Machine Learning Final Project

By: Brett Beaulieu

Topic: Regression with Energy Efficiency and Classification of Fradulent Charges.

Import all library used in the file.

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Ridge
from sklearn.linear_model import Lasso
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
from sklearn.metrics import f1_score
from sklearn.metrics import roc_curve
from sklearn.metrics import roc_auc_score
from sklearn.metrics import classification_report
from sklearn.metrics import plot_roc_curve
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import RobustScaler
from sklearn.preprocessing import Normalizer
from sklearn.tree import DecisionTreeClassifier
from imblearn.under_sampling import RandomUnderSampler
from imblearn.over_sampling import RandomOverSampler
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import GridSearchCV
from math import sqrt
```

Regression Dataset

Exploration and Preprocessing

Load in the dataset from excel.

In [2]:
```
dataset = pd.read_excel('ENB2012_data.xlsx')
dataset = dataset[dataset.columns[:-2]]
dataset
```

Out[2]:

	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2
0	0.98	514.5	294.0	110.25	7.0	2	0.0	0	15.55	21.33
1	0.98	514.5	294.0	110.25	7.0	3	0.0	0	15.55	21.33
2	0.98	514.5	294.0	110.25	7.0	4	0.0	0	15.55	21.33
3	0.98	514.5	294.0	110.25	7.0	5	0.0	0	15.55	21.33
4	0.90	563.5	318.5	122.50	7.0	2	0.0	0	20.84	28.28
...	...	...	...	...	...	...	...	...	...	...
763	0.64	784.0	343.0	220.50	3.5	5	0.4	5	17.88	21.40
764	0.62	808.5	367.5	220.50	3.5	2	0.4	5	16.54	16.88
765	0.62	808.5	367.5	220.50	3.5	3	0.4	5	16.44	17.11
766	0.62	808.5	367.5	220.50	3.5	4	0.4	5	16.48	16.61
767	0.62	808.5	367.5	220.50	3.5	5	0.4	5	16.64	16.03

768 rows × 10 columns

Change the column headers to their original names.

In [3]:
```
dataset.columns = ['Relative   Compactness', 'Surface   Area', 'Wall   Area',
                   , 'Orientation', 'Glazing   Area', 'Glazing Area   Distributi
dataset.head()
```

Out[3]:

	Relative Compactness	Surface Area	Wall Area	Roof Area	Overall Height	Orientation	Glazing Area	Glazing Area Distribution	Heating Load	C
0	0.98	514.5	294.0	110.25	7.0	2	0.0	0	15.55	
1	0.98	514.5	294.0	110.25	7.0	3	0.0	0	15.55	
2	0.98	514.5	294.0	110.25	7.0	4	0.0	0	15.55	
3	0.98	514.5	294.0	110.25	7.0	5	0.0	0	15.55	
4	0.90	563.5	318.5	122.50	7.0	2	0.0	0	20.84	

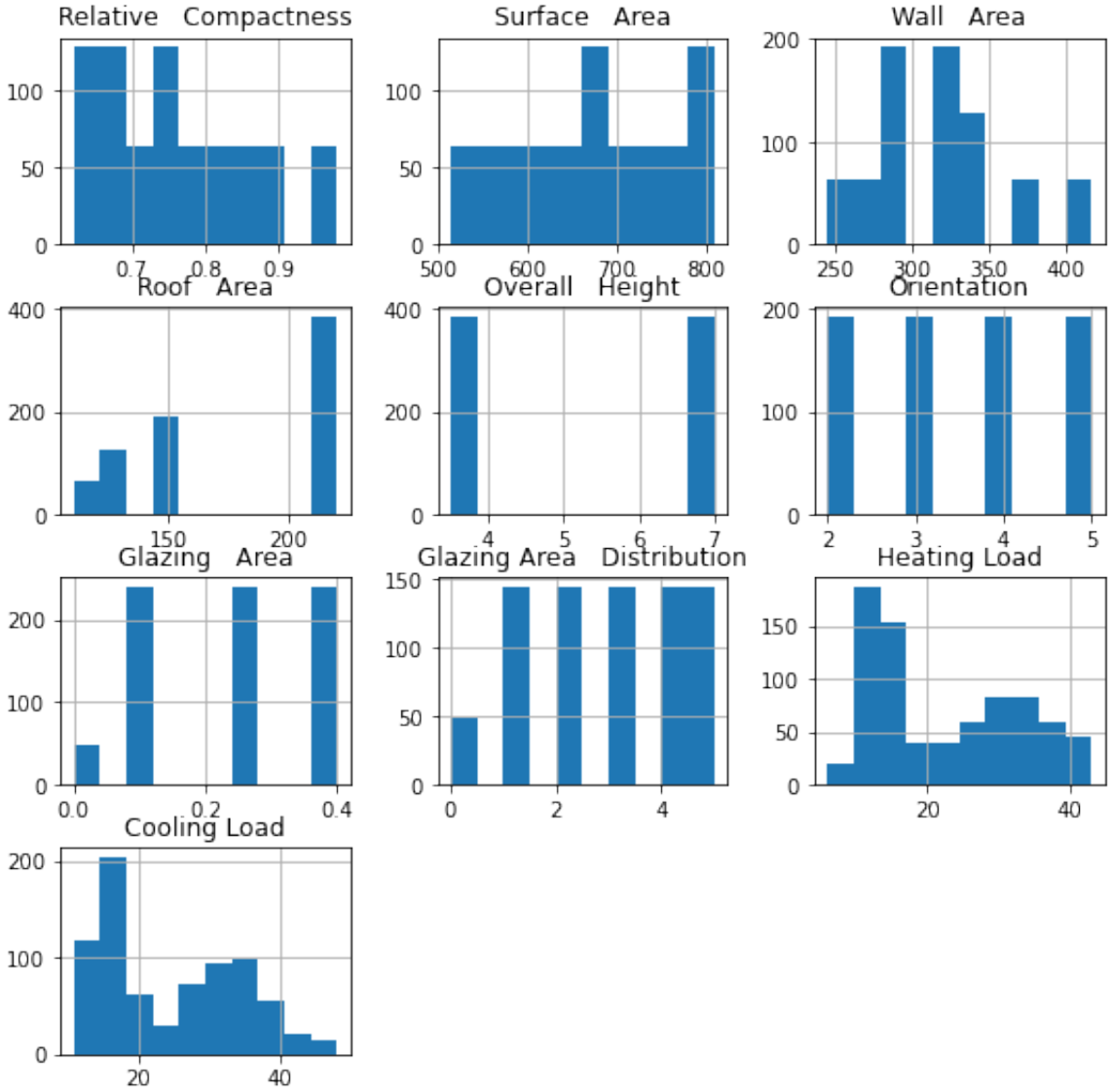
Check for null values in the dataset.

In [4]:
```
dataset.isnull().sum()
```

Out[4]:
```
Relative   Compactness          0
Surface   Area                  0
Wall   Area                     0
Roof   Area                     0
Overall   Height                0
Orientation                     0
Glazing   Area                  0
Glazing Area   Distribution     0
Heating Load                    0
Cooling Load                    0
dtype: int64
```

Visualize the data through histograms.

In [5]:
```
dataset.hist(figsize=(9,9))
plt.show()
```



Assign columns to their respective X and y. Heating and Cooling loads should be the y variable.

In [6]:
```
X = dataset.drop(['Heating Load','Cooling Load'], axis=1)
y = dataset[['Heating Load','Cooling Load']]
X
```

Out[6]:

	Relative Compactness	Surface Area	Wall Area	Roof Area	Overall Height	Orientation	Glazing Area	Glazing Area Distribution
0	0.98	514.5	294.0	110.25	7.0	2	0.0	0
1	0.98	514.5	294.0	110.25	7.0	3	0.0	0
2	0.98	514.5	294.0	110.25	7.0	4	0.0	0
3	0.98	514.5	294.0	110.25	7.0	5	0.0	0
4	0.90	563.5	318.5	122.50	7.0	2	0.0	0
...	...	...	...	...	...	...	...	...
763	0.64	784.0	343.0	220.50	3.5	5	0.4	5
764	0.62	808.5	367.5	220.50	3.5	2	0.4	5
765	0.62	808.5	367.5	220.50	3.5	3	0.4	5
766	0.62	808.5	367.5	220.50	3.5	4	0.4	5
767	0.62	808.5	367.5	220.50	3.5	5	0.4	5

768 rows × 8 columns

Split up the training and testing set with a test size of 0.33.

In [7]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.33, r
```

Normalize the training and testing data.

In [8]:
```
nr = Normalizer(copy=False)
X_train_norm = nr.fit_transform(X_train)
X_test_norm = nr.fit_transform(X_test)
```

Linear Regression

Perform Linear Regression on the dataset.

In [9]:
```
lr = LinearRegression(normalize=True)
lr.fit(X_train_norm,y_train)
```

Out[9]:
```
LinearRegression(normalize=True)
```

Using the Linear Regression model predict values based off the test data.

