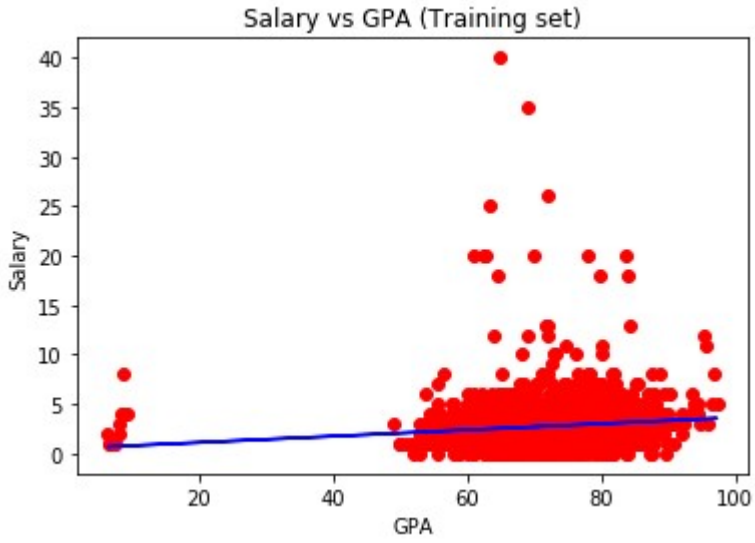
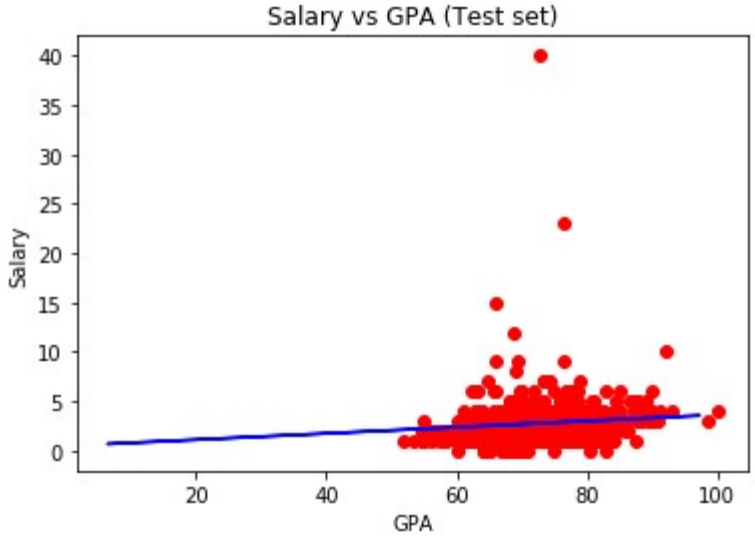
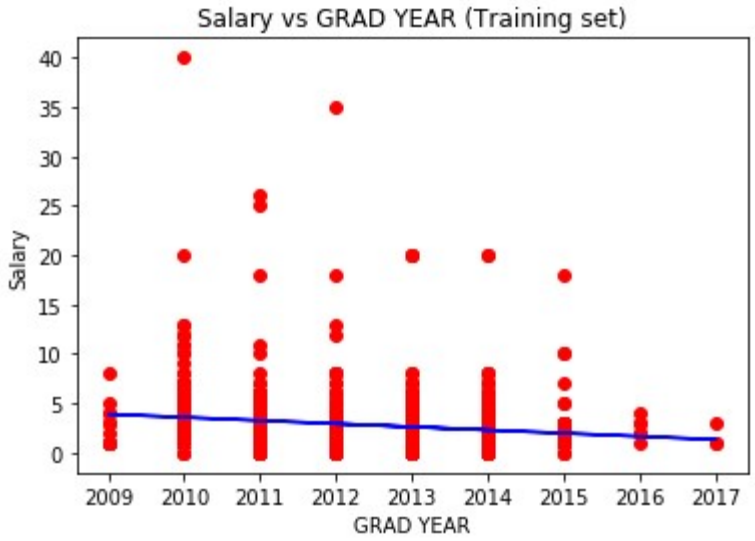
```
Python 3.6.8 |Anaconda, Inc.| (default, Feb 21 2019, 18:30:04) [MSC v.1916 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 7.3.0 -- An enhanced Interactive Python.

In [66]: import matplotlib.pyplot as plt
    ...: import pandas as pd
    ...:
    ...: df=pd.read_csv('G:\Data Analysis\output.csv')
    ...: df=df.dropna()
    ...: X = df.iloc[:,17:18].values
    ...: y = df.iloc[:, 1].values
    ...: for i in range(len(y)):
    ...:     y[i]=int(y[i])
    ...:
    ...: '''
    ...: for i in range(len(X)):
    ...:     X[i]=int(X[i])
    ...:
    ...: '''
    ...:
    ...: 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 = 0)
    ...:
    ...:
    ...: from sklearn.linear_model import LinearRegression
    ...: regressor = LinearRegression()
    ...: regressor.fit(X_train, y_train)
    ...:
    ...: y_pred = regressor.predict(X_test)
    ...: for i in range(len(y_pred)):
    ...:     y_pred[i]=int(y_pred[i])
    ...:
    ...:
    ...:
    ...:
    ...: plt.scatter(X_train, y_train, color = 'red')
    ...: plt.plot(X_train, regressor.predict(X_train), color = 'blue')
    ...: plt.title('Salary vs GPA (Training set)')
    ...: plt.xlabel('GPA')
    ...: plt.ylabel('Salary')
    ...: plt.show()
    ...:
    ...:
    ...: plt.scatter(X_test, y_test, color = 'red')
    ...: plt.plot(X_train, regressor.predict(X_train), color = 'blue')
    ...: plt.title('Salary vs GPA (Test set)')
    ...: plt.xlabel('GPA')
    ...: plt.ylabel('Salary')
    ...: plt.show()
```

