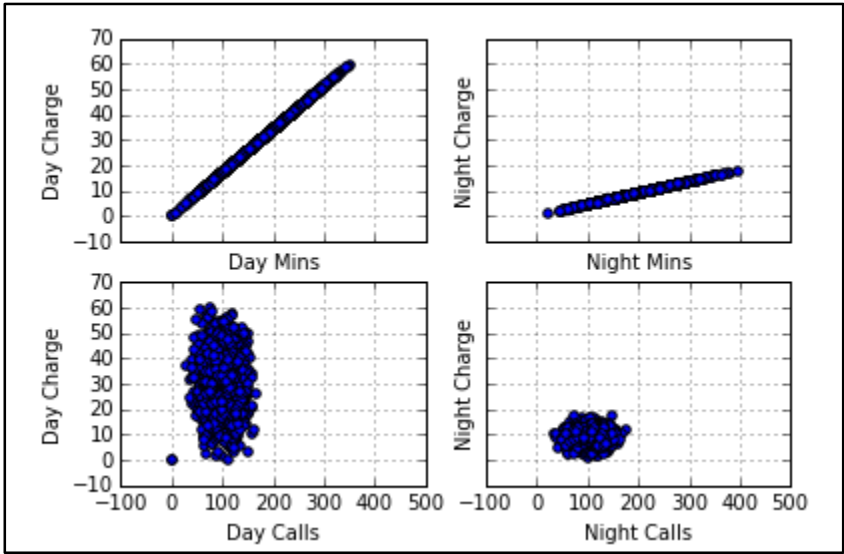
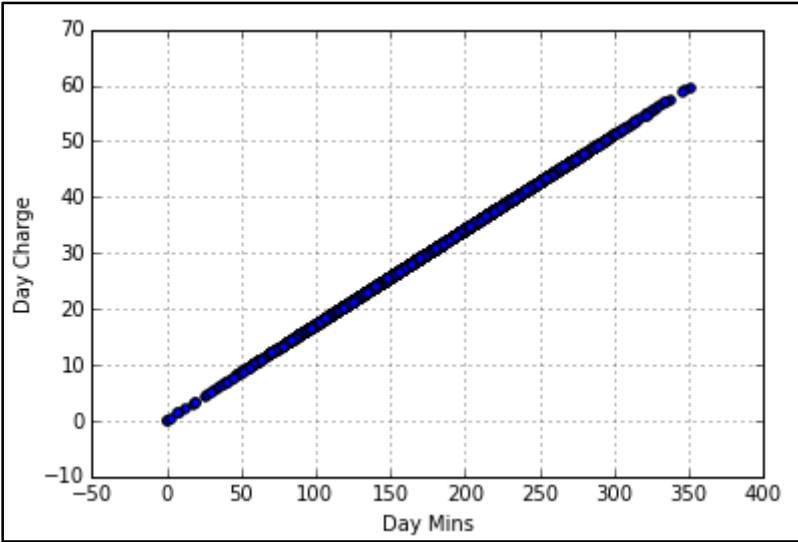
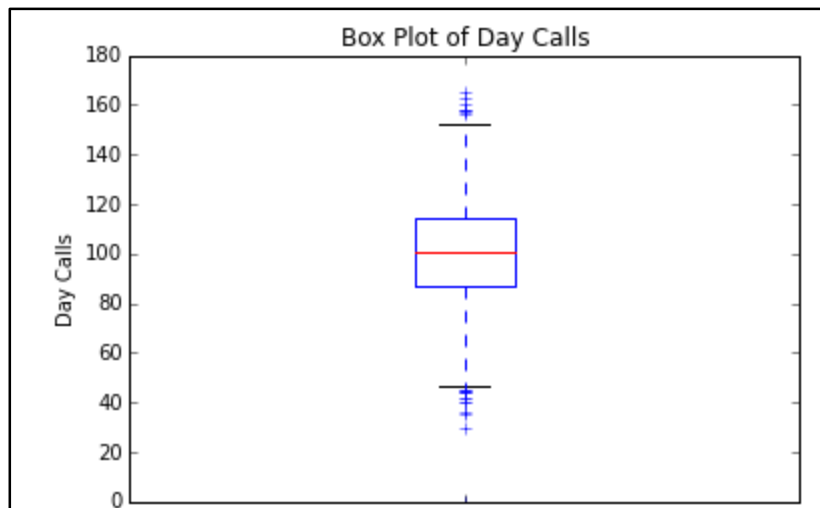
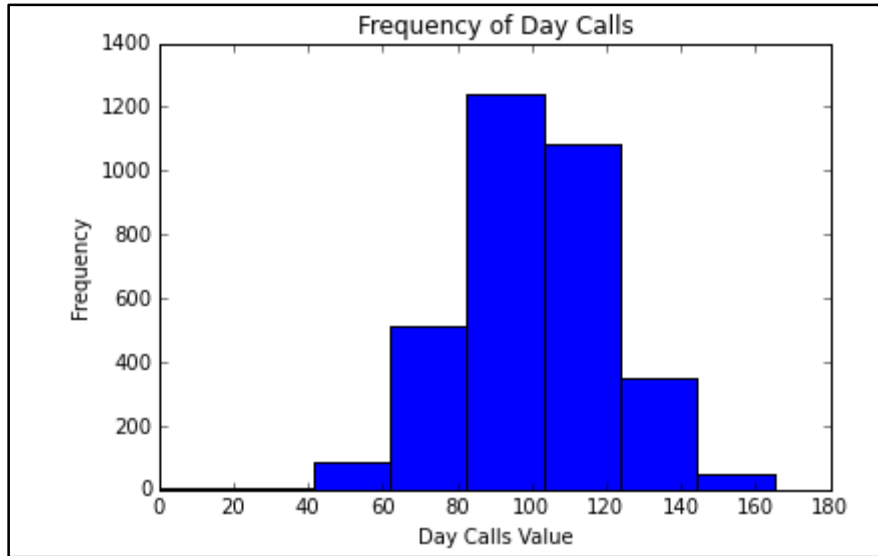
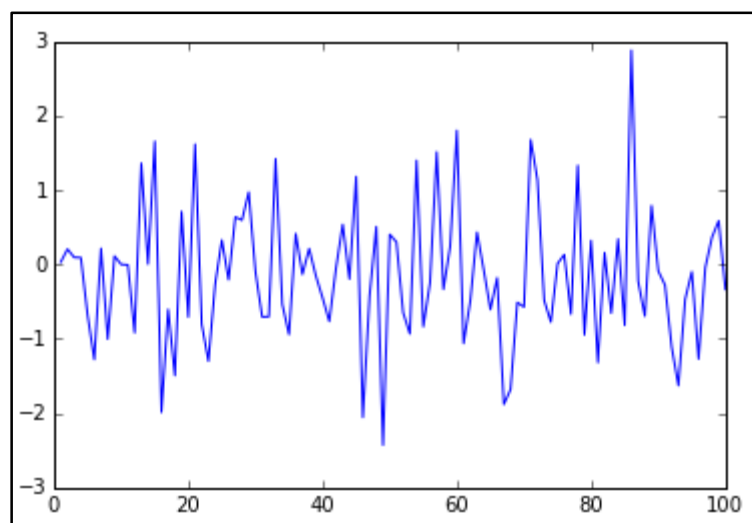
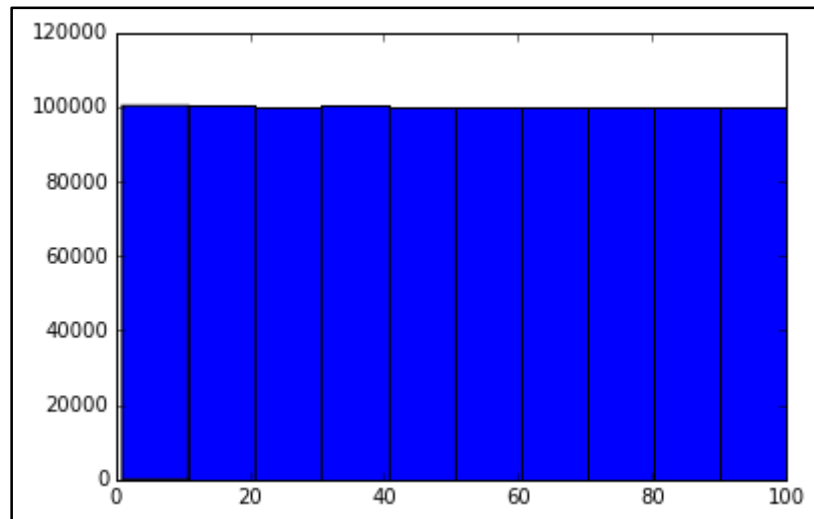
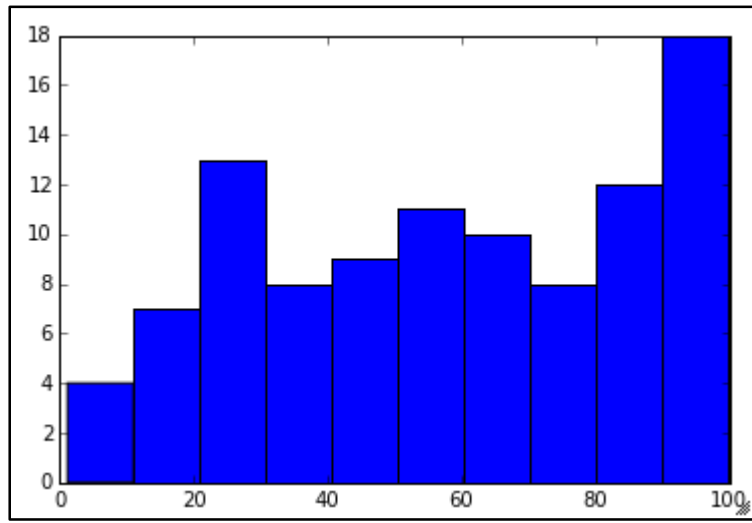