In [10]:
```
y_pred = lr.predict(X_test_norm)
y_pred
```

Out[10]:
```
array([[18.37424611, 19.66773346],
       [13.71799611, 16.44898346],
       [32.03049611, 32.48023346],
       [35.74924611, 36.63648346],
       [15.15549611, 17.35523346],
       [28.31174611, 30.82398346],
       [24.12424611, 27.01148346],
       [27.96799611, 29.60523346],
       [17.12424611, 18.91773346],
       [27.24924611, 29.57398346],
       [17.43674611, 19.29273346],
       [34.74924611, 35.82398346],
       [27.31174611, 29.98023346],
       [ 8.96799611, 12.49585846],
       [18.06174611, 19.07398346],
       [37.09299611, 37.73023346],
       [37.78049611, 38.01148346],
       [11.70237111, 14.54273346],
       [15.07737111, 17.16773346],
       [34.65549611, 35.88648346],
       [34.43674611, 35.44898346],
       [35.03049611, 36.16773346],
       [11.03049611, 14.69898346],
       [29.68674611, 31.88648346],
       [12.84299611, 15.51148346],
       [29.62424611, 32.07398346],
       [33.74924611, 34.60523346],
       [34.31174611, 35.82398346],
       [14.53049611, 17.07398346],
       [15.74924611, 18.10523346],
       [ 8.26487111, 12.23023346],
       [11.13987111, 14.68335846],
       [14.96799611, 17.23023346],
       [30.68674611, 32.94898346],
       [28.68674611, 32.01148346],
       [27.84299611, 29.82398346],
       [31.49924611, 32.69898346],
       [28.78049611, 31.98023346],
       [33.71799611, 34.73023346],
       [ 9.03049611, 13.12085846],
       [28.78049611, 30.07398346],
       [11.76487111, 14.44898346],
       [ 6.81174611, 11.71460846],
       [32.81174611, 34.51148346],
       [ 7.76487111, 12.18335846],
       [ 6.76487111, 11.79273346],
       [ 8.90549611, 13.41773346],
       [ 9.04612111, 13.02710846],
       [28.78049611, 31.88648346],
       [31.03049611, 32.54273346],
       [34.06174611, 34.73023346],
       [14.62424611, 16.66773346],
```

```
[13.81174611, 16.73023346],
[32.56174611, 34.48023346],
[12.74924611, 15.49585846],
[12.62424611, 15.54273346],
[11.17112111, 14.40210846],
[31.68674611, 33.41773346],
[40.46799611, 39.91773346],
[37.37424611, 37.85523346],
[16.24924611, 17.91773346],
[18.78049611, 19.79273346],
[14.53049611, 16.88648346],
[28.49924611, 31.04273346],
[28.71799611, 30.19898346],
[29.49924611, 32.26148346],
[29.49924611, 32.10523346],
[11.06174611, 14.48023346],
[13.46799611, 16.35523346],
[11.43674611, 14.52710846],
[26.49924611, 29.66773346],
[32.65549611, 32.54273346],
[13.57737111, 15.74585846],
[27.68674611, 30.79273346],
[32.53049611, 34.41773346],
[11.15549611, 14.52710846],
[34.93674611, 35.79273346],
[31.21799611, 33.10523346],
[31.37424611, 32.32398346],
[22.99924611, 27.63648346],
[24.84299611, 27.04273346],
[32.81174611, 34.10523346],
[ 9.48362111, 13.24585846],
[15.78049611, 17.82398346],
[11.49924611, 14.63648346],
[15.82737111, 17.33960846],
[26.03049611, 29.57398346],
[36.81174611, 37.79273346],
[28.84299611, 31.82398346],
[14.43674611, 17.35523346],
[12.15549611, 15.29273346],
[30.90549611, 33.26148346],
[ 9.82737111, 13.62085846],
[14.43674611, 16.60523346],
[37.24924611, 37.88648346],
[14.34299611, 16.79273346],
[26.68674611, 29.60523346],
[30.06174611, 32.07398346],
[ 9.81174611, 13.12085846],
[ 9.09299611, 12.91773346],
[39.78049611, 39.79273346],
[27.49924611, 29.79273346],
[ 8.01487111, 12.12085846],
[27.78049611, 29.66773346],
[29.46799611, 31.94898346],
[18.09299611, 19.54273346],
[27.34299611, 30.98023346],
[18.53049611, 19.69898346],
[15.76487111, 17.93335846],
```

```
[10.88987111, 14.48023346],
[34.18674611, 35.91773346],
[33.56174611, 34.29273346],
[23.78049611, 26.88648346],
[16.81174611, 18.63648346],
[11.38987111, 14.68335846],
[15.06174611, 17.27710846],
[10.48362111, 13.93335846],
[ 8.51487111, 12.26148346],
[13.59299611, 15.88648346],
[23.81174611, 28.04273346],
[27.49924611, 29.63648346],
[10.73362111, 14.04273346],
[31.65549611, 34.41773346],
[15.49924611, 18.07398346],
[26.90549611, 29.60523346],
[17.90549611, 19.38648346],
[16.01487111, 17.90210846],
[29.78049611, 32.29273346],
[24.71799611, 27.13648346],
[39.74924611, 39.88648346],
[20.87424611, 25.13648346],
[34.56174611, 36.10523346],
[15.56174611, 17.94898346],
[14.51487111, 16.60523346],
[15.01487111, 17.35523346],
[17.71799611, 19.26148346],
[33.40549611, 34.63648346],
[35.62424611, 36.69898346],
[14.10862111, 16.85523346],
[17.34299611, 19.01148346],
[ 8.92112111, 13.23023346],
[40.46799611, 40.16773346],
[28.53049611, 29.88648346],
[31.12424611, 32.44898346],
[34.31174611, 34.73023346],
[11.24924611, 14.54273346],
[ 8.67112111, 12.46460846],
[26.43674611, 29.69898346],
[34.62424611, 34.85523346],
[32.15549611, 34.32398346],
[33.40549611, 34.66773346],
[29.24924611, 31.63648346],
[28.96799611, 31.57398346],
[12.62424611, 15.38648346],
[27.53049611, 30.63648346],
[17.24924611, 19.13648346],
[29.18674611, 31.82398346],
[29.93674611, 32.23023346],
[11.65549611, 14.94898346],
[30.34299611, 31.98023346],
[33.34299611, 34.79273346],
[31.87424611, 34.16773346],
[18.46799611, 19.79273346],
[11.84299611, 15.35523346],
[14.74924611, 17.16773346],
[12.53049611, 15.66773346],
```

```
       [27.93674611, 30.98023346],
       [33.93674611, 35.66773346],
       [34.96799611, 35.85523346],
       [ 9.95237111, 13.85523346],
       [11.06174611, 14.69898346],
       [33.90549611, 35.79273346],
       [ 9.03049611, 12.40210846],
       [32.28049611, 32.63648346],
       [13.74924611, 16.32398346],
       [35.56174611, 36.32398346],
       [11.21799611, 14.82398346],
       [14.99924611, 16.82398346],
       [34.37424611, 35.88648346],
       [34.49924611, 35.79273346],
       [ 6.96799611, 11.52710846],
       [15.62424611, 17.55835846],
       [33.59299611, 34.82398346],
       [14.96799611, 16.85523346],
       [30.03049611, 31.91773346],
       [12.21799611, 15.19898346],
       [28.62424611, 29.73023346],
       [13.88987111, 16.65210846],
       [11.82737111, 15.32398346],
       [35.21799611, 36.48023346],
       [23.90549611, 27.94898346],
       [16.49924611, 18.63648346],
       [37.09299611, 37.88648346],
       [ 7.99924611, 12.21460846],
       [32.24924611, 33.98023346],
       [32.96799611, 34.29273346],
       [24.43674611, 27.04273346],
       [13.49924611, 15.90210846],
       [27.46799611, 29.88648346],
       [11.31174611, 14.51148346],
       [14.73362111, 17.46460846],
       [33.90549611, 34.94898346],
       [18.37424611, 19.91773346],
       [29.43674611, 32.04273346],
       [10.03049611, 13.94898346],
       [34.46799611, 35.57398346],
       [27.74924611, 30.73023346],
       [28.87424611, 29.94898346],
       [25.62424611, 29.10523346],
       [12.46799611, 15.77710846],
       [34.71799611, 35.57398346],
       [17.18674611, 18.76148346],
       [17.87424611, 19.57398346],
       [31.34299611, 32.38648346],
       [ 5.74924611, 10.60523346],
       [13.34299611, 15.77710846],
       [27.62424611, 29.98023346],
       [29.71799611, 32.13648346],
       [30.81174611, 33.35523346],
       [13.65549611, 16.73023346],
       [27.65549611, 30.91773346],
       [ 9.74924611, 13.66773346],
       [23.62424611, 27.16773346],
```

```
[32.71799611, 34.26148346],
[17.96799611, 19.26148346],
[15.46799611, 17.58960846],
[17.28049611, 19.19898346],
[27.34299611, 29.82398346],
[32.31174611, 34.23023346],
[14.53049611, 17.16773346],
[30.62424611, 33.16773346],
[15.40549611, 17.91773346],
[12.12424611, 15.51148346],
[31.21799611, 32.38648346],
[34.87424611, 36.51148346],
[26.43674611, 30.38648346],
[35.24924611, 36.38648346],
[13.09299611, 15.98023346],
[13.68674611, 16.60523346],
[ 8.93674611, 12.54273346],
[31.90549611, 32.66773346],
[ 9.71799611, 13.77710846],
[27.43674611, 30.73023346],
[32.96799611, 34.16773346],
[34.87424611, 35.88648346],
[ 6.88987111, 11.60523346],
[15.28049611, 17.98023346],
[39.84299611, 40.07398346],
[14.81174611, 17.19898346],
[36.74924611, 37.94898346],
[15.81174611, 17.91773346],
[11.40549611, 14.76148346],
[32.59299611, 34.29273346],
[34.21799611, 35.01148346]])
```

Output the error scores and scores of the Linear Regression predicition.

