Project Title :

"Price Prediction of European Motorbikes using Linear Regression and Standard Scaler"

Hypothesis:

"The price of European motorbikes is influenced by various factors such as mileage, power, fuel type, offer type, and gear system. By applying Linear Regression, we hypothesize that these features will help predict the price accurately."

Method:

1. **Data Collection**:
    - The dataset includes features such as mileage, power, fuel type, offer type, and gear system.
    - The target variable is the **price** of the motorbike.

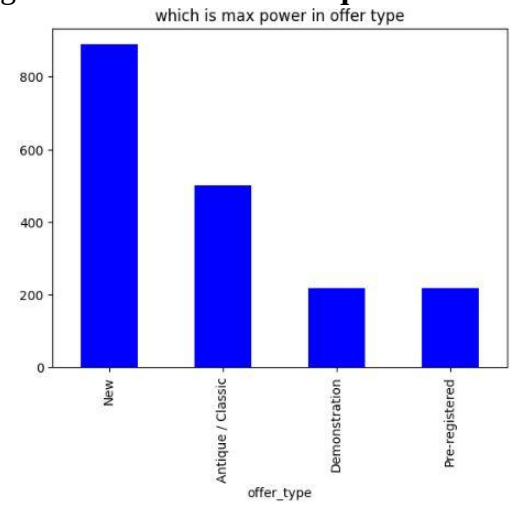
```
df.head()
```

	price	mileage	power	make_model	date	fuel	gear	offer_type	version	link
0	23990	150	218.0	Honda	03/2020	Gasoline	Manual	Demonstration	CBR1000RR-R Fireblade SP	/offers/honda-others-cbr1000rr-r-fireblade-sp-...
1	7500	2871	90.0	BMW F 800 GT	09/2018	Gasoline	Manual	Used	NaN	/offers/bmw-f-800-gt-gasoline-white-f65273c6-6...
2	800	1700	3.0	Nova Motors Retro Star	10/2019	Gasoline	NaN	Used	NaN	/offers/nova-motors-retro-star-gasoline-red-f8...
3	14990	24345	NaN	Aprilia RSV4	03/2016	Gasoline	NaN	Used	RF	/offers/aprilia-rsv4-rf-gasoline-silver-1b51fe...
4	6200	25000	128.0	Kawasaki Ninja ZX-6R	08/2009	Gasoline	NaN	Used	NaN	/offers/kawasaki-ninja-zx-6r-gasoline-blue-2f8...

2. **Data Preprocessing**:
    - Missing data handling, if necessary.
    - Convert categorical variables (like fuel type, offer type) into numerical form using techniques like one-hot encoding.
    - Features like **mileage** and **power** vary in scale, so they are standardized using **Standard Scaler**.
3. **Model Selection**:
    - **Linear Regression** is chosen as the model because it can map the relationship between multiple independent variables and the dependent variable (price) in a linear fashion.
4. **Model Training**:
    - After preprocessing, the dataset is split into **training** and **test** sets.
    - Linear Regression is trained on the training set, with standardized features, to learn the relationship between the predictors and the price.
5. **Evaluation**:
    - The model's performance is evaluated using metrics such as **Mean Squared Error (MSE)** and **R-squared** on the test

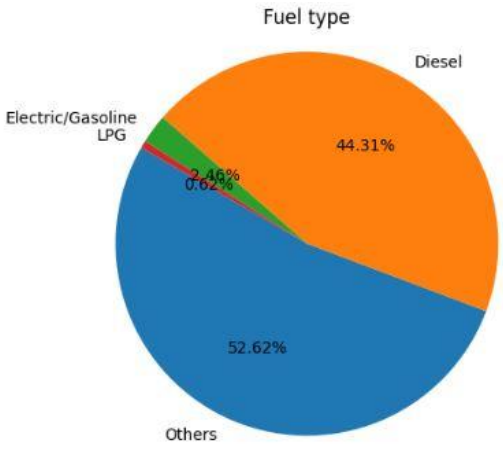
About the Figures:

1. **Bar Plot (offertype vs MaxPower)**:



- .

2. **Pie Chart(Fuel type Distribution)**:



Results:

1. **Model Performance**:
   - The **Linear Regression model** with **Standard Scaler** resulted in an **R-squared** value of indicating that the model explains of the variability in the motorbike prices.
   - The **Mean Squared Error (MSE)** on the test set was which shows the average squared difference between the actual price and the predicted price.
2. **Conclusion**:
   - The model predicts motorbike prices reasonably well based on mileage, power, fuel type, offer type, and gear.
   - Features like **power** and **fuel type** showed a strong positive correlation with price, while **mileage** showed a negative correlation.
   - Further improvements could be made by incorporating more complex models or additional features like brand or year of manufacture

```
from sklearn.metrics import mean_absolute_error,mean_squared_error,r2_score
def modelresults(predictions):
    print("Mean error on model is {}".format(mean_absolute_error(y_test,predictions)))
    print("Root mean squared error on model is{}".format(np.sqrt(mean_squared_error(y_test,predictions))))
    print("r2scoreis{}".format (r2_score(y_test,predictions)))
```

✓ 0.0s