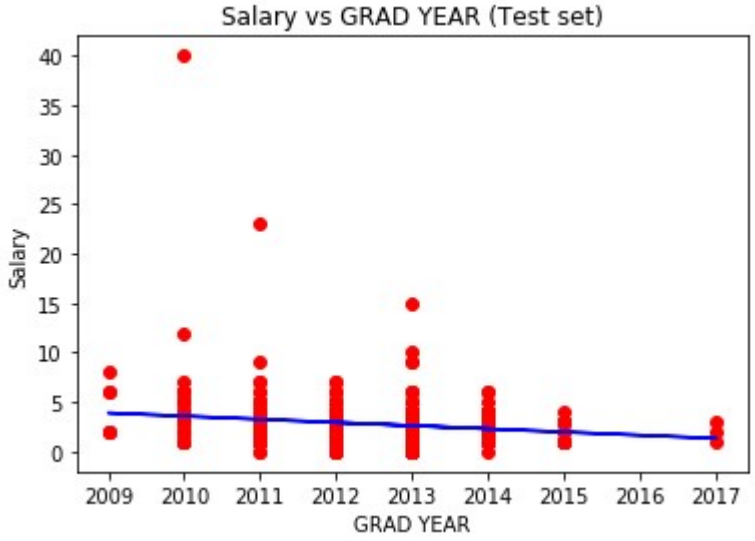





```
In [67]: import matplotlib.pyplot as plt
    ...: import pandas as pd
    ...:
    ...: df=pd.read_csv('G:\Data Analysis\output.csv')
    ...: df=df.dropna()
    ...: X = df.iloc[:,21:22].values
    ...: y = df.iloc[:, 1].values
    ...: for i in range(len(y)):
    ...:     y[i]=int(y[i])
    ...:
    ...: '''
    ...: for i in range(len(X)):
    ...:     X[i]=int(X[i])
    ...:
    ...: '''
    ...:
    ...: 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 = 0)
    ...:
```