In [11]:

```
print('Mean Absolute Error', str(mean_absolute_error(y_test, y_pred)).rjust(3
print('Root Mean Squared Error', str(sqrt(mean_squared_error(y_test, y_pred))
print('Mean Squared Absolute Error',str(mean_squared_error(y_test, y_pred)).r
print('R2 Score', str(r2_score(y_test, y_pred)).rjust(41))
```

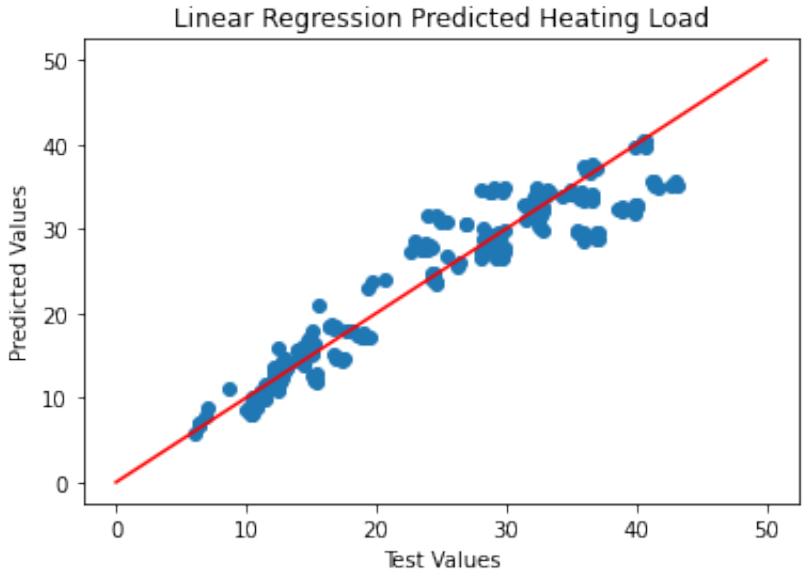
```
Mean Absolute Error             2.1419015824626975
Root Mean Squared Error         3.064763987733124
Mean Squared Absolute Error     9.39277830050584
R2 Score                        0.9036148603886349
```

Plot a graph of the predicted Linear Regression data for Heating Load vs the real test data.

In [12]:

```
plt.scatter(y_test.iloc[:,0], y_pred[:,0])
plt.plot([0,50], [0, 50],"r-")
plt.title('Linear Regression Predicted Heating Load')
plt.xlabel('Test Values')
plt.ylabel('Predicted Values')
```

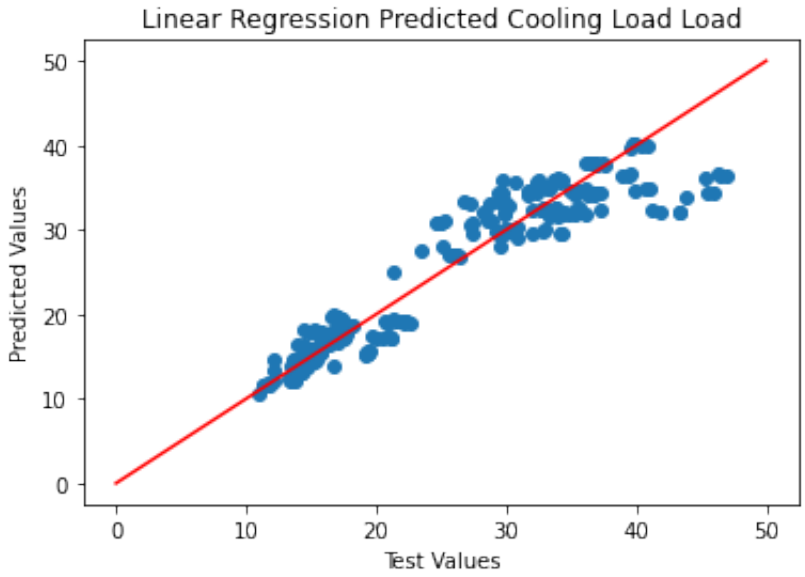
Out[12]: Text(0, 0.5, 'Predicted Values')



Plot a graph of the predicted Linear Regression data for Cooling Load vs the real test data.

In [13]:
```
plt.scatter(y_test.iloc[:,1], y_pred[:,1])
plt.plot([0,50], [0, 50],"r-")
plt.title('Linear Regression Predicted Cooling Load Load')
plt.xlabel('Test Values')
plt.ylabel('Predicted Values')
```

Out[13]: Text(0, 0.5, 'Predicted Values')



Ridge Regression

Perform Ridge Regression on the normalized data.

In [14]:
```
ridge_reg = Ridge(alpha=0.2)
ridge_reg.fit(X_train, y_train)
```

Out[14]:
```
Ridge(alpha=0.2)
```

Predict the target elements basae on the testing data.

