Data Science with Python Career Program

Capstone Project

Hello Learners,

We are excited to finally announce the Capstone Project. This project is your opportunity to present your Data Science skills. The project includes CAR DETAILS Dataset.

You have to use your Data Science skills and perform below mentioned tasks using various Tools and Technologies you have learned.

Present your project using **Slides/PPT** with reference to the below mentioned actionables and make appropriate use of relevant tools and technologies to derive Insights from data and Perform EDA and Deploy a ML Model using Streamlit & GitHub.

Problem Statement:

> **Perform EDA and derive Insights from the CAR DETAILS dataset using Various Data Analysis and Data Visualization libraries of Python such as Pandas, Matplotlib & Seaborn. Create and Deploy a ML Model Which can be accessed by all, using Streamlit and GitHub.**
>
> **Download the CAR DETAILS Dataset from here.**

Actionables:

1. Explore the Data using Excel . understand the data and prepare a short summary about the dataset in the PPT.

2. Download the CAR DETAILS dataset and perform Data cleaning and Data Pre-Processing if Necessary.

3. Use the various methods such as Handling null values, One-Hot Encoding, Imputation and Scaling of Data Pre-Processing where necessary.

4. Perform Exploratory data analysis (EDA) on the Data and perform Graphical Analysis on the Data. Include the graphs with conclusions from the Graphical Analysis.

5. Prepare the Data for Machine Learning modeling.

6. Apply various Machine Learning techniques such as Regression or classification ,Bagging, Ensemble techniques and find out the best model using various Machine Learning model evaluation metrics.

7. Save the best model and Load the model.

8. Take the original data set and make another dataset by randomly picking 20 data points from the CAR DETAILS dataset and apply the saved model on the same Dataset and test the model.

9. Make a GitHub Account by visiting the GitHub Website. Create a repository named Data Science Capstone Project and upload the model with the dataset, code file.

10. Create a Streamlit Account by visiting the Streamlit Website. Connect your GitHub account with streamlit.
11. Create an **app.py** file and other dependencies files for ***Streamlit*** app to be deployed on ***Streamlit Cloud***. Make a simple website and deploy your ML model on Streamlit, Make the website public.

12. Share the Streamlit website and GitHub repository links in the Project PPT.

Please Submit Your Capstone Project using a PPT and insert required files in it , For reference please refer Capstone Guide.

Make a folder and zip it and Upload the following files in the folder:

- **Dataset**
- **Sample Dataset**
- **Graphical Analysis code file**
- **Data Cleaning & Pre-Processing with ML Model code file.**
- **PPT**

References:-

- **Making a Streamlit APP**
- **Streamlit Components Documentation**
- **Create a New Repo and Upload Files on GitHub**

We have also released a **"Capstone Guide"** to help you submit the project.

This Capstone project is of 100 marks ,so plz perform every step of the project carefully and properly document it in the PPT.