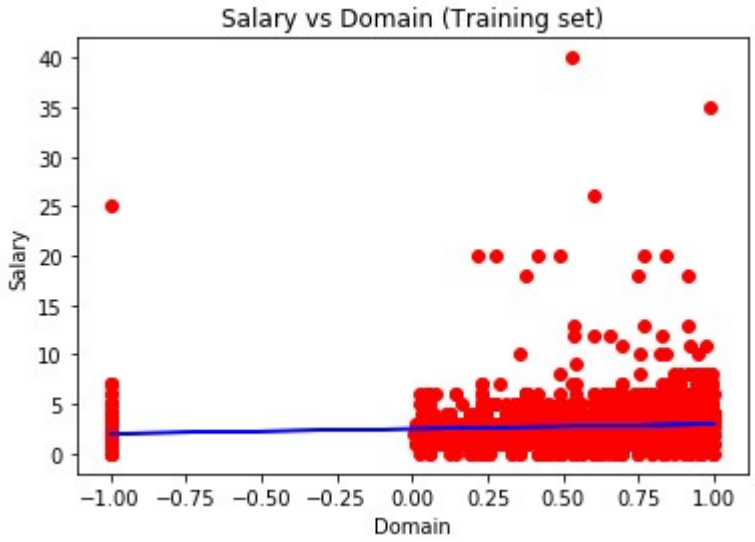
```
...:
...: from sklearn.linear_model import LinearRegression
...: regressor = LinearRegression()
...: regressor.fit(X_train, y_train)
...:
...: y_pred = regressor.predict(X_test)
...: for i in range(len(y_pred)):
...:     y_pred[i]=int(y_pred[i])
...:
...:
...:
...:
...: plt.scatter(X_train, y_train, color = 'red')
...: plt.plot(X_train, regressor.predict(X_train), color = 'blue')
...: plt.title('Salary vs GRAD YEAR (Training set)')
...: plt.xlabel('GRAD YEAR')
...: plt.ylabel('Salary')
...: plt.show()
...:
...:
...: plt.scatter(X_test, y_test, color = 'red')
...: plt.plot(X_train, regressor.predict(X_train), color = 'blue')
...: plt.title('Salary vs GRAD YEAR (Test set)')
...: plt.xlabel('GRAD YEAR')
...: plt.ylabel('Salary')
...: plt.show()
```

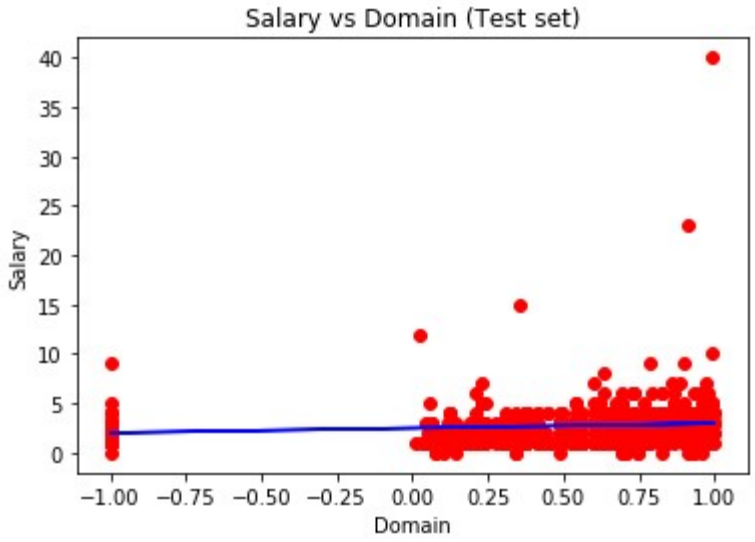





```
In [68]: import matplotlib.pyplot as plt
    ...: import pandas as pd
    ...:
    ...: df=pd.read_csv('G:\Data Analysis\output.csv')
    ...: df=df.dropna()
    ...: X = df.iloc[:,30:31].values
    ...: y = df.iloc[:, 1].values
    ...: for i in range(len(y)):
    ...:     y[i]=int(y[i])
    ...:
    ...: '''
    ...: for i in range(len(X)):
    ...:     X[i]=int(X[i])
    ...:
    ...: '''
    ...:
    ...: 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 = 0)
    ...:
    ...:
    ...: from sklearn.linear_model import LinearRegression
    ...: regressor = LinearRegression()
    ...: regressor.fit(X_train, y_train)
    ...:
    ...: y_pred = regressor.predict(X_test)
    ...: for i in range(len(y_pred)):
    ...:     y_pred[i]=int(y_pred[i])
    ...:
    ...:
    ...:
    ...:
    ...: plt.scatter(X_train, y_train, color = 'red')
    ...: plt.plot(X_train, regressor.predict(X_train), color = 'blue')
    ...: plt.title('Salary vs Domain (Training set)')
    ...: plt.xlabel('Domain')
    ...: plt.ylabel('Salary')
    ...: plt.show()
    ...:
```

