```
# ml model
# (1).supervised ml models---> in this, we will trained our labelled data
# (a).descrete ---> logisticregression , decisionreeclassifier, randomforestclassifier,naive bayes.
# (b).countinuous ---> linearregression , d.t.regressor, r.f.regressor
# (2). unsupervised ml models ----> in this, we will trained our unlabelled data
# kmeans, pca, apriori.
```

```
# (1) supervised ml models
# (2) linearregression
```

```
# workflow
# data ingestion ---> data cleaning and preprocessing ----> select ml model based on
# target column ----> model training ---> model predition---> model evaluation
```

```
import numpy as np
import pandas as pd
```

```
df=pd.read_csv("/content/insurance - insurance.csv")
df.head(2)
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.90	0	yes	southwest	16884.9240
1	18	male	33.77	1	no	southeast	1725.5523

Next steps: (Generate code with df) (New interactive sheet)

```
x=df.drop(columns=['charges'])
y=df['charges']
```

```
from sklearn.preprocessing import LabelEncoder
```

```
lb=LabelEncoder()
```

```
x['sex']=lb.fit_transform(x['sex'])
x['smoker']=lb.fit_transform(x['smoker'])
x['region']=lb.fit_transform(x['region'])
```

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=1)
```

```
from sklearn.linear_model import LinearRegression
```

```
lr=LinearRegression()
```

```
lr.fit(x_train,y_train)
```

```
▾ LinearRegression   ⓘ ⓘ
LinearRegression()
```

```
lr.fit(x_train,y_train)
```

```
▾ LinearRegression   ⓘ ⓘ
LinearRegression()
```

```
y_pred=lr.predict(x_test)
```

```
from sklearn.metrics import r2_score
```

```
r2_score(y_test,y_pred)
```

```
0.7623224022563388
```

```python
import numpy as np
import pandas as pd
```

```python
df=pd.read_csv("/content/Churn_Modelling - Churn_Modelling.csv")
df.head(2)
```

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveN
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	

Next steps: (Generate code with `df`) (New interactive sheet)

```python
x=df.drop(columns=['EstimatedSalary'])
y=df['EstimatedSalary']
```

```python
from sklearn.preprocessing import LabelEncoder
```

```python
lb=LabelEncoder()
```

```python
x['Surname']=lb.fit_transform(x['Surname'])
x['Geography']=lb.fit_transform(x['Geography'])
x['Gender']=lb.fit_transform(x['Gender'])
```

```python
from sklearn.model_selection import train_test_split
```

```python
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=1)
```

```python
from sklearn.linear_model import LinearRegression
```

```python
lr=LinearRegression()
```

```python
lr.fit(x_train,y_train)
```

```
▼ LinearRegression   ⓘ ⓘ
LinearRegression()
```

```python
lr.fit(x_train,y_train)
```

```
▼ LinearRegression   ⓘ ⓘ
LinearRegression()
```

```python
y_pred=lr.predict(x_test)
```

```python
from sklearn.metrics import r2_score
```

```python
r2_score(y_test,y_pred)
```

```
-0.0012813698592946121
```