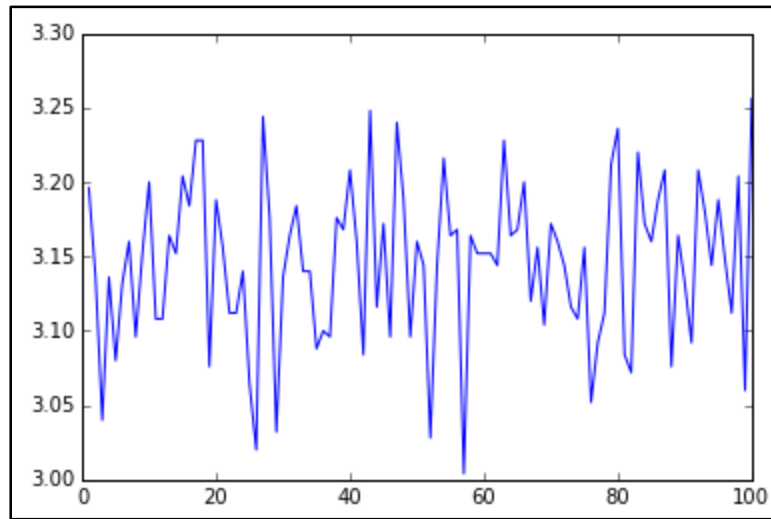
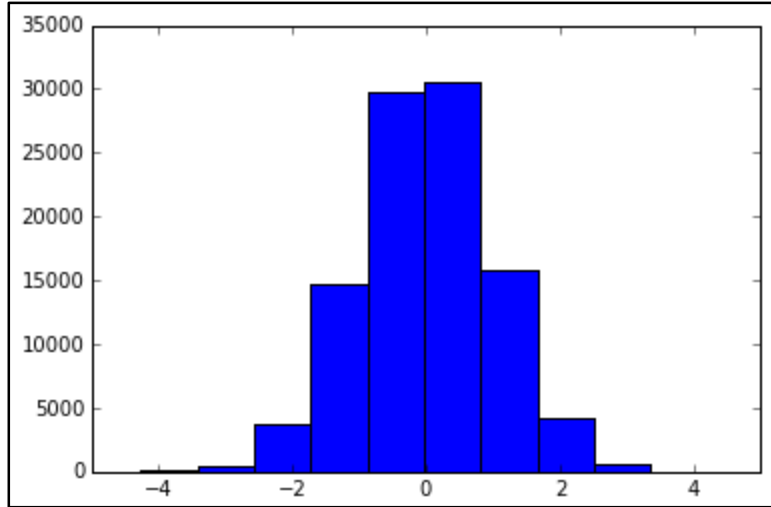
# Chapter 2: Data Cleaning



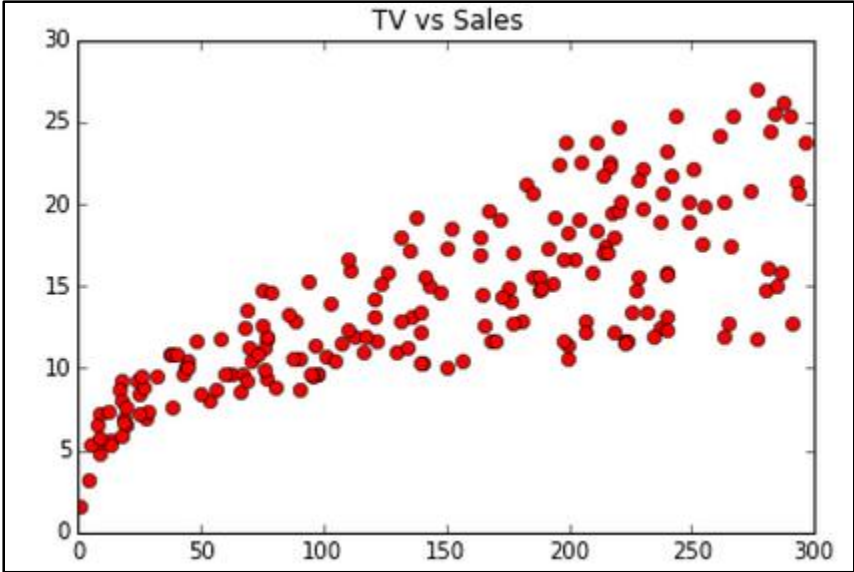
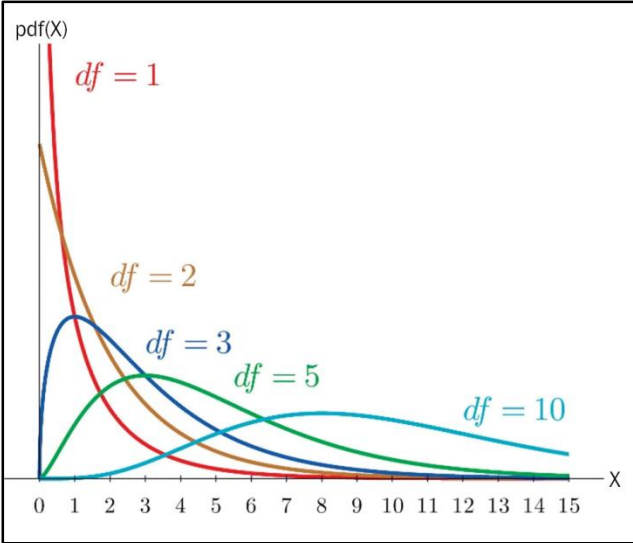


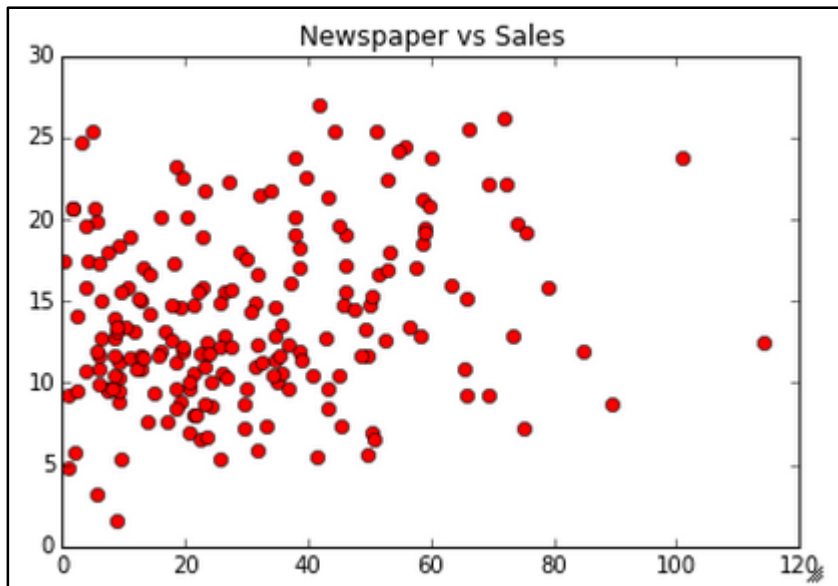
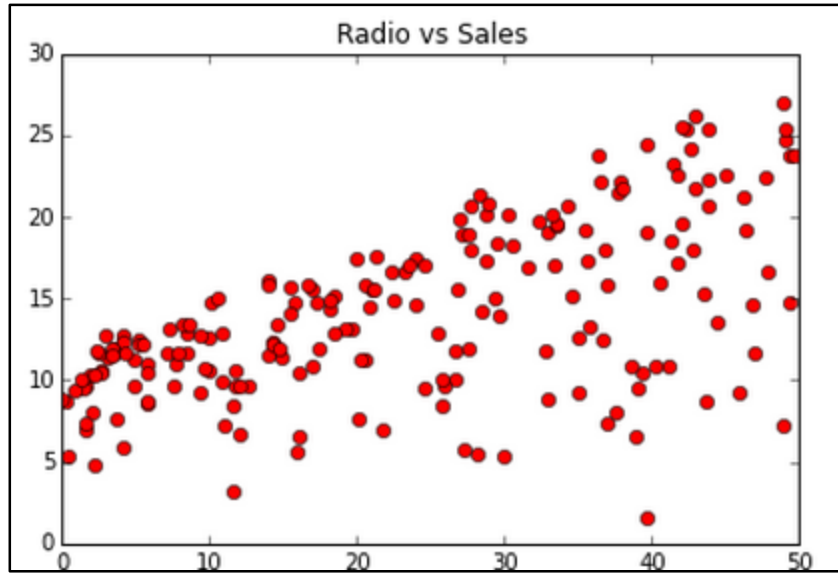
## Chapter 3: Data Wrangling



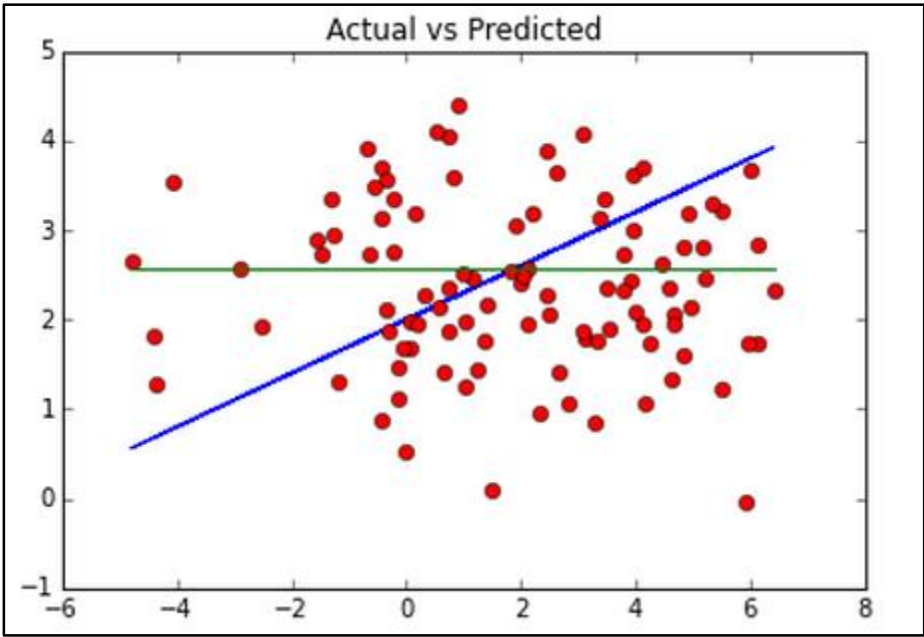
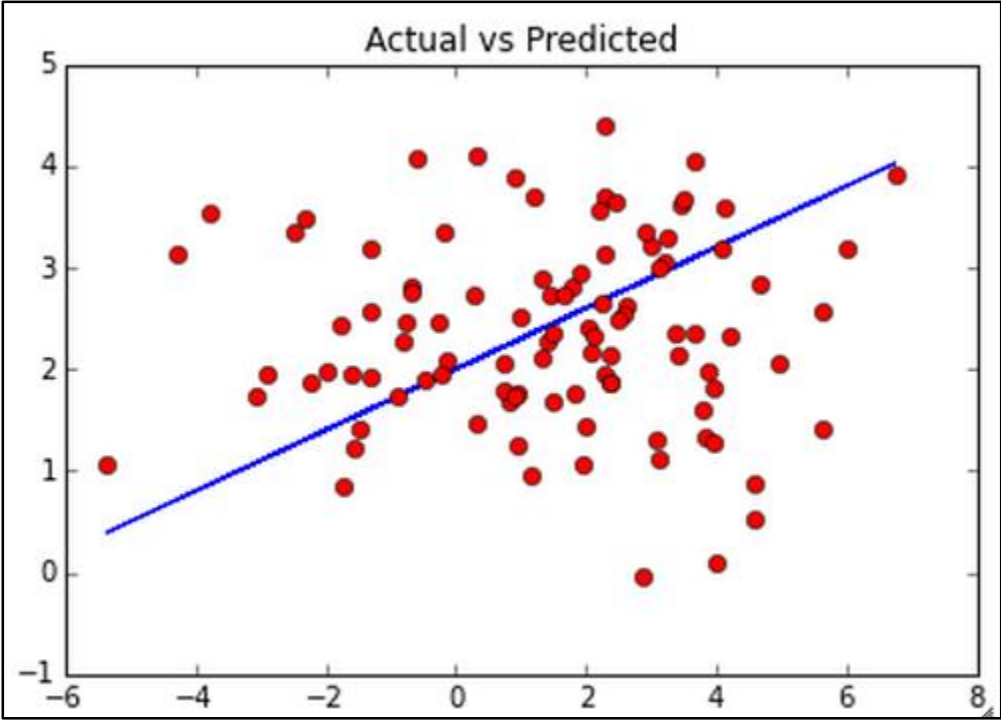


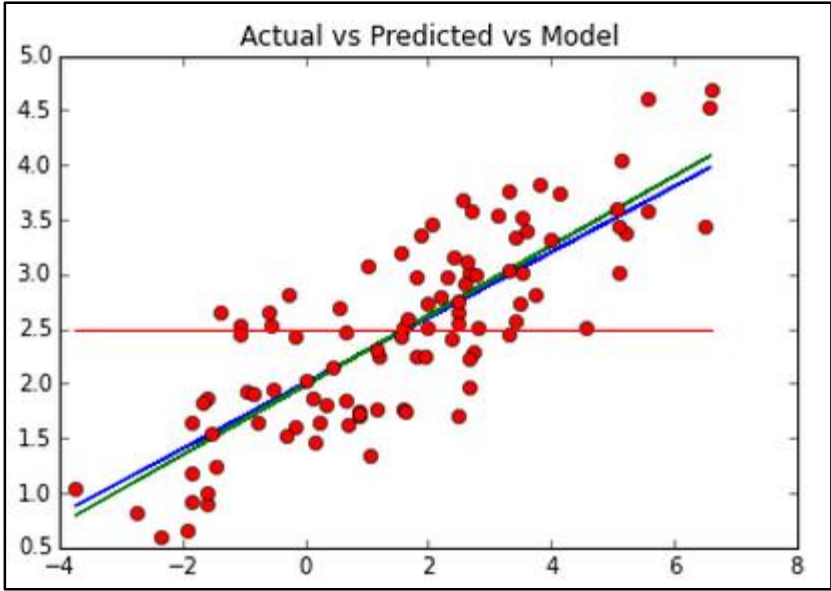
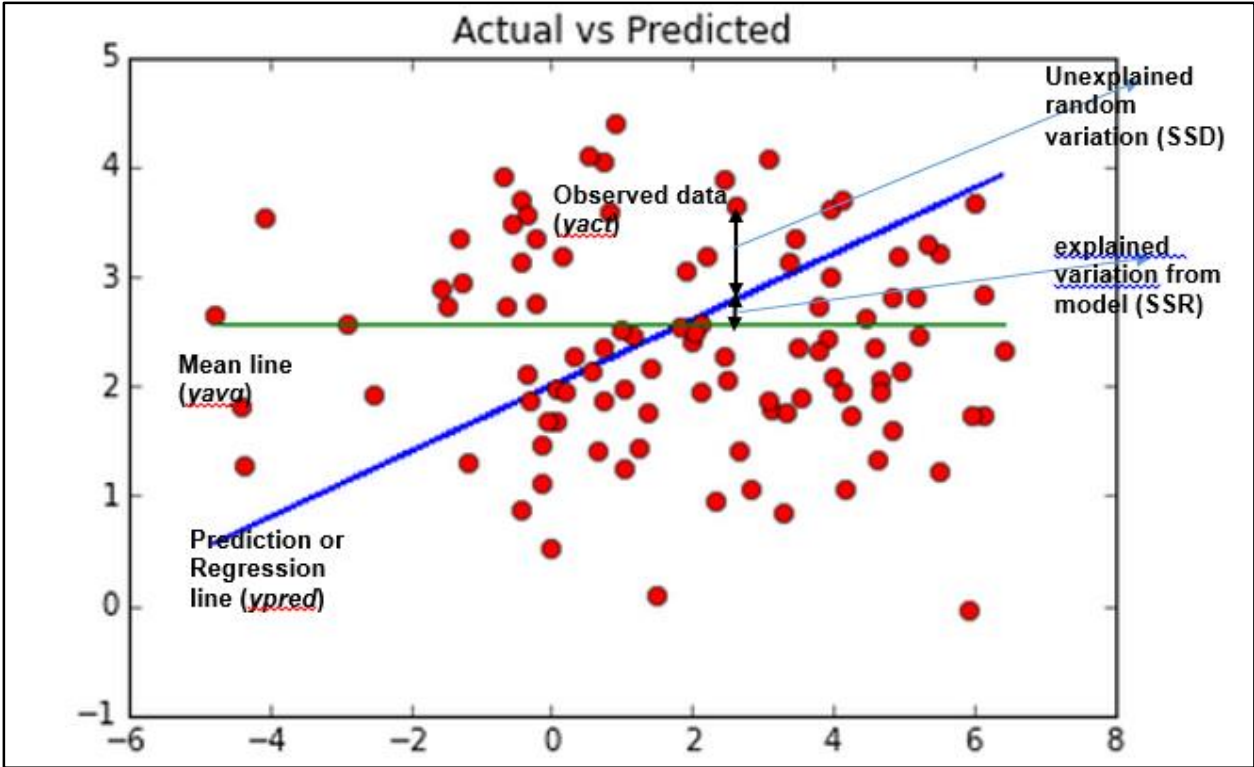
# Chapter 4: Statistical Concepts for Predictive Modelling



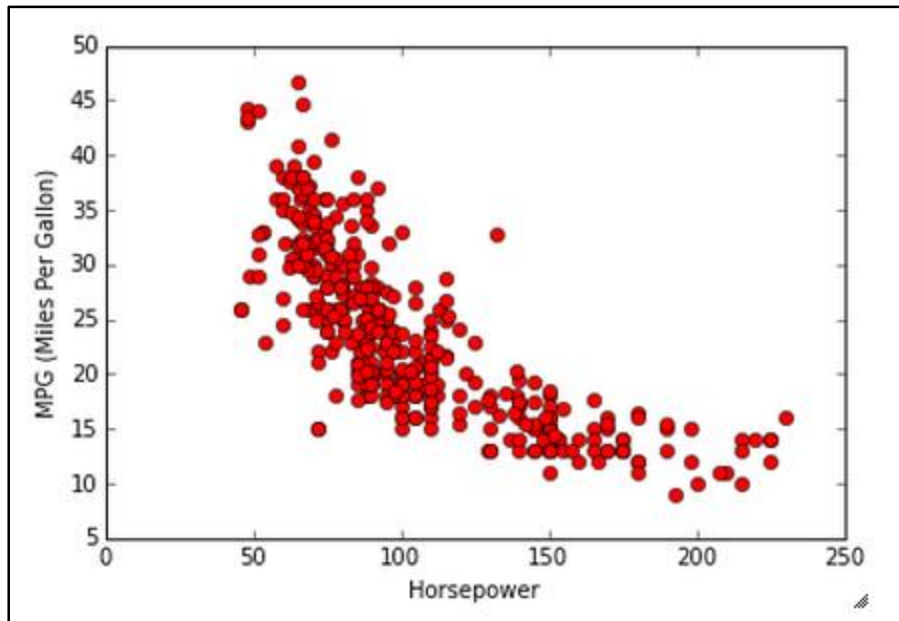
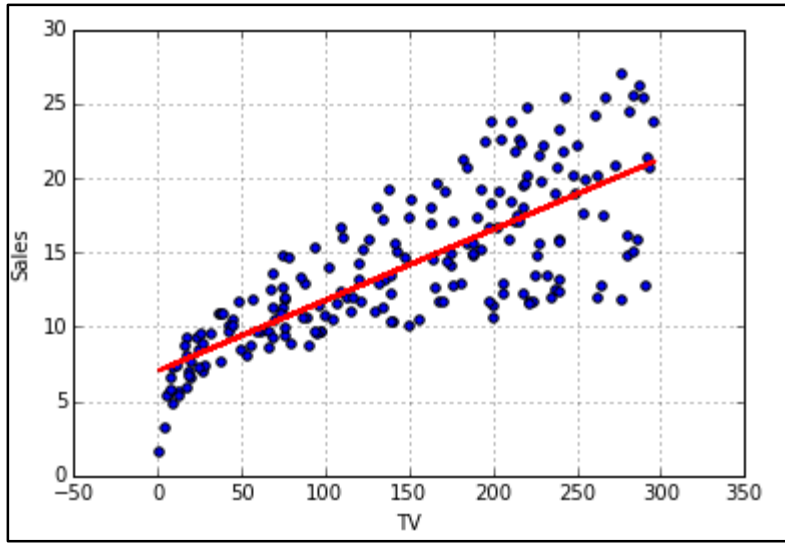