In [15]:
```
y_pred_ridge = ridge_reg.predict(X_test)
y_pred_ridge
```

Out[15]:
```
array([[19.5807136 , 20.50538015],
       [13.82020669, 16.47584962],
       [31.76753812, 32.46170392],
       [35.70669872, 36.5528955 ],
       [15.722287  , 17.67189224],
       [28.63857479, 31.09305205],
       [25.59483348, 28.25551284],
       [28.56962092, 30.27529177],
       [17.90176306, 19.47597074],
       [27.67519609, 29.96023556],
       [18.30604857, 19.90397025],
       [33.36148488, 34.50903725],
       [27.81437781, 30.31633492],
       [ 8.18420115, 11.77535555],
       [18.96217822, 19.76357132],
       [35.76409071, 36.47974896],
       [36.51742806, 36.79018247],
       [10.85191077, 13.81796741],
       [14.58293578, 16.73529935],
       [33.31951086, 34.60377032],
       [34.23909782, 35.25017591],
       [34.99533539, 36.14772892],
       [10.80344523, 14.55080376],
       [29.92501994, 32.13440535],
       [13.05457738, 15.62928038],
       [29.88304592, 32.22913842],
       [33.2337451 , 34.21209249],
       [34.11317575, 35.53437513],
       [14.96894965, 17.36145874],
       [16.52197812, 18.67506736],
       [ 7.4308638 , 11.46492205],
       [10.23775514, 13.86363327],
       [15.45718321, 17.5999921 ],
       [30.77607683, 32.99186362],
       [29.00378649, 32.20290413],
       [28.48567287, 30.46475791],
       [31.37651226, 32.67400299],
       [29.04576052, 32.10817106],
       [33.19177108, 34.30682557],
       [ 8.36825711, 12.62384135],
       [29.28098424, 30.68045834],
       [10.89388479, 13.72323434],
       [ 6.14794438, 11.14226653],
```

```
[32.6890284 , 34.33727766],
[ 7.33866136, 11.93314258],
[ 6.10597036, 11.2369996 ],
[ 8.54115847, 13.16396202],
[ 8.41023114, 12.52910827],
[29.22600741, 32.07896536],
[31.01229502, 32.50274707],
[33.49884889, 34.28399264],
[14.88062185, 16.76345019],
[13.95938841, 16.83194898],
[32.73191128, 34.53311657],
[12.31951166, 15.120687  ],
[12.74749956, 15.65211331],
[10.32170319, 13.67416712],
[31.75254394, 33.46893034],
[39.22711171, 38.73806126],
[36.02919451, 36.55164911],
[16.44304206, 18.02607505],
[20.11092119, 20.64918043],
[14.75469978, 17.04764941],
[28.81973053, 31.35441834],
[29.23901022, 30.77519142],
[29.79909787, 32.41860456],
[29.84107189, 32.32387149],
[10.88739327, 14.36133761],
[13.8542685 , 16.6324555 ],
[10.58680698, 13.74606727],
[27.06104046, 30.00590141],
[32.2977457 , 32.6055042 ],
[13.5941767 , 15.7220969 ],
[28.06639318, 31.04398483],
[32.42392461, 34.26537751],
[10.27972916, 13.7689002 ],
[34.68535735, 35.58344235],
[31.30628441, 33.13566391],
[31.23733054, 32.31790363],
[24.11485529, 28.38814547],
[26.16701509, 28.30458005],
[32.85783335, 34.24891735],
[ 8.89846469, 12.76764163],
[16.64790019, 18.39086814],
[11.37562683, 14.59987098],
[15.37824715, 16.95099979],
[26.82340158, 30.10825365],
[35.4570129 , 36.50258189],
[29.08773454, 32.01343799],
[14.84302758, 17.64565795],
[12.259266  , 15.41357995],
[30.95723257, 33.25322991],
[ 9.5589741 , 13.41471739],
[14.61551806, 16.69155005],
[35.98722048, 36.64638218],
[14.53157001, 16.8810162 ],
[27.39558318, 30.15732086],
[30.3283966 , 32.27183287],
[ 9.24751653, 12.65007563],
[ 8.45220516, 12.4343752 ],
```

```
[38.43180034, 38.52236082],
[28.22056908, 30.39285777],
[ 7.16576001, 11.3930219 ],
[28.20540368, 30.10403584],
[29.79818901, 32.12803258],
[19.31560981, 20.43348   ],
[27.71734134, 31.16155083],
[19.84581739, 20.57728029],
[15.87086046, 17.97700783],
[10.01462537, 13.69700005],
[32.74732925, 34.5547031 ],
[33.34515806, 34.19687873],
[25.32972969, 28.18361269],
[17.19927963, 18.92362806],
[10.50285893, 13.93553341],
[14.54096175, 16.83003243],
[10.27033742, 13.81988397],
[ 7.69596759, 11.53682219],
[13.55220267, 15.81682997],
[24.79875371, 28.61905967],
[27.94029989, 30.0321357 ],
[10.53544121, 13.89178411],
[30.21993871, 33.09596122],
[16.25687433, 18.60316722],
[27.36811828, 29.98306849],
[18.83625615, 20.04777053],
[16.17793827, 17.9541749 ],
[30.06420166, 32.49050471],
[26.12504106, 28.39931312],
[38.38982631, 38.6170939 ],
[23.06744293, 26.79677095],
[33.23556281, 34.79323646],
[16.29884835, 18.50843414],
[13.56159441, 15.7658462 ],
[14.49898773, 16.9247655 ],
[18.61312638, 19.88113732],
[33.21923598, 34.48107794],
[35.66472469, 36.64762857],
[13.17056855, 15.97814527],
[18.12489283, 19.64260396],
[ 8.32628309, 12.71857442],
[39.14316366, 38.9275274 ],
[29.05785448, 30.51382513],
[31.05426904, 32.40801399],
[33.76395269, 34.35589278],
[11.11052304, 14.52797083],
[ 7.87712333, 11.79818848],
[27.01906644, 30.10063449],
[34.02905648, 34.42779293],
[32.2017037 , 34.38931628],
[32.92666729, 34.23492542],
[29.61703327, 31.86666629],
[29.35192948, 31.79476615],
[12.0963819 , 14.95405378],
[27.84326341, 30.87735162],
[18.0829188 , 19.73733703],
[29.53308522, 32.05613243],
```

```
[29.99385388, 32.12041342],
[11.55678258, 14.86123726],
[30.34290572, 32.00284742],
[32.88469326, 34.3296585 ],
[31.97857393, 34.22268306],
[19.76186935, 20.76674644],
[11.95218819, 15.43641288],
[15.19207941, 17.52809195],
[12.70552554, 15.74684638],
[28.28952295, 31.21061805],
[32.52419949, 34.38806989],
[33.62658868, 34.58093739],
[ 9.74012984, 13.67608368],
[10.96456646, 14.65098104],
[32.48222546, 34.48280296],
[ 8.22617517, 11.68062248],
[31.99066789, 32.62833713],
[14.16134631, 16.60962257],
[35.52554297, 36.29152921],
[11.02657499, 14.71743698],
[14.04982797, 16.00437956],
[33.05440707, 34.53187017],
[33.09638109, 34.4371371 ],
[ 6.23189243, 10.95280039],
[15.73167874, 17.62090847],
[33.14979705, 34.40155864],
[14.00785395, 16.09911263],
[30.07780192, 31.93094727],
[12.30124003, 15.31884688],
[29.0998285 , 30.41909206],
[12.94743878, 15.81151206],
[11.2590965 , 14.83308642],
[35.17649113, 36.40909521],
[24.84072773, 28.5243266 ],
[16.89220182, 18.94646098],
[35.72211669, 36.57448204],
[ 7.12378598, 11.48775498],
[32.32762577, 34.10511706],
[33.08096312, 34.41555057],
[25.85993727, 28.32741298],
[13.51022865, 15.91156304],
[28.17859506, 30.48759084],
[11.15249707, 14.43323776],
[14.19190991, 16.94759843],
[33.41490085, 34.47345878],
[19.71989532, 20.86147951],
[29.75621499, 32.22276565],
[ 9.77384948, 13.86010499],
[34.1971238 , 35.34490899],
[28.1083672 , 30.94925176],
[29.32295827, 30.58572527],
[26.41911607, 29.68025414],
[12.40783946, 15.71869554],
[34.46222759, 35.41680913],
[17.72948722, 19.06742834],
[19.05050602, 20.36157986],
[31.19535652, 32.4126367 ],
```

```
[ 4.94544727,  9.91144709],
[13.28709888, 15.74492983],
[28.0794816 , 30.38823506],
[30.02131878, 32.29466579],
[30.91525854, 33.34796298],
[13.69428461, 16.76004884],
[28.02441916, 31.13871791],
[ 9.51700007, 13.50945046],
[25.24578164, 28.37307884],
[32.81585933, 34.34365042],
[18.87823018, 19.95303746],
[14.98722129, 17.16329886],
[18.04094478, 19.8320701 ],
[27.85635184, 30.22160185],
[32.20079484, 34.0987443 ],
[14.92697562, 17.45619181],
[30.69212878, 33.18132977],
[16.03374456, 18.436534  ],
[12.17531795, 15.60304609],
[31.09624306, 32.31328092],
[34.86941332, 36.43192814],
[26.82544779, 30.62659625],
[35.21846516, 36.31436214],
[13.44998299, 16.20445599],
[13.73625864, 16.66531576],
[ 8.14222712, 11.87008862],
[31.68359007, 32.65117006],
[ 9.47502605, 13.60418354],
[27.80128939, 30.97208469],
[32.81495047, 34.05307844],
[34.64338333, 35.67817542],
[ 6.1899184 , 11.04753346],
[15.99177054, 18.53126707],
[38.57098205, 38.87846018],
[14.27585796, 16.75813228],
[35.37306485, 36.69204804],
[16.60592617, 18.48560121],
[11.33365281, 14.69460405],
[32.46589863, 34.17064444],
[33.68000464, 34.54535893]])
```

Output the error scores and scores of the Ridge Regression predicition.

In [16]:

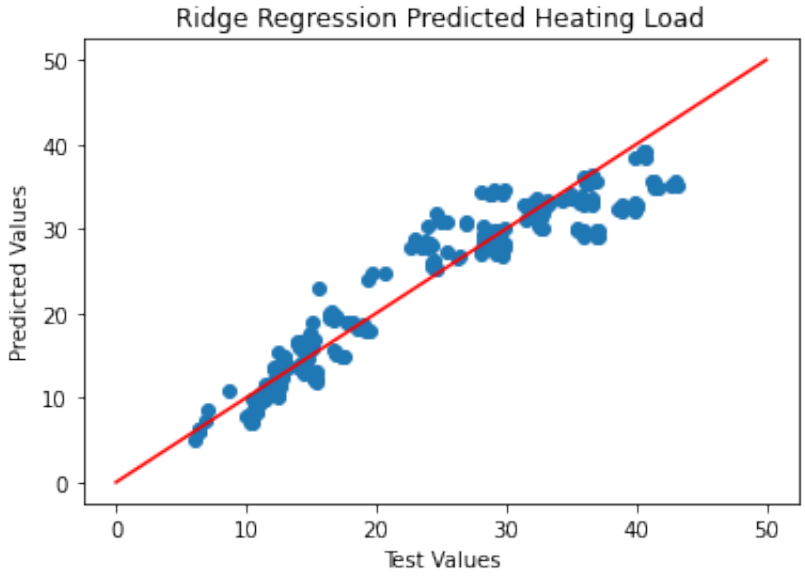
```
print('Mean Absolute Error', str(mean_absolute_error(y_test, y_pred_ridge)).r
print('Root Mean Squared Error', str(sqrt(mean_squared_error(y_test, y_pred_r
print('Mean Squared Absolute Error',str(mean_squared_error(y_test, y_pred_rid
print('R2 Score', str(r2_score(y_test, y_pred_ridge)).rjust(41))
```

```
Mean Absolute Error             2.2628656618312797
Root Mean Squared Error         3.130251468619064
Mean Squared Absolute Error     9.798474256791806
R2 Score                        0.8995346011629067
```

Plot a graph of the predicted Ridge Regression data for Heating Load vs the real test data.

In [17]:
```
plt.scatter(y_test.iloc[:,0], y_pred_ridge[:,0])
plt.plot([0,50], [0, 50],"r-")
plt.title('Ridge Regression Predicted Heating Load')
plt.xlabel('Test Values')
plt.ylabel('Predicted Values')
```

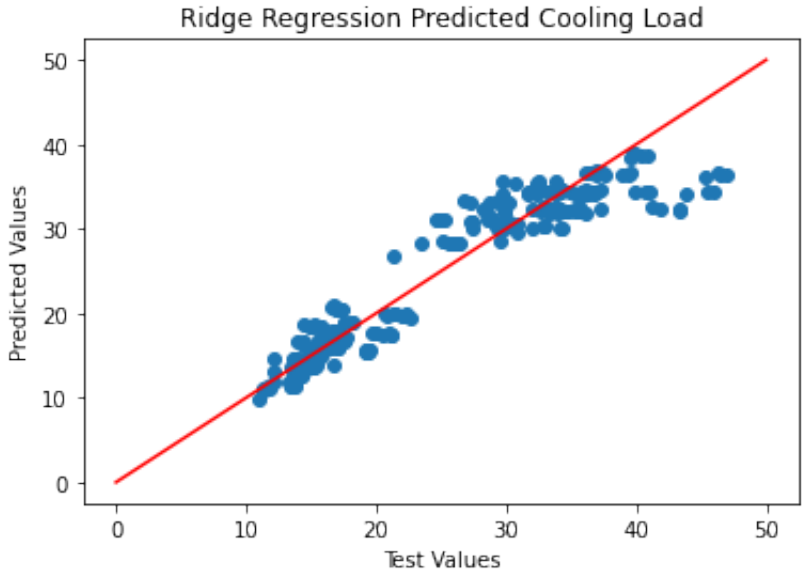
Out[17]:
```
Text(0, 0.5, 'Predicted Values')
```



Plot a graph of the predicted Rdige Regression data for Cooling Load vs the real test data.