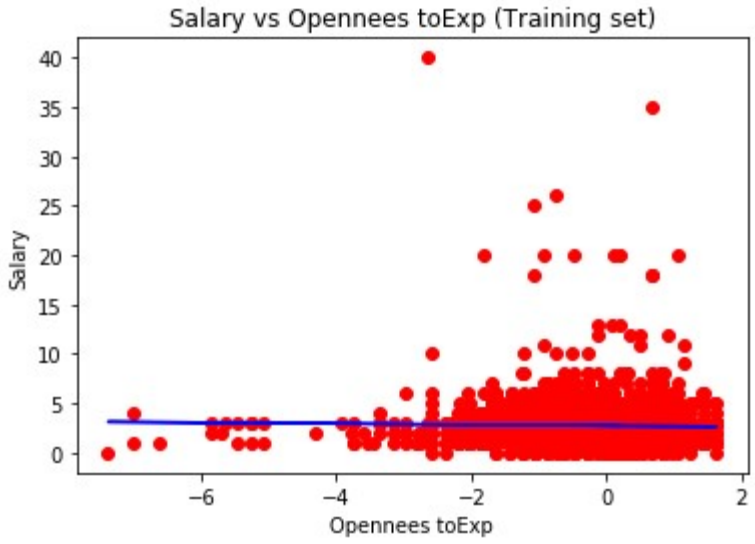
```
...:
...: plt.scatter(X_test, y_test, color = 'red')
...: plt.plot(X_train, regressor.predict(X_train), color = 'blue')
...: plt.title('Salary vs Domain (Test set)')
...: plt.xlabel('Domain')
...: plt.ylabel('Salary')
...: plt.show()
```

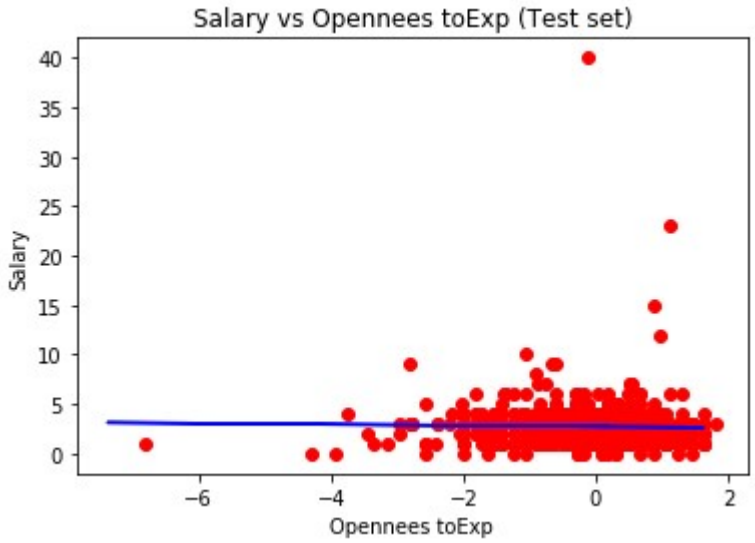





```
In [69]: import matplotlib.pyplot as plt
    ...: import pandas as pd
    ...:
    ...: df=pd.read_csv('G:\Data Analysis\output.csv')
    ...: df=df.dropna()
    ...: X = df.iloc[:,26:27].values
    ...: y = df.iloc[:, 1].values
    ...: for i in range(len(y)):
    ...:     y[i]=int(y[i])
    ...:
    ...: '''
    ...: for i in range(len(X)):
```

```
    ...:     X[i]=int(X[i])
    ...:
    ...: '''
    ...:
    ...: 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 = 0)
    ...:
    ...:
    ...: from sklearn.linear_model import LinearRegression
    ...: regressor = LinearRegression()
    ...: regressor.fit(X_train, y_train)
    ...:
    ...: y_pred = regressor.predict(X_test)
    ...: for i in range(len(y_pred)):
    ...:     y_pred[i]=int(y_pred[i])
    ...:
    ...:
    ...:
    ...:
    ...: plt.scatter(X_train, y_train, color = 'red')
    ...: plt.plot(X_train, regressor.predict(X_train), color = 'blue')
    ...: plt.title('Salary vs Opennees toExp (Training set)')
    ...: plt.xlabel('Opennees toExp')
    ...: plt.ylabel('Salary')
    ...: plt.show()
    ...:
    ...:
    ...: plt.scatter(X_test, y_test, color = 'red')
    ...: plt.plot(X_train, regressor.predict(X_train), color = 'blue')
    ...: plt.title('Salary vs Opennees toExp (Test set)')
    ...: plt.xlabel('Opennees toExp')
    ...: plt.ylabel('Salary')
    ...: plt.show()
```




Salary vs Opennees toExp (Test set)
Salary
Opennees toExp
40
35
30
25
20
15
10
5
0
−6
−4
−2
0
2

In [70]: