A MINI PROJECT REPORT ON

# “UBER DATA ANALYSIS”

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF

BACHELOR OF COMPUTER ENGINEERING

BY

### GAURAV GEND VARSHA PANSARE PRATIKSHA CHIKANE

UNDER GUIDANCE OF

### PROF . JYOTI GAIKWAD



**UNIVERSITY OF MUMBAI DEPARTMENT OF COMPUTER ENGINEERING**

**DATTA MEGHE COLLEGE OF ENGINEERING** *PLOT NO.98 SECTOR-3, AIROLI, NAVI MUMBAI* **ACADAMIC YEAR 2017-18**



**DATTA MEGHE COLLEGE OF ENGINEERING**

**AIROLI, NAVI MUMBAI**

**CERTIFICATE**

This is to certify that the project entitled “**UBER DATA ANALYSIS**” is bona fide work of “**Gaurav Gend , Varsha Pansare , Pratiksha Chikane**“ submitted to the University of Mumbai in partial fulfilment of the requirement for the award of the degree of **“Undergraduate”** in **“Computer Engineering”**.

A. B. C. Prof. A.P.PandeDr.S.D.SawarkarProjectGuide Head of the Department Principal



## DATTA MEGHE COLLEGE OF ENGINEERING

**AIROLI, NAVI MUMBAI**

# PROJECT APPROVAL

This project report entitled “**UBER DATA ANALYSIS**” of the students“**GAURAV GEND , VARSHA PANSARE , PRATIKSHA CHIKANE”**approved for the degree of **COMPUTER ENGINEERING.**

A. B. C. X. Y. Z.

Internal Examiner External Examiner

Date: Date:

Place: Place:

# DECLARATION

We declare that, this written submission represents our ideas in our own words and where others' ideas or words have been included;we have adequately cited and referenced the original sources. We also declare that,we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of the Students Signature

GAURAV GEND VARSHA PANASARE PRATIKSHA CHIKANE

# ACKNOWLEDGEMENT

Motivation and guidance are the keys towards success. I would like to extend my thanks to all the sources of motivation.

We would like to grab this opportunity to thank **Dr.S.D.Sawarkar, Principal** for encouragement and support he has given for our project.

We express our deep gratitude to**Prof. A.P.Pande, Head of the Department**who has been the constant driving force behind the completion of this project.

We wish to express our heartfelt appreciation and deep sense of gratitude to my project guide **Prof. Joyti Gaikwad**for his/her encouragement, invaluable support, timely help, lucid suggestions and excellent guidance which helped us to understand and achieve the project goal. His/Her concrete directions and critical views have greatly helped us in successful completion of this work.

We extend our sincere appreciation to allProfessors for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individually.

We are also thankful to all those who helped us directly or indirectly in completion of this work.

Place: Name of the student

Date:

### INDEX

|  |  |  |
| --- | --- | --- |
| **1.** | **INTRODUCTION** | |
|  | 1.1 | Introduction ix |
| **2.** | **LITERATURE SURVEY** | |
|  | 2.1 | Literature study x |
| **3.** | **EXISTING SYSTEM** | |
| **4.** | **PROPOSED SYSTEM** | |
|  | 4.1 | Architecture/ Framework xi |
| 4.2 | Algorithm and Process Design xii |
|  | 4.3 | Technology used xii |
|  | 4.4 | DetailsOf Hardware and Software xiii |
| **5.** | **Analysis of Algorithm** | |
|  | 5.1 | DFD /sequence diagram xiii-xiv |
| 5.2 | PERT/Gantt chart xiv |
| **6.** | **RESULTS AND DISCUSSIONS** xv | |
| **7.** | **CONCLUSIONS AND FUTURE SCOPE** xv | |
| **8.** | **REFERENCES** | |
|  | 8.1 | Annexure xv |

**LISTOFFIGURES**

|  |  |  |
| --- | --- | --- |
| **SR.NO.** | **NAMEOFTHEFIGURE** | **PAGENO.** |
| 1 | Framework | Xi |
| 2 | Data storytelling | Xi |
| 3 | Data flow diagram | xii-xiii |
| 4 | Gantt chart | Xiv |

**ABSTRACT**

The project can be used to perform data visualization on the uber data. The dataset contains millions of uber pickups in the city.This much data needs to be represented beautifully in order to analyze the rides so that further improvements in the business can be made. As Data analysis is one of the best platform to do analysis as per requirementand todays peoples were really need of uber so to fast result and easy way we use data analysis.

