```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

saldf=pd.read_csv("/content/drive/MyDrive/cardekho.csv")

saldf.head()
```


```
{"summary":"{\n  \"name\": \"saldf\",\n  \"rows\": 8128,\n
\"fields\": [\n    {\n      \"column\": \"name\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2058,\n        \"samples\": [\n
\"Volkswagen Ameo 1.5 TDI Trendline\",\n          \"Chevrolet Sail
Hatchback LS ABS\",\n          \"Hyundai i20 Asta (o)\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"year\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 4,\n        \"min\": 1983,\n
\"max\": 2020,\n        \"num_unique_values\": 29,\n
\"samples\": [\n          1997,\n          2019,\n          2002\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"selling_price\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
806253,\n        \"min\": 29999,\n        \"max\": 10000000,\n
\"num_unique_values\": 677,\n        \"samples\": [\n
944999,\n          665000,\n          484999\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"km_driven\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
56550,\n        \"min\": 1,\n        \"max\": 2360457,\n
\"num_unique_values\": 921,\n        \"samples\": [\n          6757,\n
26766,\n          15858\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"fuel\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          \"Petrol\",\n          \"CNG\",\n          \"Diesel\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"seller_type\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 3,\n        \"samples\": [\n
\"Individual\",\n          \"Dealer\",\n          \"Trustmark
Dealer\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"transmission\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          \"Automatic\",\n          \"Manual\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"owner\",\n      \"properties\": {\
n        \"dtype\": \"category\",\n        \"num_unique_values\": 5,\n
\"samples\": [\n          \"Second Owner\",\n          \"Test Drive
```

```
Car\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"mileage(km/ltr/kg)\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 4.0371450567755565,\n        \"min\":
0.0,\n        \"max\": 42.0,\n        \"num_unique_values\": 381,\n
\"samples\": [\n          18.3,\n          25.8\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"engine\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
503.9163029923002,\n        \"min\": 624.0,\n        \"max\": 3604.0,\
n        \"num_unique_values\": 121,\n        \"samples\": [\n
2497.0,\n          2362.0\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"max_power\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 320,\n
\"samples\": [\n          \"56.3\",\n          \"198.5\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"seats\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 0.95958752023394,\n
\"min\": 2.0,\n        \"max\": 14.0,\n        \"num_unique_values\":
9,\n        \"samples\": [\n          14.0,\n          4.0\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"saldf"}
```


```
saldf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8128 entries, 0 to 8127
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   name                8128 non-null   object
 1   year                8128 non-null   int64
 2   selling_price       8128 non-null   int64
 3   km_driven           8128 non-null   int64
 4   fuel                8128 non-null   object
 5   seller_type         8128 non-null   object
 6   transmission        8128 non-null   object
 7   owner               8128 non-null   object
 8   mileage(km/ltr/kg)  7907 non-null   float64
 9   engine              7907 non-null   float64
 10  max_power           7913 non-null   object
 11  seats               7907 non-null   float64
dtypes: float64(3), int64(3), object(6)
memory usage: 762.1+ KB
```

```
saldf.isnull().sum()
```

```
name                     0
year                     0
selling_price            0
km_driven                0
fuel                     0
seller_type              0
transmission             0
owner                    0
mileage(km/ltr/kg)     221
engine                 221
max_power              215
seats                  221
dtype: int64
```

```
inp=saldf[["year"]]
out=saldf["selling_price"]
```

```
LR=LinearRegression()
```

```
Train_data=pd.concat([inp,out],axis=1)
train=Train_data.dropna()
```

```
inp1=saldf[["year"]]
out1=saldf["selling_price"]
```

```
LR.fit(inp1,out1)
```

```
LinearRegression()
```

```
LR.predict([[5]])
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/utils/
validation.py:2739: UserWarning: X does not have valid feature names,
but LinearRegression was fitted with feature names
  warnings.warn(

array([-1.65193584e+08])
```