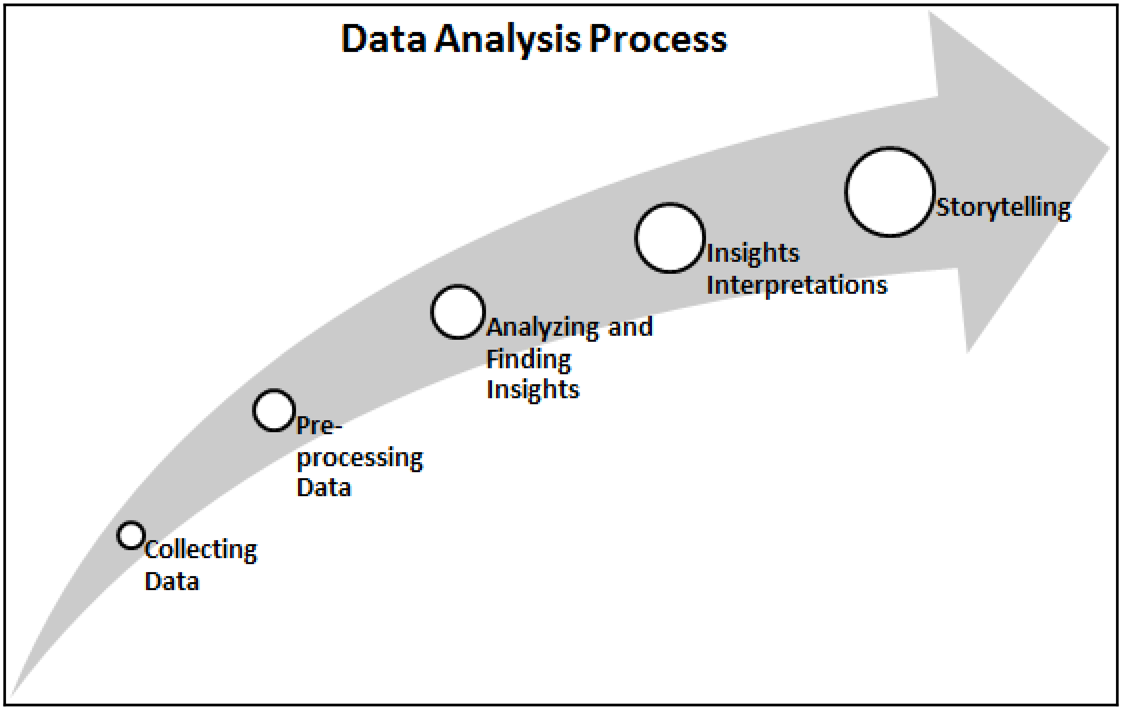
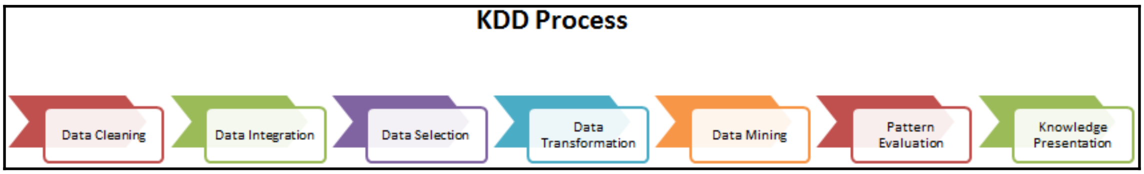
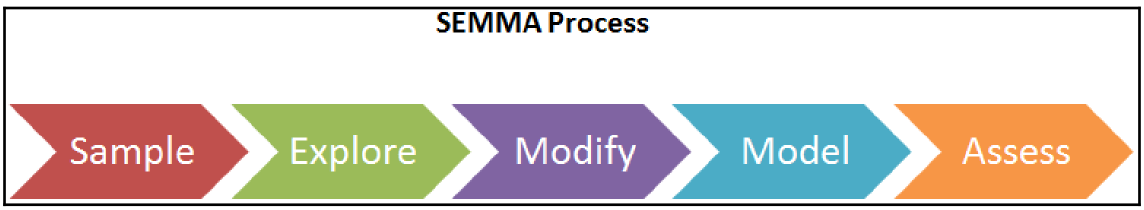
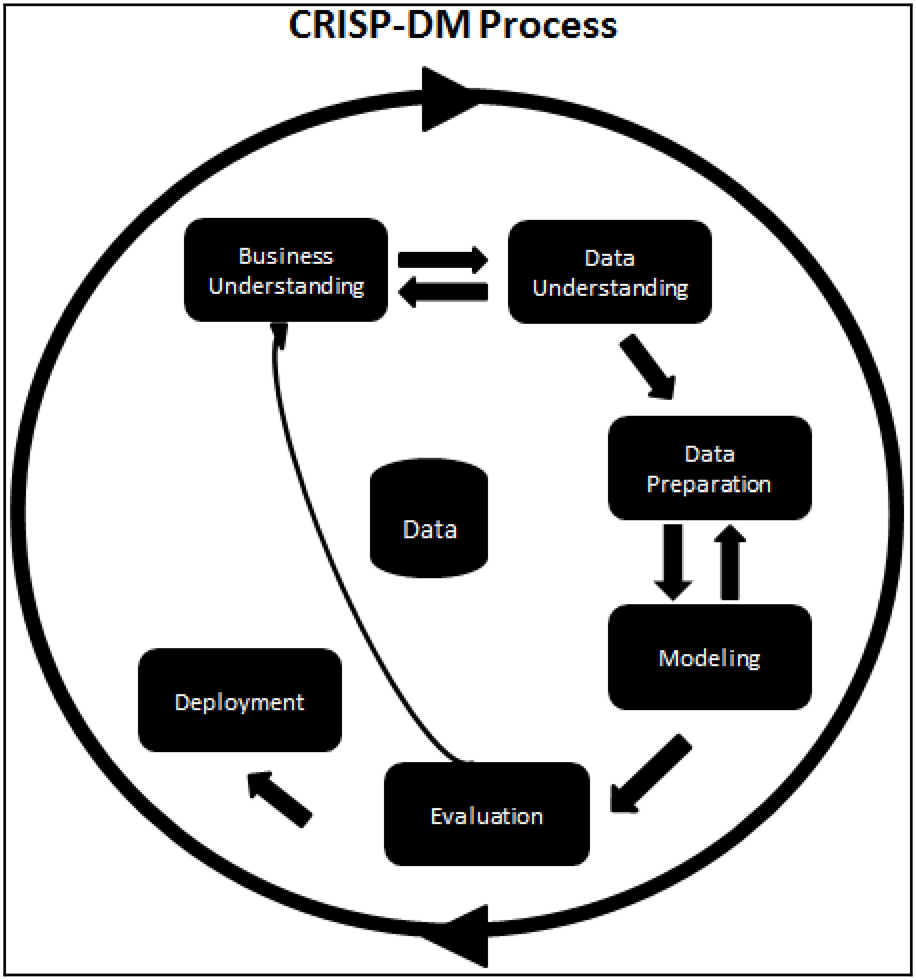
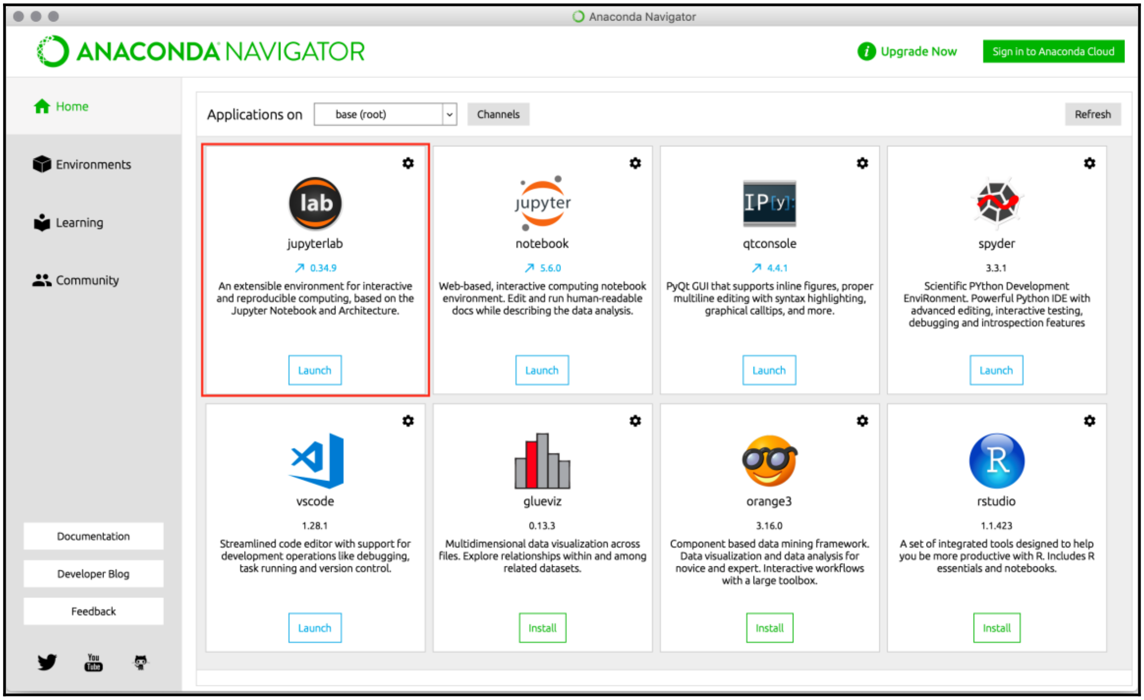
Chapter 1: Getting Started with Python Libraries

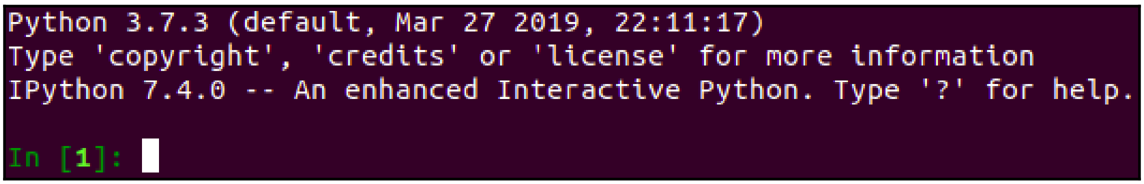







CRISP-DM Process
Business Understanding
Data Understanding
Data Preparation
Data
Modeling
Deployment
Evaluation


Anaconda Navigator
ANACONDA NAVIGATOR
Upgrade Now
Sign in to Anaconda Cloud
Home
Environments
Learning
Community
Applications on
base (root)
Channels
Refresh
jupyterlab
0.34.9
An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.
Launch
notebook
5.6.0
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
Launch
qtconsole
4.4.1
PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.
Launch
spyder
3.3.1
Scientific PYthon Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features
Launch
vscode
1.28.1
Streamlined code editor with support for development operations like debugging, task running and version control.
Launch
glueviz
0.13.3
Multidimensional data visualization across files. Explore relationships within and among related datasets.
Install
orange3
3.16.0
Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox.
Install
rstudio
1.1.423
A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.
Install
Documentation
Developer Blog
Feedback


Python 3.7.3 (default, Mar 27 2019, 22:11:17)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.4.0 -- An enhanced Interactive Python. Type '?' for help.
In [1]:

```
In [1]: x=5

In [2]: x=x+5

In [3]: y=x+5

In [4]: x,y
Out[4]: (10, 15)

In [5]: history
x=5
x=x+5
y=x+5
x,y
history
```

```
$ ipython3
Python 3.7.3 (default, Mar 27 2019, 22:11:17)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.4.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: !pwd
/home/avinash
```

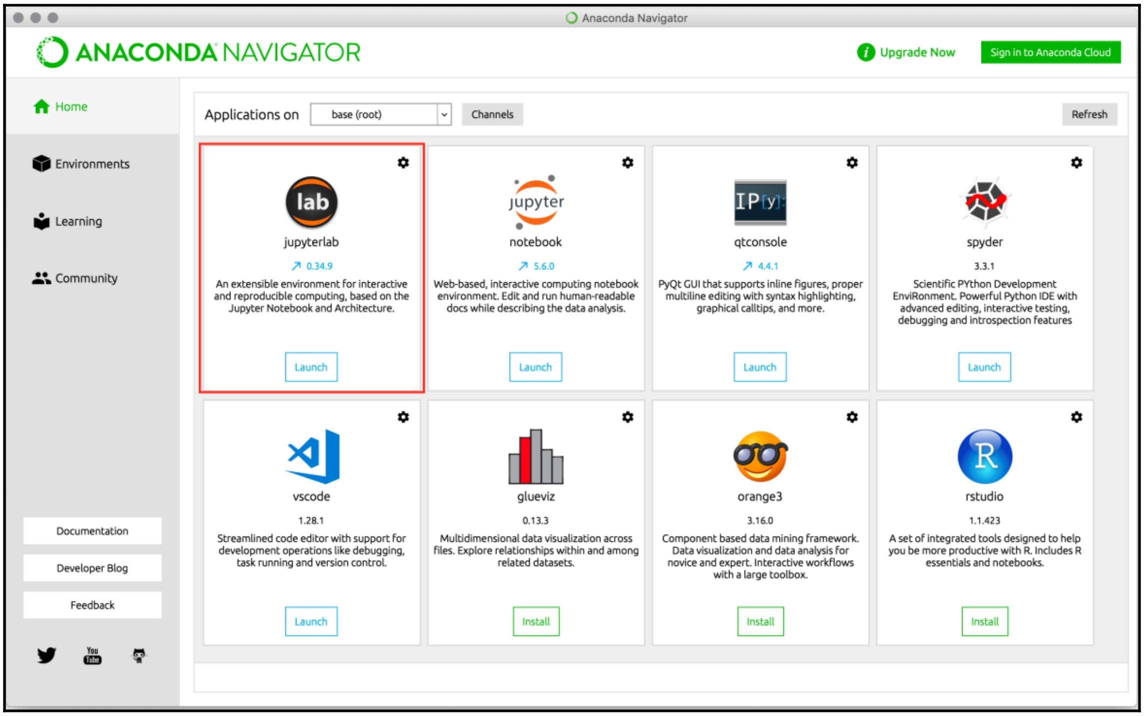
```
In [6]: def helloworld():
   ...:     print("Hello Everyone!")
   ...:

In [7]: helloworld()
Hello Everyone!
```

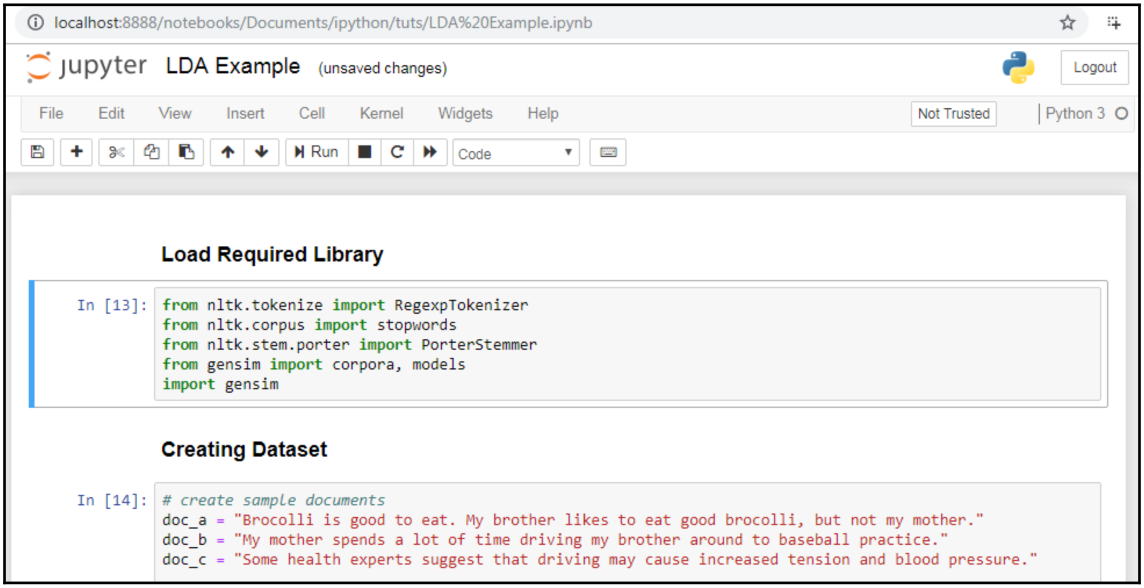
```
$ ipython3
Python 3.7.3 (default, Mar 27 2019, 22:11:17)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.4.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: exit()
```

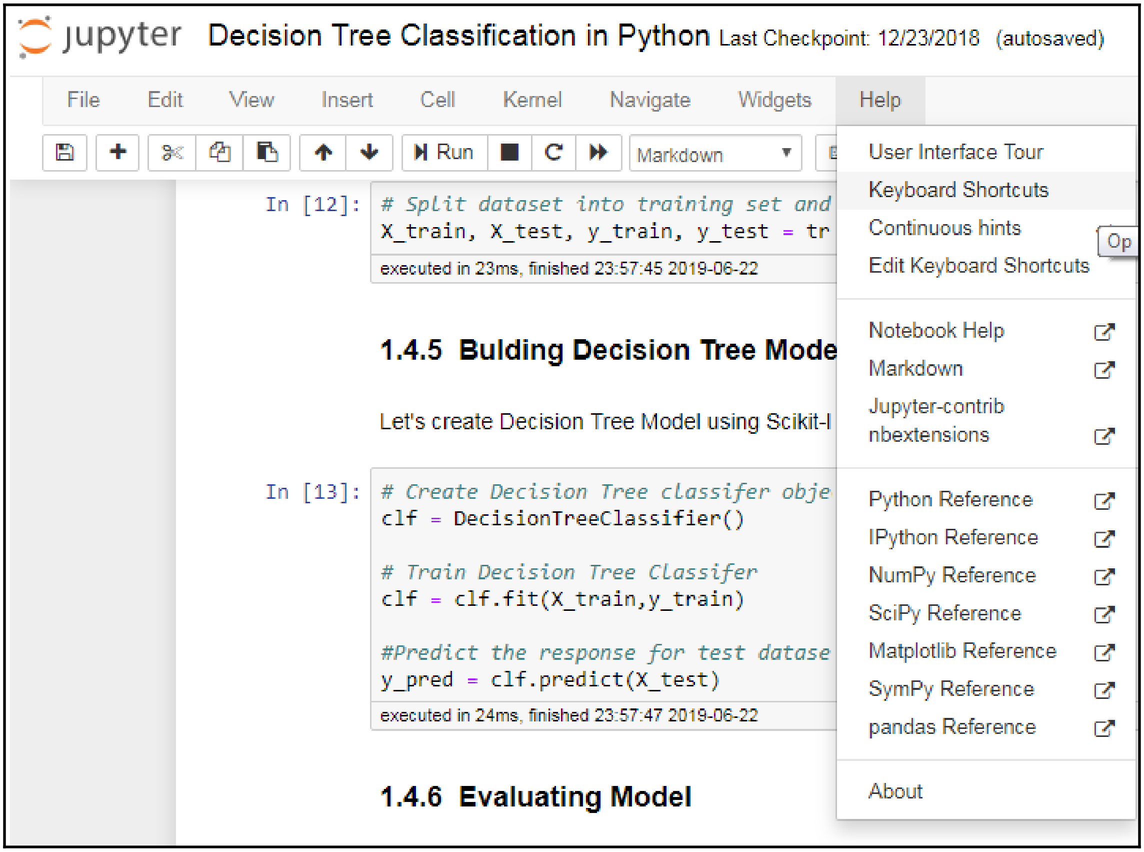
```
$ ipython3
Python 3.7.3 (default, Mar 27 2019, 22:11:17)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.4.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: quit()
```