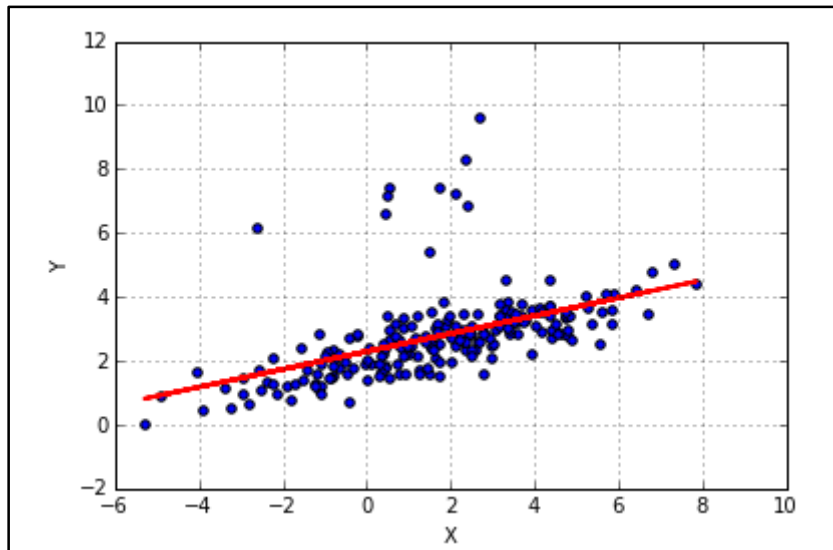
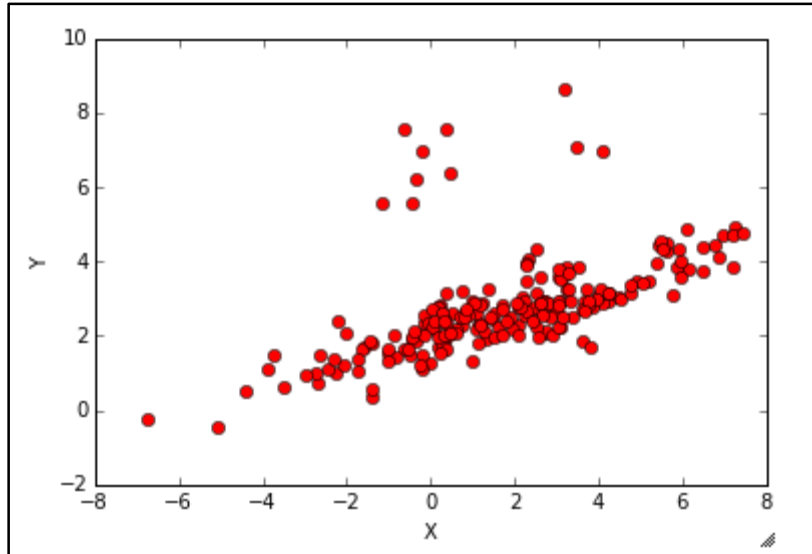
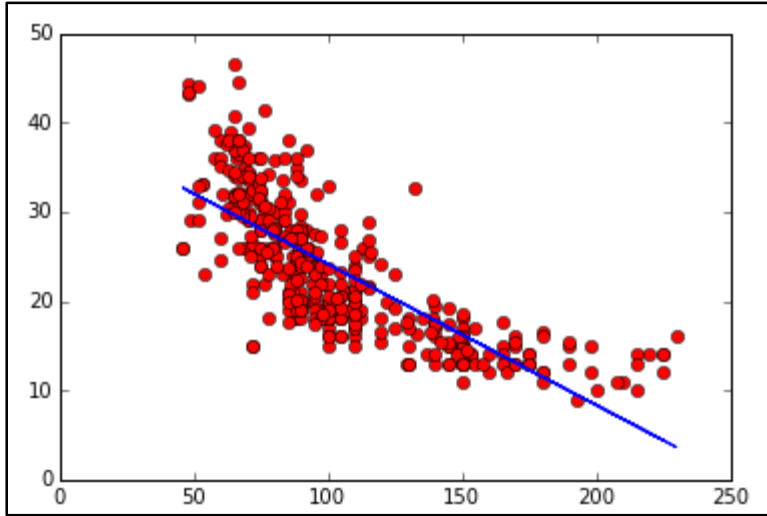
# Chapter 5: Linear Regression with Python

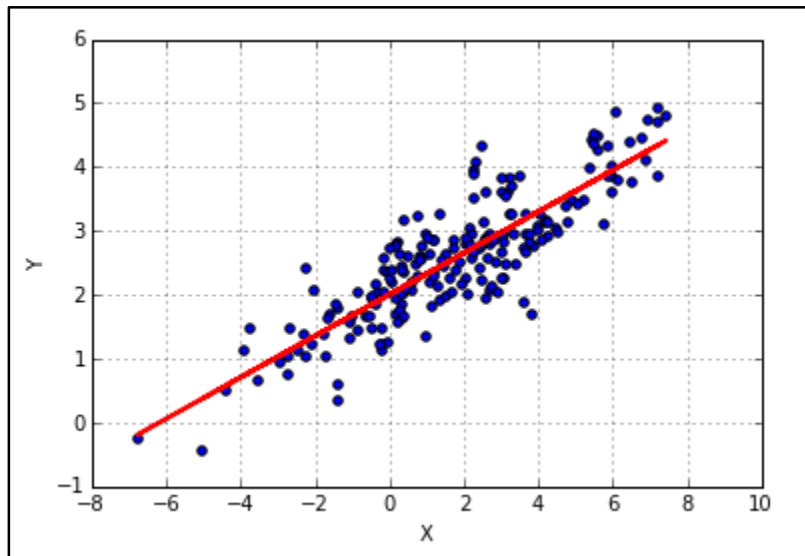
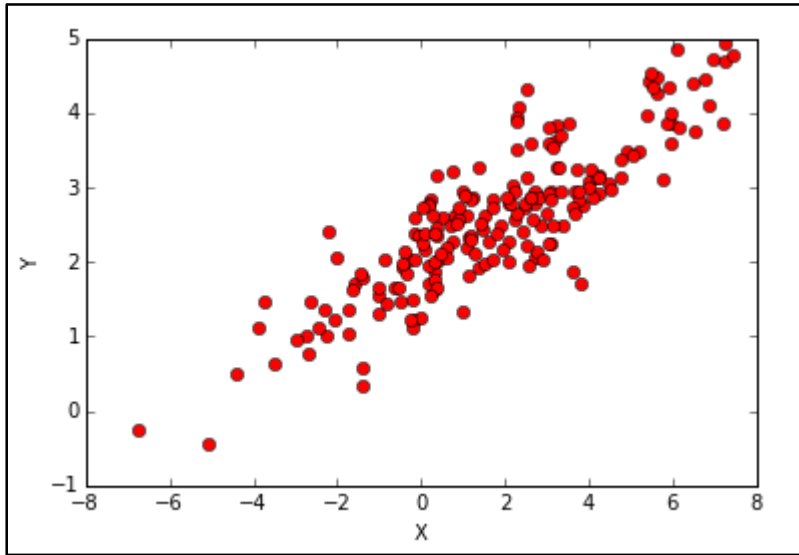


