Project 2 - Linear Regression-Multivariate

1.Understanding the problem statement and client requirement

Problem Statement

Find the price of a house for the given details:area in sq.ft,no.of.bedrooms,how old is that house.

2.Collect the data

▾ 3.Data Analysis

```
import pandas as pd
```

```
df = pd.read_csv("/content/prjct2-dataset .csv")
df
```

	area	bedroom	age	price
0	2200	3.0	20	8400000
1	2600	2.0	15	13000000
2	3000	4.0	14	15000000
3	3200	NaN	25	12000000
4	3600	3.0	8	19000000
5	4000	5.0	10	22000000

▾ Handling the missing datas

```
pd.isnull(df['bedroom'])
```

```
0     False
```

```
1    False
2    False
3     True
4    False
5    False
Name: bedroom, dtype: bool
```

```
import
```

```
median = df['bedroom'].median()
```

▾ filling the missing data using fillna()

```
#filling the missing values in the bedroom column with the median data and assigning that
df['bedroom']= df['bedroom'].fillna(median)
print(df)
```

```
    area  bedroom  age     price
0   2200      3.0   20   8400000
1   2600      2.0   15  13000000
2   3000      4.0   14  15000000
3   3200      3.0   25  12000000
4   3600      3.0    8  19000000
5   4000      5.0   10  22000000
```

▾ 4.Build ML model

```
from sklearn import linear_model
reg = linear_model.LinearRegression()
x=df.drop('price',axis=1)
x
```

	area	bedroom	age
0	2200	3.0	20
1	2600	2.0	15
2	3000	4.0	14
3	3200	3.0	25
4	3600	3.0	8
5	4000	5.0	10

```
y = df['price']
y
```

```
0     8400000
1    13000000
2    15000000
3    12000000
4    19000000
5    22000000
Name: price, dtype: int64
```

```python
#Trainning the model
reg.fit(x,y)
```

```
LinearRegression()
```

```python
reg.predict([[2000,2,2]])
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:451: UserWarning: X does not
  "X does not have valid feature names, but"
array([13565323.45989718])
```

5.Downloading the model

```python
import pickle
```

```python
with open('prjct2model','wb') as f:
  pickle.dump(reg,f)
```

✓ 0s   completed at 2:04 PM   ● ✕