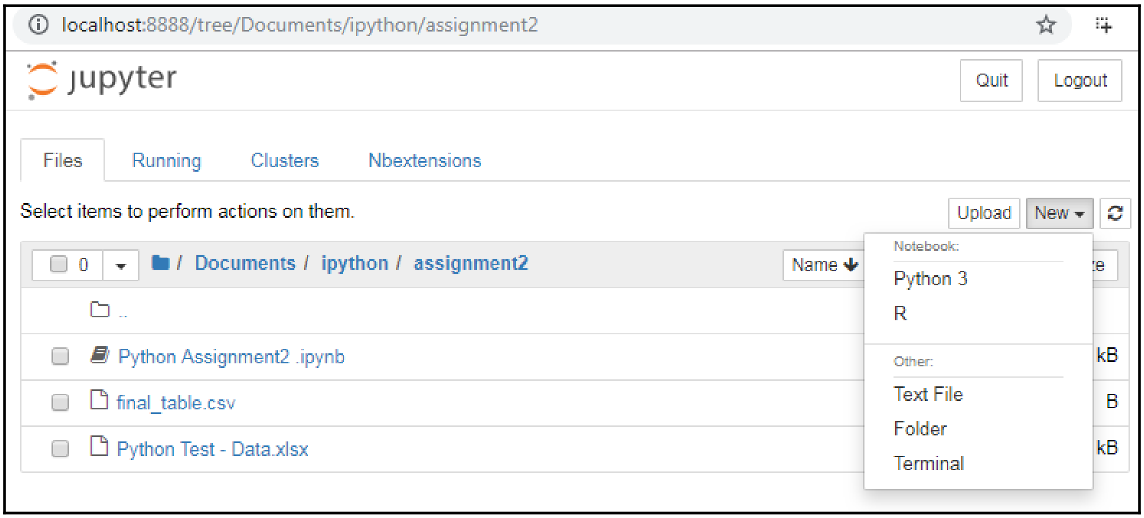
```
In [8]: help(numpy.ar
  arange()       arcsinh        argmax()       argwhere()     array_equal()
  arccos         arctan         argmin()       around()       array_equiv()
  arccosh        arctan2        argpartition() array()        array_repr()   >
  arcsin         arctanh        argsort()      array2string() array_split()
```

```
In [1]: import numpy

In [2]: numpy.arange?
```

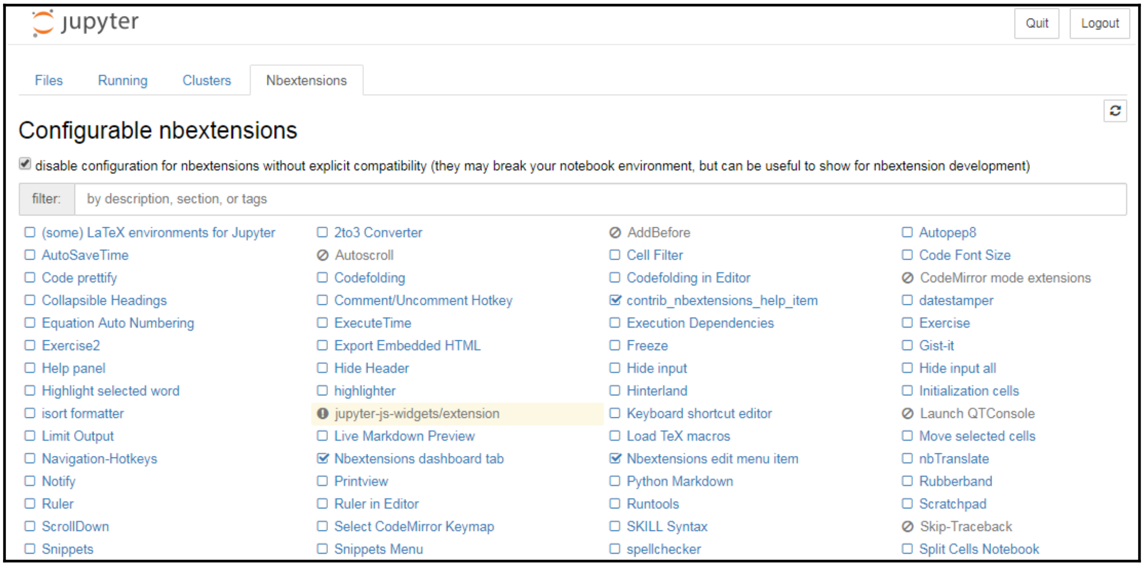


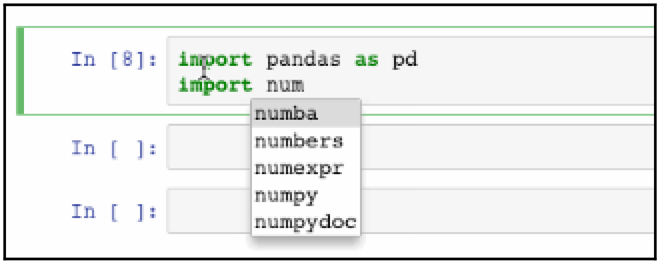

localhost:8888/notebooks/Documents/ipython/tuts/LDA%20Example.ipynb
jupyter LDA Example (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
Load Required Library
In [13]:
from nltk.tokenize import RegexpTokenizer
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from gensim import corpora, models
import gensim
Creating Dataset
In [14]:
# create sample documents
doc_a = "Brocolli is good to eat. My brother likes to eat good brocolli, but not my mother."
doc_b = "My mother spends a lot of time driving my brother around to baseball practice."
doc_c = "Some health experts suggest that driving may cause increased tension and blood pressure."

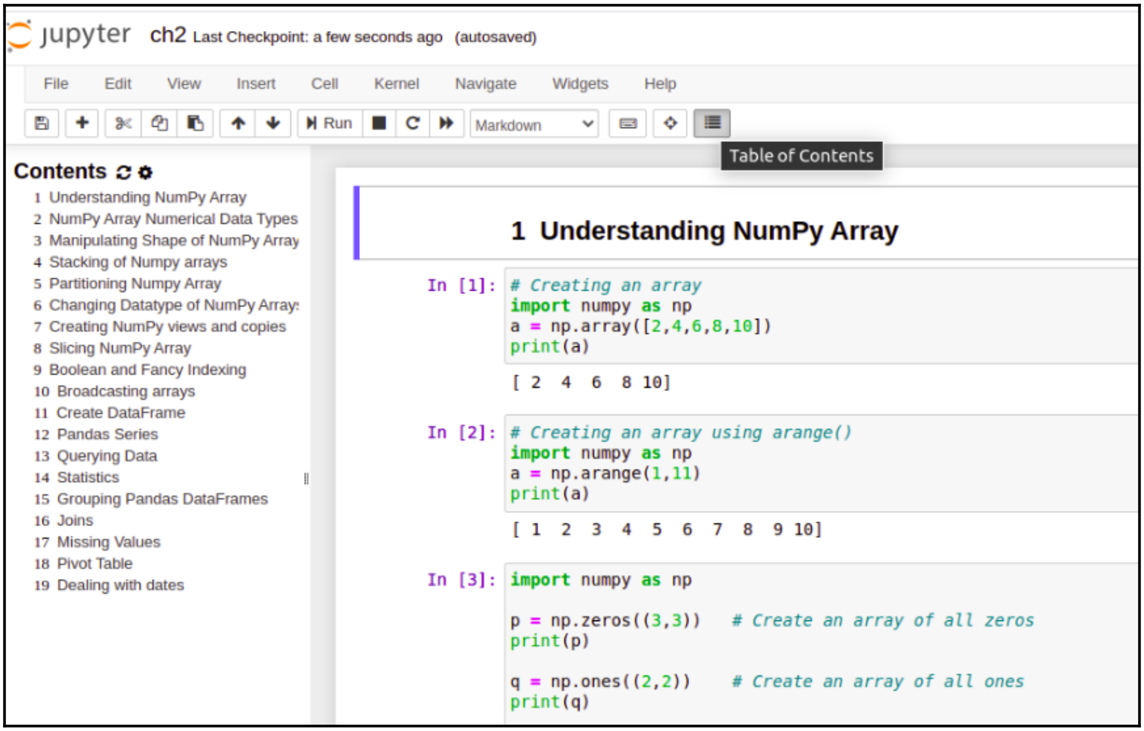

jupyter
Decision Tree Classification in Python Last Checkpoint: 12/23/2018 (autosaved)
File
Edit
View
Insert
Cell
Kernel
Navigate
Widgets
Help
Run
Markdown
User Interface Tour
Keyboard Shortcuts
Continuous hints
Edit Keyboard Shortcuts
Notebook Help
Markdown
Jupyter-contrib nbextensions
Python Reference
IPython Reference
NumPy Reference
SciPy Reference
Matplotlib Reference
SymPy Reference
pandas Reference
About
In [12]:
Split dataset into training set and
X_train, X_test, y_train, y_test = tr
executed in 23ms, finished 23:57:45 2019-06-22
1.4.5 Bulding Decision Tree Mode
Let's create Decision Tree Model using Scikit-l
In [13]:
# Create Decision Tree classifer obje
clf = DecisionTreeClassifier()
# Train Decision Tree Classifer
clf = clf.fit(X_train,y_train)
#Predict the response for test datase
y_pred = clf.predict(X_test)
executed in 24ms, finished 23:57:47 2019-06-22
1.4.6 Evaluating Model


localhost:8888/tree/Documents/ipython/assignment2
jupyter
Quit
Logout
Files
Running
Clusters
Nbextensions
Select items to perform actions on them.
Upload
New
Documents / ipython / assignment2
Name
Notebook:
Python 3
R
Other:
Text File
Folder
Terminal
Python Assignment2 .ipynb
final_table.csv
Python Test - Data.xlsx

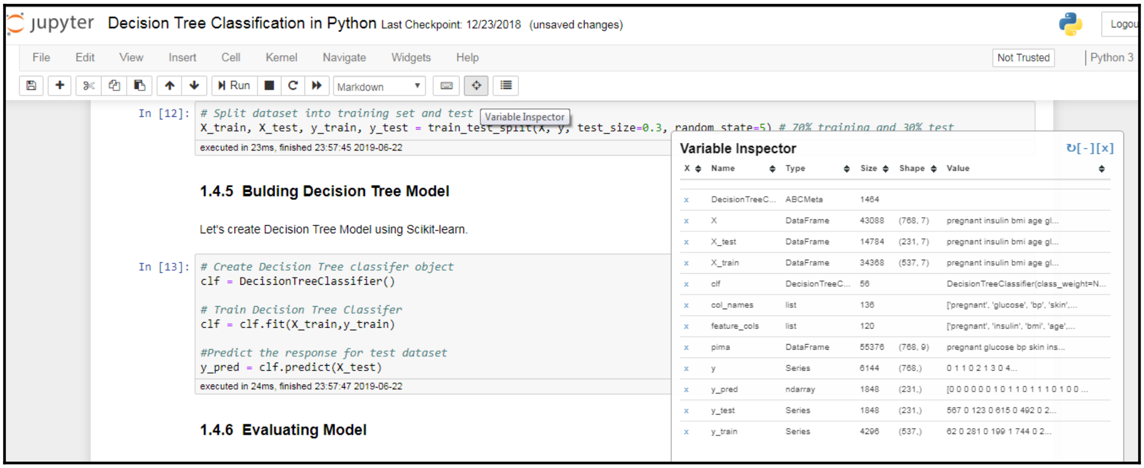
```
In [1]: !dir

 Volume in drive C has no label.
 Volume Serial Number is D0C9-975F

 Directory of C:\Users\Aadmin\Documents\ipython\tuts

22-06-2019  23:34    <DIR>          .
22-06-2019  23:34    <DIR>          ..
11-05-2019  16:19         1,147,737  Inventory Model Simulation with Spreadsheet.ipynb
22-06-2019  22:56    <DIR>          .ipynb_checkpoints
24-02-2019  20:10            23,475 AB Testing.ipynb
02-05-2019  21:15           150,385 AdaBoost Classifier in Python-Copy1.ipynb
19-03-2019  09:21           191,553 AdaBoost Classifier in Python.ipynb
10-07-2018  15:54        25,337,978 articles.txt
20-12-2018  15:26           163,127 Cohort Analysis.ipynb
18-03-2019  09:15    <DIR>          computer_vision
22-06-2019  23:34           209,831 Customer Life Time Value (final version).ipynb
11-05-2019  10:03            81,155 Customer life time value V1.ipynb
11-05-2019  10:01           431,692 Customer Lifetime Value - Predictive Modeling.ipynb
```

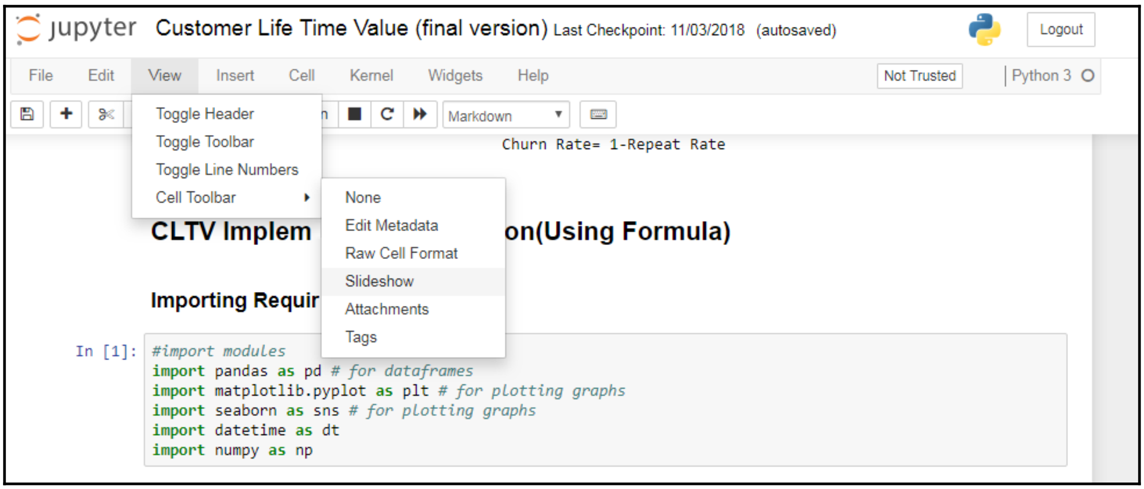


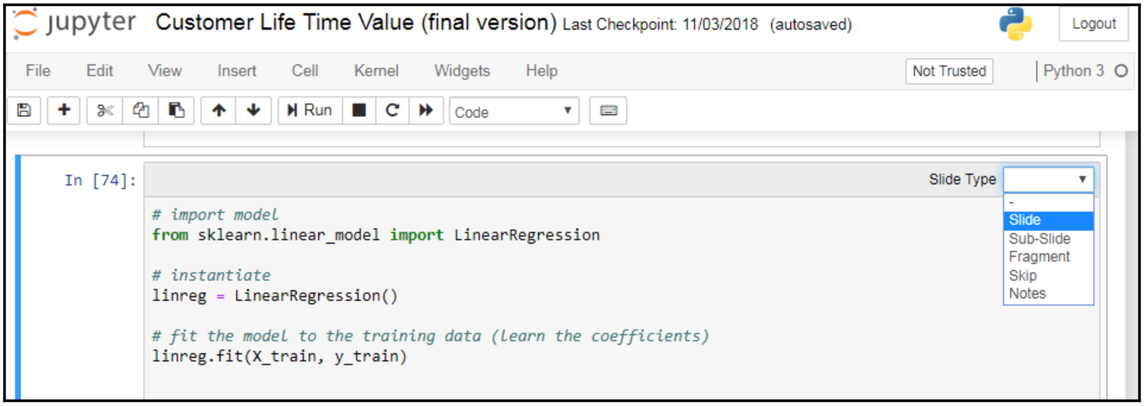

In [8]: import pandas as pd
import num
numba
numbers
numexpr
numpy
numpydoc
In [ ]:
In [ ]:

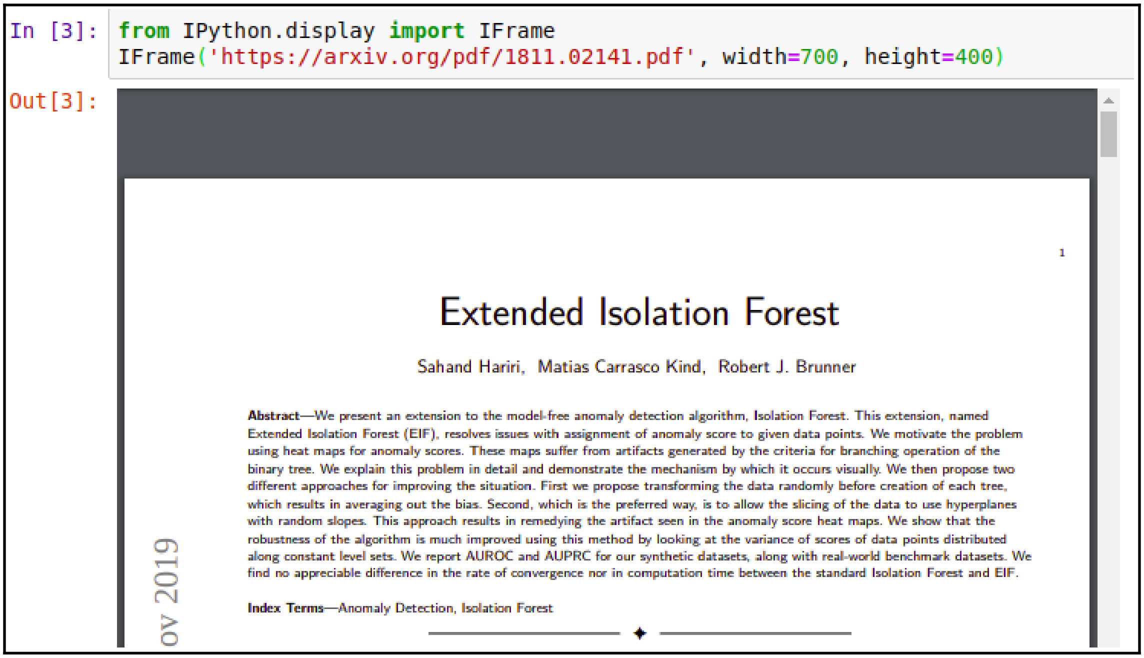

jupyter
ch2 Last Checkpoint: a few seconds ago (autosaved)
File
Edit
View
Insert
Cell
Kernel
Navigate
Widgets
Help
Run
Markdown
Table of Contents
Contents
1 Understanding NumPy Array
2 NumPy Array Numerical Data Types
3 Manipulating Shape of NumPy Array
4 Stacking of Numpy arrays
5 Partitioning Numpy Array
6 Changing Datatype of NumPy Array
7 Creating NumPy views and copies
8 Slicing NumPy Array
9 Boolean and Fancy Indexing
10 Broadcasting arrays
11 Create DataFrame
12 Pandas Series
13 Querying Data
14 Statistics
15 Grouping Pandas DataFrames
16 Joins
17 Missing Values
18 Pivot Table
19 Dealing with dates
1 Understanding NumPy Array
In [1]: # Creating an array
import numpy as np
a = np.array([2,4,6,8,10])
print(a)
[2 4 6 8 10]
In [2]: # Creating an array using arange()
import numpy as np
a = np.arange(1,11)
print(a)
[1 2 3 4 5 6 7 8 9 10]
In [3]: import numpy as np
p = np.zeros((3,3)) # Create an array of all zeros
print(p)
q = np.ones((2,2)) # Create an array of all ones
print(q)

```
In [13]: # Create Decision Tree classifer object
         clf = DecisionTreeClassifier()

         # Train Decision Tree Classifer
         clf = clf.fit(X_train,y_train)

         #Predict the response for test dataset
         y_pred = clf.predict(X_test)

         executed in 24ms, finished 23:57:47 2019-06-22
```






jupyter Customer Life Time Value (final version) Last Checkpoint: 11/03/2018 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Code
In [74]:
Slide Type
-
Slide
Sub-Slide
Fragment
Skip
Notes
# import model
from sklearn.linear_model import LinearRegression
# instantiate
linreg = LinearRegression()
# fit the model to the training data (learn the coefficients)
linreg.fit(X_train, y_train)

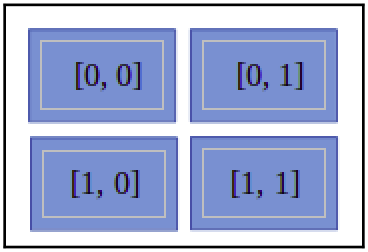

In [3]: from IPython.display import IFrame
IFrame('https://arxiv.org/pdf/1811.02141.pdf', width=700, height=400)
Out[3]:
1
Extended Isolation Forest
Sahand Hariri, Matias Carrasco Kind, Robert J. Brunner
Abstract—We present an extension to the model-free anomaly detection algorithm, Isolation Forest. This extension, named Extended Isolation Forest (EIF), resolves issues with assignment of anomaly score to given data points. We motivate the problem using heat maps for anomaly scores. These maps suffer from artifacts generated by the criteria for branching operation of the binary tree. We explain this problem in detail and demonstrate the mechanism by which it occurs visually. We then propose two different approaches for improving the situation. First we propose transforming the data randomly before creation of each tree, which results in averaging out the bias. Second, which is the preferred way, is to allow the slicing of the data to use hyperplanes with random slopes. This approach results in remedying the artifact seen in the anomaly score heat maps. We show that the robustness of the algorithm is much improved using this method by looking at the variance of scores of data points distributed along constant level sets. We report AUROC and AUPRC for our synthetic datasets, along with real-world benchmark datasets. We find no appreciable difference in the rate of convergence nor in computation time between the standard Isolation Forest and EIF.
Index Terms—Anomaly Detection, Isolation Forest
ov 2019

In [2]:

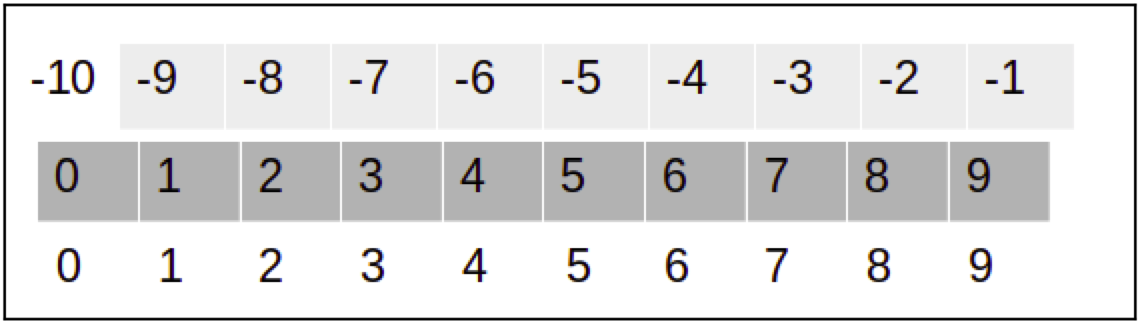
```
from IPython.display import YouTubeVideo
YouTubeVideo('ukzFI9rgwfU', width=800, height=300)
```

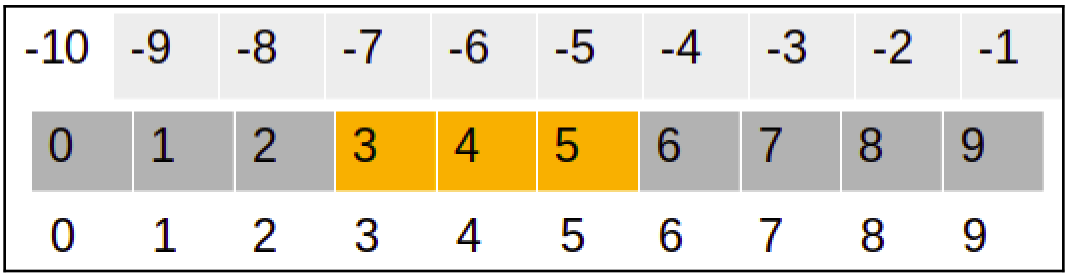
Out[2]:

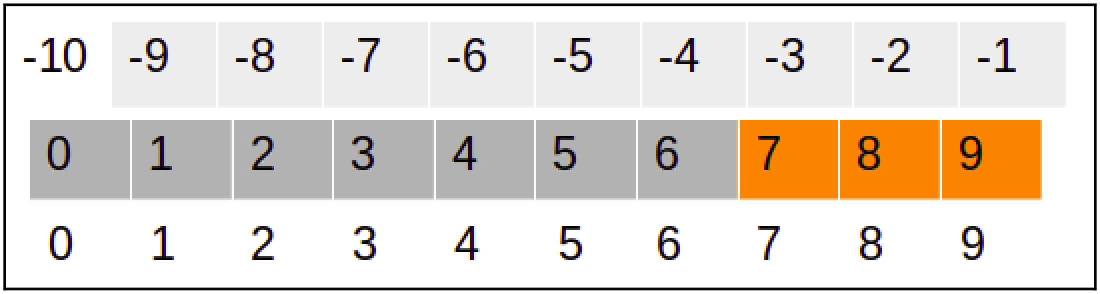
Chapter 2: NumPy and pandas

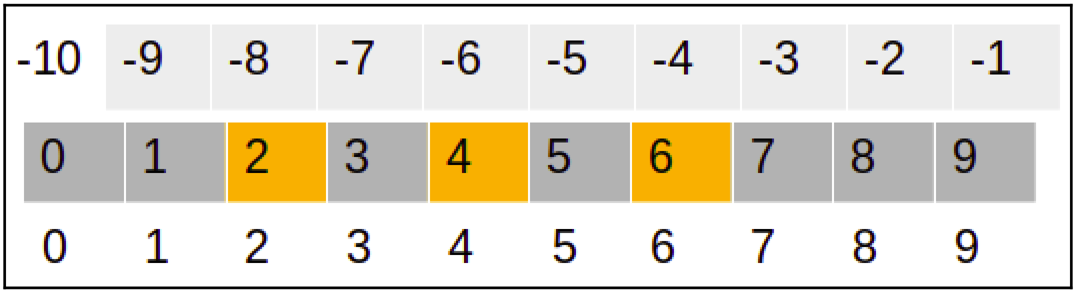


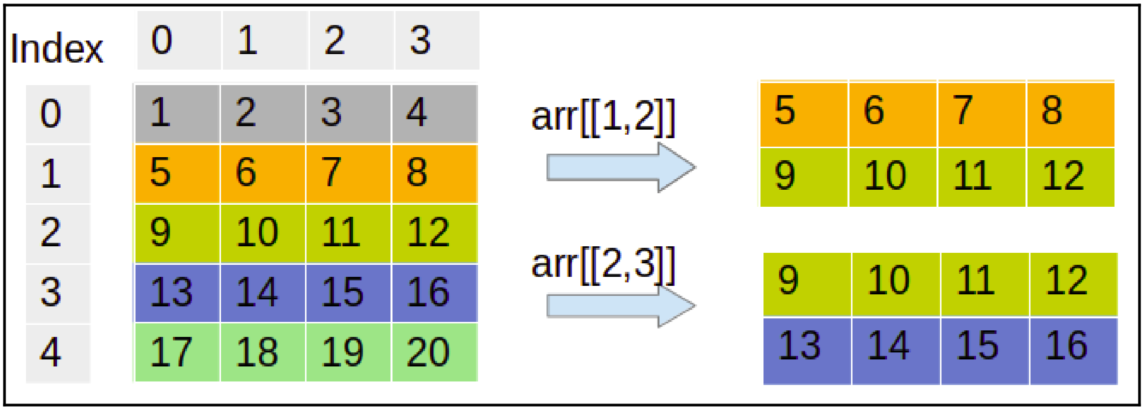

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-29-61a3a50e24b1> in <module>
----> 1 np.int(42.0 + 1.j)

TypeError: can't convert complex to int
```











	Country	CountryID	Continent	Adolescent fertility rate (%)	Adult literacy rate (%)	Gross national income per capita (PPP international $)	Net primary school enrolment ratio female (%)	Net primary school enrolment ratio male (%)	Population (in thousands) total
0	Afghanistan	1	1	151.0	28.0	NaN	NaN	NaN	26088.0
1	Albania	2	2	27.0	98.7	6000.0	93.0	94.0	3172.0
2	Algeria	3	3	6.0	69.9	5940.0	94.0	96.0	33351.0
3	Andorra	4	2	NaN	NaN	NaN	83.0	83.0	74.0
4	Angola	5	3	146.0	67.4	3890.0	49.0	51.0	16557.0

Date	Yearly Mean Total Sunspot Number	Yearly Mean Standard Deviation	Number of Observations	Definitive/Provisional Indicator
1700-12-31	8.3	NaN	NaN	1.0
1701-12-31	18.3	NaN	NaN	1.0
1702-12-31	26.7	NaN	NaN	1.0
1703-12-31	38.3	NaN	NaN	1.0
1704-12-31	60.0	NaN	NaN	1.0

Date	Yearly Mean Total Sunspot Number	Yearly Mean Standard Deviation	Number of Observations	Definitive/Provisional Indicator
2015-12-31	69.8	6.4	8903.0	1.0
2016-12-31	39.8	3.9	9940.0	1.0
2017-12-31	21.7	2.5	11444.0	1.0
2018-12-31	7.0	1.1	12611.0	1.0
2019-12-31	3.6	0.5	12401.0	0.0

Date	Yearly Mean Total Sunspot Number	Definitive/Provisional Indicator
1700-12-31	8.3	1.0
1701-12-31	18.3	1.0
1702-12-31	26.7	1.0
1703-12-31	38.3	1.0
1704-12-31	60.0	1.0

Date	Yearly Mean Total Sunspot Number	Yearly Mean Standard Deviation	Number of Observations	Definitive/Provisional Indicator
2002-12-31	163.6	9.8	6588.0	1.0
2003-12-31	99.3	7.1	7087.0	1.0
2004-12-31	65.3	5.9	6882.0	1.0
2005-12-31	45.8	4.7	7084.0	1.0
2006-12-31	24.7	3.5	6370.0	1.0
2007-12-31	12.6	2.7	6841.0	1.0
2008-12-31	4.2	2.5	6644.0	1.0
2009-12-31	4.8	2.5	6465.0	1.0
2010-12-31	24.9	3.4	6328.0	1.0
2011-12-31	80.8	6.7	6077.0	1.0
2012-12-31	84.5	6.7	5753.0	1.0
2013-12-31	94.0	6.9	5347.0	1.0

Date	Yearly Mean Total Sunspot Number	Yearly Mean Standard Deviation	Number of Observations	Definitive/Provisional Indicator
1705-12-31	96.7	NaN	NaN	1.0
1717-12-31	105.0	NaN	NaN	1.0
1718-12-31	100.0	NaN	NaN	1.0
1726-12-31	130.0	NaN	NaN	1.0
1727-12-31	203.3	NaN	NaN	1.0
1728-12-31	171.7	NaN	NaN	1.0
1729-12-31	121.7	NaN	NaN	1.0
1736-12-31	116.7	NaN	NaN	1.0
1737-12-31	135.0	NaN	NaN	1.0
1738-12-31	185.0	NaN	NaN	1.0
1739-12-31	168.3	NaN	NaN	1.0
1740-12-31	121.7	NaN	NaN	1.0

	CountryID	Continent	Adolescent fertility rate (%)	Adult literacy rate (%)	Gross national income per capita (PPP international $)	Net primary school enrolment ratio female (%)	Net primary school enrolment ratio male (%)	Population (in thousands) total
count	202.000000	202.000000	177.000000	131.000000	178.000000	179.000000	179.000000	1.890000e+02
mean	101.500000	3.579208	59.457627	78.871756	11250.112360	84.033520	85.698324	3.409964e+04
std	58.456537	1.808263	49.105286	20.415760	12586.753417	17.788047	15.451212	1.318377e+05
min	1.000000	1.000000	0.000000	23.600000	260.000000	6.000000	11.000000	2.000000e+00
25%	51.250000	2.000000	19.000000	68.400000	2112.500000	79.000000	79.500000	1.328000e+03
50%	101.500000	3.000000	46.000000	86.500000	6175.000000	90.000000	90.000000	6.640000e+03
75%	151.750000	5.000000	91.000000	95.300000	14502.500000	96.000000	96.000000	2.097100e+04
max	202.000000	7.000000	199.000000	99.800000	60870.000000	100.000000	100.000000	1.328474e+06

```
Country                                                  202
CountryID                                                202
Continent                                                202
Adolescent fertility rate (%)                            177
Adult literacy rate (%)                                  131
Gross national income per capita (PPP international $)   178
Net primary school enrolment ratio female (%)            179
Net primary school enrolment ratio male (%)              179
Population (in thousands) total                          189
dtype: int64
```

```
CountryID                                                 101.5
Continent                                                   3.0
Adolescent fertility rate (%)                              46.0
Adult literacy rate (%)                                    86.5
Gross national income per capita (PPP international $)   6175.0
Net primary school enrolment ratio female (%)              90.0
Net primary school enrolment ratio male (%)                90.0
Population (in thousands) total                          6640.0
dtype: float64
```

```
CountryID                                                   58.456537
Continent                                                    1.808263
Adolescent fertility rate (%)                               49.105286
Adult literacy rate (%)                                     20.415760
Gross national income per capita (PPP international $)   12586.753417
Net primary school enrolment ratio female (%)               17.788047
Net primary school enrolment ratio male (%)                 15.451212
Population (in thousands) total                         131837.708677
dtype: float64
```

Continent	CountryID	Adolescent fertility rate (%)	Adult literacy rate (%)	Gross national income per capita (PPP international $)	Net primary school enrolment ratio female (%)	Net primary school enrolment ratio male (%)	Population (in thousands) total
1	110.238095	37.300000	76.900000	14893.529412	85.789474	88.315789	16843.350000
2	100.333333	20.500000	97.911538	19777.083333	92.911111	93.088889	17259.627451
3	99.354167	111.644444	61.690476	3050.434783	67.574468	72.021277	16503.195652
4	56.285714	49.600000	91.600000	24524.000000	95.000000	94.400000	73577.333333
5	94.774194	77.888889	87.940909	7397.142857	89.137931	88.517241	15637.241379
6	121.228571	39.260870	87.607143	12167.200000	89.040000	89.960000	25517.142857
7	80.777778	57.333333	69.812500	2865.555556	85.444444	88.888889	317683.666667

```
Continent
1     76.900000
2     97.911538
3     61.690476
4     91.600000
5     87.940909
6     87.607143
7     69.812500
Name: Adult literacy rate (%), dtype: float64
```

	EmpNr	Dest
0	5	The Hague
1	3	Amsterdam
2	9	Rotterdam

	EmpNr	Amount
0	5	10.0
1	9	5.0
2	7	2.5

	EmpNr	Dest	Amount
0	5	The Hague	10.0
1	9	Rotterdam	5.0

	EmpNr	Dest	Amount
0	5	The Hague	10.0
1	3	Amsterdam	NaN
2	9	Rotterdam	5.0
3	7	NaN	2.5

	EmpNr	Dest	Amount
0	5	The Hague	10.0
1	9	Rotterdam	5.0
2	7	NaN	2.5

	EmpNr	Dest	Amount
0	5	The Hague	10.0
1	3	Amsterdam	NaN
2	9	Rotterdam	5.0

```
Country                                                    0
CountryID                                                  0
Continent                                                  0
Adolescent fertility rate (%)                             25
Adult literacy rate (%)                                   71
Gross national income per capita (PPP international $)    24
Net primary school enrolment ratio female (%)             23
Net primary school enrolment ratio male (%)               23
Population (in thousands) total                           13
dtype: int64
```

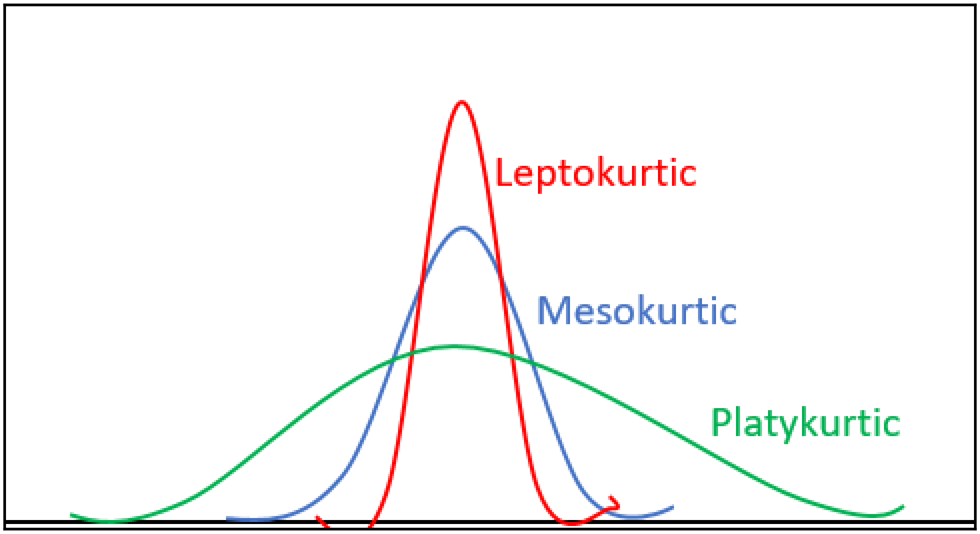
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 118 entries, 1 to 200
Data columns (total 9 columns):
Country                                                   118 non-null object
CountryID                                                 118 non-null int64
Continent                                                 118 non-null int64
Adolescent fertility rate (%)                             118 non-null float64
Adult literacy rate (%)                                   118 non-null float64
Gross national income per capita (PPP international $)    118 non-null float64
Net primary school enrolment ratio female (%)             118 non-null float64
Net primary school enrolment ratio male (%)               118 non-null float64
Population (in thousands) total                           118 non-null float64
dtypes: float64(6), int64(2), object(1)
memory usage: 9.2+ KB
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 202 entries, 0 to 201
Data columns (total 9 columns):
Country                                                   202 non-null object
CountryID                                                 202 non-null int64
Continent                                                 202 non-null int64
Adolescent fertility rate (%)                             202 non-null float64
Adult literacy rate (%)                                   202 non-null float64
Gross national income per capita (PPP international $)    202 non-null float64
Net primary school enrolment ratio female (%)             202 non-null float64
Net primary school enrolment ratio male (%)               202 non-null float64
Population (in thousands) total                           202 non-null float64
dtypes: float64(6), int64(2), object(1)
memory usage: 14.3+ KB
```

	Weather	Food	Price	Number
0	cold	soup	3.745401	8
1	hot	soup	9.507143	8
2	cold	icecream	7.319939	8
3	hot	chocolate	5.986585	8
4	cold	icecream	1.560186	8
5	hot	icecream	1.559945	8
6	cold	soup	0.580836	8

