Linear Regression

Diamond price prediction

Abstract:

- Have you ever asked yourself, how are diamonds priced? Well, this project talks about the diamonds price prediction based on their cut, color, clarity & other attributes and it also covers the building a simple linear regression model using Python.

Design:

- This project is one of the T5 Data Science Boot Camp requirements. Data provided by Kaggle . In this module, we will be laying the foundation for our analysis by processing and exploring a large amount of data on diamond datasets. This dataset has been made available thanks to Kaggle which is the home for many such datasets and competitions.

Understanding the dataset:

This dataset considered the classic Diamonds dataset which contains the prices and other attributes of almost 54,000 diamonds and this dataset is hosted on Kaggle. The dataset contains 53940 rows and 10 variables. Before jumping into building the model, let's have a look into the variables & their definitions.

- **Fields include**
- **Price** is in US dollars
- **Carat** weight of the diamond
- **Cut** quality of the cut (Fair, Good, Very Good, Premium, Ideal)
- **color** diamond color, from J (worst) to D (best)
- **clarity** a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
- **x** length in mm
- **y** width in mm
- **z** depth in m
- **depth**: The height of a diamond
- **table**: The width of the diamond's table expressed as a percentage of its average diameter

Algorithms :

1. Import Required Packages
2. Load the dataset
3. Perform the exploratory data analysis (EDA)
4. Prepare the dataset for training
5. Create a linear regression model
6. Train the model to fit the data
7. Make predictions using the trained model

Tools:

- Pandas for data manipulation
- IQR for discover outliers
- Remove Duplicate or unnecessary data
- Matplotlib for plotting
- Seaborn for visualizations
- Sklearn Linear Regression library

Communication:

- The slides will be provided here, Feel free to any pull requests besides details are provided at the readme of the project.