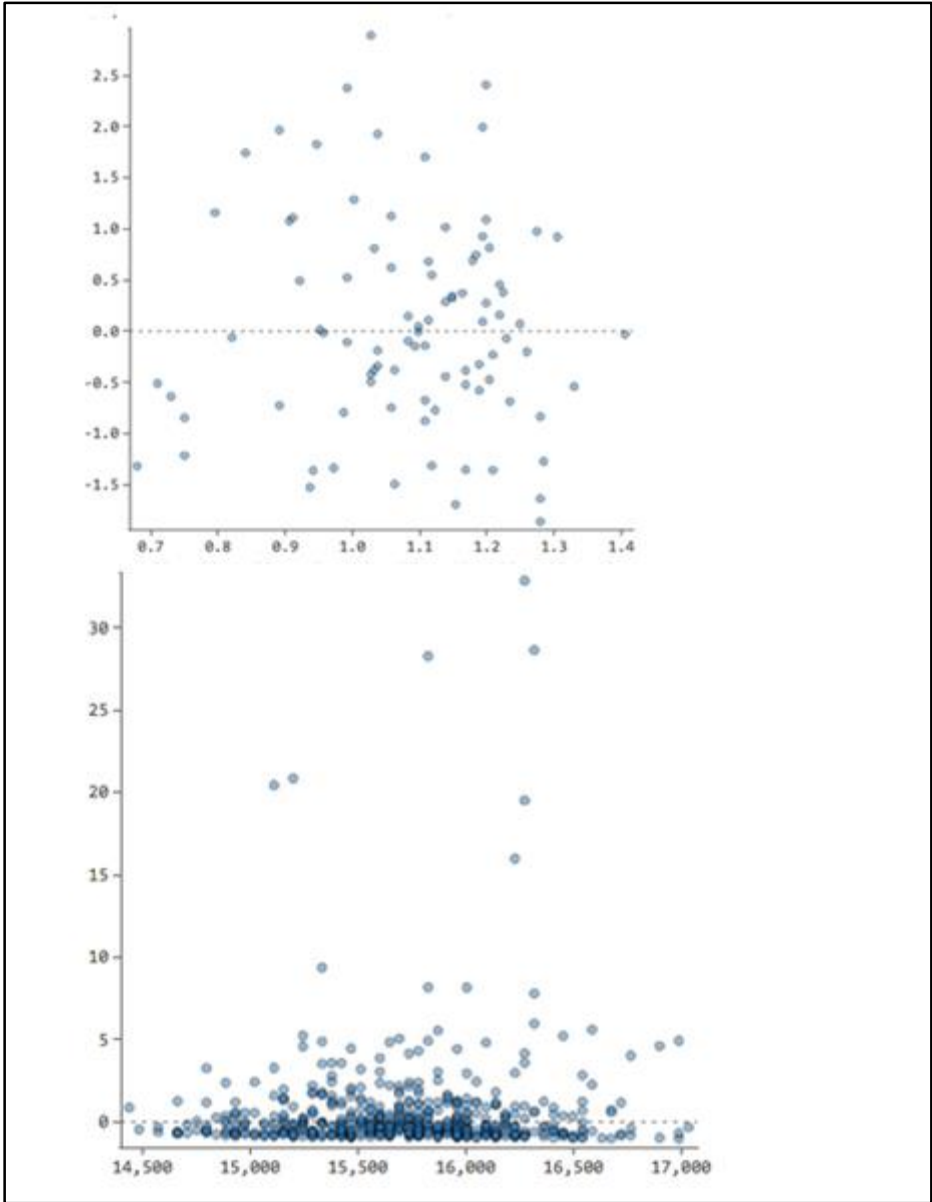




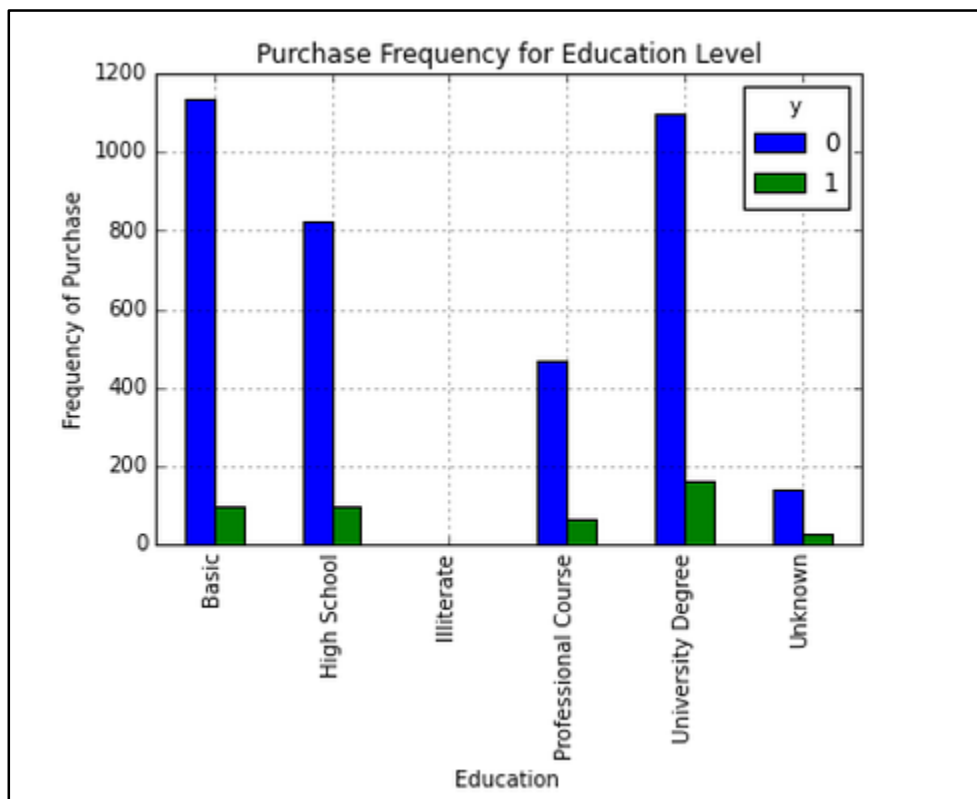
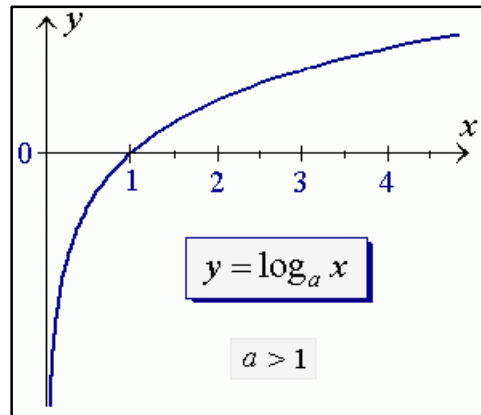


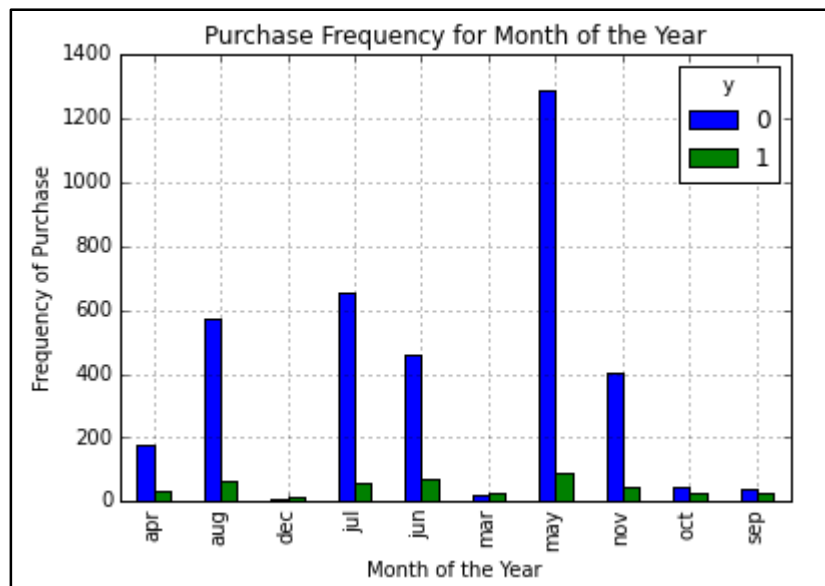
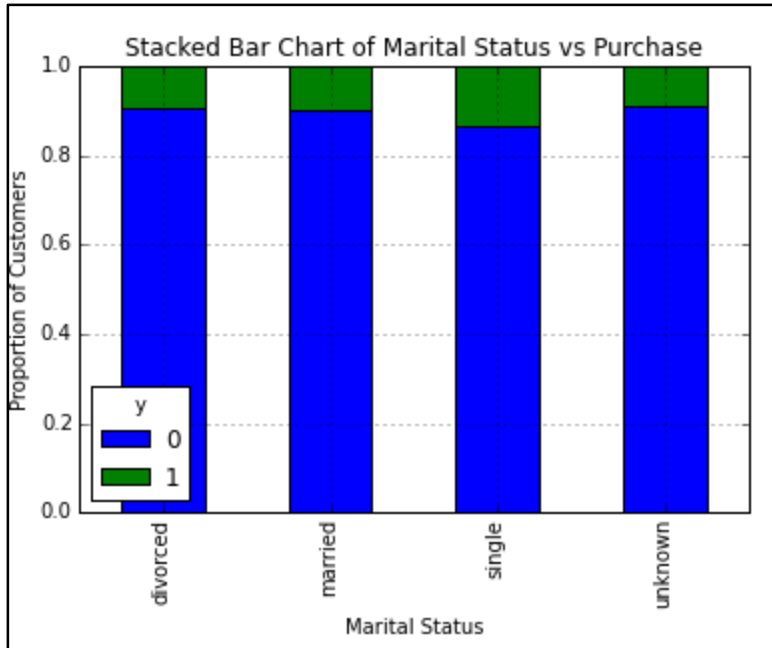


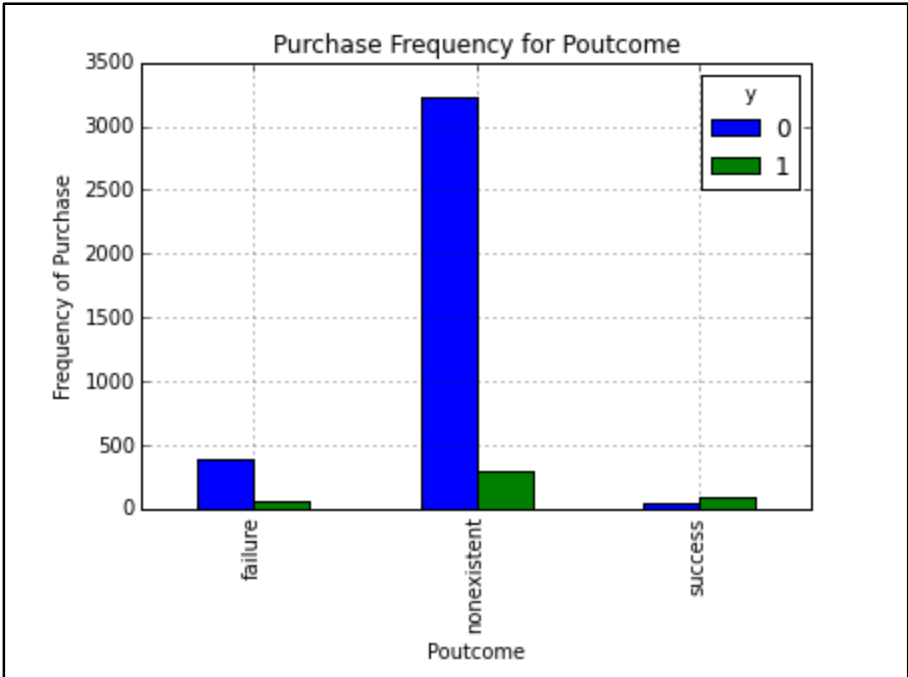
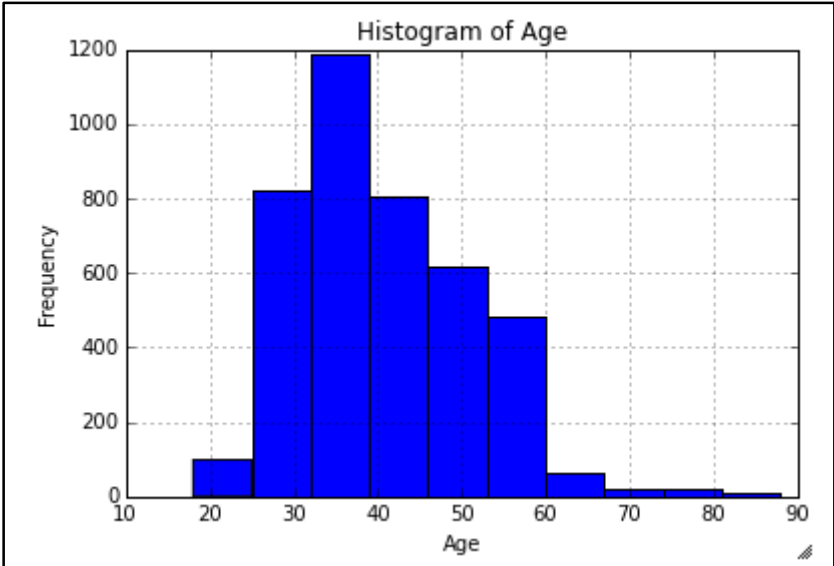