In this proposal there is will be mainly focus on statistical model according to stops, miles ,purpose, days ,date ,time etc

**Keywords: visualization technique, statisticals models,storytelling.**

## INTRODUCTION.

Looking at Data find that the data is increasing day by day and approx 2.5 quintillion bytes of data generate every day. Now, from this data analysis and get useful information which is most important and to understand that here we perform data analysis on UBER data using data analysis. From the dataset, we are comparing business vs personal trips, the frequency for the purpose of the trip, checking how many are the round trips, evaluating the frequency of the trip in each month, and so on. Looking at Data find that the data is increasing day by day and approx

2.5 quintillion bytes of data generate every day. Now, from this data analysis and get useful information which is most important and to understand that here we perform data analysis on UBER data using data analysis. We have to do analysis of uber need to overcome that all thing like on which date receiving the most ride requests followed by visualization ,for what purpose the uber use to persons for ex: meeting

,party etc. and how many time pick up and destination point repeat for increasing car rate at particular location how many cars how many miles cars cover accordingly changing a rates to be done etc and we can also get to know which category type people use uber.

The remainder of this paper is organized as follows: 2 part is of discussion about literature study .3 existing system to know what the existing system is exist for this proposal.4 proposed system in this there will be descriptive look of architecture, technology used, and details of hardware and software used.5analysis of any one of the algorithm.6 Result and discussion is to be done on project. 7 brief discussion on conclusion and future scope .8 at end the references.

## LITERATURE STUDY.

[1] We propose new algorithms for approximate nearest neighbor matching and evaluate and compare them with previous algorithms. In order to scale to very large data sets that would otherwise not fit in the memory of a single machine, we propose a distributed nearest neighbor matching framework that can be used with any of the algorithms described in the paper..[2] how exact unsupervised work and help analysis using different algorithm.[3]how R language is Effective for executing uber data analysis.

### EXISTING SYSTEM.

**Step-1** Importing libraries and read the data.

**Step-2** Cleaning the data.

**Step-3** Getting an hour, day, days of the week, a month from the date of the trip.

**Step-4** Visualizing the data. Different categories of data. From data, we can see most of the people use UBER for business purposes.

### This system is existing with each will easily come to what the existing is about.

1. **PROPOSED SYSTEM.**
   1. **ARCHITECTURE/FRAMEWORK**



## 1-s2.0-S037872060100091X-gr1.gifAlgorithm and process design.

FIG:Datastorytelling

1. The process of cleaning, transforming, manipulating data into useful information that is Data analysis. When we take a particular decision based on previous data that is data analysis. We can make future decisions using data analysis.
2. Visualization techniques.
3. storytelling concept which is important component of machine learning.
4. K-protype algorithm .

### DATA SET USED.

**My uber.csv**

## TECHNOLOGY USED.

Here we use data analysis various technique like visualization

.When we take a particular decision based on previous data that is data analysis. We can make future decisions using data analysis All the business has lots of data. To grow business, sometimes data analysis required. By analyzing data we get important topics on which work out and make our plan for the future through which made perfect future decisions. Most of the businesses going online where the data generate increases day by day. To grow business with this competitive environment data analysis is necessary.Storytelling which is important component of machine learning.

### ) DETAILS OF HARDWARE USED. HARDWARE-

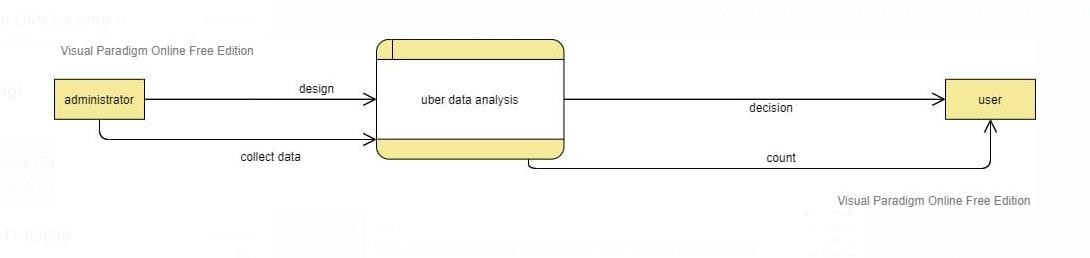
* + 1. LAPTOP/PC
    2. RAM>=4GB

### SOFTWARE-

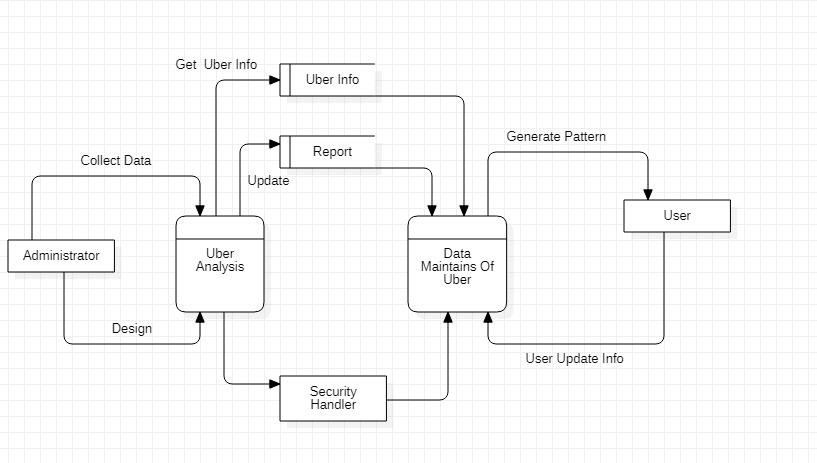
1. MYSQL
2. PYTHON.
3. Google collab.

### ANALYSIS OF ALGORITHM.

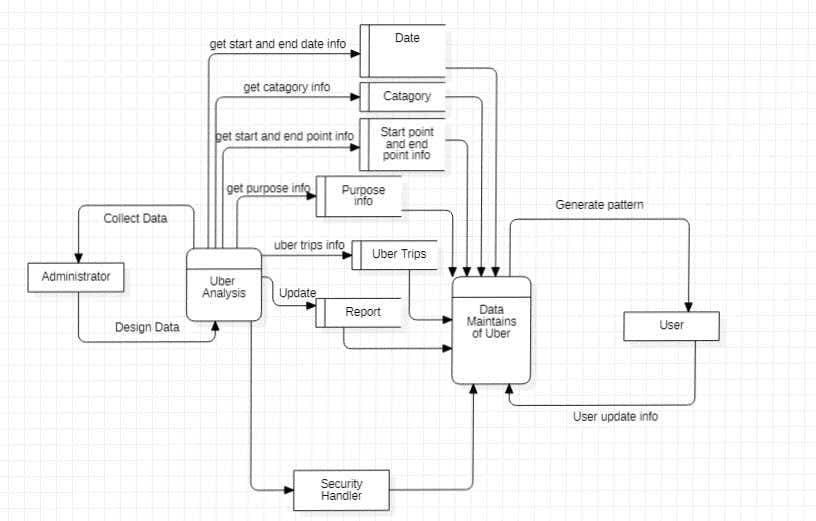
* 1. **DFD Level 0 DFD**



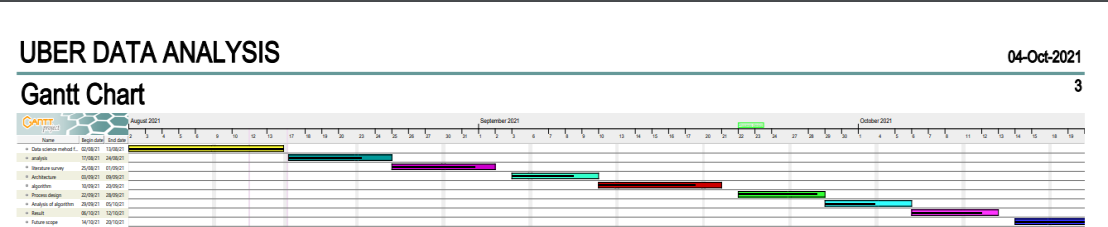
## Level 1 DFD



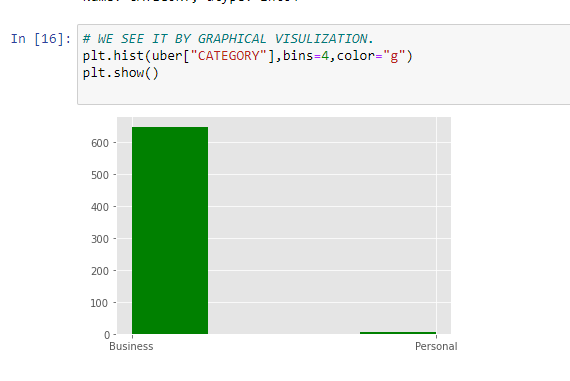
**Level 2 DFD**

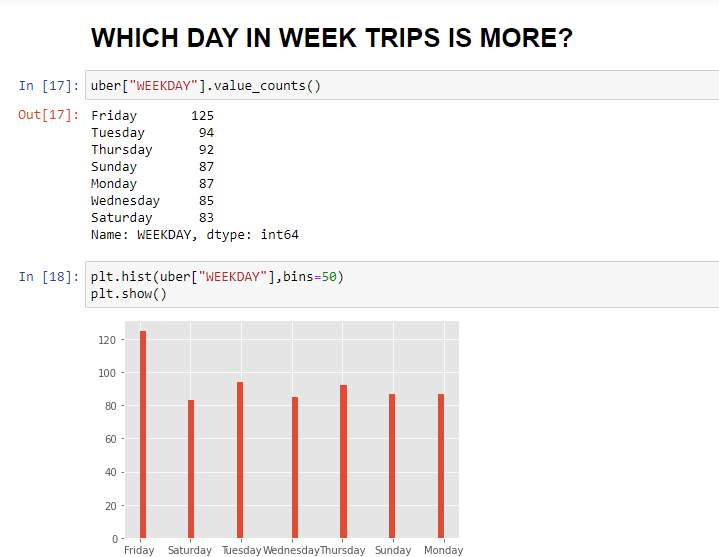


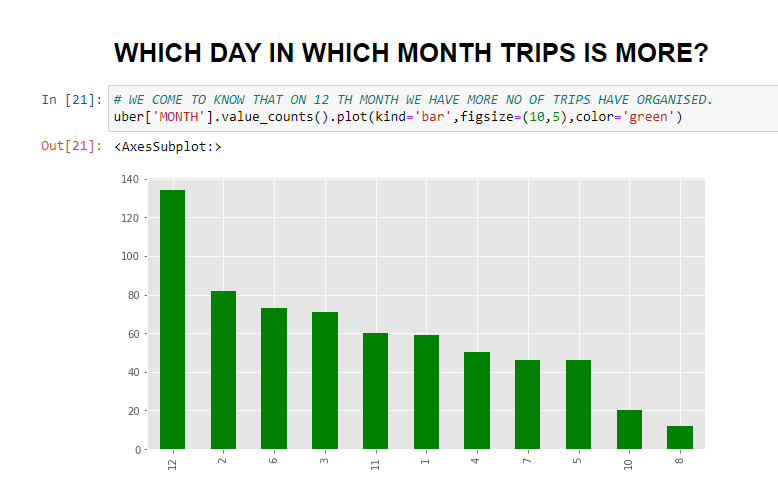
## GNATT CHART.



