**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Individual Project done By-:**  **NAME-** MONICA PATEL  **EMAIL –** [Monicapatel826@gmail.com](mailto:Monicapatel826@gmail.com)  **CONTRIBUTION -**  DATA CLEANING, EXPLORATORY DATA ANALYSIS, DATA VISUALISATION, REPORT GENERATION,PPT, TECHNICAL DOCUMENT,CLUSTERING , SUMMARY  DATA INSPECTION CHECKING THE NULL VALUES CHECKING DUPLICATES : DROPPING ID AND DRIVING LICENSE COLUMN. ▪EDA  Data summary Explore data summary Eda Create Model Linear regression Decision tree  Random Forest Conclusion. ▪ CLASSIFICATION MODELS USED : LOGISTIC REGRESSION : DECISION TREE |
| **Please paste the GitHub Repo link.**  Github Link:- https://github.com/1993monica |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**  PROBLEM STATEMENT: NYC TAXI TRIP TIME PREDICTION. **The dataset is based on the 2016 NYC Yellow Cab trip record data made available in Big Query on Google Cloud Platform. The data was originally published by the NYC Taxi and Limousine Commission (TLC). The data was sampled and cleaned for the purposes of this project. Based on individual trip attributes, you should predict the duration of each trip in the test set.****My task is to build a model that predicts the total ride duration of taxi trips in New York City. Mine's primary task dataset is one released by the NYC Taxi and Limousine Commission, which includes pickup time, geo-coordinates, number of passengers, and several other variables.** As the first step, perform data wrangling over the raw data, after wrangling , I just divide the complete project into four main parts -In first part I analyze data, in second part I do some visualization and In third part I made models and in the fourth part I got some conclusion.  In the first part, I overview the whole data, import some libraries , and overview columns one by one and also make some new columns for better understanding the data . It's quite a challenging part.  In the second part, I make some statements to analyze and try our best to represent our statements in graphs or charts. To complete this part we just take help from google and youtube.  In the third part, I made some models like Linear regression, random forest, decision tree and checked them individually and got a perfect model.  In the fourth part, after analysis we get some observations and try to give the best possibility theory for our statement and finally we write them as our conclusion. This is the most challenging part of this project. |