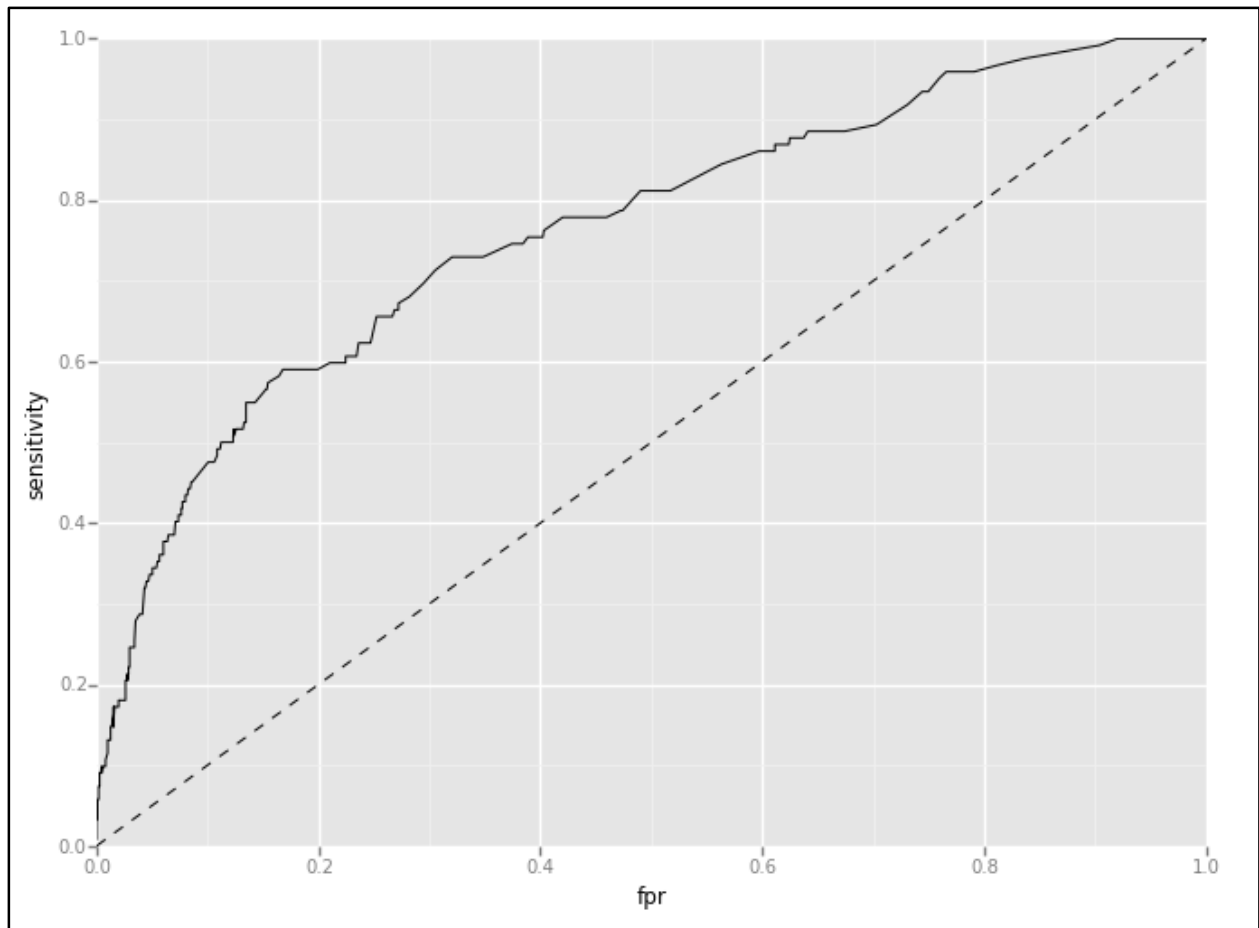
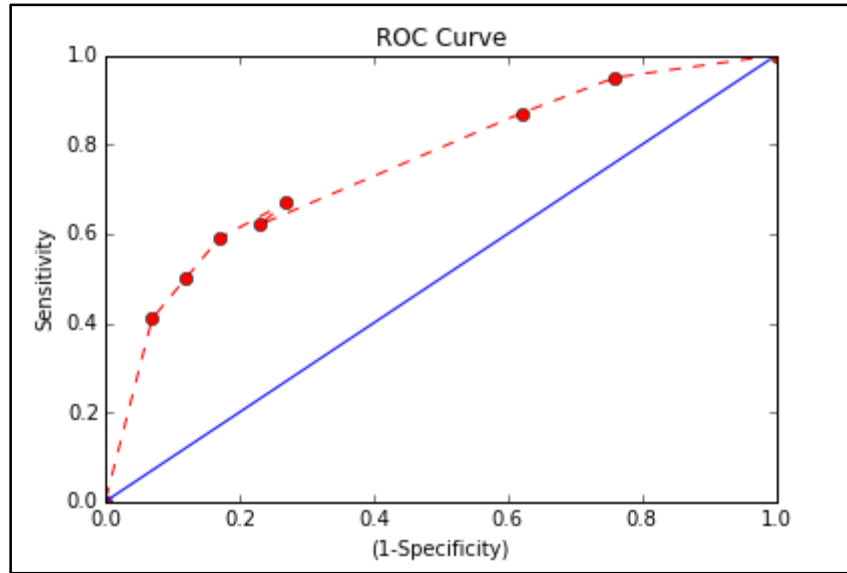


