplotlib.pyplot.
2. Importing Module named datetime to convert time_stamp to date and time.
3. Merged both our Uber dataset and weather dataset using .concat() function.
4. Adding two attributes of day and hour.
5. Going forward we used pandas fillna method to fill NAN values with fillna (0,inplace=true).
6. Using the groupby() function we grouped our destination and generated a heatmap.
7. Then we have plotted bar graph and histogram of different attributes.
8. With the help of graphs and histogram possible predictions were made.

After successful prediction, we have linked it with our GUI which will take distance and weather input and show the respected rate prediction.

IV. RESULTS AND DISCUSSION

This examination predicts the pace of Uber as indicated by the climate or condition which will assist the client with picking carefully between nearby transportation and Uber administrations for their everyday use. We can likewise anticipate the motivation behind most Uber taken over the course of the day and number of distance voyaged which will furnish us with the data that which day is more involved or occupied.

V. CONCLUSION AND FUTURE WORK

Conclusion:

- Our framework will give better and productive answer for current circumstance of recruiting a Uber while saving the costs of individuals who are everyday client.
- Isolation of information and investigation is finished involving python module to give exactness in dataset utilized.
- The value of this venture is that we find out about the working how Uber taxis are allocated and how are they utilized for traveler in view of solo calculation and furthermore make sense of the critical ideas of AI.

Future Scope:

The application can be additionally stretched out to areas like digital forms of money, share market, Telecom, Healthcare, E-trade and public area as information science will before long cover other space and will furnish us with legitimate business experiences and secret patterns of organizations

REFERENCES

[1] https://www.uber.com/us/en/careers/teams/datascience/
[2] https://iedu.us/tag/project-in-r-uber-dataanalysis-project/
[3] https://growvation.com/paritoshsankhla/project/uber-data-analysis/5e95ee80-9455- 4473-acaf-b670fe2abc8b
[4] F. Kooti, M. Grbovic, L. M. Aiello, N. Djuric, V. Radosavlejevic, and K. Lerman, "Analyzing uber's ride-sharing economy," in 26th Conference on World Wide Web Companion, no. January, 2017, pp. 574–582.
[5] Uber Technologies Inc., "Introduction to the api," 2018, https://developer.uber.com/docs/riders/ride-requests/tutorials/api/introduction, Last accessed on 2018-12-01.
[6] J. W. Tukey, Exploratory data analysis. Reading, Mass., 1977, vol. 2.