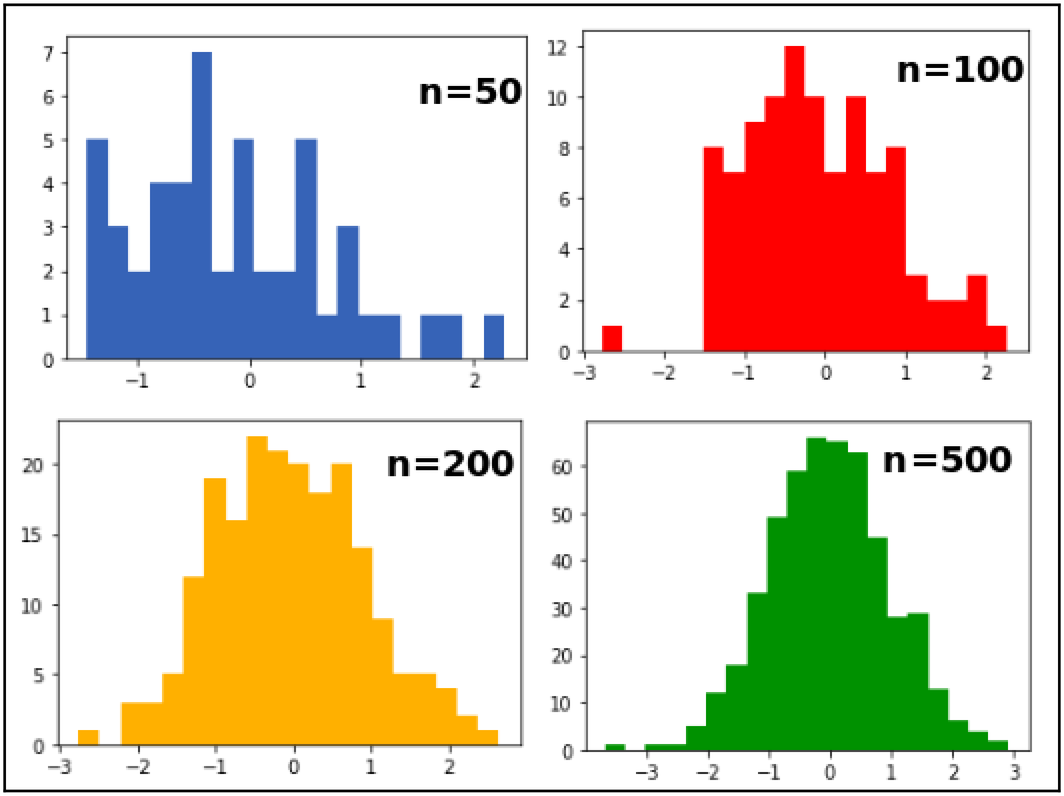
Food	chocolate	icecream	soup
Weather			
cold	NaN	16.0	16.0
hot	8.0	8.0	8.0

```
DatetimeIndex(['2000-01-01', '2000-01-02', '2000-01-03', '2000-01-04',
               '2000-01-05', '2000-01-06', '2000-01-07', '2000-01-08',
               '2000-01-09', '2000-01-10', '2000-01-11', '2000-01-12',
               '2000-01-13', '2000-01-14', '2000-01-15', '2000-01-16',
               '2000-01-17', '2000-01-18', '2000-01-19', '2000-01-20',
               '2000-01-21', '2000-01-22', '2000-01-23', '2000-01-24',
               '2000-01-25', '2000-01-26', '2000-01-27', '2000-01-28',
               '2000-01-29', '2000-01-30', '2000-01-31', '2000-02-01',
               '2000-02-02', '2000-02-03', '2000-02-04', '2000-02-05',
               '2000-02-06', '2000-02-07', '2000-02-08', '2000-02-09',
               '2000-02-10', '2000-02-11', '2000-02-12', '2000-02-13',
               '2000-02-14'],
              dtype='datetime64[ns]', freq='D')
```

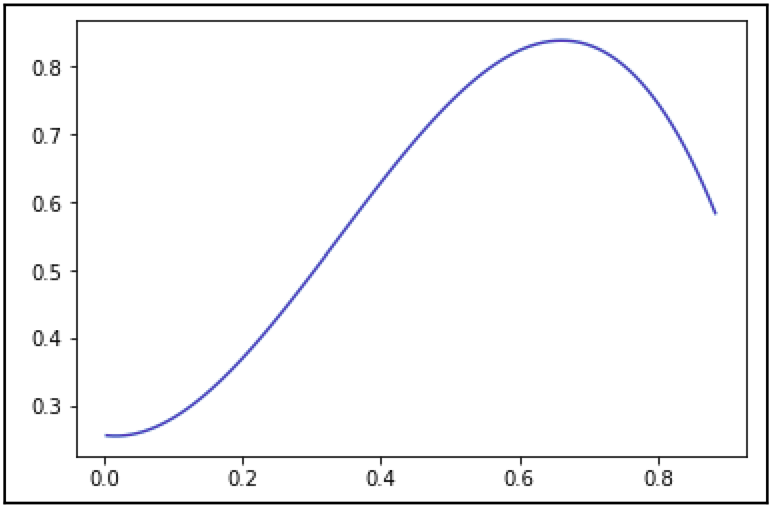


	communcation_skill_score	quantitative_skill_score
communcation_skill_score	69.20	-6.55
quantitative_skill_score	-6.55	34.20

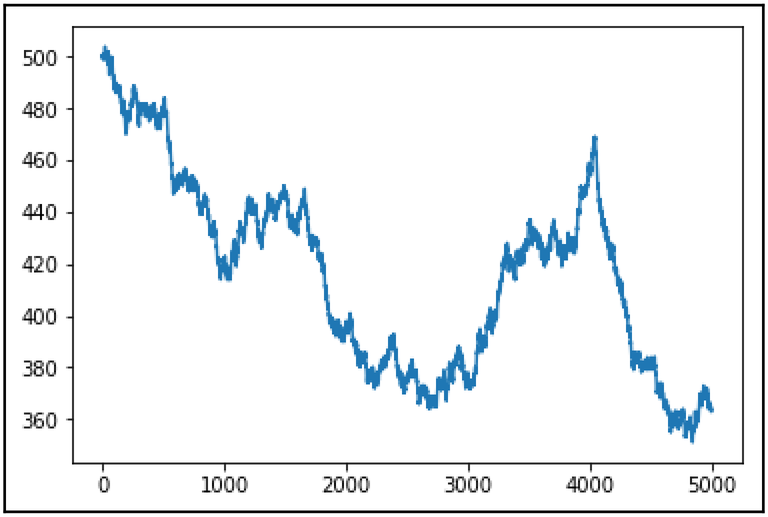
	communcation_skill_score	quantitative_skill_score
communcation_skill_score	1.00000	-0.13464
quantitative_skill_score	-0.13464	1.00000

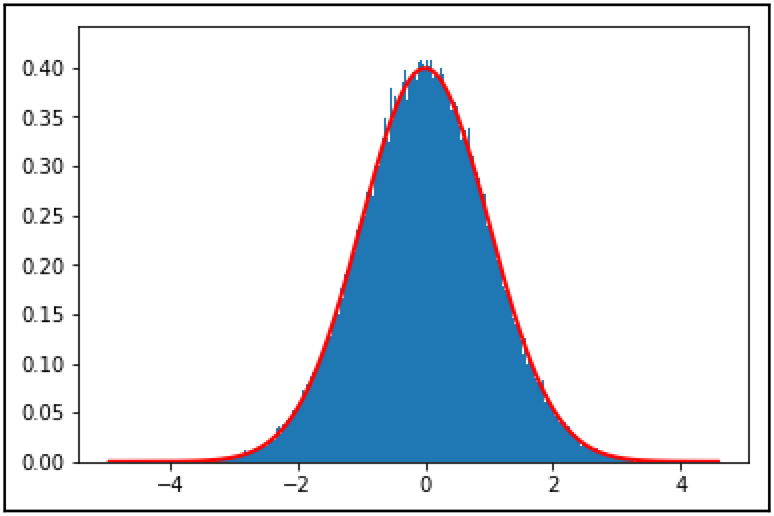

n=50
n=100
n=200
n=500

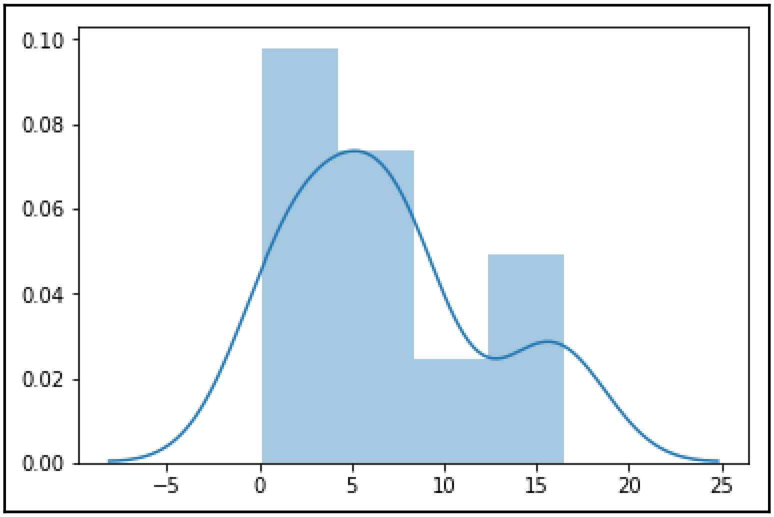
Chapter 4: Linear Algebra

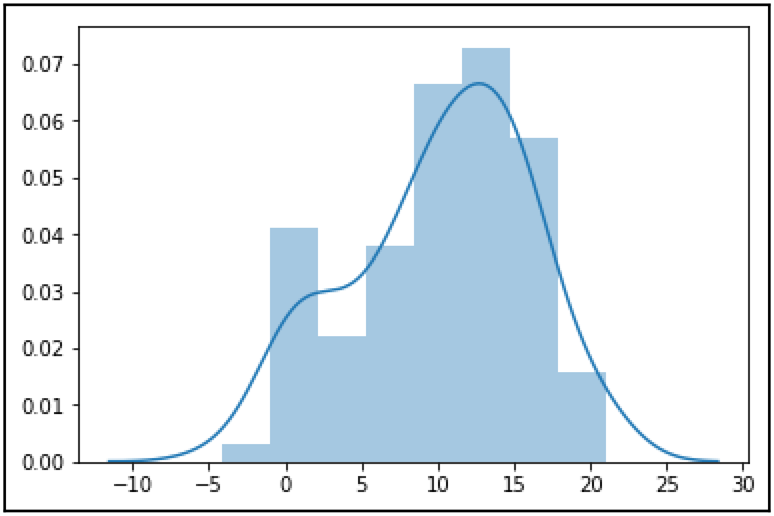


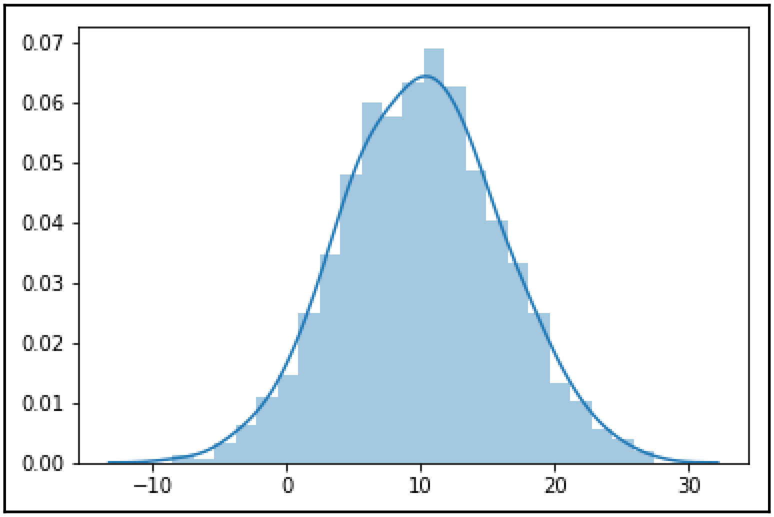
$$P(X) = \frac{n!}{(n-X)!\,X!} \cdot (p)^X \cdot (q)^{n-X}$$

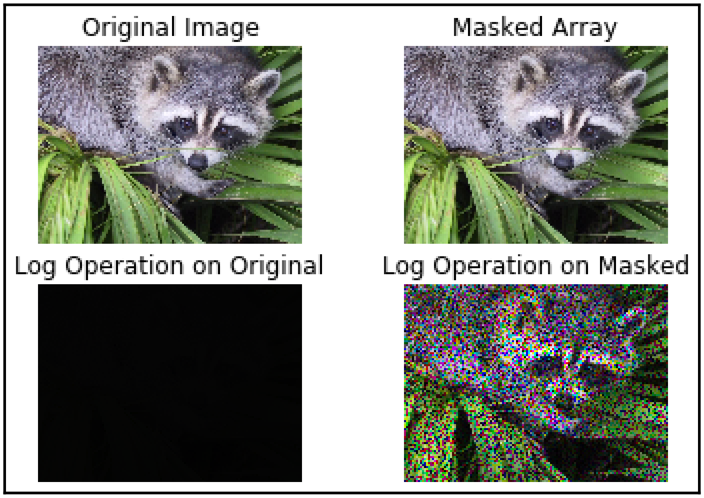


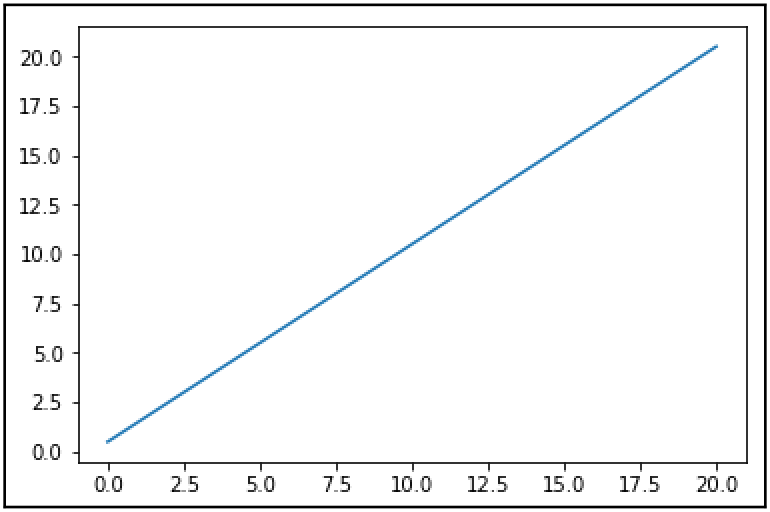

0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−4
−2
0
2
4

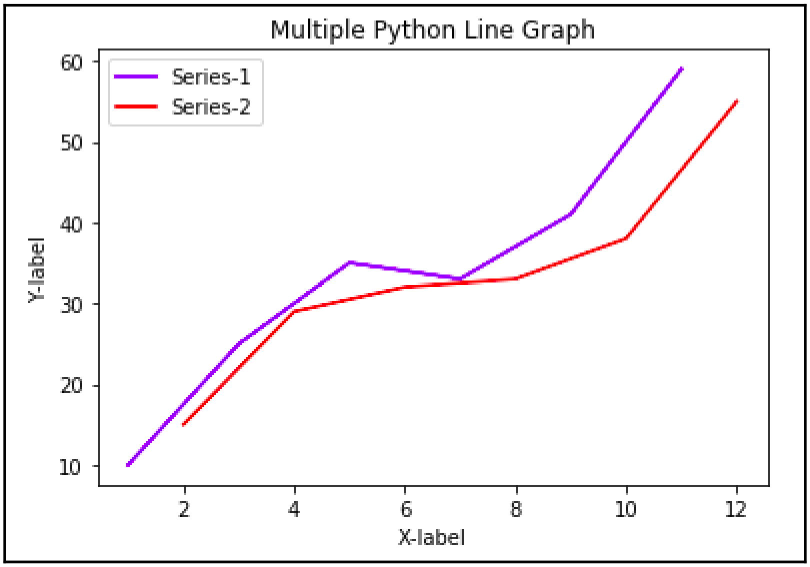

0.10
0.08
0.06
0.04
0.02
0.00
−5
0
5
10
15
20
25

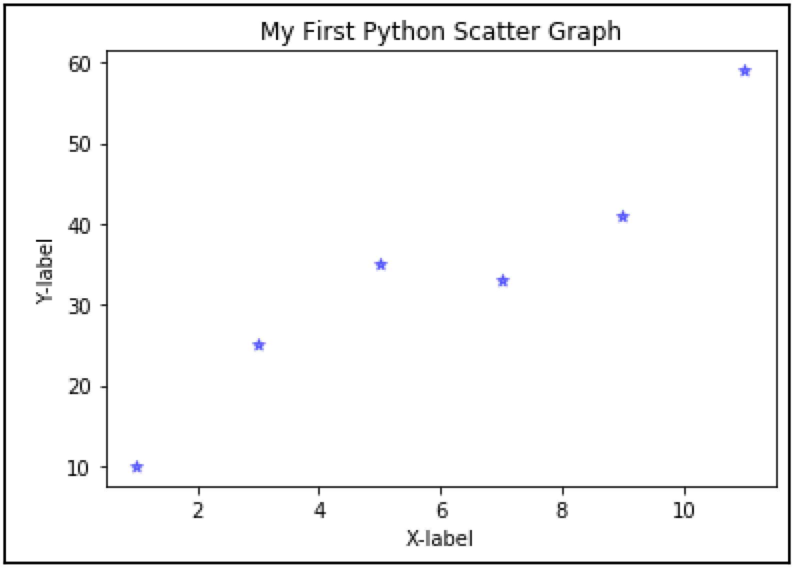

0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
−10
−5
0
5
10
15
20
25
30

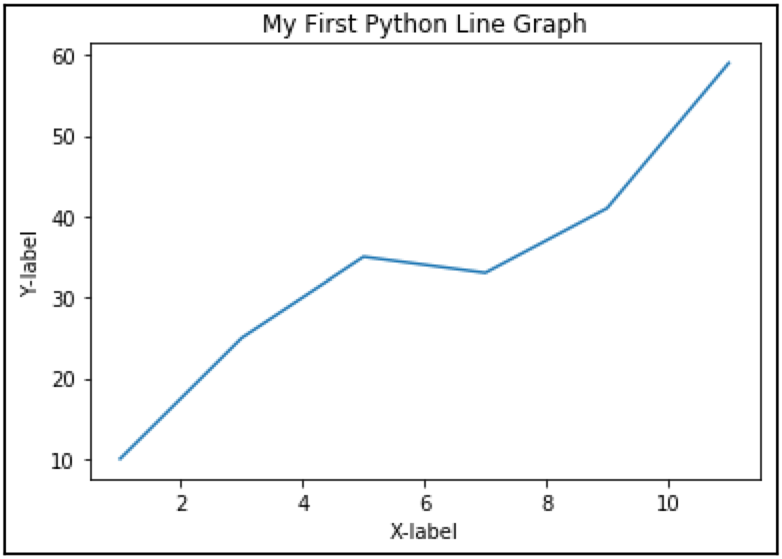

0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
−10
0
10
20
30

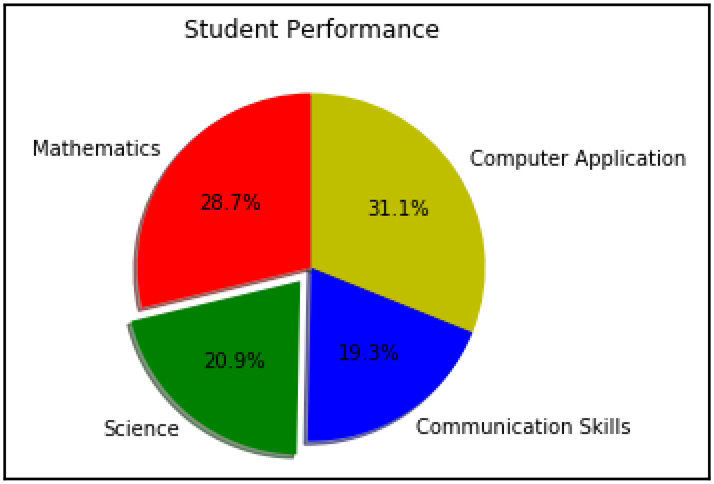

Original Image
Masked Array
Log Operation on Original
Log Operation on Masked

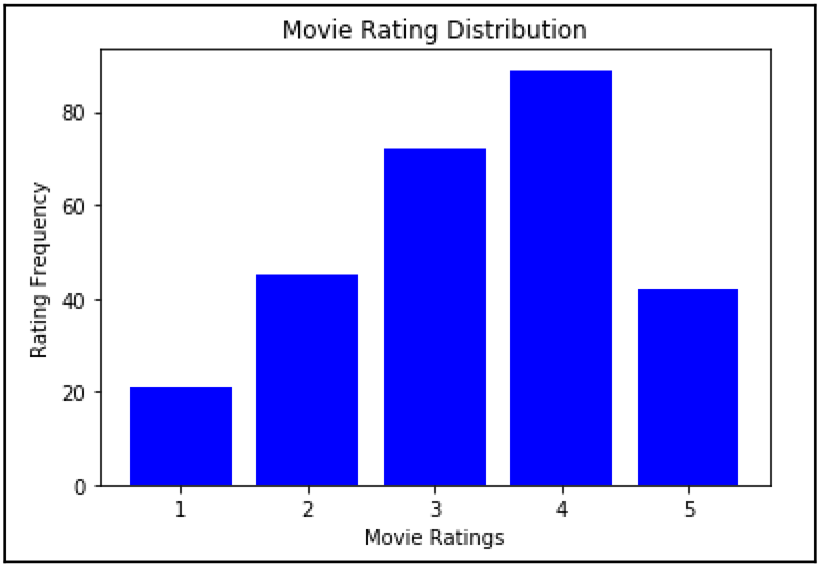

20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

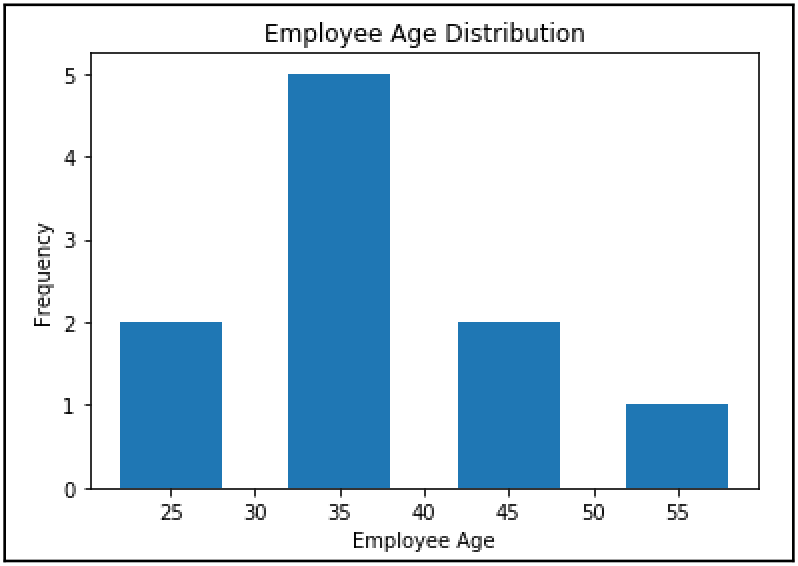

Multiple Python Line Graph
Series-1
Series-2
Y-label
X-label
60
50
40
30
20
10
2
4
6
8
10
12

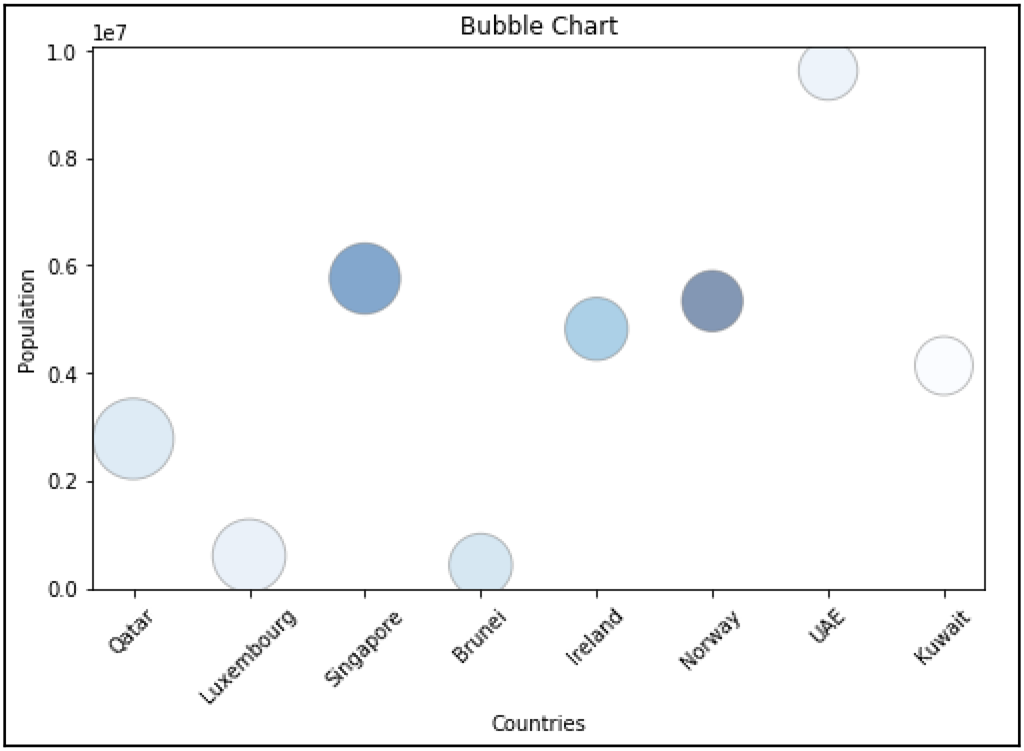

My First Python Scatter Graph
60
50
40
30
20
10
Y-label
2
4
6
8
10
X-label

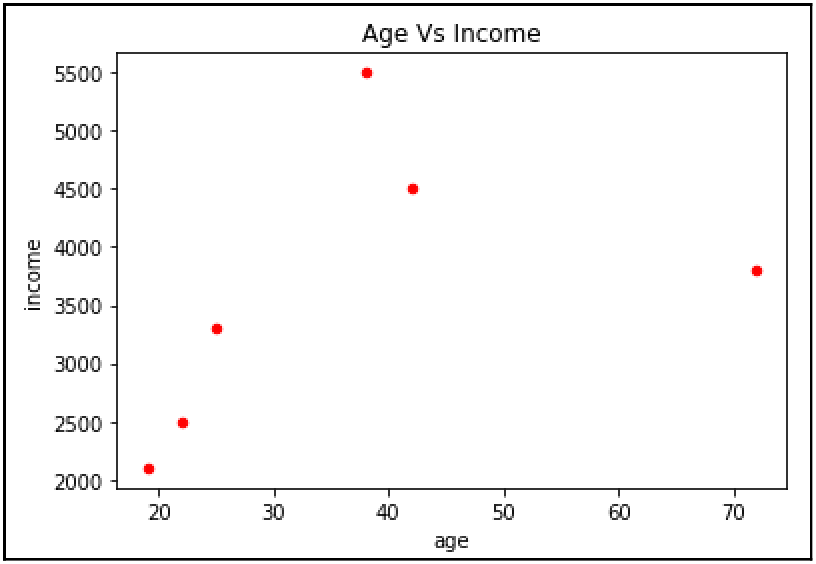

My First Python Line Graph
60
50
40
30
20
10
Y-label
2
4
6
8
10
X-label

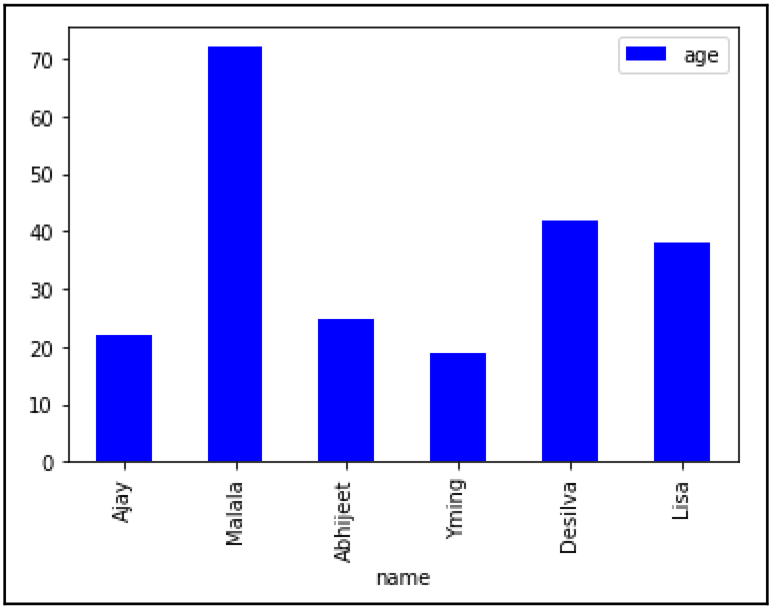

Student Performance
Mathematics
28.7%
Computer Application
31.1%
20.9%
19.3%
Science
Communication Skills

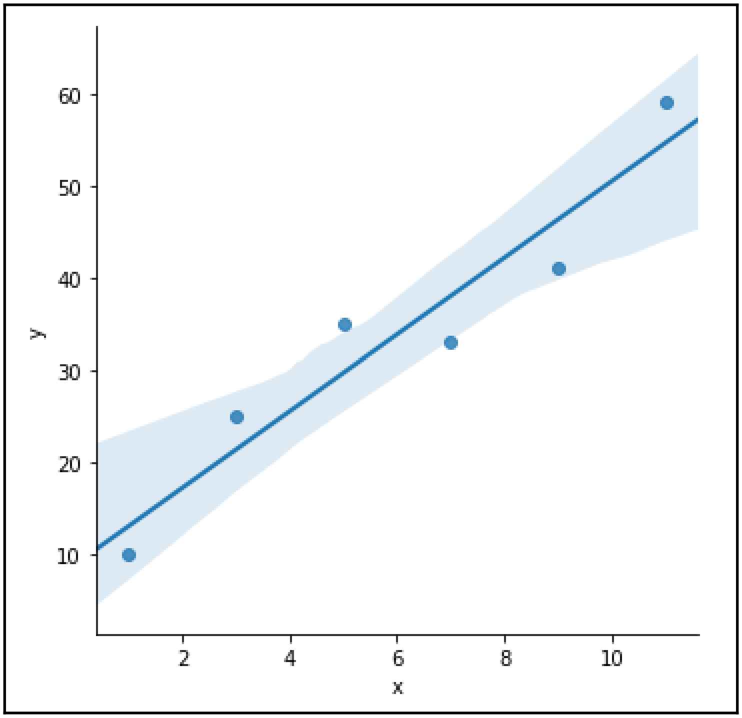

Movie Rating Distribution
Rating Frequency
0
20
40
60
80
1
2
3
4
5
Movie Ratings


Employee Age Distribution
Frequency
0
1
2
3
4
5
25
30
35
40
45
50
55
Employee Age

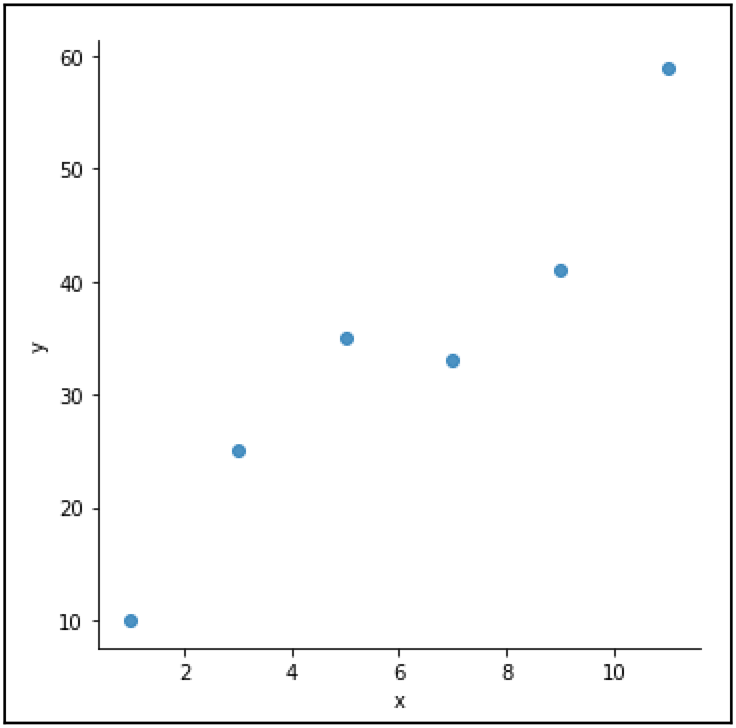

Bubble Chart
1e7
1.0
0.8
0.6
0.4
0.2
0.0
Population
Qatar
Luxembourg
Singapore
Brunei
Ireland
Norway
UAE
Kuwait
Countries

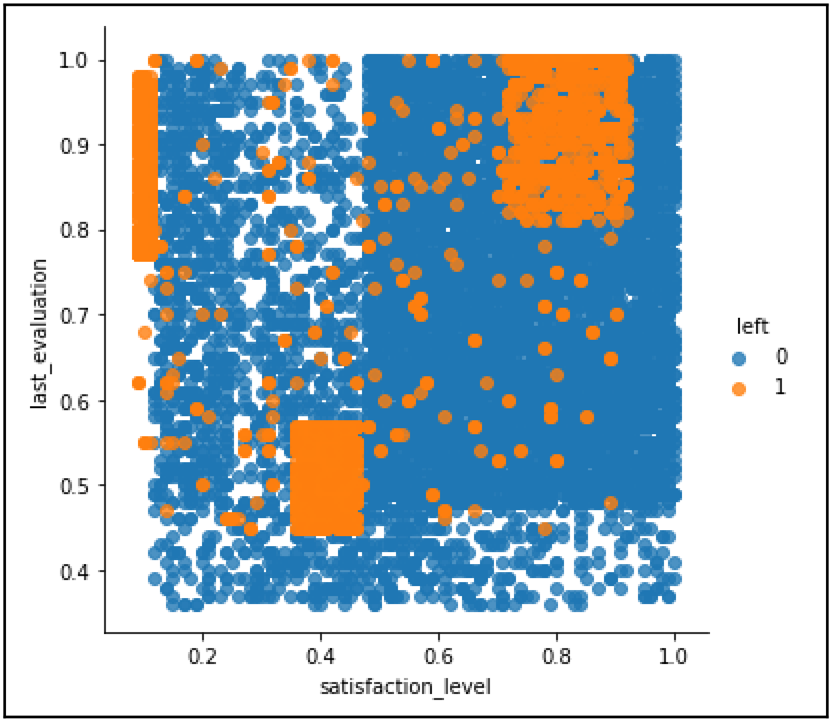

Age Vs Income
income
5500
5000
4500
4000
3500
3000
2500
2000
20
30
40
50
60
70
age

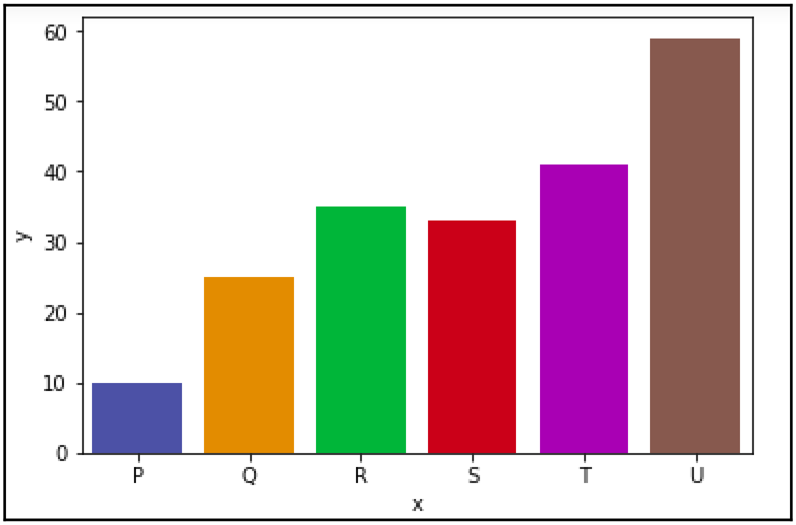

age
70
60
50
40
30
20
10
0
Ajay
Malala
Abhijeet
Yming
Desilva
Lisa
name

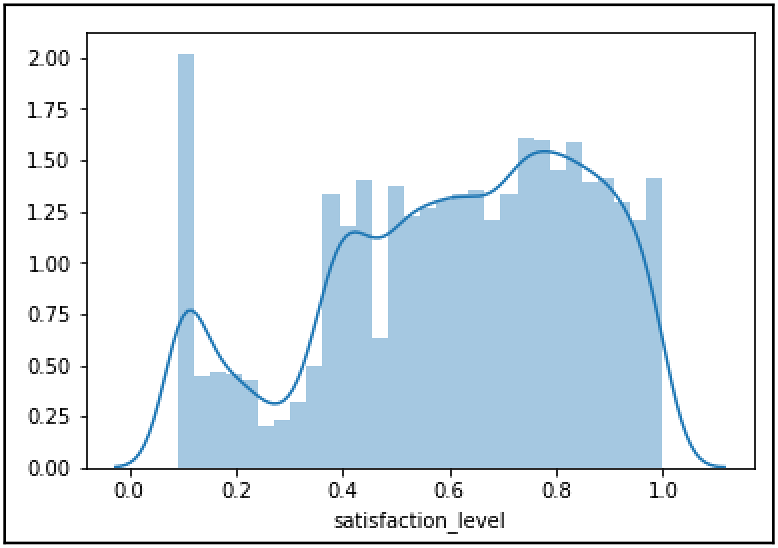


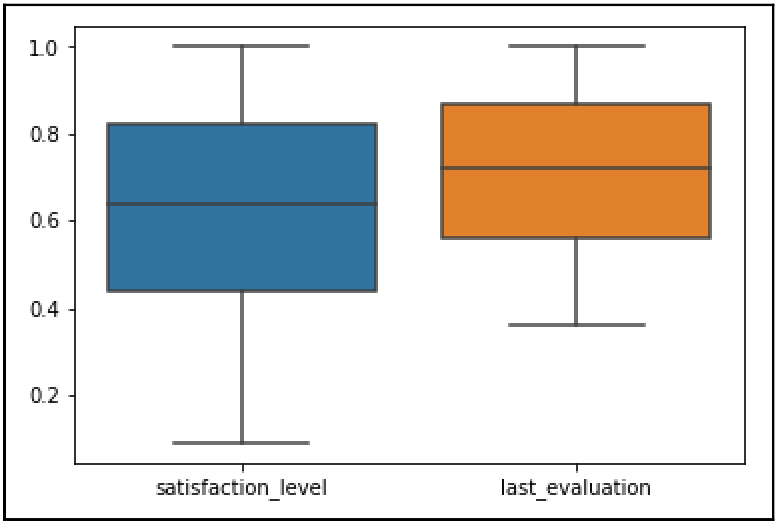
60
50
40
30
20
10
y
2
4
6
8
10
x

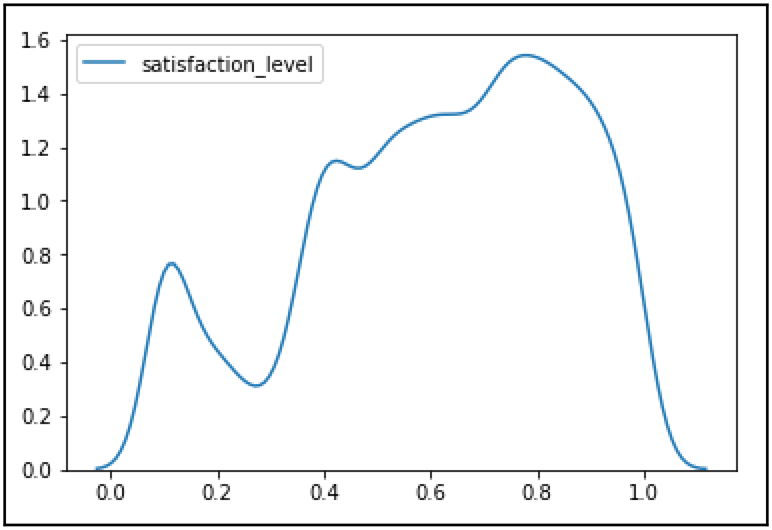

60
50
40
30
20
10
y
2
4
6
8
10
x

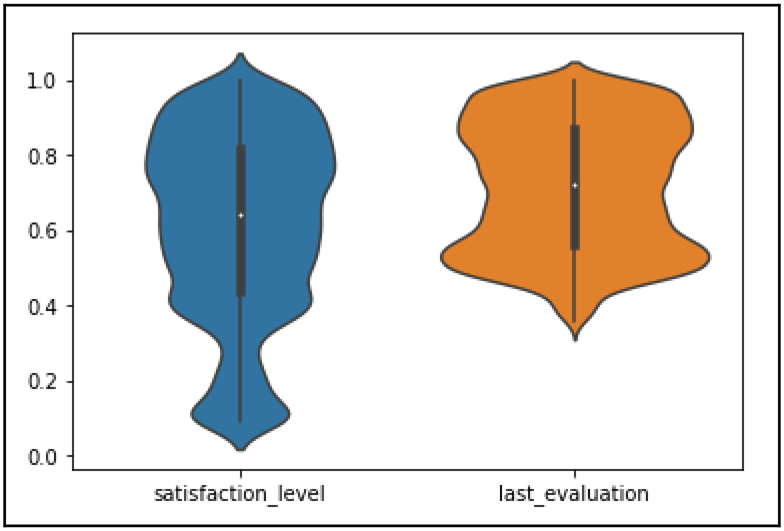

1.0
0.9
0.8
0.7
0.6
0.5
0.4
last_evaluation
0.2
0.4
0.6
0.8
1.0
satisfaction_level
left
0
1