## Chapter 6: Logistic Regression with Python

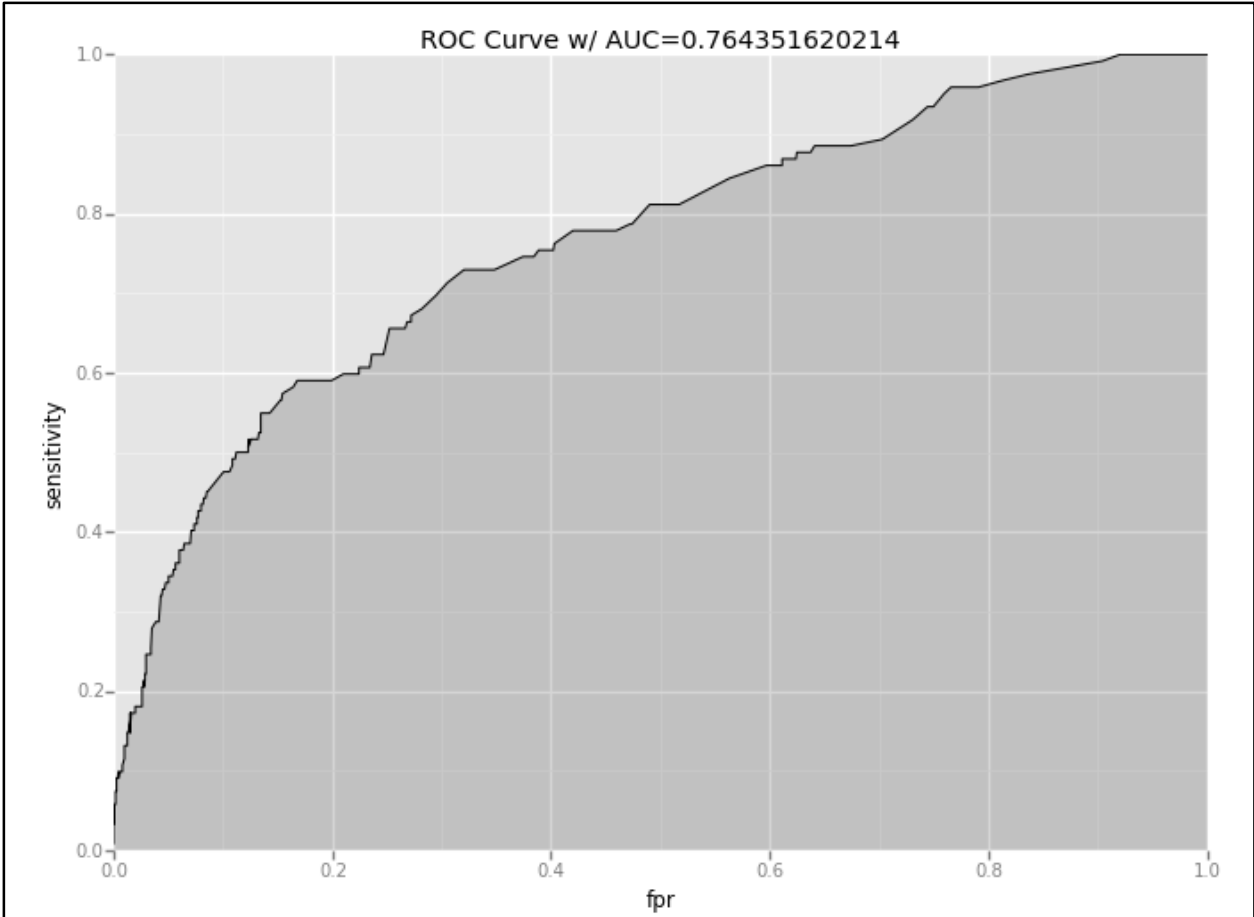




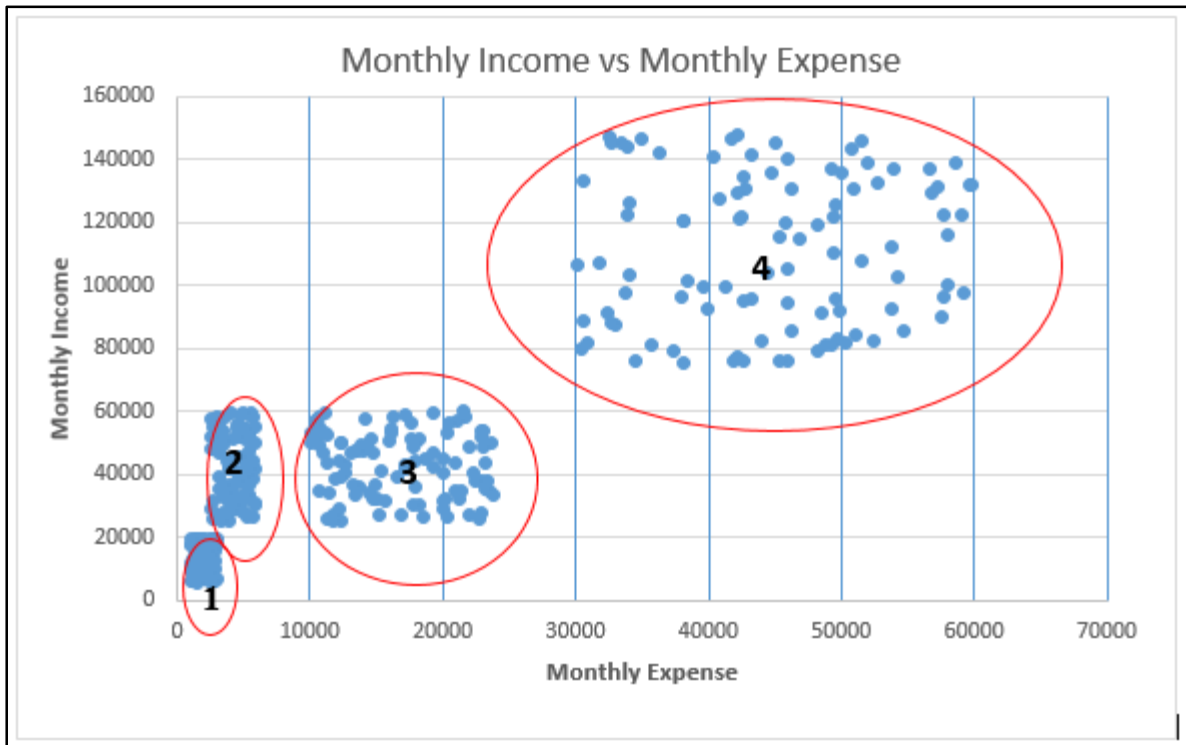








## Chapter 7: Clustering with Python

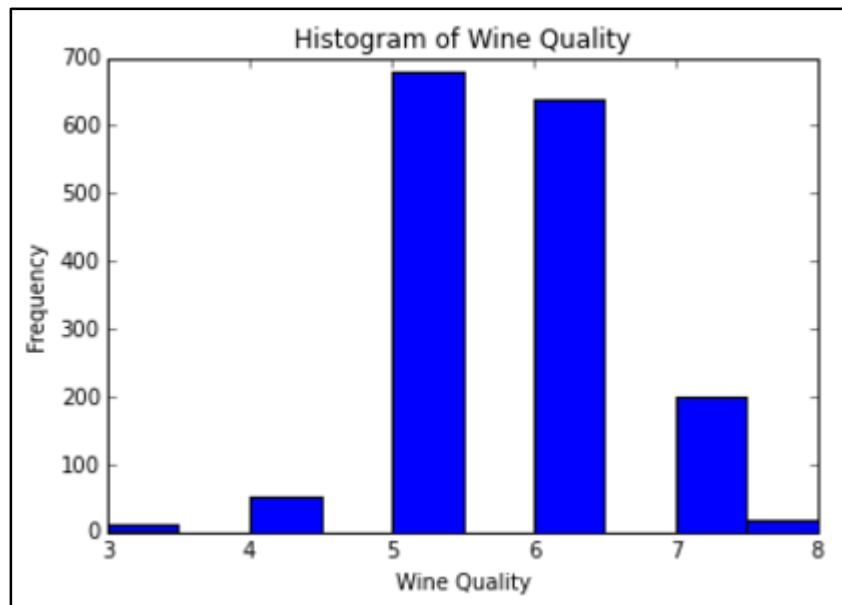


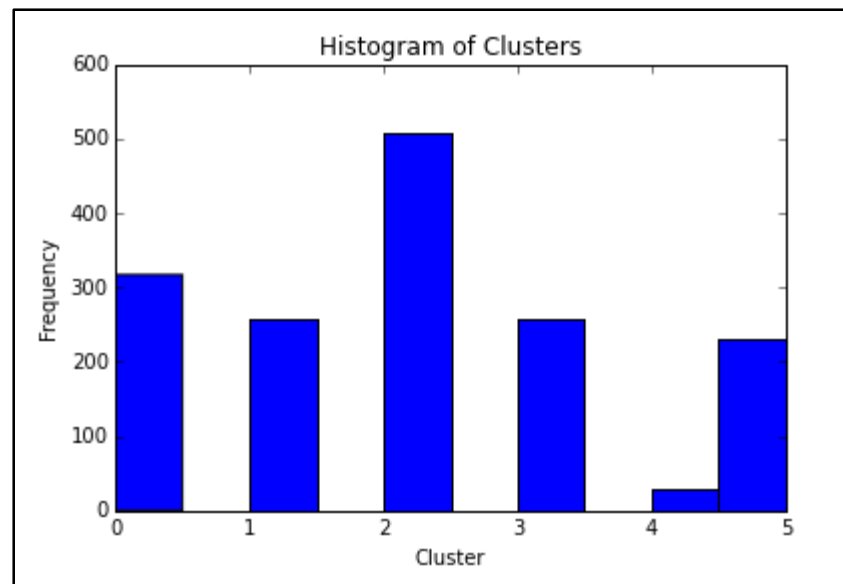
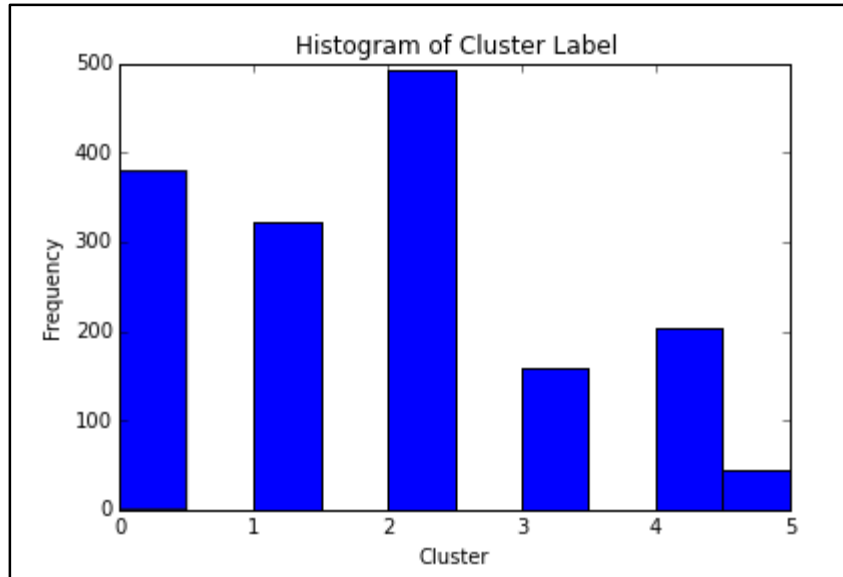
	1	2	3	4	5	6	7	8	9	10
1	0	0.67	1.01	0.63	0.33	0.45	0.72	0.49	0.37	0.04
2	0.67	0	0.76	1.06	0.55	0.47	0.06	0.36	0.81	0.66
3	1.01	0.76	0	0.88	1.08	0.62	0.76	1	0.84	0.97
4	0.63	1.06	0.88	0	0.89	0.65	1.1	1.01	0.35	0.6
5	0.33	0.55	1.08	0.89	0	0.6	0.6	0.23	0.68	0.35
6	0.45	0.47	0.62	0.65	0.6	0	0.5	0.58	0.38	0.41
7	0.72	0.06	0.76	1.1	0.6	0.5	0	0.41	0.85	0.71
8	0.49	0.36	1	1.01	0.23	0.58	0.41	0	0.78	0.5
9	0.37	0.81	0.84	0.35	0.68	0.38	0.85	0.78	0	0.34
10	0.04	0.66	0.97	0.6	0.35	0.41	0.71	0.5	0.34	0