In [18]:
```
plt.scatter(y_test.iloc[:,1], y_pred_ridge[:,1])
plt.plot([0,50], [0, 50],"r-")
plt.title('Ridge Regression Predicted Cooling Load')
plt.xlabel('Test Values')
plt.ylabel('Predicted Values')
```

Out[18]: Text(0, 0.5, 'Predicted Values')



Lasso Regression

Perform Lasso Regression on the normalized data.

In [19]:
```
lasso_reg = Lasso(alpha=0.2)
lasso_reg.fit(X_train, y_train)
```

Out[19]: Lasso(alpha=0.2)

Predict the target elements based on the testing data.

In [20]:
```
y_pred_lasso = lasso_reg.predict(X_test)
y_pred_lasso
```

Out[20]:
```
array([[18.05086054, 19.19248427],
       [13.49281914, 16.33850139],
       [31.02031791, 31.92330173],
       [33.23124127, 33.98456156],
       [15.87402926, 17.90830114],
       [29.39937251, 32.07206987],
       [28.08212622, 30.84946965],
       [29.36528043, 31.29499317],
       [16.04153368, 17.80546956],
       [29.4888763 , 31.99889449],
       [16.78530021, 18.17103963],
       [34.9230009 , 36.2674581 ],
       [29.86075956, 32.18167953],
       [ 9.90051697, 13.55978409],
       [17.52906673, 18.5366097 ],
       [35.46238859, 36.34741155],
       [36.57803838, 36.89576666],
```

```
[10.81178792, 13.82252257],
[12.61673594, 14.92392067],
[34.9230009 , 36.2674581 ],
[31.59379767, 32.78033188],
[32.11559148, 33.43620645],
[12.20966492, 15.89297787],
[30.66493285, 33.09351451],
[14.96275831, 17.64556265],
[30.66493285, 33.09351451],
[31.68330146, 32.70715651],
[31.59379767, 32.78033188],
[14.75837947, 17.35994604],
[16.76770633, 18.74696075],
[ 8.78486718, 13.01142898],
[10.06802139, 13.4569525 ],
[15.502146  , 17.72551611],
[29.56687694, 31.96923828],
[29.54928306, 32.5451594 ],
[29.36528043, 31.29499317],
[30.64843464, 31.7405167 ],
[29.54928306, 32.5451594 ],
[31.68330146, 32.70715651],
[ 9.67854425, 13.85008859],
[30.48093022, 31.84334828],
[10.81178792, 13.82252257],
[ 8.45122485, 13.37028787],
[30.32823733, 31.75888724],
[ 9.71678519, 14.39173251],
[ 8.45122485, 13.37028787],
[10.98234552, 15.41317715],
[ 9.67854425, 13.85008859],
[28.3013166 , 30.94779364],
[31.51579703, 32.80998809],
[32.05518472, 32.88994154],
[14.98035219, 17.06964154],
[13.8647024 , 16.52128643],
[31.94808706, 33.53903803],
[12.44923151, 15.02675225],
[14.59087505, 17.46277762],
[10.06802139, 13.4569525 ],
[31.05440999, 32.70037843],
[37.48930932, 37.15850514],
[35.83427185, 36.53019658],
[14.99794606, 16.49372042],
[18.79462707, 19.55805434],
[14.98035219, 17.06964154],
[29.77125577, 32.2548549 ],
[30.48093022, 31.84334828],
[30.66493285, 33.09351451],
[30.66493285, 33.09351451],
[12.20966492, 15.89297787],
[15.85643538, 18.48422226],
[10.43990465, 13.63973754],
[28.74510977, 31.63332442],
[31.76408443, 32.2888718 ],
[13.71479185, 16.04819689],
[28.65560599, 31.7064998 ],
```

```
[29.95635407, 31.5761022 ],
[10.06802139, 13.4569525 ],
[32.3375642 , 33.14590195],
[30.31064346, 32.33480836],
[30.27655138, 31.55773166],
[25.79084299, 30.0224694 ],
[28.82589275, 31.21503972],
[31.94808706, 33.53903803],
[10.42231078, 14.21565866],
[16.76770633, 18.74696075],
[12.95343145, 16.25854794],
[13.73238572, 15.47227578],
[27.76192891, 30.86784019],
[35.09050533, 36.16462651],
[29.54928306, 32.5451594 ],
[14.75837947, 17.35994604],
[13.84710852, 17.09720755],
[29.9387602 , 32.15202332],
[10.94410459, 14.87153323],
[14.60846892, 16.8868565 ],
[35.83427185, 36.53019658],
[14.60846892, 16.8868565 ],
[28.50569544, 31.23341026],
[29.78884965, 31.67893378],
[10.79419404, 14.39844369],
[ 9.67854425, 13.85008859],
[36.37365954, 36.61015004],
[28.99339717, 31.11220814],
[ 8.41298392, 12.82864394],
[30.23264282, 32.36446456],
[29.04508312, 31.31336371],
[17.67897728, 19.00969924],
[28.28372272, 31.52371476],
[18.42274381, 19.37526931],
[14.25417954, 16.12815035],
[ 9.69613813, 13.27416747],
[34.17923438, 35.90188803],
[31.07200386, 32.12445731],
[27.71024296, 30.66668461],
[15.51973987, 17.14959499],
[10.43990465, 13.63973754],
[12.61673594, 14.92392067],
[12.05975438, 15.41988833],
[ 9.15675044, 13.19421402],
[13.71479185, 16.04819689],
[27.51779037, 31.1535237 ],
[29.86075956, 32.18167953],
[12.43163764, 15.60267337],
[32.58003171, 35.23930227],
[16.39582307, 18.56417571],
[29.11699303, 31.81610946],
[17.52906673, 18.5366097 ],
[14.6260628 , 16.31093538],
[31.03681611, 33.27629954],
[28.82589275, 31.21503972],
[36.37365954, 36.61015004],
[26.48292356, 30.18688389],
```

```
[34.9230009 , 36.2674581 ],
[16.39582307, 18.56417571],
[11.72305886, 14.08526106],
[12.61673594, 14.92392067],
[17.15718347, 18.35382467],
[31.07200386, 32.12445731],
[33.23124127, 33.98456156],
[11.3511756 , 13.90247603],
[16.41341694, 17.9882546 ],
[ 9.67854425, 13.85008859],
[37.48930932, 37.15850514],
[30.10904696, 31.66056324],
[31.51579703, 32.80998809],
[32.42706798, 33.07272658],
[12.58154819, 16.07576291],
[ 9.5286337 , 13.37699905],
[28.74510977, 31.63332442],
[32.79895124, 33.25551161],
[31.20432053, 33.17346796],
[31.31141819, 32.52437147],
[28.67319986, 31.13057868],
[28.3013166 , 30.94779364],
[12.07734825, 14.84396722],
[28.28372272, 31.52371476],
[16.41341694, 17.9882546 ],
[28.67319986, 31.13057868],
[30.02826398, 32.07884795],
[13.32531471, 16.44133298],
[30.40014724, 32.26163298],
[31.31141819, 32.52437147],
[30.83243727, 32.99068293],
[18.42274381, 19.37526931],
[13.47522526, 16.91442251],
[15.13026273, 17.54273107],
[14.59087505, 17.46277762],
[29.02748925, 31.88928483],
[33.80735112, 35.71910299],
[35.29488417, 36.45024313],
[11.31598785, 15.05431826],
[13.51346619, 17.45606644],
[33.80735112, 35.71910299],
[ 9.90051697, 13.55978409],
[31.39220117, 32.10608677],
[16.22831865, 18.66700729],
[32.85935801, 33.80177652],
[12.58154819, 16.07576291],
[12.46682539, 14.45083113],
[34.55111764, 36.08467306],
[34.55111764, 36.08467306],
[ 8.45122485, 13.37028787],
[13.88229627, 15.94536531],
[31.68330146, 32.70715651],
[12.46682539, 14.45083113],
[30.02826398, 32.07884795],
[13.84710852, 17.09720755],
[30.10904696, 31.66056324],
[10.97929234, 13.71969099],
```

```
[10.96169846, 14.29561211],
[32.48747474, 33.61899149],
[27.51779037, 31.1535237 ],
[15.14785661, 16.96680995],
[35.46238859, 36.34741155],
[ 8.41298392, 12.82864394],
[31.20432053, 33.17346796],
[32.31997032, 33.72182307],
[28.45400948, 31.03225468],
[13.71479185, 16.04819689],
[28.99339717, 31.11220814],
[12.58154819, 16.07576291],
[12.24485267, 14.74113563],
[32.05518472, 32.88994154],
[18.42274381, 19.37526931],
[29.04508312, 31.31336371],
[12.24790586, 16.43462179],
[31.59379767, 32.78033188],
[28.65560599, 31.7064998 ],
[30.48093022, 31.84334828],
[27.01816239, 30.50227012],
[12.2272588 , 15.31705675],
[31.96568093, 32.96311691],
[16.2635064 , 17.51516506],
[17.30709402, 18.8269142 ],
[30.27655138, 31.55773166],
[ 7.18566451, 12.34884323],
[13.34290859, 15.86541186],
[30.23264282, 32.36446456],
[29.41696639, 31.49614875],
[29.9387602 , 32.15202332],
[13.49281914, 16.33850139],
[28.65560599, 31.7064998 ],
[10.94410459, 14.87153323],
[27.71024296, 30.66668461],
[31.94808706, 33.53903803],
[17.52906673, 18.5366097 ],
[13.36050246, 15.28949074],
[16.41341694, 17.9882546 ],
[29.86075956, 32.18167953],
[29.58447081, 31.39331717],
[14.75837947, 17.35994604],
[29.56687694, 31.96923828],
[16.02393981, 18.38139068],
[13.84710852, 17.09720755],
[31.51579703, 32.80998809],
[32.11559148, 33.43620645],
[28.32196366, 32.06535868],
[32.48747474, 33.61899149],
[15.11266886, 18.11865219],
[13.49281914, 16.33850139],
[ 9.90051697, 13.55978409],
[31.02031791, 31.92330173],
[10.94410459, 14.87153323],
[28.28372272, 31.52371476],
[30.32823733, 31.75888724],
[32.3375642 , 33.14590195],
```

```
[ 8.45122485, 13.37028787],
[16.02393981, 18.38139068],
[36.7455428 , 36.79293507],
[12.24485267, 14.74113563],
[35.09050533, 36.16462651],
[16.76770633, 18.74696075],
[12.95343145, 16.25854794],
[29.95635407, 31.5761022 ],
[32.42706798, 33.07272658]])
```