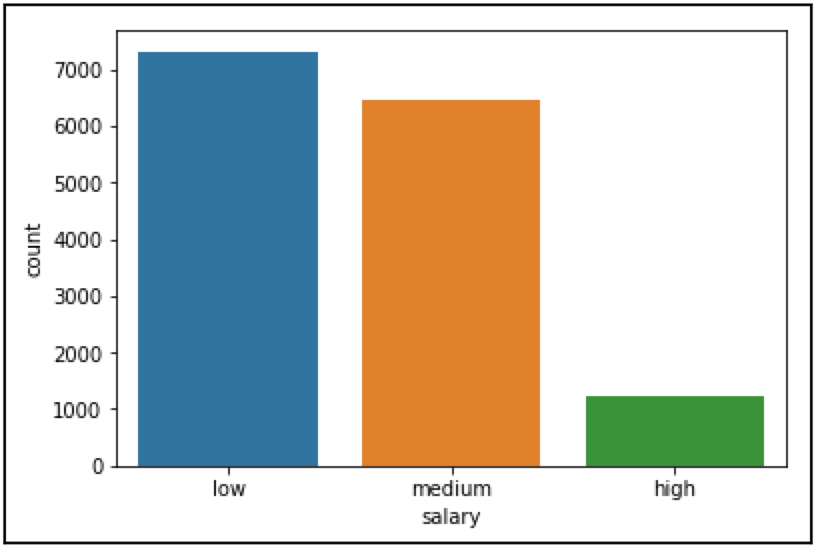

60
50
40
30
20
10
0
y
P
Q
R
S
T
U
x

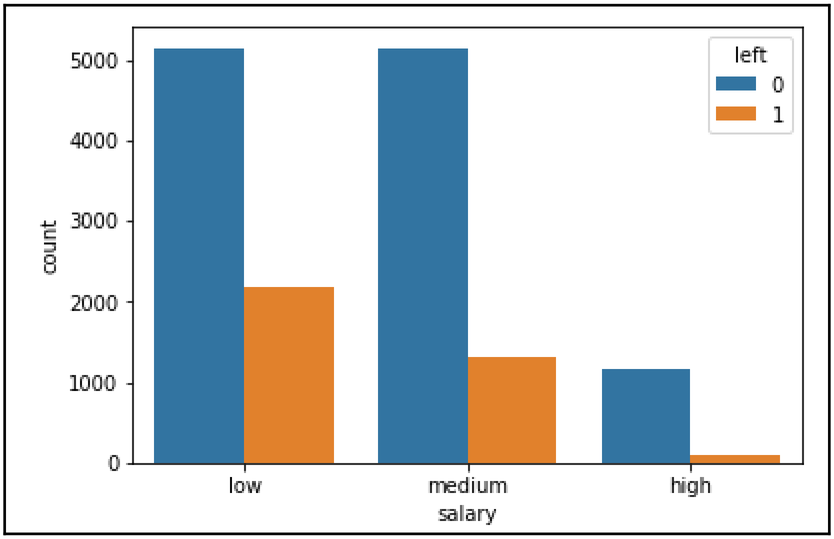

2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0
satisfaction_level

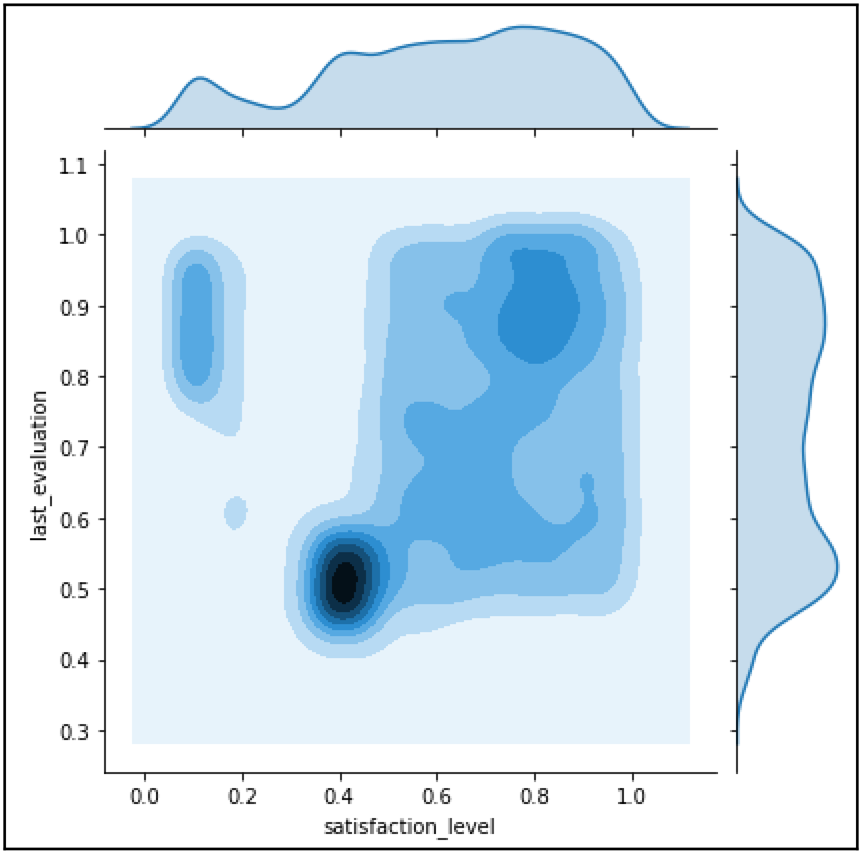

1.0
0.8
0.6
0.4
0.2
satisfaction_level
last_evaluation

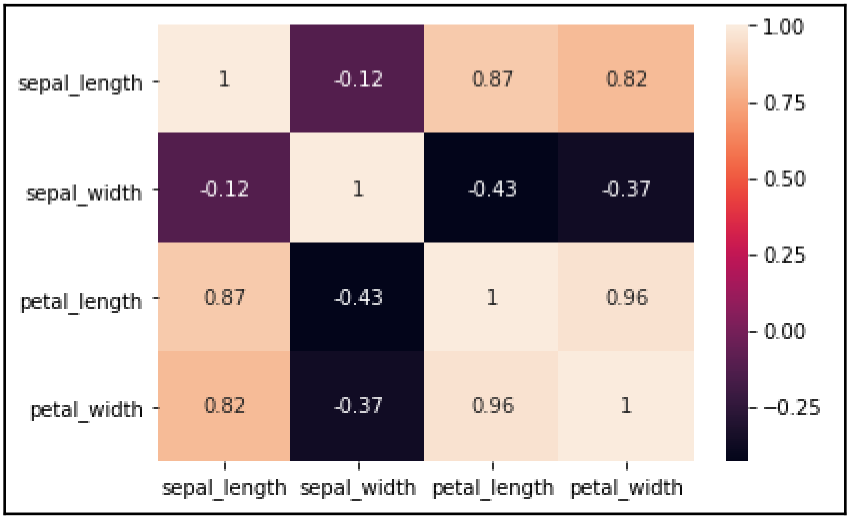

satisfaction_level
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

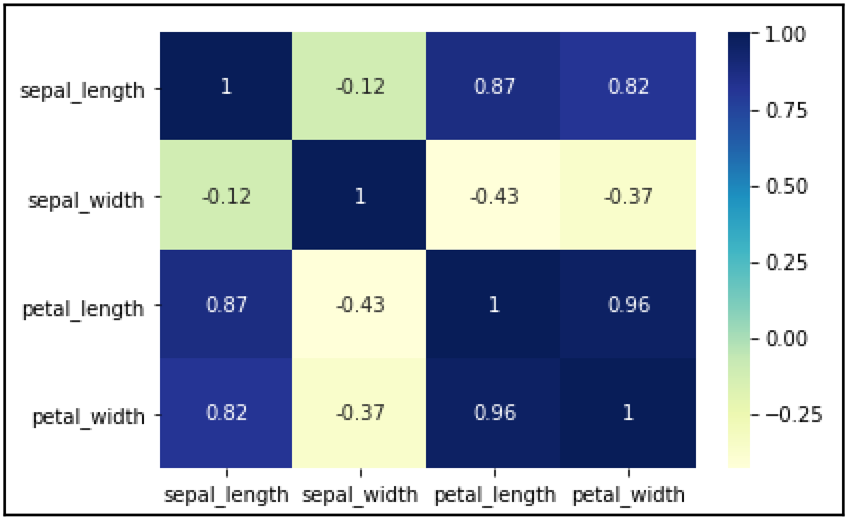

1.0
0.8
0.6
0.4
0.2
0.0
satisfaction_level
last_evaluation


7000
6000
5000
4000
3000
2000
1000
0
count
low
medium
high
salary

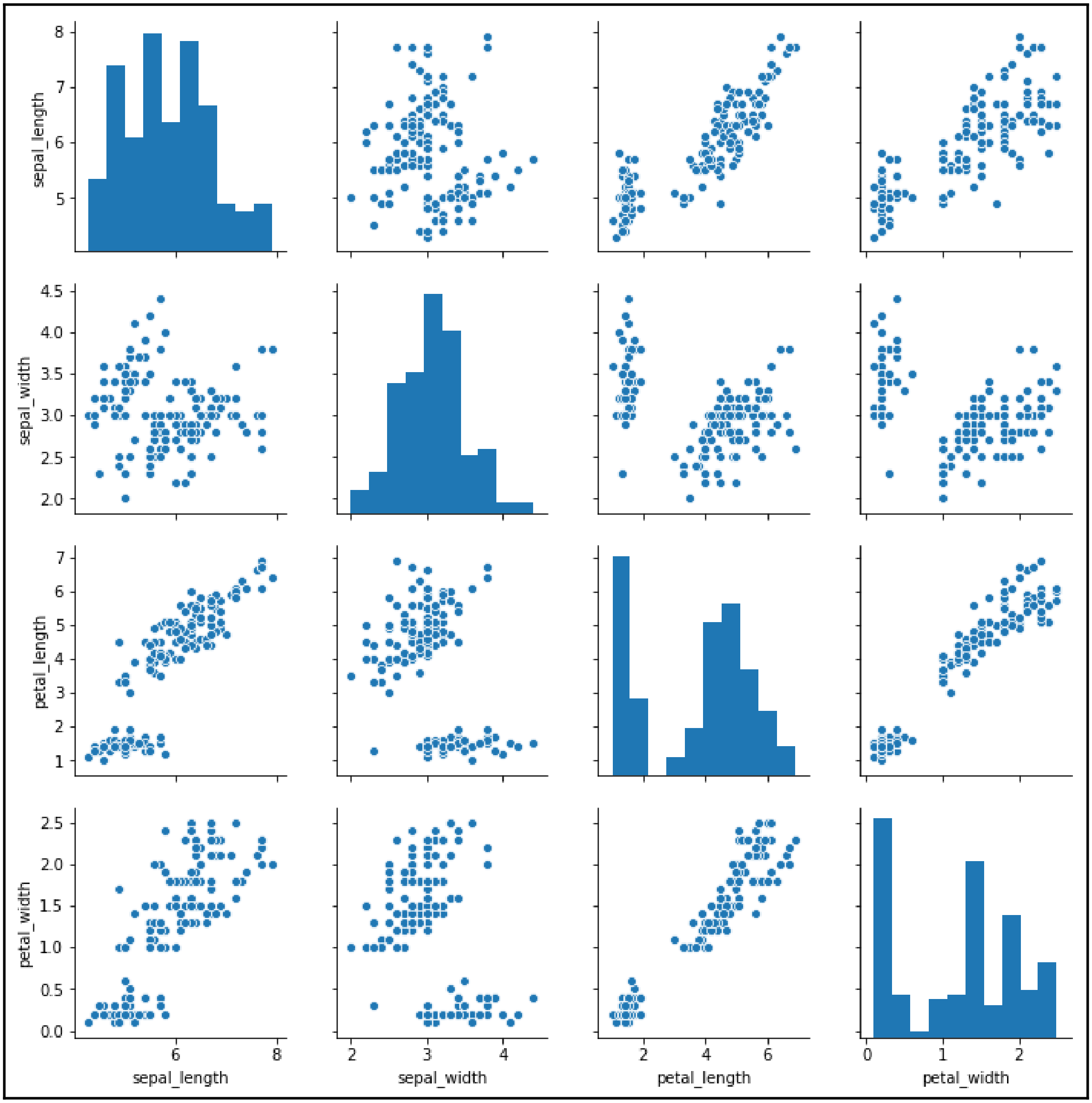

5000
4000
3000
2000
1000
0
count
left
0
1
low
medium
high
salary

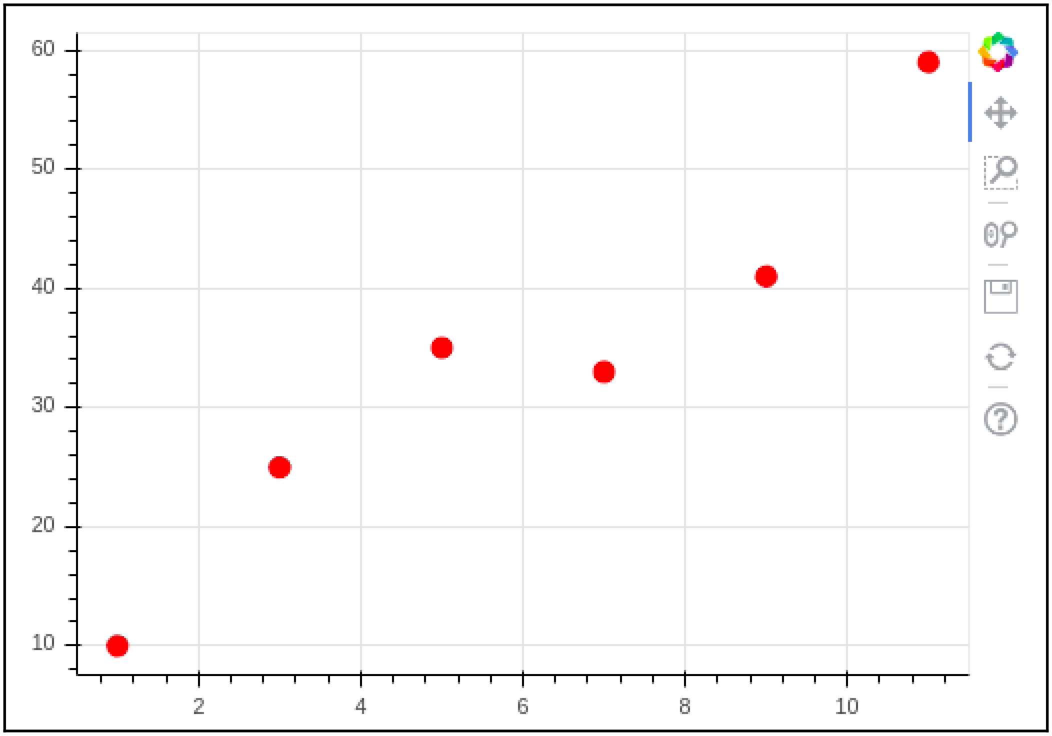

1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
last_evaluation
0.0
0.2
0.4
0.6
0.8
1.0
satisfaction_level

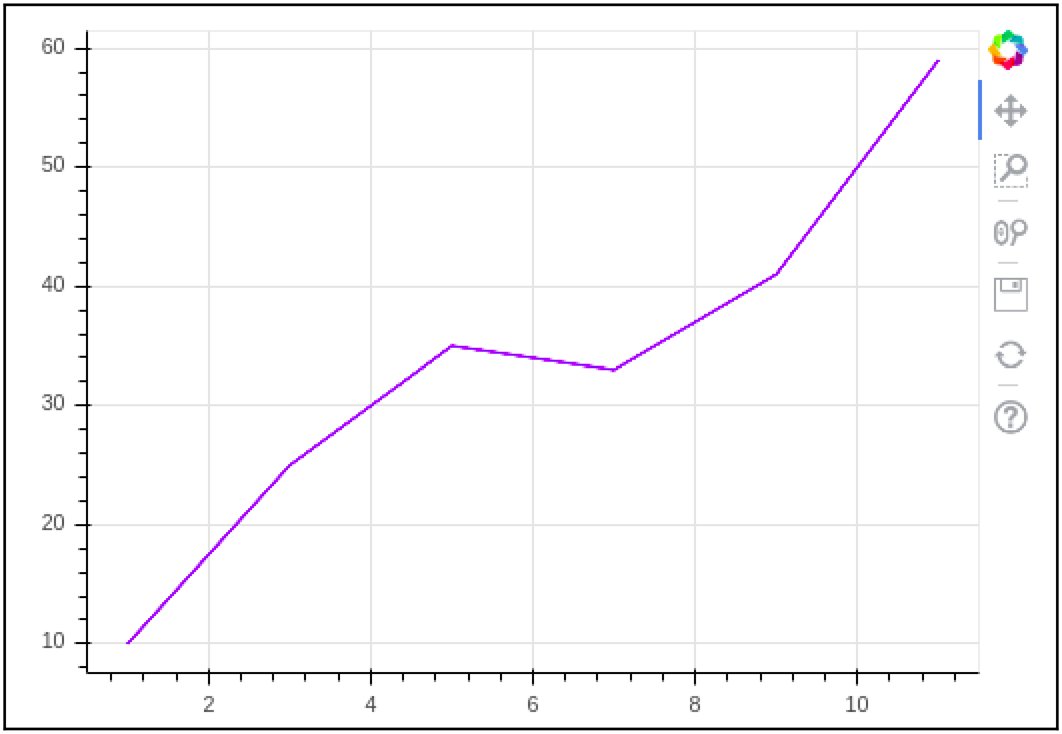

sepal_length
sepal_width
petal_length
petal_width
1
-0.12
0.87
0.82
-0.12
1
-0.43
-0.37
0.87
-0.43
1
0.96
0.82
-0.37
0.96
1
sepal_length
sepal_width
petal_length
petal_width
1.00
0.75
0.50
0.25
0.00
−0.25