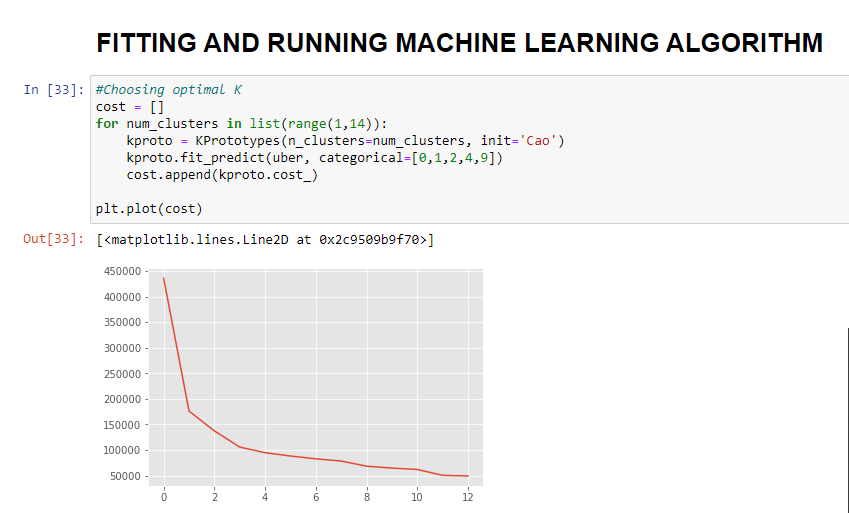
1. **RESULTS AND DISCUSSIONS.**

****

****

****

****

****

K-Prototype is **a clustering method based on partitioning**. Its algorithm is an improvement of the K-Means and K-Mode clustering algorithm to handle clustering with the mixed data types. According to elbow method there is bend at approximate value 3 so there will be three clusters. from these we can get useful information as there is three clusters, from first cluster we get the miles greater than 100 ,and another approximately less 100 and last one approximately less than 50 .and accordingly we have analyze each with needful requirements.And with this we can get help of it for taking further decisions.

## CONCLUSION :

We successfully take a particular decision based on previous data that with the help of data analysis. And also make future decisions using data analysis.. We have done analysis of Uber need to overcome that all thing like on what time there should not be holiday, for what purpose the uber use to persons for ex meeting ,party etc. all this thing is done by simple and portable computer language .The above shows how data analysis could help Uber perform optimal pricing and the optimal positioning of cars in order to serve their customer faster and grow their business. But by running this dataset with machine learning algorithm we getsome needful conclusion , By elbow method we get K=3 as per concept of it. According to need we can take further future decision .

1. **References:**

[1].Lowe, D.G., & Muja, M. (2014) IEEE Trans. Pattern Anal. Mach. Intell., (cited 324 times, HIC:11,CV:69).

[2].Memoona khanum Tahira mahboob ,Pakistan International Journal of Computer Applications (0975 – 8887)

[3].Mrunal patil, asst.prof umakant mandawkar ,sandip university nashik. [4].Referencelink:[https://www.kdnuggets.com/2020/07/clustering-rideshare-data-](https://www.kdnuggets.com/2020/07/clustering-rideshare-data-uber.html) [uber.html.](https://www.kdnuggets.com/2020/07/clustering-rideshare-data-uber.html)

[5].Referencelink :[https://www.codespeedy.com/analyse-uber-data-in-python-using-](https://www.codespeedy.com/analyse-uber-data-in-python-using-machine-learning/) [machine-learning/](https://www.codespeedy.com/analyse-uber-data-in-python-using-machine-learning/)