In [21]:

```
print('Mean Absolute Error', str(mean_absolute_error(y_test, y_pred_lasso)).r
print('Root Mean Squared Error', str(sqrt(mean_squared_error(y_test, y_pred_l
print('Mean Squared Absolute Error',str(mean_squared_error(y_test, y_pred_las
print('R2 Score', str(r2_score(y_test, y_pred_lasso)).rjust(41))
```

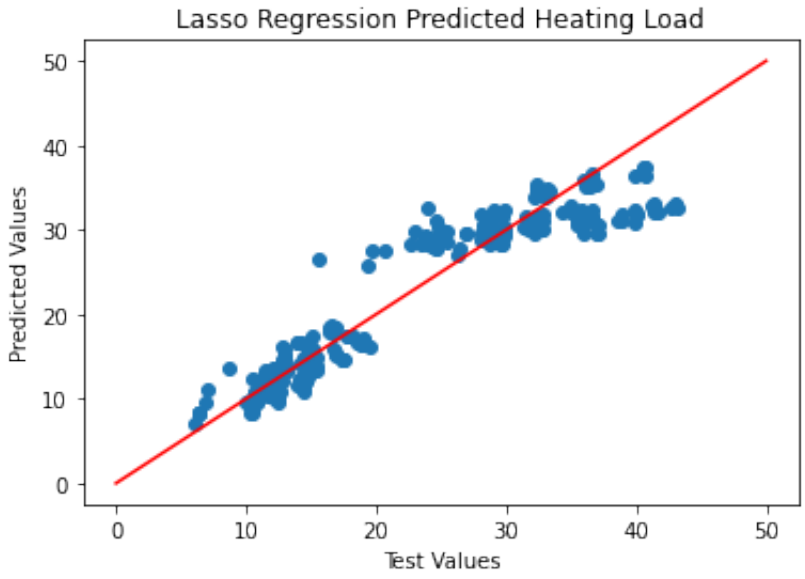
```
Mean Absolute Error              2.6244742250819693
Root Mean Squared Error            3.55975563075367
Mean Squared Absolute Error      12.671860150682459
R2 Score                         0.8700113674552772
```

Plot a graph of the predicted Lasso Regression data for Heating Load vs the real test data.

In [22]:

```
plt.scatter(y_test.iloc[:,0], y_pred_lasso[:,0])
plt.plot([0,50], [0, 50],"r-")
plt.title('Lasso Regression Predicted Heating Load')
plt.xlabel('Test Values')
plt.ylabel('Predicted Values')
```

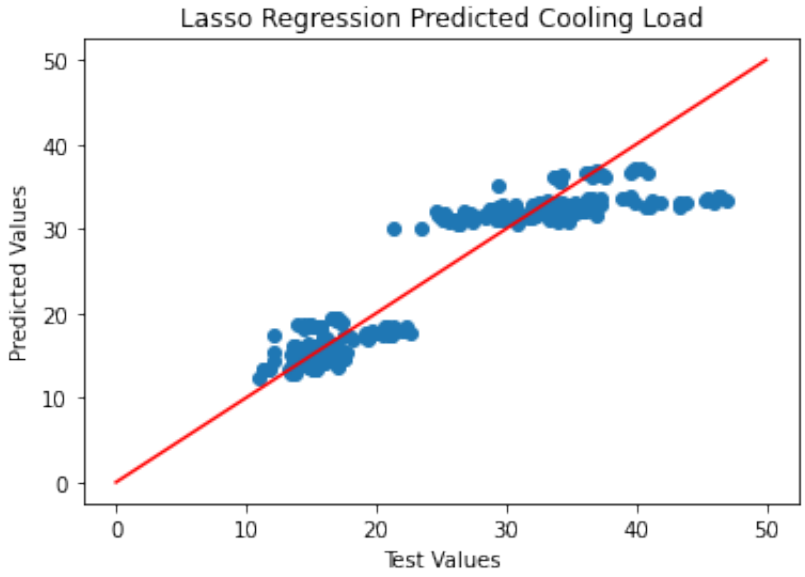
Out[22]: Text(0, 0.5, 'Predicted Values')



Plot a graph of the predicted Lasso Regression data for Cooling Load vs the real test data.

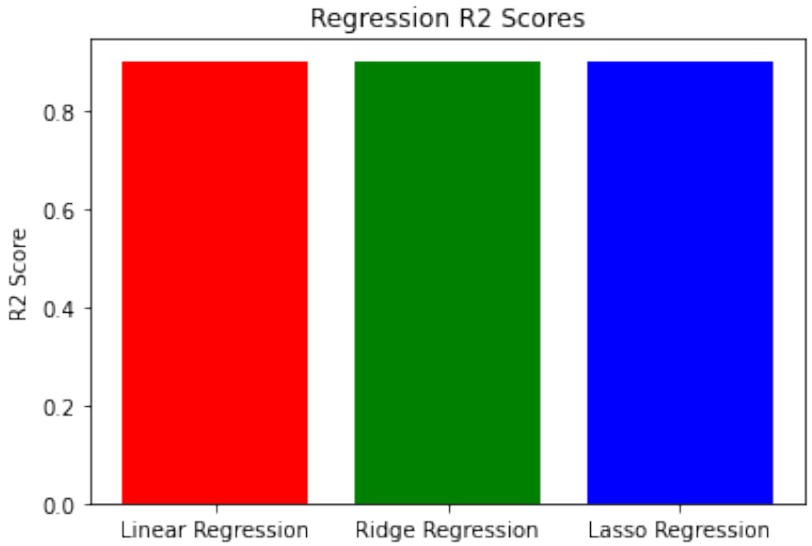
In [23]:
```
plt.scatter(y_test.iloc[:,1], y_pred_lasso[:,1])
plt.plot([0,50], [0, 50],"r-")
plt.title('Lasso Regression Predicted Cooling Load')
plt.xlabel('Test Values')
plt.ylabel('Predicted Values')
```

Out[23]: Text(0, 0.5, 'Predicted Values')



In [24]:
```
r2_linear=r2_score(y_test, y_pred)
r2_ridge=r2_score(y_test, y_pred_ridge)
r2_lasso=r2_score(y_test, y_pred_lasso)
r2=[r2_linear,r2_linear,r2_linear]
plt.bar(['Linear Regression','Ridge Regression', 'Lasso Regression'], r2,colo
plt.title('Regression R2 Scores')
plt.ylabel('R2 Score')
```

Out[24]: Text(0, 0.5, 'R2 Score')



Classification Dataset

Exploration and Preprocessing

In [25]:
```
df = pd.read_csv('creditcard.csv')
df
```

Out[25]:

	Time	V1	V2	V3	V4	V5	V6	
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.2395
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.0788
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.7914
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.2376
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.5929
...	...	...	...	...	...	...	...	...
284802	172786.0	-11.881118	10.071785	-9.834783	-2.066656	-5.364473	-2.606837	-4.9182
284803	172787.0	-0.732789	-0.055080	2.035030	-0.738589	0.868229	1.058415	0.0243
284804	172788.0	1.919565	-0.301254	-3.249640	-0.557828	2.630515	3.031260	-0.2968
284805	172788.0	-0.240440	0.530483	0.702510	0.689799	-0.377961	0.623708	-0.6861
284806	172792.0	-0.533413	-0.189733	0.703337	-0.506271	-0.012546	-0.649617	1.5770