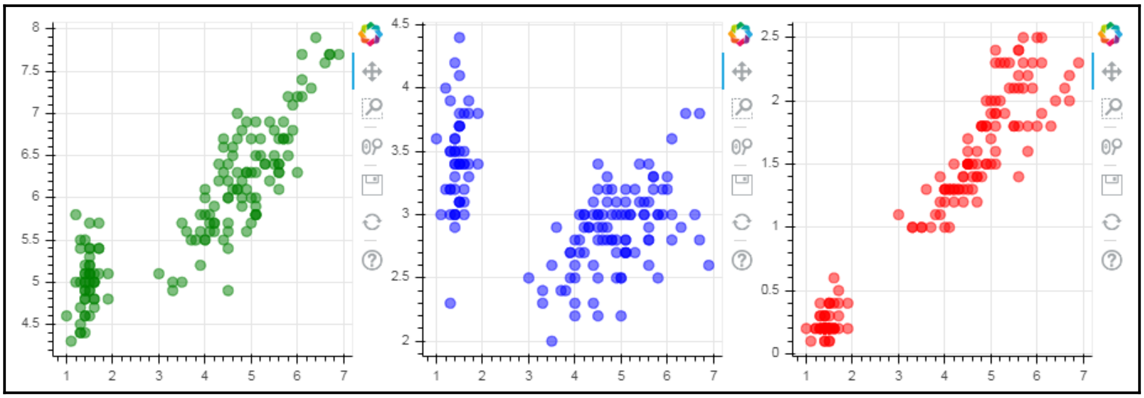


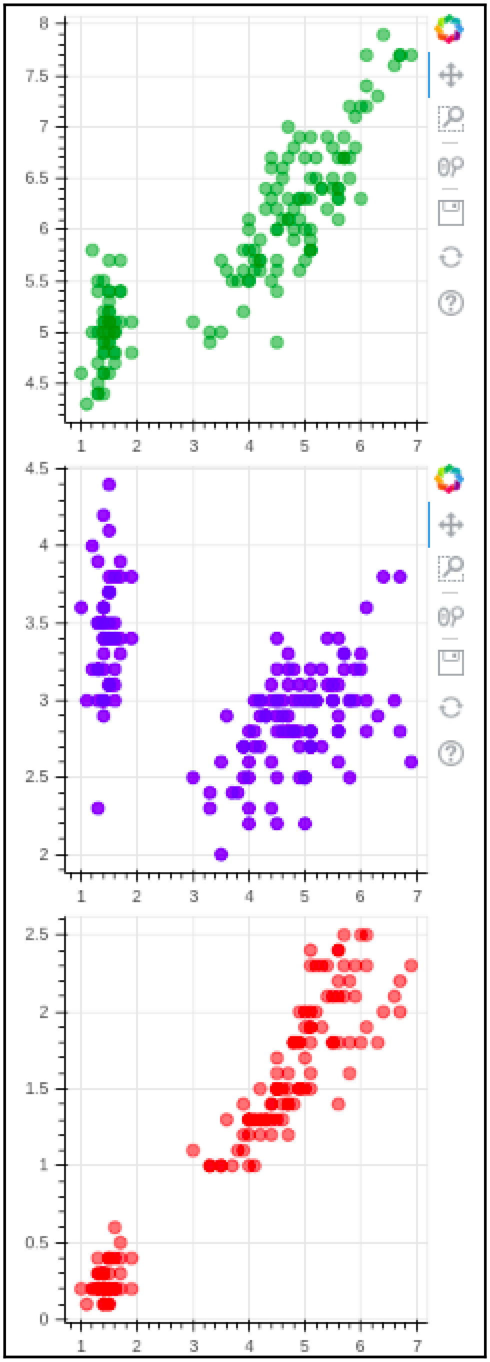
sepal_length
sepal_width
petal_length
petal_width
1
-0.12
0.87
0.82
-0.12
1
-0.43
-0.37
0.87
-0.43
1
0.96
0.82
-0.37
0.96
1
sepal_length
sepal_width
petal_length
petal_width
1.00
0.75
0.50
0.25
0.00
−0.25

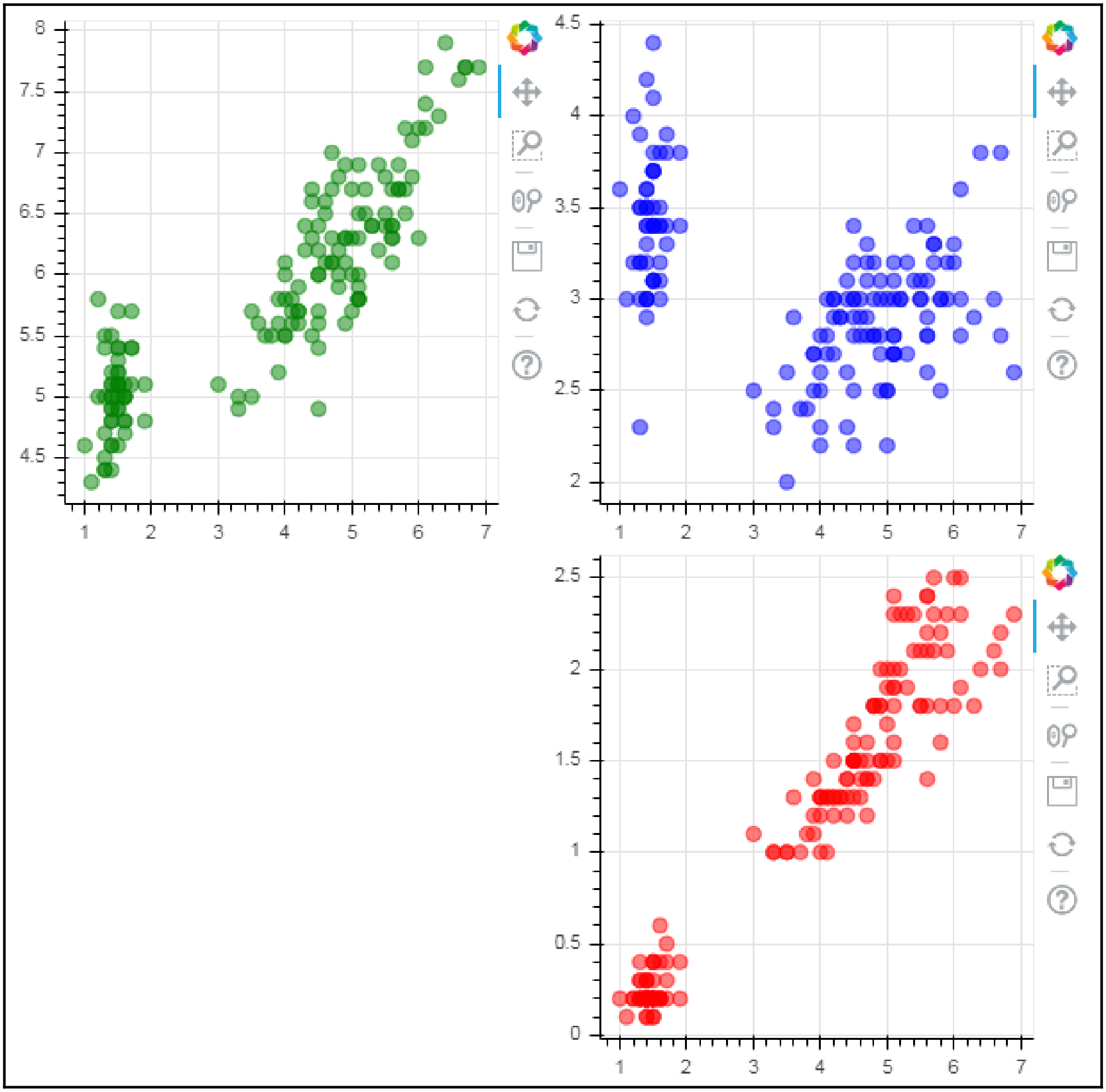

sepal_length
sepal_width
petal_length
petal_width
sepal_length
sepal_width
petal_length
petal_width

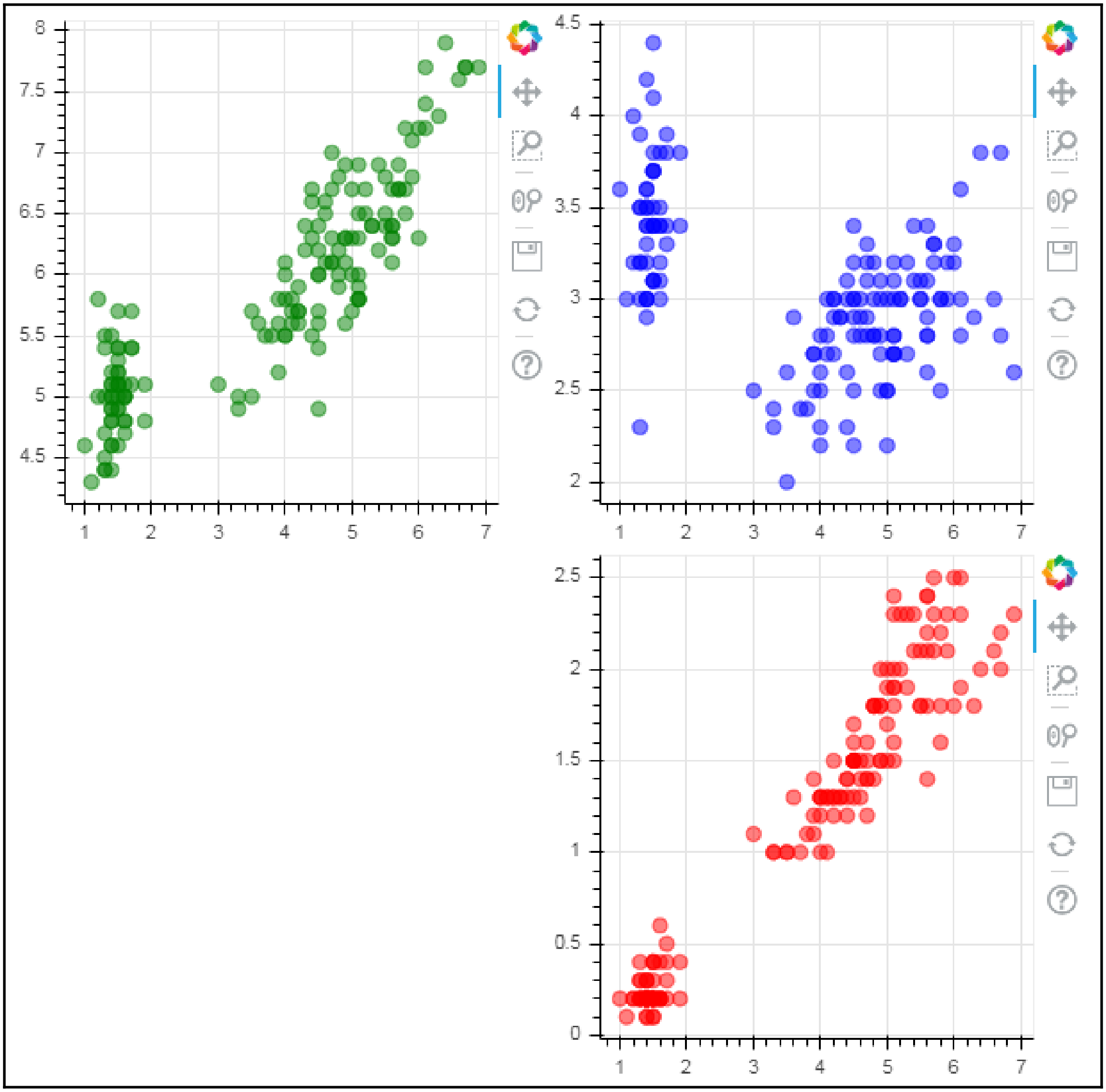

60
50
40
30
20
10
2
4
6
8
10

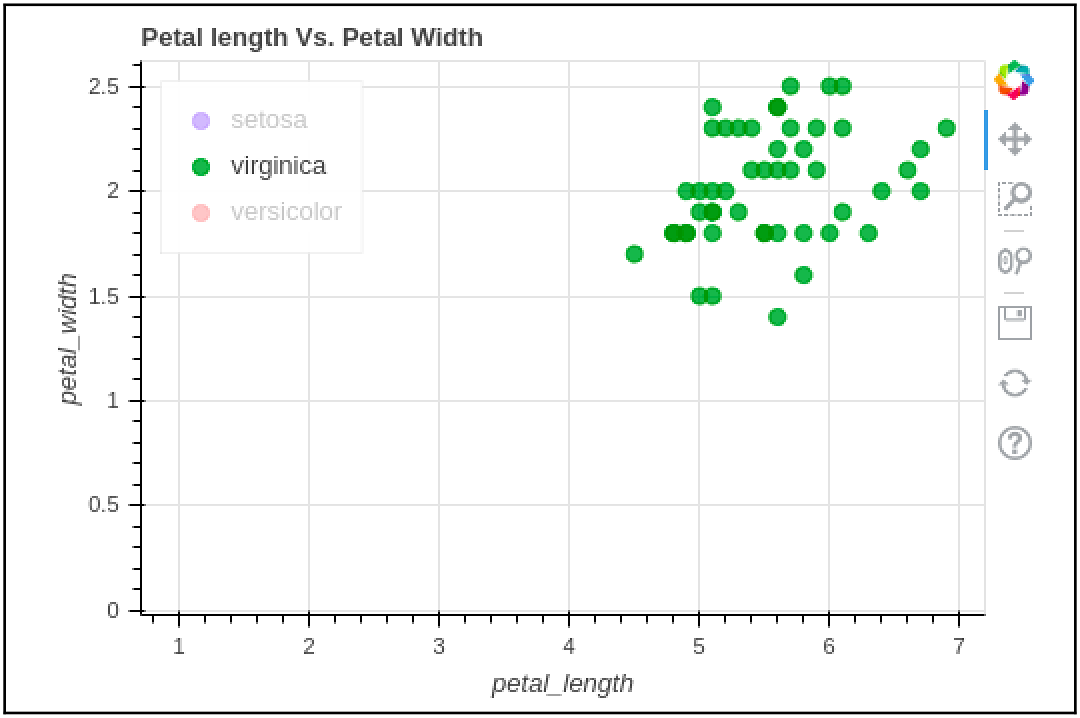

60
50
40
30
20
10
2
4
6
8
10

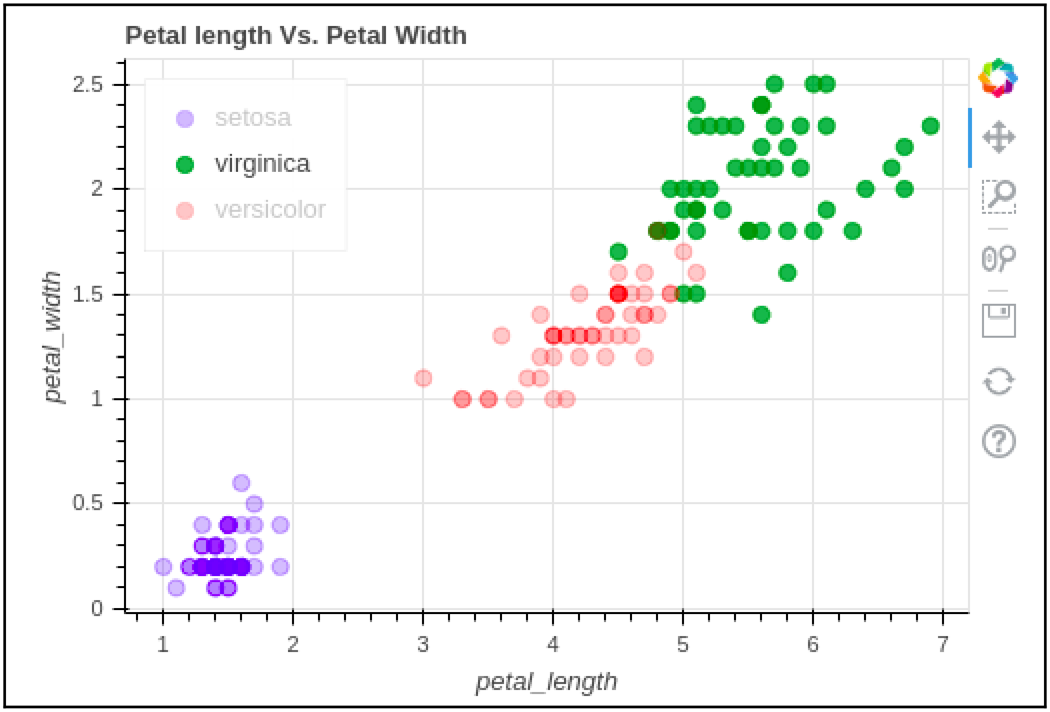


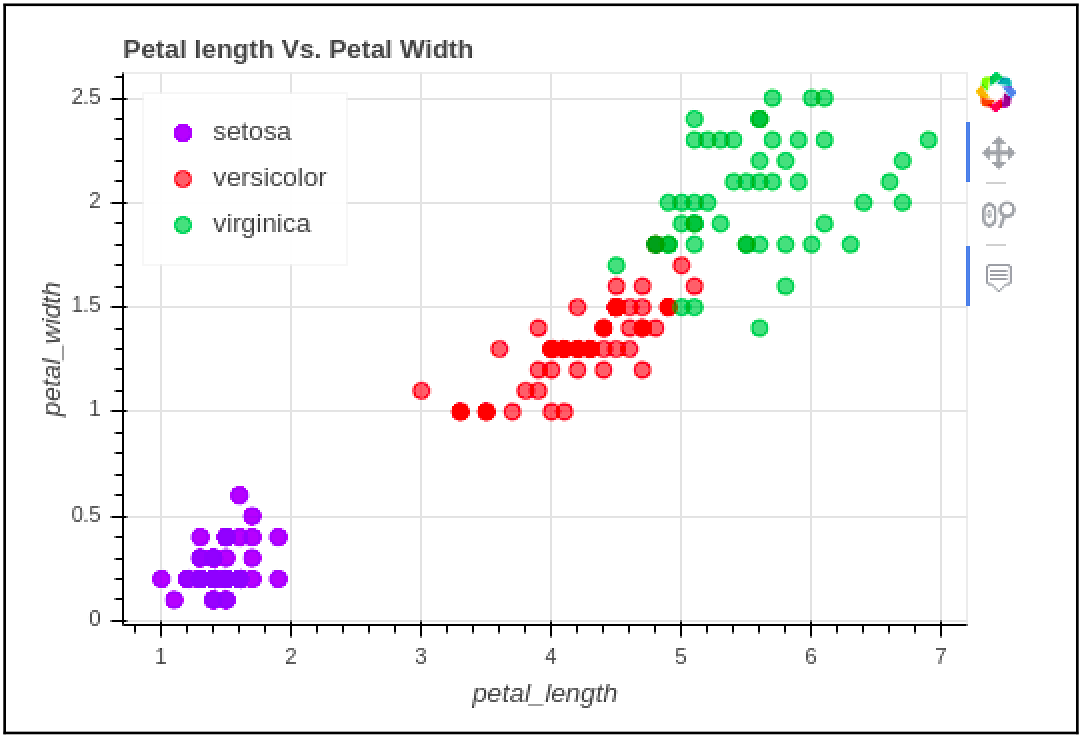

8
7.5
7
6.5
6
5.5
5
4.5
1
2
3
4
5
6
7
4.5
4
3.5
3
2.5
2
1
2
3
4
5
6
7
2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7

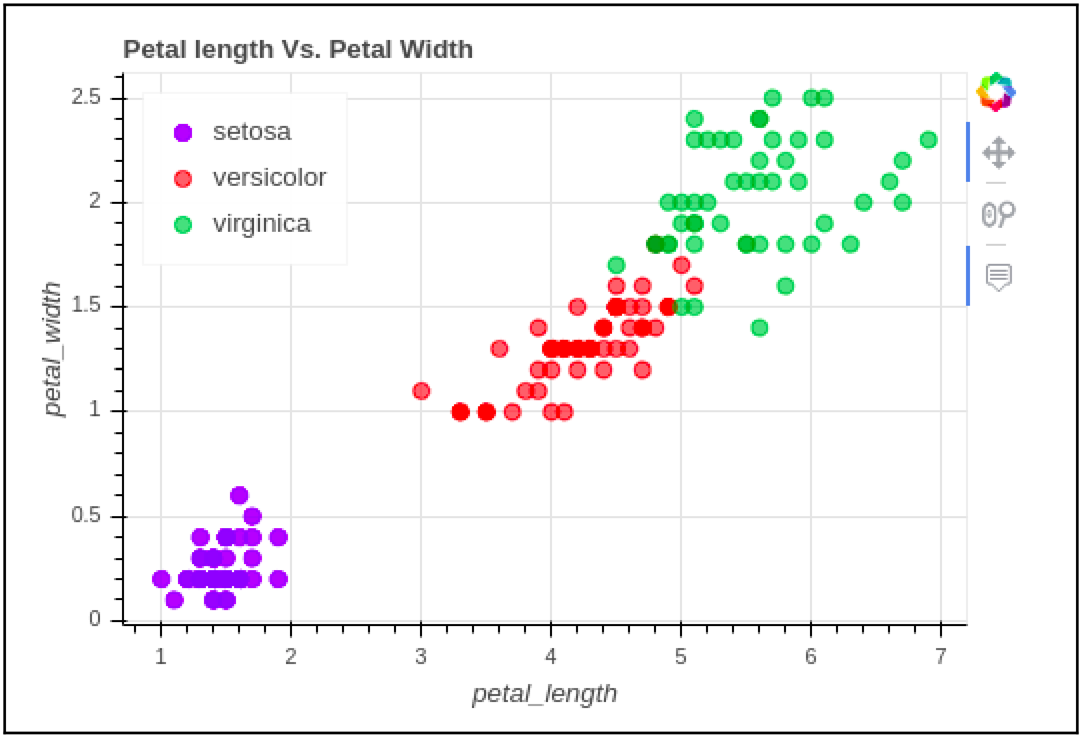

8
7.5
7
6.5
6
5.5
5
4.5
1 2 3 4 5 6 7
4.5
4
3.5
3
2.5
2
1 2 3 4 5 6 7
2.5
2
1.5
1
0.5
0
1 2 3 4 5 6 7

