Model Notebook for Coding in SK-LEARN

In [1]:
```
# import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.linear_model import LinearRegression
```

In [2]:
```
# import the data
df = sns.load_dataset('tips')
df.head()
```

Out[2]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

In [3]:
```
# # preprocess the data
# scalar = StandardScaler()
# df[['total_bill', 'tip']] = scalar.fit_transform(df[['total_bill', 'tip']])
```

In [4]:
```
# separate the features X and the target/labels y
X = df[['total_bill']]
y = df['tip']
```

In [5]:
```
# train test split the data
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.8)
```

In [6]:
```
# call the model
model = LinearRegression()
```

In [7]:
```
# train the model
model.fit(X_train, y_train)
```

Out[7]:
```
LinearRegression()
```

In [8]:
```
model.predict([[15]])
```

```
C:\Users\ustb\.anaconda\anwaar\Lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature name
s, but LinearRegression was fitted with feature names
  warnings.warn(
```

Out[8]:
```
array([2.47845878])
```

In [9]:
```
# prediction
y_pred = model.predict(X_test)
```

In [10]:
```
# evaluation the model
from sklearn.metrics import mean_squared_error, r2_score
print('MSE: ', mean_squared_error(y_test, y_pred))
print('R2: ', r2_score(y_test, y_pred))
```

```
MSE:  1.0242636186029668
R2:  0.5169238433173655
```

Save and load a model

In [12]:
```
# save the model
#import pickle
#pickle.dump(model, open('./saved_models/model_01.pkl', 'wb'))
```

In [13]:
```
# load the model
#import pickle
```

```
#model_load = pickle.load(open('./saved_models/model_01.pkl', 'rb'))
#model_load.predict([[15]])
```

In [ ]: