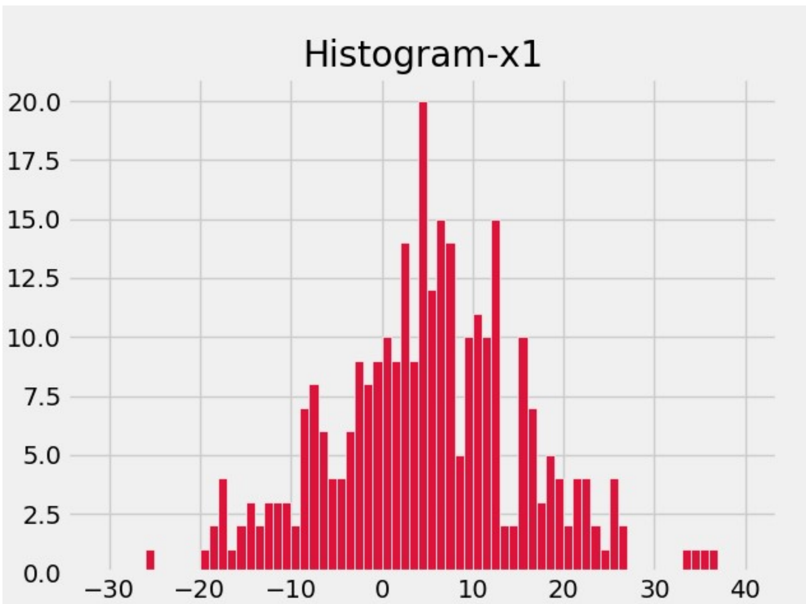
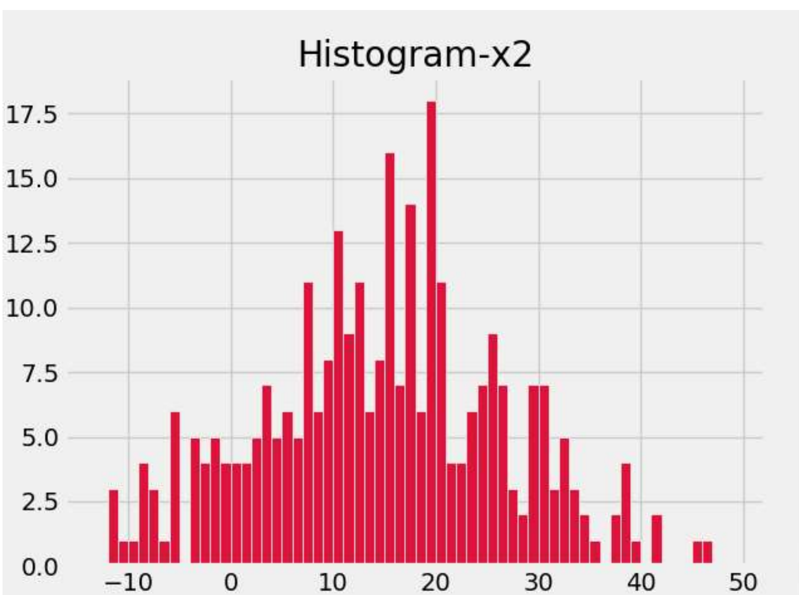
# ECE 592 005 – IOT Analytics

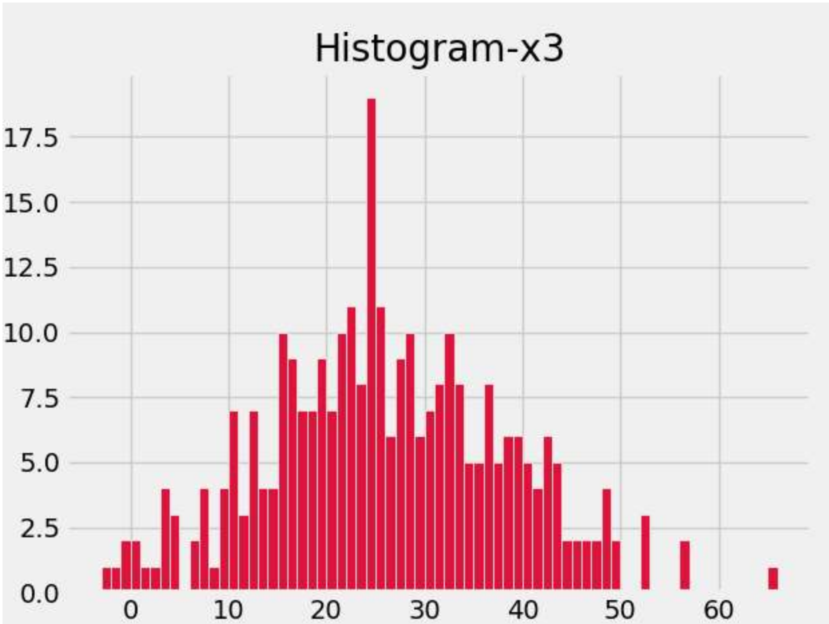
# Project 2 – Regression

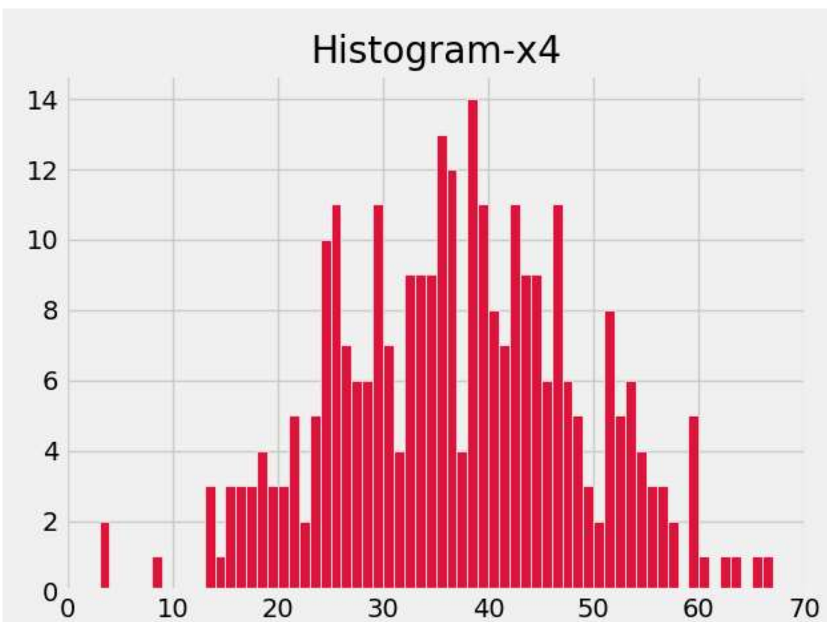
**Task-1:**

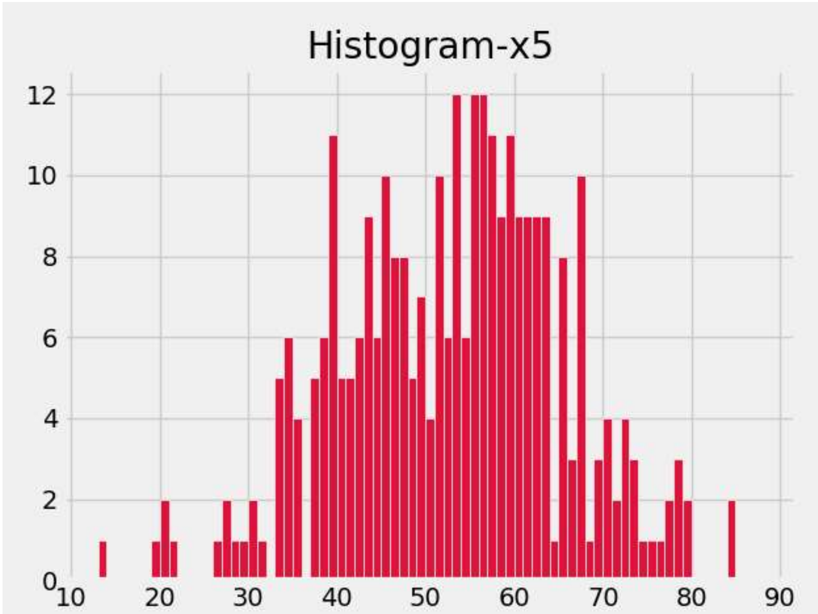
**1.1 Histograms for each column $x_i$:**




Histogram-x3
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
10
20
30
40
50
60


Histogram-x4
14
12
10
8
6
4
2
0
0
10
20
30
40
50
60
70



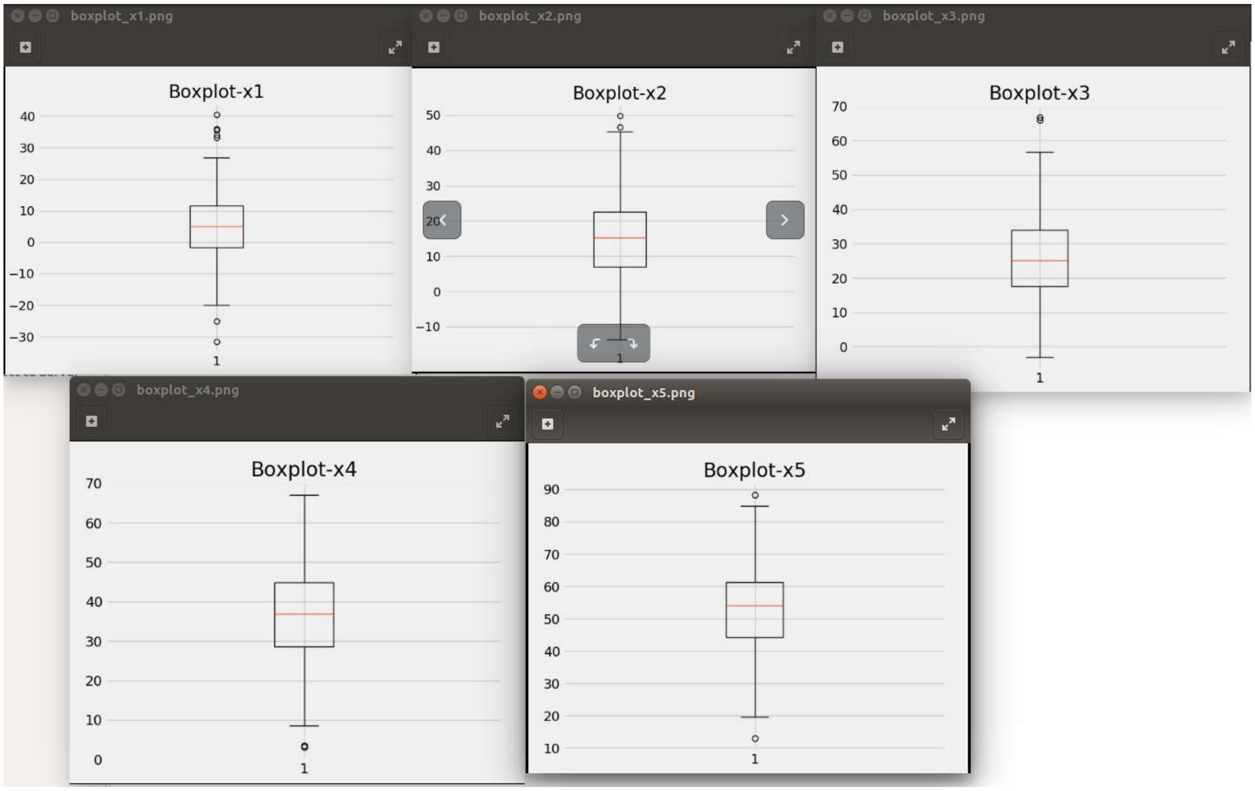

**Mean and Variance for each column $x_i$:**

| Column | Mean | Variance |
|---|---|---|
| $X_1$ | 4.990267333333332 | 119.3076884422069 |
| $X_2$ | 14.763730076666667 | 145.05252952563043 |
| $X_3$ | 26.085963666666668 | 154.44265868083923 |
| $X_4$ | 37.007495000000034 | 139.459924134075 |
| $X_5$ | 53.059329999999996 | 160.07008254776667 |

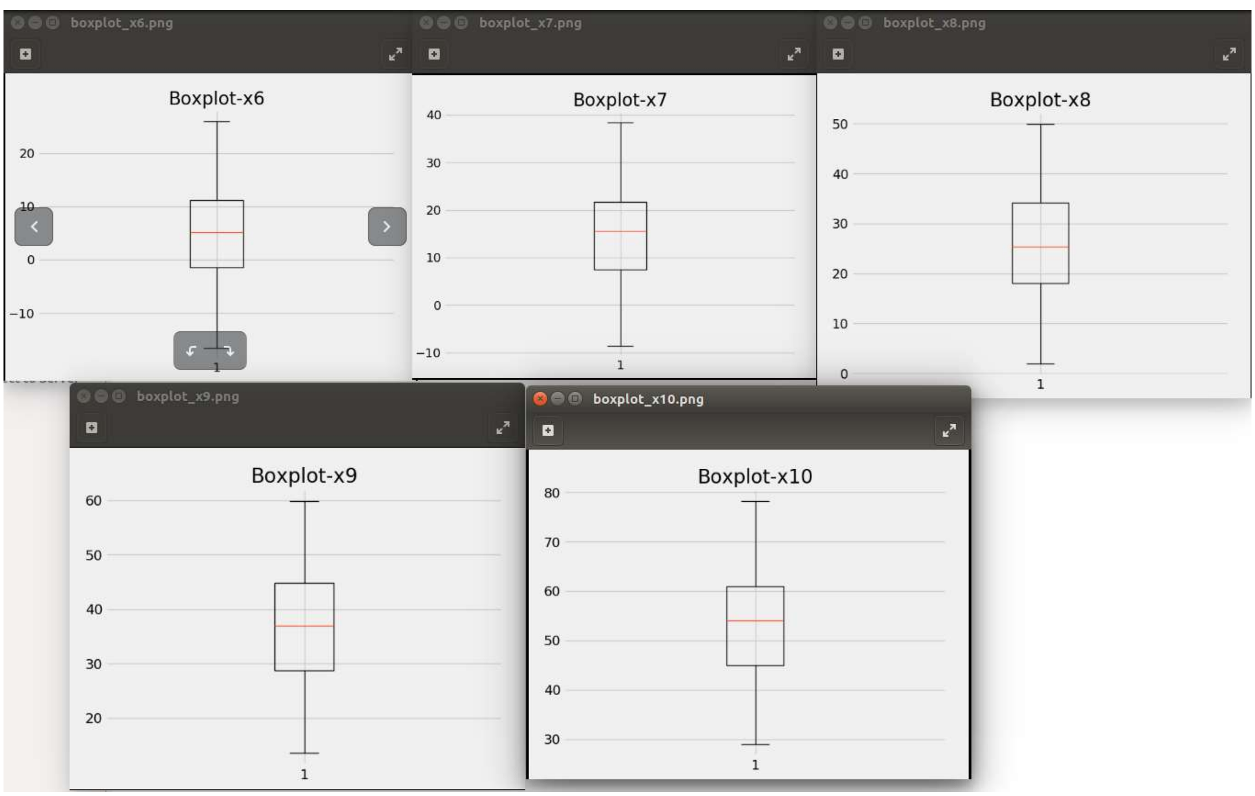
### 1.2 Detecting and Removing Outliers from the Dataset:

We use Box Plot to detect the outliers in the dataset, remove it using z-score and the dataset after removal of the outliers has reduced from 300 rows to 239 rows.

**Box-Plot with Outliers:**




**Box -Plot without Outliers:**



In the figure, x6 = x1 , x7 = x2 , x8 = x3 , x9 = x4 and x10 = x5.

**1.3 Correlation Matrix for all Variables:**

The correlation matrix found for y,x1,x2,x3,x4,x5 is,

| Y | X1 | X2 | X3 | X4 | X5 |
|---|---|---|---|---|---|
| 1. | 0.48361383 | -0.00954987 | 0.11076584 | 0.29588932 | 0.24886194 |
| 0.48361383 | 1. | -0.09212281 | -0.03527944 | -0.07825077 | -0.09368099 |
| -0.00954987 | -0.09212281 | 1. | -0.02991405 | 0.0579247 | 0.03714248 |
| 0.11076584 | -0.03527944 | -0.02991405 | 1. | -0.03479931 | -0.06683803 |
| 0.29588932 | -0.07825077 | 0.0579247 | -0.03479931 | 1. | 0.01418369 |
| 0.24886194 | -0.09368099 | 0.03714248 | -0.06683803 | 0.01418369 | 1. |

A correlation matrix has is a matrix that is has values between -1 and 1 and gives the linear relationship between the variables in the two dimensional matrix. The correlation co-efficient ie) each $x_{ij}$ gives the strength of the linear relationship between the two variables considered.
We draw conclusions based on the following idea,
(i)Positive correlation – the variables move in the same direction
(ii) Negative correlation – if one variable increases , the other variable decreases
(iii) If the correlation is greater than or equal to +/- 0.80 , the association between the variables are very strong and they have a high degree of correlation
(iv) Values between +/- 0.5 to +/- 0.8 means a sufficient degree of correlation
(v) If the correlation value is less than +/- 0.5 , the association between the variables are very weak and they have a lower degree of correlation

(i)Positive correlation – (y,x1) , (y,x3) , (y,x4) , (y,x5) , (x2,x4) , (x2,x5) , (x4,x5)
(ii) Negative correlation – (y,x2) , (x1,x2) , (x1,x3) , (x1,x4) , (x1,x5) , (x2,x3) , (x3,x4) , (x3,x5)

1.4 :

(i)On observing the histogram, we see that there are several outliers present in every $x_i$. The outliers should be removed in order to produce better regression results.
(ii) Using a z-score of 3 still did not remove some outliers and hence the z-score was changed to be 2. Hence all data that have a z-score < 2 are retained in the dataset.
(iii) We see that almost all independent variables have a weak / lower degree of correlation.
On observing the correlation between Y and a $x_i$ , we see that only x1 has an average degree of correlation while the others have a lower degree of correlation. These variables are not going to have much of an effect while performing regression.

**Task-2:**

Task2 is run on the dataset which has no outliers. Outliers were removed using z-score in the previous task and the new dataset is stored in a separate csv for this task.

Statsmodel package in python was used to perform simple Linear Regression on Y and the column x1.

ie) the model $Y = a_0 + a_1X_1 + \varepsilon$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.234
Model:                            OLS   Adj. R-squared:                  0.231
Method:                 Least Squares   F-statistic:                     72.05
Date:                Sun, 28 Oct 2018   Prob (F-statistic):           2.35e-15
Time:                        18:08:22   Log-Likelihood:                -1641.3
No. Observations:                 238   AIC:                             3287.
Df Residuals:                     236   BIC:                             3294.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1823.0435     17.674    103.149      0.000    1788.225    1857.862
x1            14.3174      1.687      8.488      0.000      10.994      17.640
==============================================================================
Omnibus:                       26.471   Durbin-Watson:                   2.033
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               31.857
Skew:                           0.829   Prob(JB):                     1.21e-07
Kurtosis:                       3.679   Cond. No.                         11.9
==============================================================================

('Intercept/a0:', 1823.043461212659)
('MSE/variance value:', 75020.29037354182)
('MSE of residuals:', 57717.90493298641)
('R-Square:', 0.2338823379254178)
('F-Value:', 72.04667700902738)
('Chi-square Results:', Power_divergenceResult(statistic=-2.7525142541622645e+19, pvalue=1.0))
```

**2.1 Estimates for a0, a1 and variance:**

a0 = 1823.0435

a1 = 14.3174

variance / MSE of residuals = 57717.90493298641

variance / MSE of model = 75020.29037354182

**2.2 p-value, F-value and $R^2$ value:**

('R-Square:', 0.2338823379254178)

('F-Value:', 72.04667700902738)

p-value for a0 = 0.0

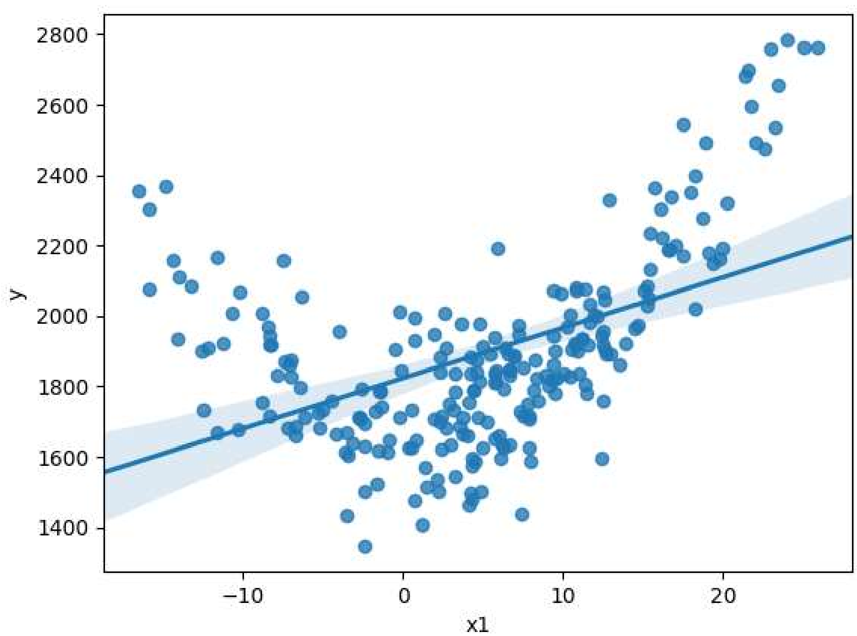
p-value for a1 = 0.0

The p-values are used to accept / reject the null hypothesis. The p-values for both a0 and a1 are zero and hence both the co-efficient is significant in the linear regression model.

The $R^2$ gives a measure of how well the linear model fits the data. The closer R-square is to 1 , the better the fit. The R-square value is around 0.233 in our case and hence we do not see a good fit. We can say that about 23% of the sample variability observed in the results drawn from the model can be due to the predictors in the model itself.

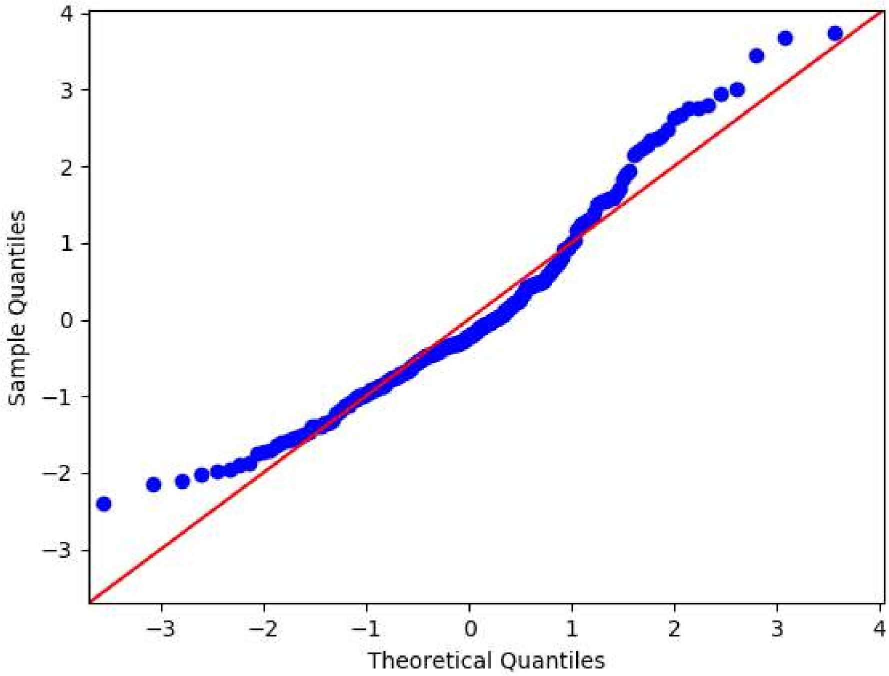
The F-value tests the overall significance of the regression model. Its values generally varies from 0 to any arbitrary value. A high F-value means that the data does not support the null hypothesis.

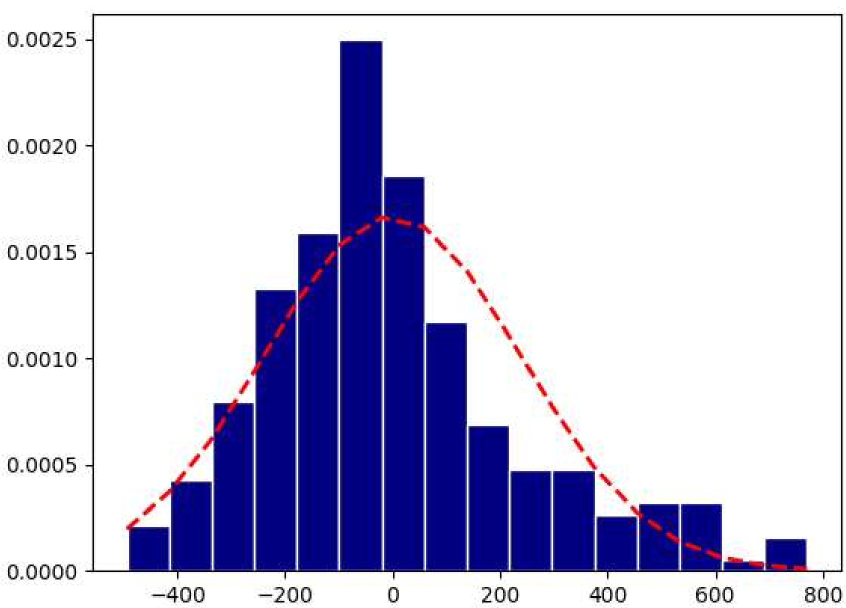
**2.3 Regression line against the data:**




**2.4 Residuals Analysis:**

**(a) Q-Q Plot and Histogram:**



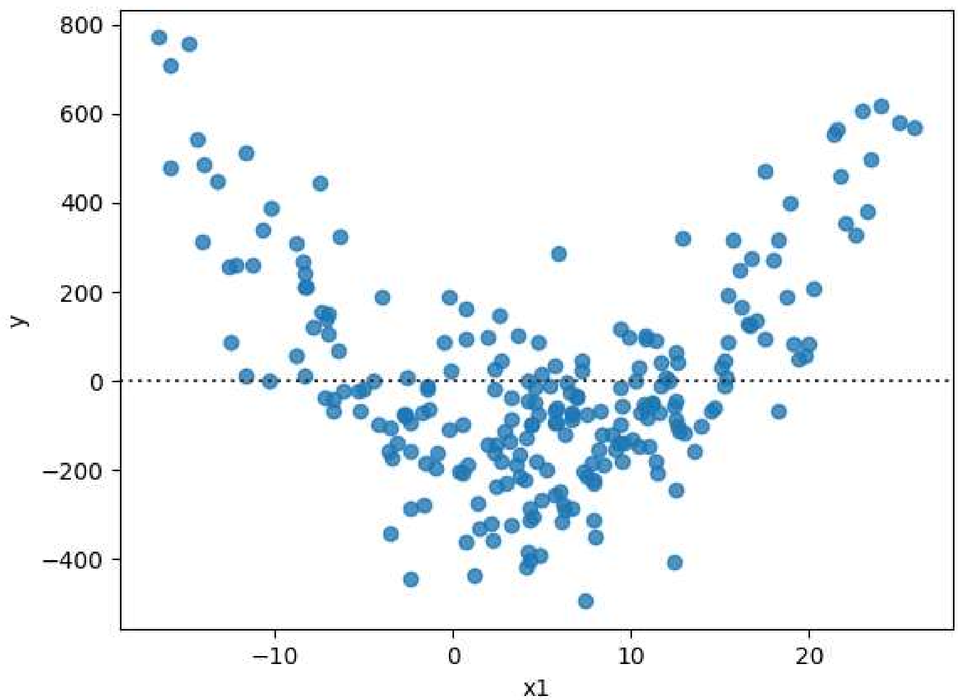

We observe that a greater percentage of data falls in the line except a few points and the histogram is almost normally distributed.

The chi-square test gives a value,

('Chi-square Results:', Power_divergenceResult(statistic=-2.7525142541622645e+19, pvalue=1.0))

The p-value indicates that the residuals follow a normal distribution.

**(b) Scatter Plot:**




From the scatter plot of the residuals we do not see any trends. It is good that there are no trends in the scatter plot as a linear pattern would suggest dependency.

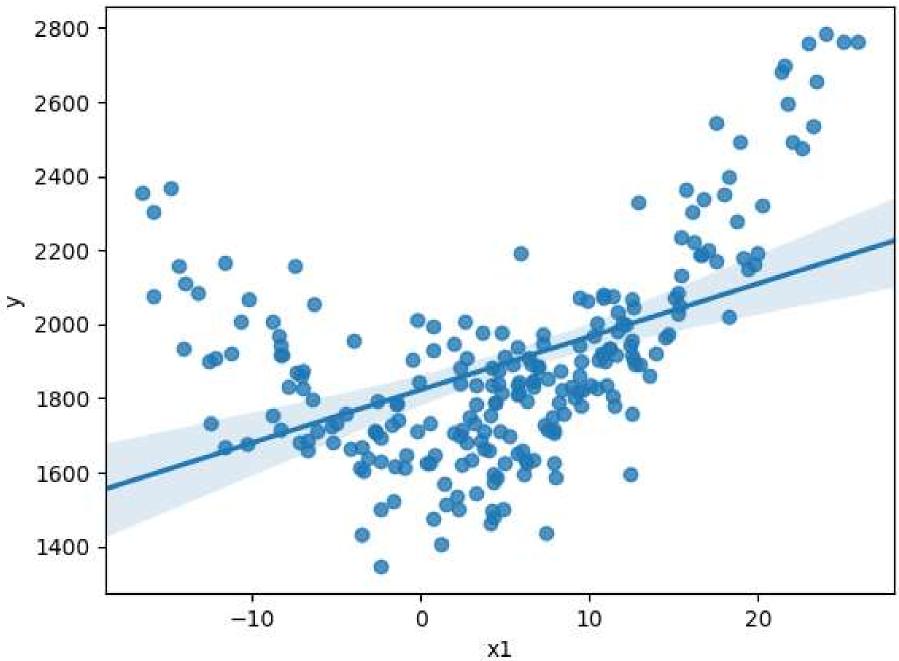
**2.7 Higher Order Polynomial Regression:**

In the regression plot and residual analysis, we see that the data is slightly not fit to the model. We do a higher order polynomial regression ie ) $Y = a_0 + a_1X + a_2X_2 + \varepsilon$ and calculate all values again, plot the regression line do residual analysis etc

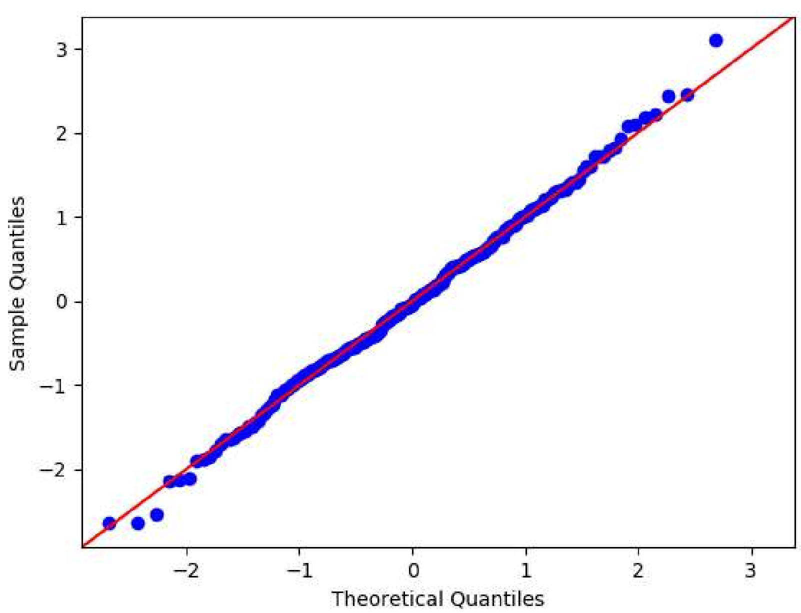
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.731
Model:                            OLS   Adj. R-squared:                  0.729
Method:                 Least Squares   F-statistic:                     319.3
Date:                Sun, 28 Oct 2018   Prob (F-statistic):           9.85e-68
Time:                        18:08:24   Log-Likelihood:                -1516.8
No. Observations:                 238   AIC:                             3040.
Df Residuals:                     235   BIC:                             3050.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1703.4184     11.962    142.403      0.000    1679.852    1726.985
x1            -1.3376      1.252     -1.068      0.286      -3.804       1.129
x2             1.7962      0.086     20.840      0.000       1.626       1.966
==============================================================================
Omnibus:                        0.262   Durbin-Watson:                   1.892
Prob(Omnibus):                  0.877   Jarque-Bera (JB):                0.148
Skew:                           0.058   Prob(JB):                        0.929
Kurtosis:                       3.039   Cond. No.                         224.
==============================================================================

('Intercept/a0:', 1703.4183726969652)
('MSE/variance value:', 75020.29037354182)
('MSE of residuals:', 20351.19689007222)
('R-Square:', 0.7310133467699154)
('F-Value:', 319.32464757644817)
('Chi-square Results:', Power_divergenceResult(statistic=-7.080683452758312e+18, pvalue=1.0))
```

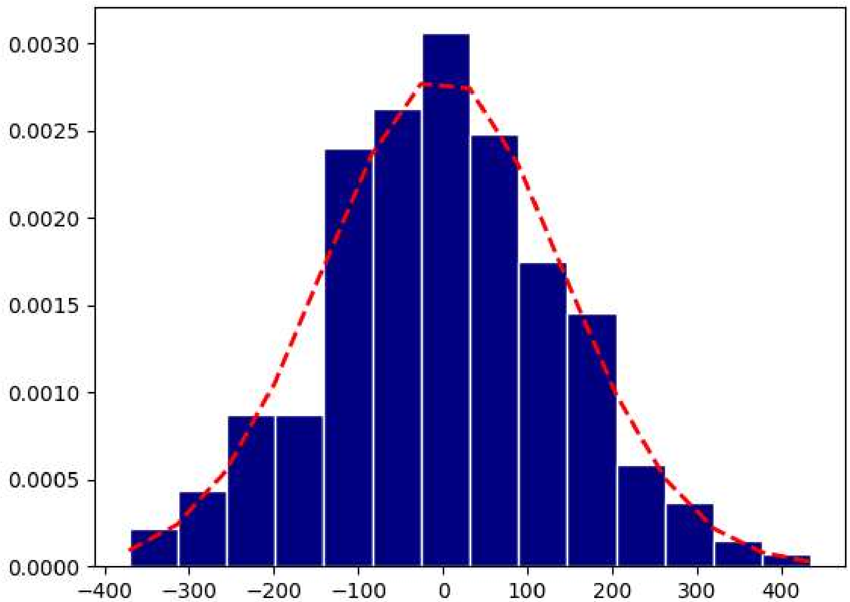
We see a higher R-square value of 0.731 in the higher order regression. This infers that the linear model now better fits the data. The MSE of the residuals have also greatly reduced. Chi-square results still say that the residuals are normally distributed.

The regression plot against the data,




The q-q plot and histogram for the residuals,






Hence the higher order polynomial regression tend to produce better results as observed from the p-vales, R-square, the residuals plot and a lowered variance / MSE value.

2.8:

For a simple linear regression analysis, the model explains that the variables are always linearly related such that the outcome population mean for any value considered in the $x_i$ column is a0+a1x

The null hypothesis for the simple linear regression is $H_0$ : a1 = 0 and the alternate hypothesis is

$H_1$ : a1 != 0 (two-tailed) or a1>0 / a1 < 0 (one-tailed). If the null hypothesis is true, then the coefficient a1 has no significance on the y value predicted.

From our results, we get p-value for a1 as 0.00, leading to reject the null hypothesis or accepting the alternative hypothesis concluding that the a1 value does have significance on the predicted y value.

The Cond.No. obtained in the regression model results has a value of 11.0, indicating that there is less multi-colinearity. Even though the R-square value indicates that the model is not a good fit for the data, the residuals on plotting and a chi-square test carried out says that the residuals are normally distributed. We do not see any trends in the residuals scatter plot which is a good sign.

On doing a higher order polynomial regression, we get the p-values for a1 to be 0.286 and a2 to be 0.0 which are still lower values and fall within the 95$^{th}$ percentile (0.05). Hence, these a1 and a2 co-efficient also have significance on the predicted y value.

The higher R-square value observed is a sign of model being a better fit for the data. The MSE value is considerably seen lowered for the higher order regression. The q-q plot and the histogram and the chi-square test indicate that the residuals are normally distributed.

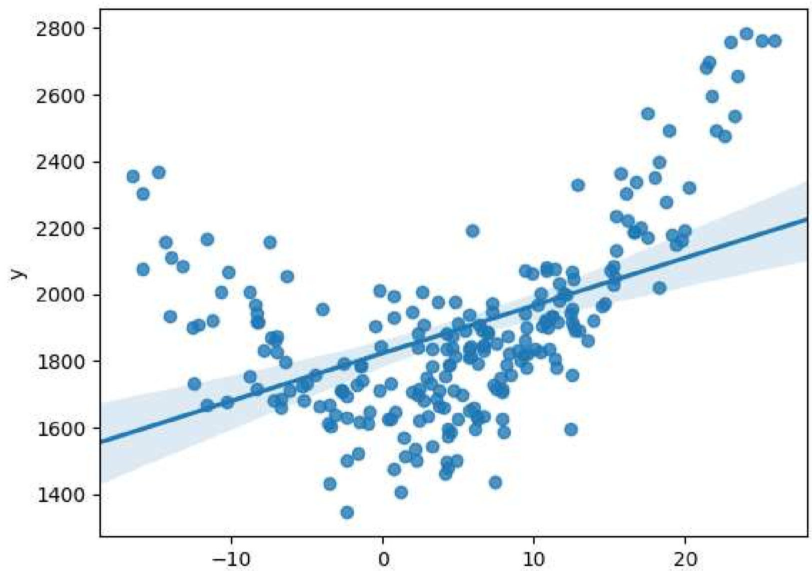
**Task3:**

### 3.1 Multi variable Regression on all Independent variables:

Task3 is run on the dataset which has no outliers. Outliers for all $x_i$ s were removed using z-score in the previous task and the new dataset is stored in a separate csv for this task.

Statsmodel package in python was used to perform simple Linear Regression on Y and the column x1,x2,x3,x4,x5.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.458
Model:                            OLS   Adj. R-squared:                  0.447
Method:                 Least Squares   F-statistic:                     39.24
Date:                Sun, 28 Oct 2018   Prob (F-statistic):           3.99e-29
Time:                        19:28:57   Log-Likelihood:                -1600.1
No. Observations:                 238   AIC:                             3212.
Df Residuals:                     232   BIC:                             3233.
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        985.5359     90.164     10.931      0.000     807.892    1163.180
x1            16.1601      1.448     11.159      0.000      13.307      19.013
x2             0.3671      1.222      0.300      0.764      -2.040       2.774
x3             4.0470      1.207      3.352      0.001       1.668       6.426
x4             8.5004      1.218      6.982      0.000       6.102      10.899
x5             7.5418      1.202      6.277      0.000       5.174       9.909
==============================================================================
Omnibus:                       66.347   Durbin-Watson:                   2.106
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              117.234
Skew:                           1.514   Prob(JB):                     3.49e-26
Kurtosis:                       4.628   Cond. No.                         495.
==============================================================================

('Intercept/a0:', 985.5358972637231)
('MSE/variance value:', 75020.29037354182)
('MSE of residuals:', 41519.99041687105)
('R-Square:', 0.45822602059279005)
('F-Value:', 39.24457091639811)
('Chi-square Results:', Power_divergenceResult(statistic=-2.364638783783354e+18, pvalue=1.0))

[[ 1.          0.48361383 -0.00954987  0.11076584  0.29588932  0.24886194]
 [ 0.48361383  1.         -0.09212281 -0.03527944 -0.07825077 -0.09368099]
 [-0.00954987 -0.09212281  1.         -0.02991405  0.0579247   0.03714248]
 [ 0.11076584 -0.03527944 -0.02991405  1.         -0.03479931 -0.06683803]
 [ 0.29588932 -0.07825077  0.0579247  -0.03479931  1.          0.01418369]
 [ 0.24886194 -0.09368099  0.03714248 -0.06683803  0.01418369  1.        ]]
```




a0 =  985.5359

a1 =  16.1601

a2 = 0.3671

a3 = 4.0470

a4 = 8.5004

a5 = 7.5418

variance / MSE of residuals = 75020.29037354182

variance / MSE of model = 41519.99041687105

**3.2: p-value, F-value, $R^2$ value and Correlation Matrix:**

('R-Square:', 0.45822602059279005)

('F-Value:', 39.24457091639811)

p-value for a0 = 0.000

p-value for a1 = 0.000

p-value for a2 = 0.764

p-value for a3 = 0.001

p-value for a4 = 0.000

p-value for a5 = 0.000

The correlation Matrix,

| Y | X1 | X2 | X3 | X4 | X5 |
|---|---|---|---|---|---|
| 1. | 0.48361383 | -0.00954987 | 0.11076584 | 0.29588932 | 0.24886194 |
| 0.48361383 | 1. | -0.09212281 | -0.03527944 | -0.07825077 | -0.09368099 |
| -0.00954987 | -0.09212281 | 1. | -0.02991405 | 0.0579247 | 0.03714248 |
| 0.11076584 | -0.03527944 | -0.02991405 | 1. | -0.03479931 | -0.06683803 |
| 0.29588932 | -0.07825077 | 0.0579247 | -0.03479931 | 1. | 0.01418369 |
| 0.24886194 | -0.09368099 | 0.03714248 | -0.06683803 | 0.01418369 | 1. |

The p-values are used to accept / reject the null hypothesis. The p-values for both a0,a1,a3,a4,a5 are zero and hence both the co-efficient is significant in the linear regression model. We see that the p-value for a2 is 0.76 > 0.5 ($95^{th}$ %) and hence can be removed as it confirms the null hypothesis that

$H_0$ : a2 = 0, If the null hypothesis is true, then the coefficient a1 has no significance on the y value predicted. Hence, a2 can be removed.

The $R^2$ gives a measure of how well the linear model fits the data. The closer R-square is to 1 , the better the fit. The R-square value is around 0.458 in our case and hence we do not see a good fit. We can say that about 45% of the sample variability observed in the results drawn from the model can be due to the predictors in the model itself.

The F-value tests the overall significance of the regression model. Its values generally varies from 0 to any arbitrary value. A high F-value means that the data does not support the null hypothesis.

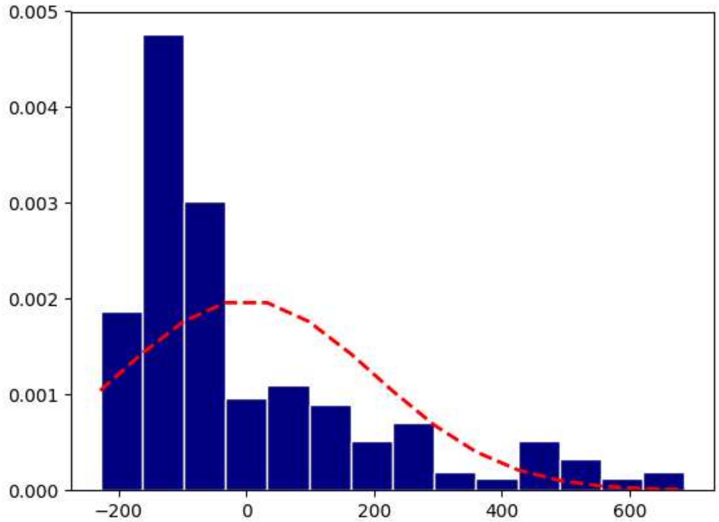
We see that almost all independent variables have a weak / lower degree of correlation.

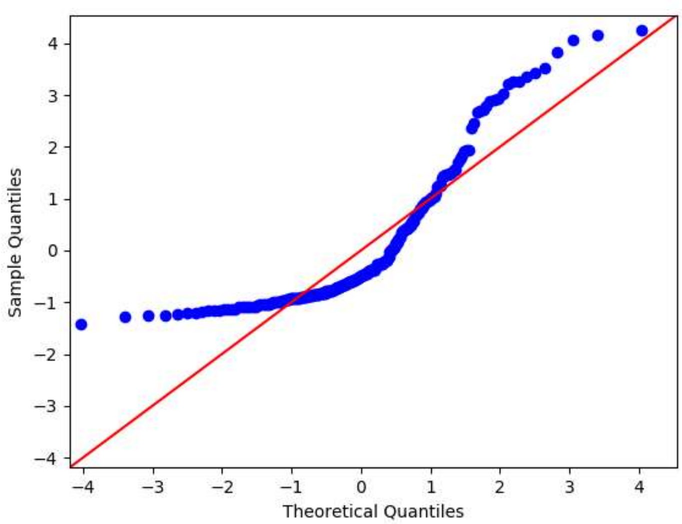
On observing the correlation between Y and a $x_i$ , we see that only x1 has an average degree of correlation while the others have a lower degree of correlation.

**3.3 Residual Analysis:**

(a**) Q-Q Plot and Histogram for Residuals:**

**With all the $x_i$ values:**



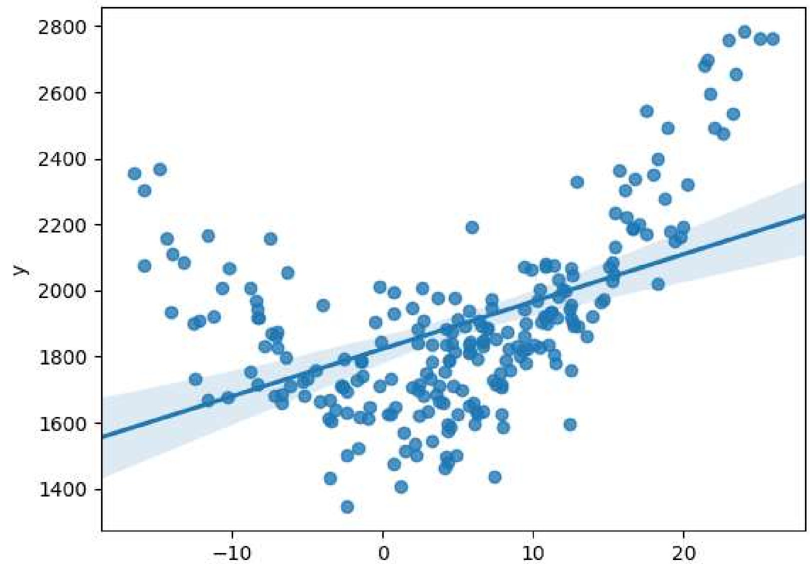




We observe that a percentage of data diverges from the line and the histogram is almost distributed toward the right and has a right tail. Hence the histogram/ data is said to be right skewed. This indicates that the data has some larger values that drives the mean upward.

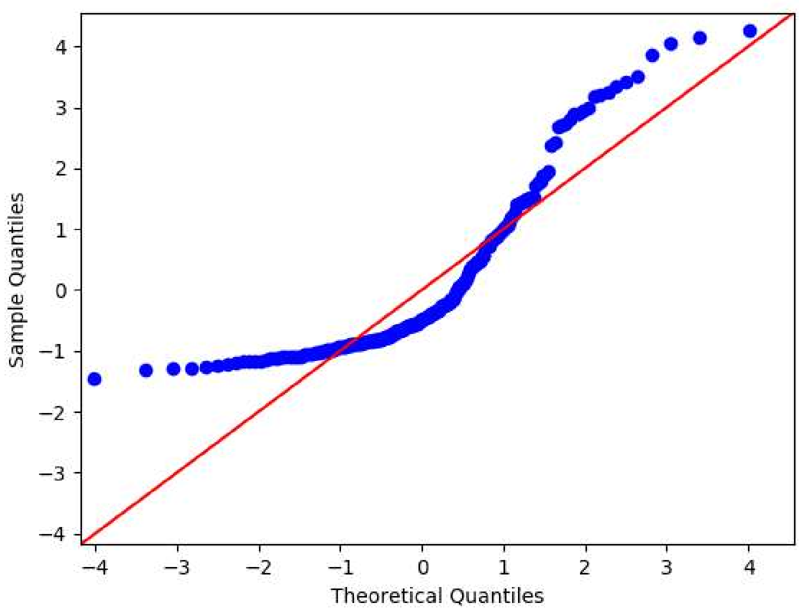
The chi-square test gives a value,

('Chi-square Results:', Power_divergenceResult(statistic=-2.364638783783354e+18, pvalue=1.0))

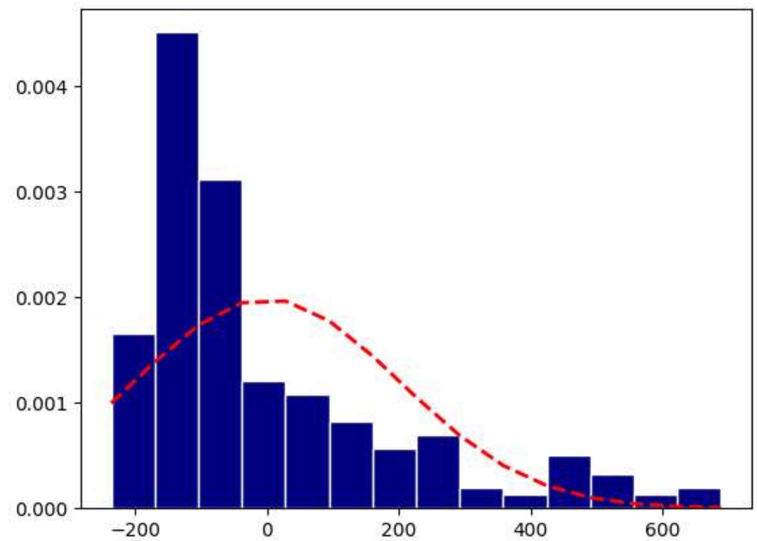
**Without a2:**



2800
2600
2400
2200
2000
1800
1600
1400
y
−10
0
10
20

Sample Quantiles
Theoretical Quantiles

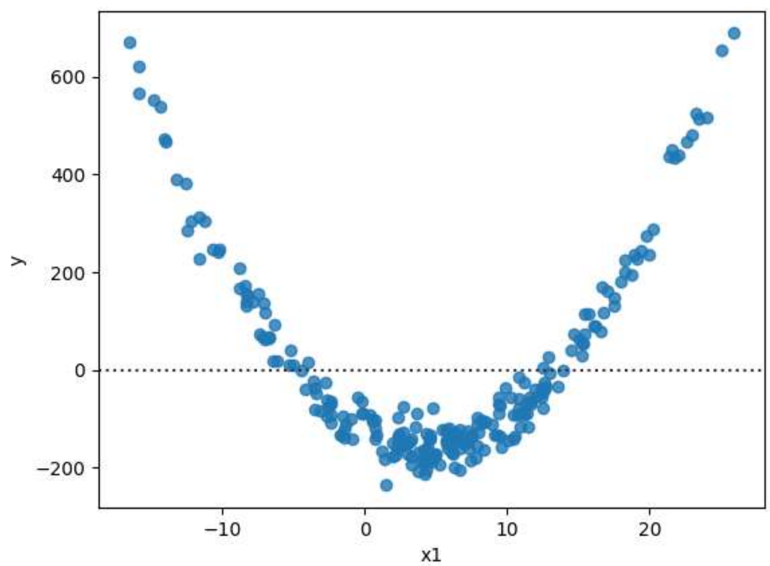