284807 rows × 31 columns

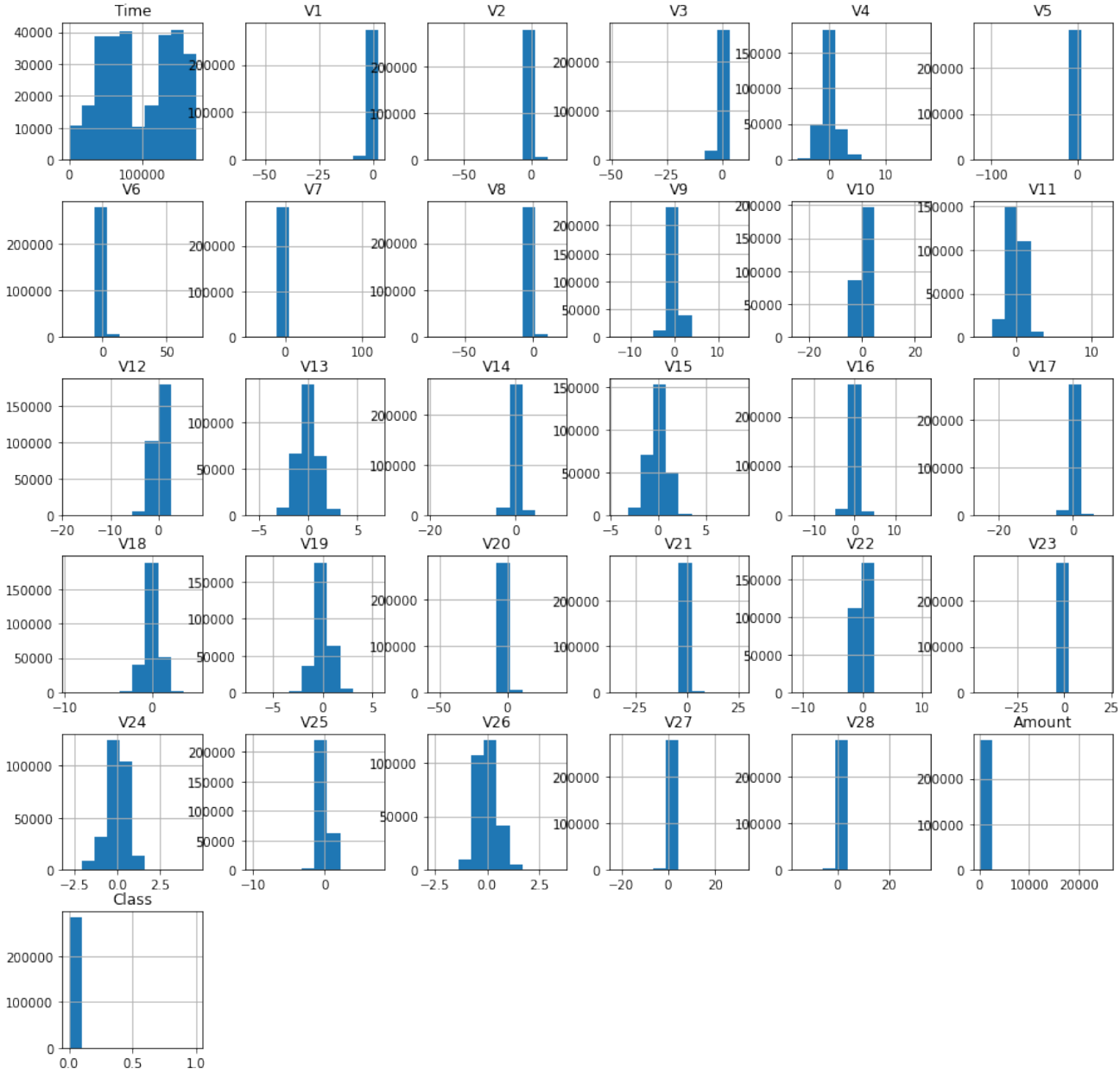
Gather infomation about the variables in the dataset.

In [26]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 284807 entries, 0 to 284806
Data columns (total 31 columns):
 #   Column  Non-Null Count   Dtype  
---  ------  --------------   -----  
 0   Time    284807 non-null  float64
 1   V1      284807 non-null  float64
 2   V2      284807 non-null  float64
 3   V3      284807 non-null  float64
 4   V4      284807 non-null  float64
 5   V5      284807 non-null  float64
 6   V6      284807 non-null  float64
 7   V7      284807 non-null  float64
 8   V8      284807 non-null  float64
 9   V9      284807 non-null  float64
 10  V10     284807 non-null  float64
 11  V11     284807 non-null  float64
 12  V12     284807 non-null  float64
 13  V13     284807 non-null  float64
 14  V14     284807 non-null  float64
 15  V15     284807 non-null  float64
 16  V16     284807 non-null  float64
 17  V17     284807 non-null  float64
 18  V18     284807 non-null  float64
 19  V19     284807 non-null  float64
 20  V20     284807 non-null  float64
 21  V21     284807 non-null  float64
 22  V22     284807 non-null  float64
 23  V23     284807 non-null  float64
 24  V24     284807 non-null  float64
 25  V25     284807 non-null  float64
 26  V26     284807 non-null  float64
 27  V27     284807 non-null  float64
 28  V28     284807 non-null  float64
 29  Amount  284807 non-null  float64
 30  Class   284807 non-null  int64  
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

Visualize the data through histograms looking for common frequencies amongst the data.

In [27]:
```
df.hist(figsize=(15,15))
plt.show()
```



Search for null values in the dataset.

In [28]:
```
df.isnull().sum()
```

Out[28]:
```
Time      0
V1        0
V2        0
V3        0
V4        0
V5        0
V6        0
V7        0
V8        0
V9        0
V10       0
V11       0
V12       0
V13       0
V14       0
V15       0
V16       0
V17       0
V18       0
V19       0
V20       0
V21       0
V22       0
V23       0
V24       0
V25       0
V26       0
V27       0
V28       0
Amount    0
Class     0
dtype: int64
```

There are no null values requiring data manipulation.

View the dispersion of the binary Class column.

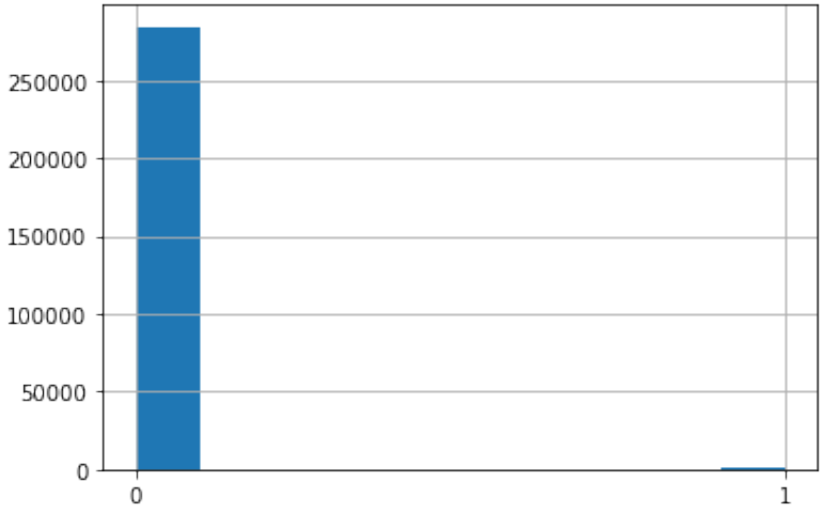
In [29]:
```
count=df.Class.value_counts()
print(count)
```

```
0    284315
1       492
Name: Class, dtype: int64
```

Histogram of the Class column.

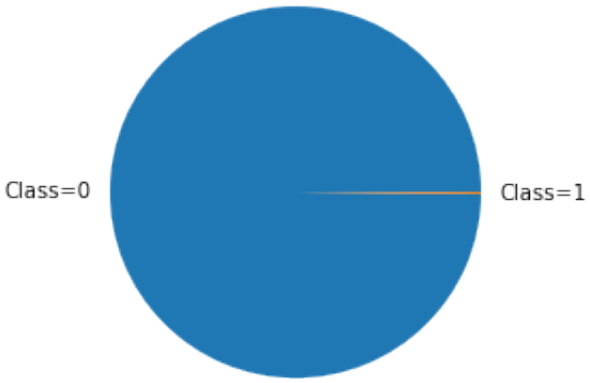
In [30]:
```
df.Class.hist()
plt.xticks([0,1])
```

Out[30]:
```
([<matplotlib.axis.XTick at 0x7f894a7228e0>,
  <matplotlib.axis.XTick at 0x7f894a7228b0>],
 [Text(0, 0, ''), Text(0, 0, '')])
```



View piechart of the data to properly show the significant skew imbalance.

In [31]:
```
l = ['Class=0','Class=1']
plt.pie(count,labels=l)
plt.show()
```



Separate the features and target variables.

In [32]:
```
X = df.drop(['Class'], axis=1)
y = df['Class']
X
```

Out[32]:

	Time	V1	V2	V3	V4	V5	V6	
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.2395
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.0788
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.7914
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.2376
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.5929
...	...	...	...	...	...	...	...	
284802	172786.0	-11.881118	10.071785	-9.834783	-2.066656	-5.364473	-2.606837	-4.9182
284803	172787.0	-0.732789	-0.055080	2.035030	-0.738589	0.868229	1.058415	0.0243
284804	172788.0	1.919565	-0.301254	-3.249640	-0.557828	2.630515	3.031260	-0.2968
284805	172788.0	-0.240440	0.530483	0.702510	0.689799	-0.377961	0.623708	-0.6861
284806	172792.0	-0.533413	-0.189733	0.703337	-0.506271	-0.012546	-0.649617	1.5770

284807 rows × 30 columns

Split the training and testing data.