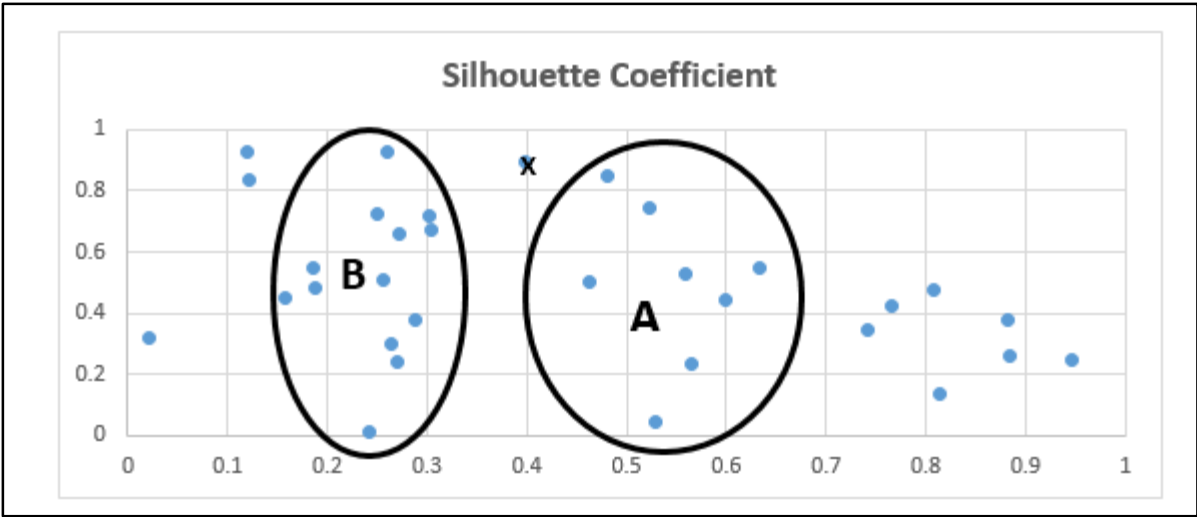
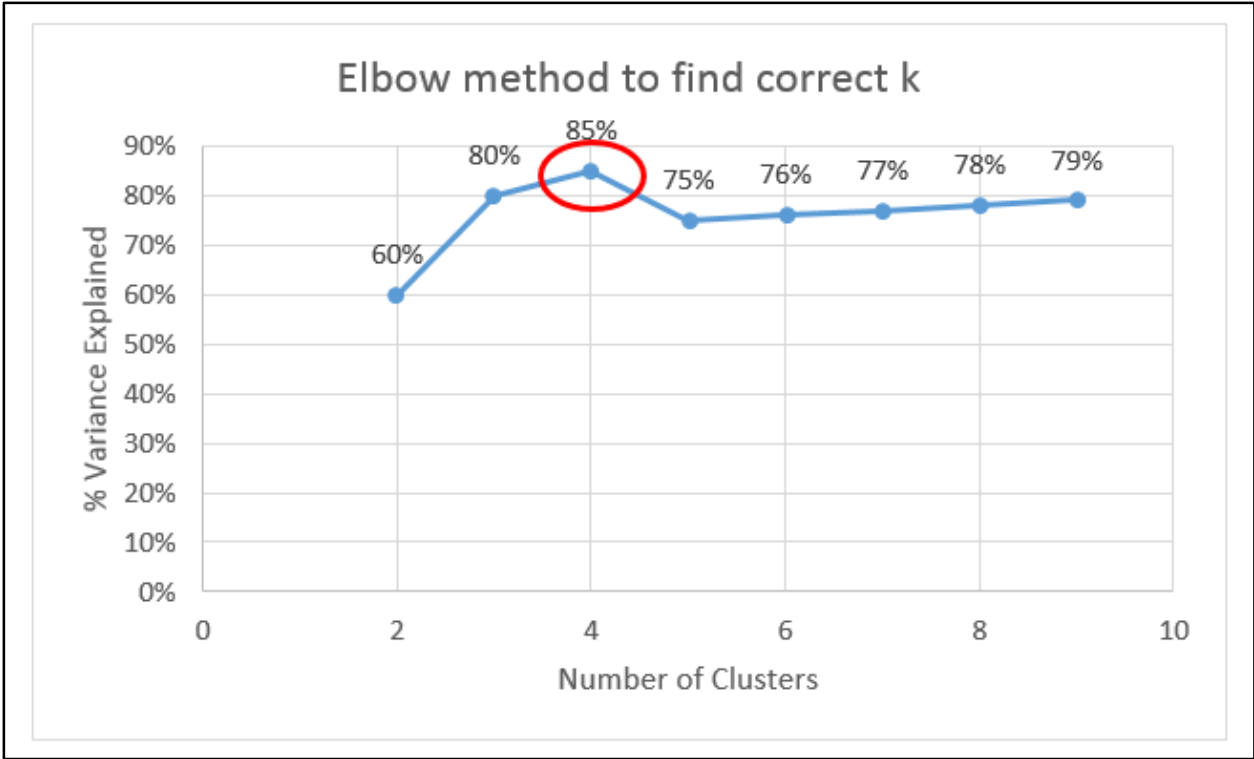
	<b>1 10</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>1 10</b>	0	0.66	0.97	0.6	0.33	0.41	0.71	0.49	0.34
<b>2</b>	0.66	0	0.76	1.06	0.55	0.47	<b>0.06</b>	0.36	0.81
<b>3</b>	0.97	0.76	0	0.88	1.08	0.62	0.76	1	0.84
<b>4</b>	0.6	1.06	0.88	0	0.89	0.65	1.1	1.01	0.35
<b>5</b>	0.33	0.55	1.08	0.89	0	0.6	0.6	0.23	0.68
<b>6</b>	0.41	0.47	0.62	0.65	0.6	0	0.5	0.58	0.38
<b>7</b>	0.71	<b>0.06</b>	0.76	1.1	0.6	0.5	0	0.41	0.85
<b>8</b>	0.49	0.36	1	1.01	0.23	0.58	0.41	0	0.78
<b>9</b>	0.34	0.81	0.84	0.35	0.68	0.38	0.85	0.78	0

	<b>1 10</b>	<b>2 7</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>8</b>	<b>9</b>
<b>1 10</b>	0	0.66	0.97	0.6	0.33	0.41	0.49	0.34
<b>2 7</b>	0.66	0	0.76	1.06	0.55	0.47	0.36	0.81
<b>3</b>	0.97	0.76	0	0.88	1.08	0.62	1	0.84
<b>4</b>	0.6	1.06	0.88	0	0.89	0.65	1.01	0.35
<b>5</b>	0.33	0.55	1.08	0.89	0	0.6	<b>0.23</b>	0.68
<b>6</b>	0.41	0.47	0.62	0.65	0.6	0	0.58	0.38
<b>8</b>	0.49	0.36	1	1.01	<b>0.23</b>	0.58	0	0.78
<b>9</b>	0.34	0.81	0.84	0.35	0.68	0.38	0.78	0

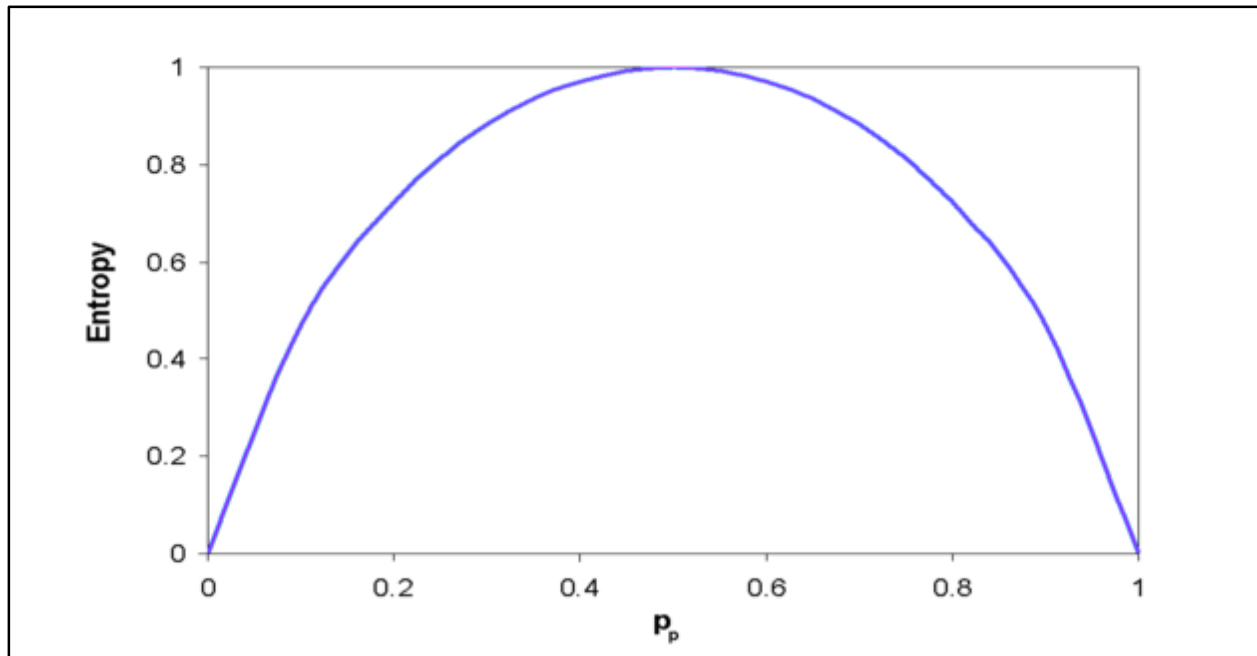
	1 10	2 7	5 8	3	4	6	9
1 10	0	0.66	0.33	0.97	0.6	0.41	0.34
2 7	0.66	0	0.36	0.76	1.06	0.47	0.81
5 8	0.33	0.36	0	1	0.89	0.58	0.68
3	0.97	0.76	1	0	0.88	0.62	0.84
4	0.6	1.06	0.89	0.88	0	0.65	0.35
6	0.41	0.47	0.58	0.62	0.65	0	0.38
9	0.34	0.81	0.68	0.84	0.35	0.38	0







## Chapter 8: Trees and Random Forests with Python



```
# appending 333 similar datasets to form a bigger dataset

import pandas as pd          # importing pandas library
filepath='E:/Personal/Learning/Predictive Modeling Book/Book Datasets/Merge and Join/lotofdata' # defining filepath variable as the folder
# which has all the small datasets
data_final=pd.read_csv('E:/Personal/Learning/Predictive Modeling Book/Book Datasets/Merge and Join/lotofdata/001.csv') # initialising the
# data-final data frame with the first dataset of the lot
data_final_size=len(data_final) # initialising the data_final_size variable which counts the number of rows in the data_final data frame
for i in range(1,333):      # defining a loop over all the 333 files
    if i<10:
        filename='0'+str(i)+'.csv' # the files are named as 001.csv, 101.csv etc. Accordingly, 3 conditions arise for the filename
        # variable. i<10 requires appending 2 zeros at the beginning.
    if 10<=i<100:
        filename='0'+str(i)+'.csv' # i<100 requires appending 1 zeros at the beginning.
    if i>=100:
        filename=str(i)+'.csv'     # i>=100 requires appending no zeros at the beginning.

    file=filepath+'/'+filename     # defining the file variable by appending filepath and filename variable. file variable contains a new
    # file in every iteration
    data=pd.read_csv(file)        # file is read as data frame called data
    data_final_size+=len(data)    # data_final_size variable is updated by adding the length of the currently read file

    data_final=pd.concat([data_final,data],axis=0) # concatenating/appending data to the data_final data frame on the axis=0 i.e. on rows
print data_final_size            # printing the final_data_size variable containing the number of rows in the final
# data frame
```