```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.458
Model:                            OLS   Adj. R-squared:                  0.449
Method:                 Least Squares   F-statistic:                     49.23
Date:                Sun, 28 Oct 2018   Prob (F-statistic):           5.56e-30
Time:                        19:54:42   Log-Likelihood:                -1600.2
No. Observations:                 238   AIC:                             3210.
Df Residuals:                     233   BIC:                             3228.
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        990.2395     88.621     11.174      0.000     815.639    1164.840
x1            16.1225      1.440     11.197      0.000      13.286      18.959
x3             4.0363      1.204      3.351      0.001       1.663       6.409
x4             8.5186      1.214      7.019      0.000       6.127      10.910
x5             7.5513      1.199      6.299      0.000       5.189       9.913
==============================================================================
Omnibus:                       65.793   Durbin-Watson:                   2.109
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              115.740
Skew:                           1.504   Prob(JB):                     7.37e-26
Kurtosis:                       4.621   Cond. No.                         477.
==============================================================================

('Intercept/a0:', 990.2394659311667)
('MSE/variance value:', 75020.29037354182)
('MSE of residuals:', 41357.8805445604)
('R-Square:', 0.45801519773149)
('F-Value:', 49.22533834194451)
('Chi-square Results:', Power_divergenceResult(statistic=-4.917963600924561e+18, pvalue=1.0))
```

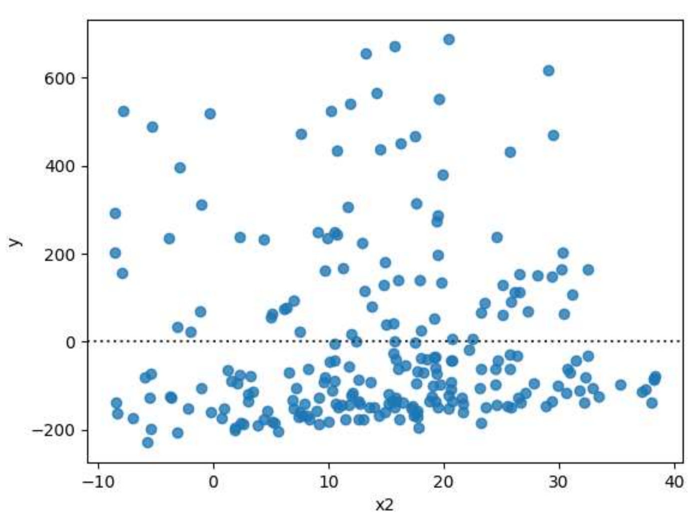
As predicted, we do not see much of a change after removing the co-efficient a2 from the list of independent variables considered to perform multi-variable regression. The R-square value, MSE , chi-square test results and the plots also indicate that a2 is not significant for predicting y.

**(b) Scatter Plot of the Residuals:**

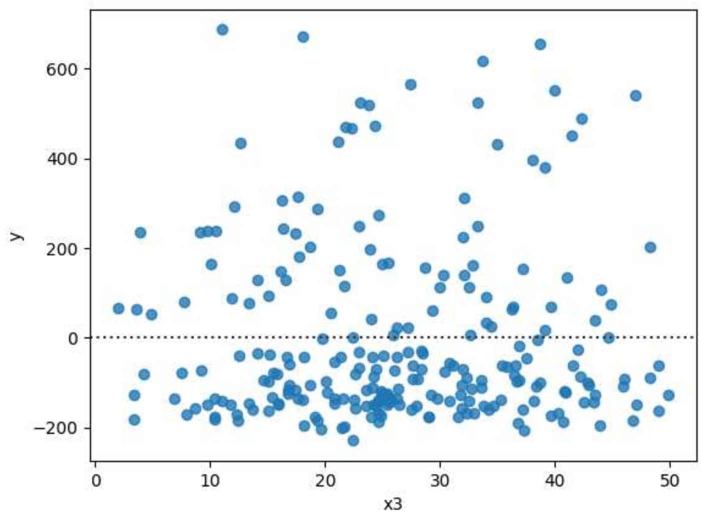
Y, x1




Y, x2:

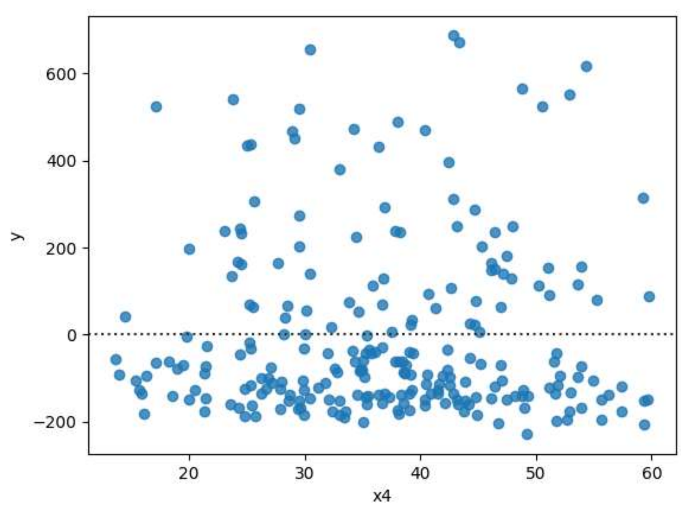



Y, x3:



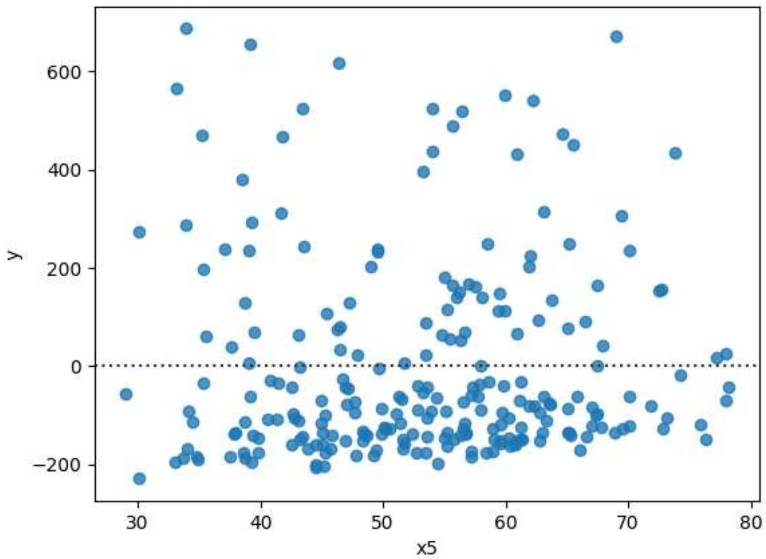
600
400
200
0
−200
y
0
10
20
30
40
50
x3


Y, x4:



600
400
200
0
−200
y
20
30
40
50
60
x4


Y, x5:




From the scatter plot of the residuals we do not see any trends. It is good that there are no trends in the scatter plot as a linear pattern would suggest dependency. For the scatter plot between y and x1 alone, we observe a parabolic curve. This clearly indicates that the two variables have a non-linear association. Since we focus of linear associations and linear regression, we do not focus more on that.

**3.4:**

The null hypothesis for the linear multivariable regression is $H_0$ : a1 = a2 = a3 = a4 = a5 = 0 and the alternate hypothesis is H1 or H2 or H3 or H4 or H5 is true

$H_1$ : a1 != 0 (two-tailed) or a1>0 / a1 < 0 (one-tailed).

$H_2$ : a2 != 0 (two-tailed) or a2>0 / a2 < 0 (one-tailed).

$H_3$ : a3 != 0 (two-tailed) or a3>0 / a3 < 0 (one-tailed).

$H_4$ : a4 != 0 (two-tailed) or a4>0 / a4 < 0 (one-tailed).

$H_5$ : a5 != 0 (two-tailed) or a5>0 / a5 < 0 (one-tailed).

If the null hypothesis is true, then none of the coefficients a1,a2,a3,a4,a4 has significance on the y value predicted.

From our results, we get p-value for a2 as 0.76, leading to reject the null hypothesis or accepting the alternative hypothesis concluding that the coefficient value do have significance on the predicted y value.

Even though the R-square value indicates that the model is not a good fit for the data, the residuals on plotting and a chi-square test carried out says that the residuals are normally distributed. We do not see any trends in the residuals scatter plot which is a good sign.

We carry out a number of experiments with various combinations of the independent variables to see which combination produces a better model with a higher R-square value. Some combinations performed very poorly with a lower R-square value and a very high MSE value.

The below screenshots are some examples,

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.250
Model:                            OLS   Adj. R-squared:                  0.244
Method:                 Least Squares   F-statistic:                     39.22
Date:                Sun, 28 Oct 2018   Prob (F-statistic):           2.01e-15
Time:                        20:22:31   Log-Likelihood:                -1638.8
No. Observations:                 238   AIC:                             3284.
Df Residuals:                     235   BIC:                             3294.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1738.6347     41.188     42.212      0.000    1657.490    1819.779
x1            14.4511      1.673      8.637      0.000      11.155      17.748
x3             3.1839      1.406      2.264      0.024       0.414       5.954
==============================================================================
Omnibus:                       25.552   Durbin-Watson:                   2.025
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               30.377
Skew:                           0.821   Prob(JB):                     2.53e-07
Kurtosis:                       3.604   Cond. No.                         77.2
==============================================================================

('Intercept/a0:', 1738.634690447658)
('MSE/variance value:', 75020.29037354182)
('MSE of residuals:', 56725.71876764182)
('R-Square:', 0.25024256186021177)
('F-Value:', 39.21735153642151)
('Chi-square Results:', Power_divergenceResult(statistic=2.4609431760888783e+19, pvalue=0.0))
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.343
Model:                            OLS   Adj. R-squared:                  0.335
Method:                 Least Squares   F-statistic:                     40.80
Date:                Sun, 28 Oct 2018   Prob (F-statistic):           3.04e-21
Time:                        20:23:48   Log-Likelihood:                -1623.0
No. Observations:                 238   AIC:                             3254.
Df Residuals:                     234   BIC:                             3268.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1315.5609     82.960     15.858      0.000    1152.118    1479.004
x1            15.3261      1.577      9.722      0.000      12.220      18.432
x3             3.7209      1.322      2.815      0.005       1.117       6.325
x5             7.5871      1.317      5.762      0.000       4.993      10.181
==============================================================================
Omnibus:                       47.078   Durbin-Watson:                   2.126
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               70.219
Skew:                           1.174   Prob(JB):                     5.65e-16
Kurtosis:                       4.252   Cond. No.                         347.
==============================================================================
('Intercept/a0:', 1315.5608503210283)
('MSE/variance value:', 75020.29037354182)
('MSE of residuals:', 49888.62504440638)
('R-Square:', 0.3434159849781413)
('F-Value:', 40.79667828557059)
('Chi-square Results:', Power_divergenceResult(statistic=-9.50935453540518e+18, pvalue=1.0))
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.366
Model:                            OLS   Adj. R-squared:                  0.358
Method:                 Least Squares   F-statistic:                     44.97
Date:                Sun, 28 Oct 2018   Prob (F-statistic):           5.49e-23
Time:                        20:25:52   Log-Likelihood:                -1618.9
No. Observations:                 238   AIC:                             3246.
Df Residuals:                     234   BIC:                             3260.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1410.0628     63.050     22.364      0.000    1285.844    1534.282
x1            15.2547      1.547      9.860      0.000      12.206      18.303
x3             3.5031      1.297      2.701      0.007       0.948       6.058
x4             8.5512      1.310      6.527      0.000       5.970      11.132
==============================================================================
Omnibus:                       38.320   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               51.725
Skew:                           1.052   Prob(JB):                     5.86e-12
Kurtosis:                       3.890   Cond. No.                         208.
==============================================================================
('Intercept/a0:', 1410.0627638281683)
('MSE/variance value:', 75020.29037354182)
('MSE of residuals:', 48193.777183761056)
('R-Square:', 0.365721871584621)
('F-Value:', 44.97444371110175)
('Chi-square Results:', Power_divergenceResult(statistic=-1.5995110178494833e+19, pvalue=1.0))
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.147
Model:                            OLS   Adj. R-squared:                  0.140
Method:                 Least Squares   F-statistic:                     20.32
Date:                Sun, 28 Oct 2018   Prob (F-statistic):           7.26e-09
Time:                        20:27:36   Log-Likelihood:                -1654.1
No. Observations:                 238   AIC:                             3314.
Df Residuals:                     235   BIC:                             3325.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       1300.7744     97.734     13.309      0.000    1108.227    1493.321
x4             7.3305      1.510      4.854      0.000       4.355      10.306
x5             6.0408      1.487      4.062      0.000       3.111       8.970
==============================================================================
Omnibus:                       83.619   Durbin-Watson:                   1.947
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              194.138
Skew:                           1.678   Prob(JB):                     6.97e-43
Kurtosis:                       5.884   Cond. No.                         391.
==============================================================================

('Intercept/a0:', 1300.7743636527864)
('MSE/variance value:', 75020.29037354182)
('MSE of residuals:', 64504.87698957433)
('R-Square:', 0.14742356077798613)
('F-Value:', 20.317554643217843)
('Chi-square Results:', Power_divergenceResult(statistic=-1.1666975348069941e+20, pvalue=1.0))
```