In [33]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.4, ra
```

Alter the imbalanced data with the SMOTE function from imblearn.

In [34]:
```
smo = SMOTE(random_state=42)
X_smo, y_smo = smo.fit_resample(X_train,y_train)
count[0]=len(X_smo)
count[1]=len(y_smo)
```

Show the updated pie chart after imbalance of the dataset was altered by the SMOTE function.

In [35]:
```
plt.pie(count,labels=l)
plt.show()
```

Split our balanced data in a training and testing set.

In [36]:
```
scaler = StandardScaler()
X_train_scale = scaler.fit_transform(X_smo)
X_test_scale = scaler.transform(X_test)
```

Logistic Regression

Use GridsearchCV to determine the best parameters for the Logistic Regression model.

In [37]:
```
parameters = {
    'C': [0.01, 0.1, 1, 10, 10],
    'solver' : ["lbfgs","liblinear"]
}
lr = LogisticRegression(max_iter=1000,random_state=42)
glr = GridSearchCV(lr, parameters, cv=3, verbose=5, n_jobs=3)
gd=glr.fit(X_train_scale, y_smo)
gd
```

```
Fitting 3 folds for each of 10 candidates, totalling 30 fits
```

Out[37]:
```
GridSearchCV(cv=3, estimator=LogisticRegression(max_iter=1000, random_state=42
),
             n_jobs=3,
             param_grid={'C': [0.01, 0.1, 1, 10, 10],
                         'solver': ['lbfgs', 'liblinear']},
             verbose=5)
```

Print the best parameters for the GridSearchCV.

In [38]:
```
print("Best parameters : %s" % gd.best_params_)
```

```
Best parameters : {'C': 10, 'solver': 'lbfgs'}
```

Train the GridSearch best parameters on the Logistic Regression model.

In [39]:
```
log_reg = LogisticRegression(solver='lbfgs', random_state = 42, max_iter = 10
log_reg.fit(X_train_scale, y_smo)
```

Out[39]:
```
LogisticRegression(C=10, max_iter=1000, random_state=42)
```

Make predictions for the Logistic Regression model.

In [40]:
```
y_pred = log_reg.predict(X_test_scale)
y_pred
```

Out[40]:
```
array([1, 0, 0, ..., 0, 0, 0])
```

Show the confusion matrix.

In [41]:
```
print(confusion_matrix(y_test, y_pred))
```

```
[[112744    988]
 [    23    168]]
```

Print the accuracy score.

In [42]:
```
print(accuracy_score(y_test,y_pred))
```

```
0.9911255848248378
```

Print the classification report for the Logisitic Regression model.

In [43]:
```
target_names = ['class=0', 'class=1']
print(classification_report(y_test,y_pred,target_names=target_names))
```

```
              precision    recall  f1-score   support

     class=0       1.00      0.99      1.00    113732
     class=1       0.15      0.88      0.25       191

    accuracy                           0.99    113923
   macro avg       0.57      0.94      0.62    113923
weighted avg       1.00      0.99      0.99    113923
```

Decision Tree Classifier

Use GridsearchCV to determine the best parameters for the Decision Tree Classifer model.

In [44]:
```
parameters = {
    'criterion' : ["gini", "entropy"],
    'max_depth' : [10,12,15,20]
}
dtc = DecisionTreeClassifier(random_state=42)
glr = GridSearchCV(dtc, parameters, cv=3, verbose=5, n_jobs=3)
gd=glr.fit(X_train_scale, y_smo)
gd
```

```
Fitting 3 folds for each of 8 candidates, totalling 24 fits
```

Out[44]:
```
GridSearchCV(cv=3, estimator=DecisionTreeClassifier(random_state=42), n_jobs=3
,
             param_grid={'criterion': ['gini', 'entropy'],
                         'max_depth': [10, 12, 15, 20]},
             verbose=5)
```

Print the best parameters for the GridSearchCV of the Decision Tree Classifier.

In [45]:
```
print("Best parameters : %s" % gd.best_params_)
```

```
Best parameters : {'criterion': 'entropy', 'max_depth': 20}
```

Train the GridsearchCV parameters on the Decision Tree Clasifier model.

In [46]:
```
dtc=DecisionTreeClassifier(criterion='entropy',max_depth=20, random_state=42)
dtc.fit(X_train_scale, y_smo)
```

Out[46]:
```
DecisionTreeClassifier(criterion='entropy', max_depth=20, random_state=42)
```

Make predicitons with the Decision Tree Classifier model.

In [47]:
```
y_pred2 = dtc.predict(X_test_scale)
y_pred2
```

Out[47]:
```
array([1, 0, 0, ..., 0, 0, 0])
```

Print out the Decision Tree Classifier confusion matrix.

In [48]:
```
cm = confusion_matrix(y_test, y_pred2)
cm
```

Out[48]:
```
array([[113521,    211],
       [    35,    156]])
```

Print out the Decision Tree Classifier accuracy score.

In [49]:
```
print(accuracy_score(y_test,y_pred2))
```

```
0.9978406467526312
```

Print out the Decision Tree Classifier classification report.

In [50]:
```
target_names = ['class=0', 'class=1']
print(classification_report(y_test,y_pred2,target_names=target_names))
```

```
              precision    recall  f1-score   support

     class=0       1.00      1.00      1.00    113732
     class=1       0.43      0.82      0.56       191

    accuracy                           1.00    113923
   macro avg       0.71      0.91      0.78    113923
weighted avg       1.00      1.00      1.00    113923
```

Random Forest Classifier

Use GridsearchCV to determine the best parameters for the Ranfom Forest Classifer model.

In [51]:
```
parameters = {
    'max_depth' : [9,10,11],
    'max_features': list(range(1,4))
}
rfc = RandomForestClassifier(random_state=42)
glr = GridSearchCV(rfc, parameters, cv=3, verbose=5, n_jobs=3)
gd=glr.fit(X_train, y_train)
gd
```

```
Fitting 3 folds for each of 9 candidates, totalling 27 fits
```

Out[51]:
```
GridSearchCV(cv=3, estimator=RandomForestClassifier(random_state=42), n_jobs=3
,
             param_grid={'max_depth': [9, 10, 11], 'max_features': [1, 2, 3]},
             verbose=5)
```

Find the best parameters from the GridSearchCV function

In [52]:
```
print("Best parameters : %s" % gd.best_params_)
```

```
Best parameters : {'max_depth': 11, 'max_features': 3}
```

Apply the best parameters to the Random Forest Classifer Model.

In [53]:
```
rfc=RandomForestClassifier(random_state=42,max_depth=11, max_features= 3)
rfc.fit(X_train, y_train)
```

Out[53]:
```
RandomForestClassifier(max_depth=11, max_features=3, random_state=42)
```

Predict data using the Random Forest Classifier Model.

In [54]:
```
y_pred3 = rfc.predict(X_test)
y_pred3
```

Out[54]:
```
array([1, 0, 0, ..., 0, 0, 0])
```

Show the confusion matrix for the model.

In [55]:
```
cm = confusion_matrix(y_test, y_pred3)
cm
```

Out[55]:
```
array([[113724,      8],
       [    50,    141]])
```

Show the accuracy score for the model.

In [56]:
```
print(accuracy_score(y_test,y_pred3))
```

```
0.9994908841937098
```

Show the classification report for the Random Forest Classifier.

In [57]:
```
target_names = ['class=0', 'class=1']
print(classification_report(y_test,y_pred3,target_names=target_names))
```

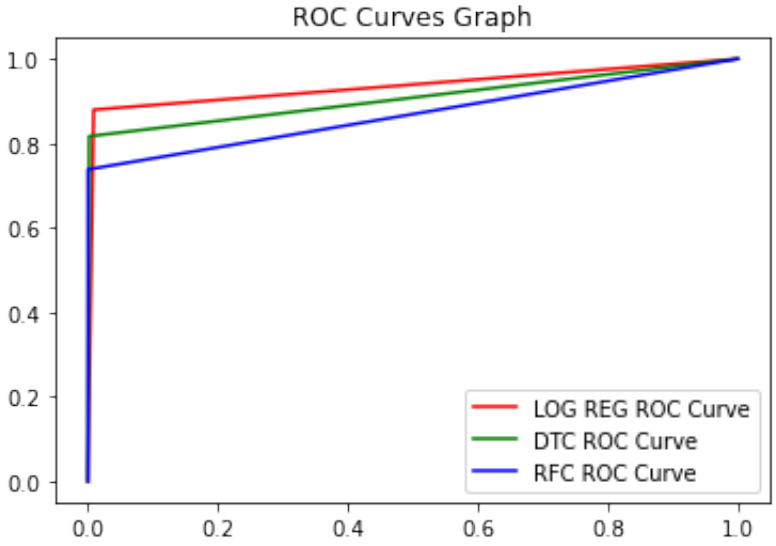
```
              precision    recall  f1-score   support

     class=0       1.00      1.00      1.00    113732
     class=1       0.95      0.74      0.83       191

    accuracy                           1.00    113923
   macro avg       0.97      0.87      0.91    113923
weighted avg       1.00      1.00      1.00    113923
```

Show the ROC cruve graph for all three classification models.

In [58]:
```
fpr_forest, tpr_forest, thresholds_forest = roc_curve(y_test, y_pred)
plt.plot(fpr_forest,tpr_forest, 'r-', label="LOG REG ROC Curve")
fpr_forest, tpr_forest, thresholds_forest = roc_curve(y_test, y_pred2)
plt.plot(fpr_forest,tpr_forest, 'g-', label="DTC ROC Curve")
fpr_forest, tpr_forest, thresholds_forest = roc_curve(y_test, y_pred3)
plt.plot(fpr_forest,tpr_forest, 'b-', label="RFC ROC Curve")
plt.title('ROC Curves Graph')
plt.legend()
plt.show()
```


ROC Curves Graph
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
LOG REG ROC Curve
DTC ROC Curve
RFC ROC Curve