**Data Modelling**

To do the data modeling, we can use Excel or SQL both to do this, but for this project I have used Python

Programming language for this. Data loaded into the python Jupiter notebook. To create the facts and dimension table I have used Google’s draw it app for its user-friendly application.

For Data modelling understanding the input table data in detail is very important. Normalizing the table in more and more detail according to the Fact table and Dimension table, it is divided based on the attributes.

Tools Involved

Google Sheets

Google Draw it App for modelling

MS Excel for Data Cleaning

SQL Server for Data Analyzing

Tableau for Visualization

Python for Predicting Future Price

PPT for Report Building

**Data Source**

Trip & Limousine Commision, Cloud data – https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page

flat file,

HR, Marketing, CRM, University, File

Cloud data - https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page

Internal Database ,

Internal data present in database I do extract data from data management,

Sometimes I do file system like MS excel as well.

My Company store data in database, sales, HR, marketing team, that data is in different data mart, I pull the data from it, sometimes I get data from clients and customers through Excel as well.

**Tools used**

DATABASE PROGRAMMING LANGUAGE DATA VIZ. TOOLS DATA CLEANING/ANALYSING

SQL SERVER PYTHON TABLEAU MS EXCEL

**Process involved**

Stage -1

Data Collection and Cleansing

Stage -2

Data Transform

Stage -3

Data Loading

Stage – 4

Data Export Through Query

Stage-5

Data Visualization

Stage-6

Future Prediction

**Aim of the Project**

To Acquire data from the various sources as Customer Data, Internal Data, Social Media Data and collate it into single database.

This can help in identifying the potential customers for the company for its future business and marketing.

Various marketing campaigns and advertising strategies can be directed towards the customer segments.

**Business requirement**

As per Business requirements at Uber, The Descriptive and Predictive Analysis is to be done on the historical data.

The data is gathered for past one year. There is total 100000 rows and 19 Columns in the table. The table is divided into various sub-sets representing more normalized data tables.

1. Predictive Analysis: Just as the name suggests, predictive analytics tries to predict what is likely to happen in the future. Predictive analytics holds a variety of statistical techniques from modeling, machine, learning, data mining theory that analyze current and historical facts to make predictions about a future event. Predictive analytics tells you what will happen

2. Descriptive: Descriptive analytics is a simple, surface-level type of analysis that looks at what has happened in the past. The two main techniques used in descriptive analytics are data aggregation and data mining—so, the data analyst first gathers the data and presents it in a summarized format (that’s the aggregation part) and then “mines” the data to discover patterns. It involves the use of data visualization tools like charts, graphs, and tables to present data in a way that's easy to understand. Descriptive analytics tells you what happened.

**Data Collection**

The data is collected through internal data source as well as external data source such as social media and customers. All the datasets

Various attributes such as pickup, dropoff location, distance, datetime, fare, tip, surcharge and payment type etc. collected through data sources and some attributes are created based on the requirements through discussing with the team.

Details of the steps followed in creating the master database is as follows:

Attributes such as ratecard type, payment type collected from TLC Gov. website into the excel sheet.