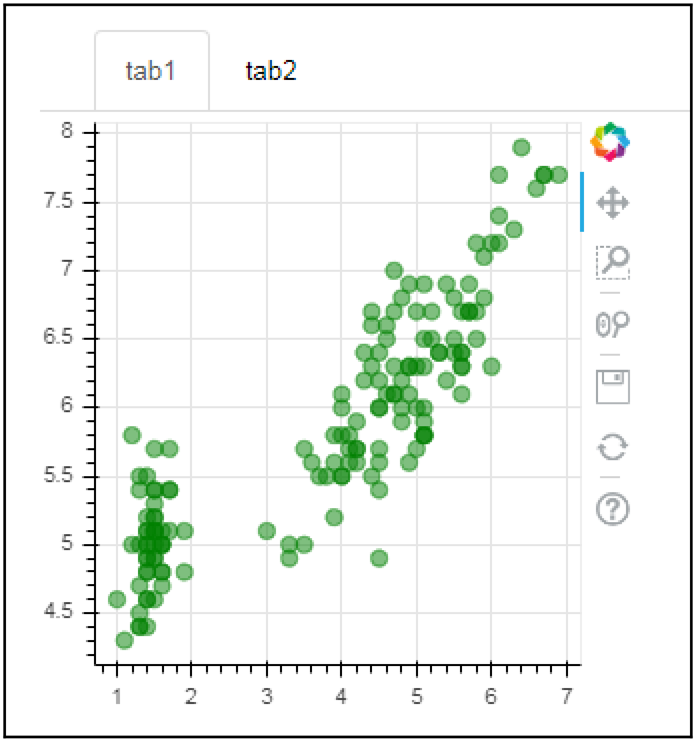

8
7.5
7
6.5
6
5.5
5
4.5
1 2 3 4 5 6 7
4.5
4
3.5
3
2.5
2
1 2 3 4 5 6 7
2.5
2
1.5
1
0.5
0
1 2 3 4 5 6 7

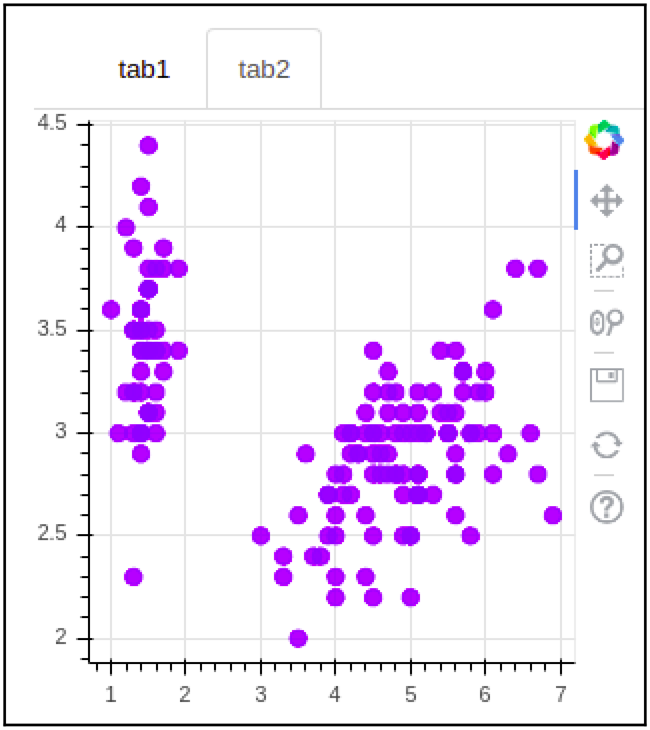

Petal length Vs. Petal Width
setosa
virginica
versicolor
petal_width
petal_length
2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7

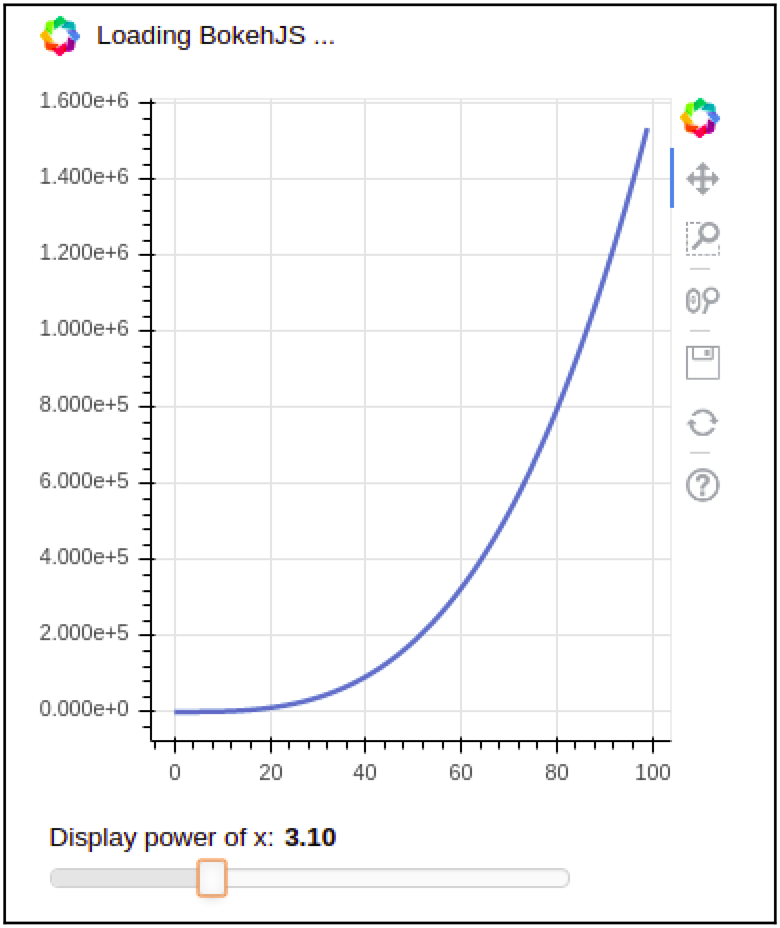

Petal length Vs. Petal Width
setosa
virginica
versicolor
petal_width
petal_length
2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7


Petal length Vs. Petal Width
setosa
versicolor
virginica
petal_width
petal_length
2.5
2
1.5
1
0.5
0
1
2
3
4
5
6
7


Petal length Vs. Petal Width
setosa
versicolor
virginica
petal_width
petal_length
0
0.5
1
1.5
2
2.5
1
2
3
4
5
6
7


tab1
tab2
8
7.5
7
6.5
6
5.5
5
4.5
1
2
3
4
5
6
7


tab1
tab2
4.5
4
3.5
3
2.5
2
1
2
3
4
5
6
7


Loading BokehJS ...
1.600e+6
1.400e+6
1.200e+6
1.000e+6
8.000e+5
6.000e+5
4.000e+5
2.000e+5
0.000e+0
0
20
40
60
80
100
Display power of x: 3.10

Chapter 6: Retrieving, Processing, and Storing Data

	0	1	2
0	14	32	33
1	24	45	26
2	27	38	39

	name	performance_score
0	Allen Smith	723
1	S Kumar	520
2	Jack Morgan	674
3	Ying Chin	556
4	Dheeraj Patel	711

	name	age	income	gender	department	grade
0	Allen Smith	45.0	NaN	None	Operations	G3
1	S Kumar	NaN	16000.0	F	Finance	G0
2	Jack Morgan	32.0	35000.0	M	Finance	G2
3	Ying Chin	45.0	65000.0	F	Sales	G3
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2

	name	age	income	gender	department	grade
0	Allen Smith	45.0	NaN	None	Operations	G3
1	S Kumar	NaN	16000.0	F	Finance	G0
2	Jack Morgan	32.0	35000.0	M	Finance	G2
3	Ying Chin	45.0	65000.0	F	Sales	G3
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2

	Flag	English short name	English long name	Domestic short name(s)	Capital	Currency	Location
0	NaN	Antigua and Barbuda[n 1]	Antigua and Barbuda	English: Antigua and Barbuda	St. John's	East Caribbean dollar	Caribbean
1	NaN	Bahamas, The[n 1]	Commonwealth of The Bahamas	English: Bahamas	Nassau	Bahamian dollar	Lucayan Archipelago
2	NaN	Barbados[n 1]	Barbados	English: Barbados	Bridgetown	Barbadian dollar	Caribbean
3	NaN	Belize[n 1][n 2]	Belize	English: Belize	Belmopan	Belize dollar	Central America
4	NaN	Canada[n 3]	Canada	English: CanadaFrench: Canada	Ottawa	Canadian dollar	Northern America

	name	age	income	gender	department	grade
0	Allen Smith	45.0	NaN	None	Operations	G3
1	S Kumar	NaN	16000.0	F	Finance	G0
2	Jack Morgan	32.0	35000.0	M	Finance	G2
3	Ying Chin	45.0	65000.0	F	Sales	G3
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2

	0	1	2
0	14	32	33
1	24	45	26
2	27	38	39

	name	age	income	gender	department	grade	performance_score
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723
1	S Kumar	NaN	16000.0	F	Finance	G0	520
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674
3	Ying Chin	45.0	65000.0	F	Sales	G3	556
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711

	name	age	income	gender	department	grade	performance_score
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711
5	Satyam Sharma	NaN	62000.0	NaN	Sales	G3	649
6	James Authur	54.0	NaN	F	Operations	G3	53
7	Josh Wills	54.0	52000.0	F	Finance	G3	901
8	Leo Duck	23.0	98000.0	M	Sales	G4	709

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9 entries, 0 to 8
Data columns (total 7 columns):
name                 9 non-null object
age                  7 non-null float64
income               7 non-null float64
gender               7 non-null object
department           9 non-null object
grade                9 non-null object
performance_score    9 non-null int64
dtypes: float64(2), int64(1), object(4)
memory usage: 584.0+ bytes
```

	age	income	performance_score
count	7.000000	7.000000	9.000000
mean	40.428571	52857.142857	610.666667
std	12.204605	26028.372797	235.671912
min	23.000000	16000.000000	53.000000
25%	31.000000	38500.000000	556.000000
50%	45.000000	52000.000000	674.000000
75%	49.500000	63500.000000	711.000000
max	54.000000	98000.000000	901.000000

	name	department
0	Allen Smith	Operations
1	S Kumar	Finance
2	Jack Morgan	Finance
3	Ying Chin	Sales
4	Dheeraj Patel	Operations
5	Satyam Sharma	Sales
6	James Authur	Operations
7	Josh Wills	Finance
8	Leo Duck	Sales

	name
0	Allen Smith
1	S Kumar
2	Jack Morgan
3	Ying Chin
4	Dheeraj Patel
5	Satyam Sharma
6	James Authur
7	Josh Wills
8	Leo Duck

	name	department
0	Allen Smith	Operations
1	S Kumar	Finance
2	Jack Morgan	Finance
3	Ying Chin	Sales
4	Dheeraj Patel	Operations
5	Satyam Sharma	Sales
6	James Authur	Operations
7	Josh Wills	Finance
8	Leo Duck	Sales

	name	age	income	gender	department	grade	performance_score
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723
1	S Kumar	NaN	16000.0	F	Finance	G0	520
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674

	name	age	income	gender	department	grade	performance_score
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674
3	Ying Chin	45.0	65000.0	F	Sales	G3	556
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711

	name	age	income	gender	department	grade	performance_score
3	Ying Chin	45.0	65000.0	F	Sales	G3	556
5	Satyam Sharma	NaN	62000.0	NaN	Sales	G3	649
8	Leo Duck	23.0	98000.0	M	Sales	G4	709

	name	age	income	gender	department	grade	performance_score
1	S Kumar	NaN	16000.0	F	Finance	G0	520
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674
3	Ying Chin	45.0	65000.0	F	Sales	G3	556
5	Satyam Sharma	NaN	62000.0	NaN	Sales	G3	649
7	Josh Wills	54.0	52000.0	F	Finance	G3	901
8	Leo Duck	23.0	98000.0	M	Sales	G4	709

	name	age	income	gender	department	grade	performance_score
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711
7	Josh Wills	54.0	52000.0	F	Finance	G3	901
8	Leo Duck	23.0	98000.0	M	Sales	G4	709

	name	age	income	gender	department	grade	performance_score
1	S Kumar	NaN	16000.0	F	Finance	G0	520
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674
3	Ying Chin	45.0	65000.0	F	Sales	G3	556
5	Satyam Sharma	NaN	62000.0	NaN	Sales	G3	649

	name	age	income	gender	department	grade	performance_score
6	James Authur	54.0	NaN	F	Operations	G3	53

	name	age	income	gender	department	grade	performance_score
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674
3	Ying Chin	45.0	65000.0	F	Sales	G3	556
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711
7	Josh Wills	54.0	52000.0	F	Finance	G3	901
8	Leo Duck	23.0	98000.0	M	Sales	G4	709

	name	age	income	gender	department	grade	performance_score
0	Allen Smith	45.000000	NaN	NaN	Operations	G3	723
1	S Kumar	40.428571	16000.0	F	Finance	G0	520
2	Jack Morgan	32.000000	35000.0	M	Finance	G2	674
3	Ying Chin	45.000000	65000.0	F	Sales	G3	556
4	Dheeraj Patel	30.000000	42000.0	F	Operations	G2	711
5	Satyam Sharma	40.428571	62000.0	NaN	Sales	G3	649
6	James Authur	54.000000	NaN	F	Operations	G3	53
7	Josh Wills	54.000000	52000.0	F	Finance	G3	901
8	Leo Duck	23.000000	98000.0	M	Sales	G4	709

	name	age	income	gender	department	grade	performance_score
0	Allen Smith	45.000000	52000.0	NaN	Operations	G3	723
1	S Kumar	40.428571	16000.0	F	Finance	G0	520
2	Jack Morgan	32.000000	35000.0	M	Finance	G2	674
3	Ying Chin	45.000000	65000.0	F	Sales	G3	556
4	Dheeraj Patel	30.000000	42000.0	F	Operations	G2	711
5	Satyam Sharma	40.428571	62000.0	NaN	Sales	G3	649
6	James Authur	54.000000	52000.0	F	Operations	G3	53
7	Josh Wills	54.000000	52000.0	F	Finance	G3	901
8	Leo Duck	23.000000	98000.0	M	Sales	G4	709

	name	age	income	gender	department	grade	performance_score
0	Allen Smith	45.000000	52000.0	F	Operations	G3	723
1	S Kumar	40.428571	16000.0	F	Finance	G0	520
2	Jack Morgan	32.000000	35000.0	M	Finance	G2	674
3	Ying Chin	45.000000	65000.0	F	Sales	G3	556
4	Dheeraj Patel	30.000000	42000.0	F	Operations	G2	711
5	Satyam Sharma	40.428571	62000.0	F	Sales	G3	649
6	James Authur	54.000000	52000.0	F	Operations	G3	53
7	Josh Wills	54.000000	52000.0	F	Finance	G3	901
8	Leo Duck	23.000000	98000.0	M	Sales	G4	709

	name	age	income	gender	department	grade	performance_score
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723
1	S Kumar	NaN	16000.0	F	Finance	G0	520
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674
3	Ying Chin	45.0	65000.0	F	Sales	G3	556
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711
5	Satyam Sharma	NaN	62000.0	NaN	Sales	G3	649
6	James Authur	54.0	NaN	F	Operations	G3	53
7	Josh Wills	54.0	52000.0	F	Finance	G3	901
8	Leo Duck	23.0	98000.0	M	Sales	G4	709

	name	age	income	gender	department	grade	performance_score
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723
1	S Kumar	NaN	16000.0	F	Finance	G0	520
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674
3	Ying Chin	45.0	65000.0	F	Sales	G3	556
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711
5	Satyam Sharma	NaN	62000.0	NaN	Sales	G3	649
8	Leo Duck	23.0	98000.0	M	Sales	G4	709

color
green
blue
red
blue

green	blue	red
1	0	0
0	1	0
0	0	1
0	1	0

	name	age	income	gender	department	grade	performance_score	F	M
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723	0	0
1	S Kumar	NaN	16000.0	F	Finance	G0	520	1	0
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674	0	1
3	Ying Chin	45.0	65000.0	F	Sales	G3	556	1	0
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711	1	0

	name	age	income	gender	department	grade	performance_score	grade_encoded
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723	2
1	S Kumar	NaN	16000.0	F	Finance	G0	520	0
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674	1
3	Ying Chin	45.0	65000.0	F	Sales	G3	556	2
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711	1

	name	age	income	gender	department	grade	performance_score	grade_encoded	performance_std_scaler
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723	2	0.505565
1	S Kumar	NaN	16000.0	F	Finance	G0	520	0	-0.408053
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674	1	0.285037
3	Ying Chin	45.0	65000.0	F	Sales	G3	556	2	-0.246032
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711	1	0.451558

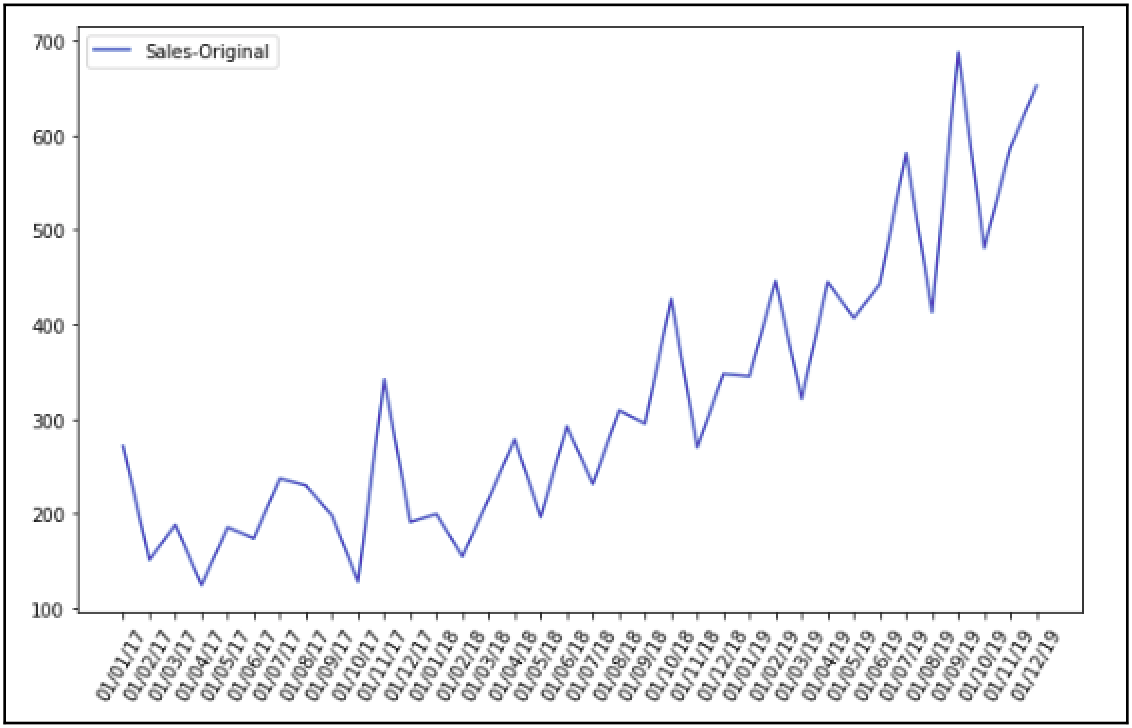
	name	age	income	gender	department	grade	performance_score	grade_encoded	performance_std_scaler	performance_minmax_scaler
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723	2	0.505565	0.790094
1	S Kumar	NaN	16000.0	F	Finance	G0	520	0	-0.408053	0.550708
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674	1	0.285037	0.732311
3	Ying Chin	45.0	65000.0	F	Sales	G3	556	2	-0.246032	0.593160
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711	1	0.451558	0.775943

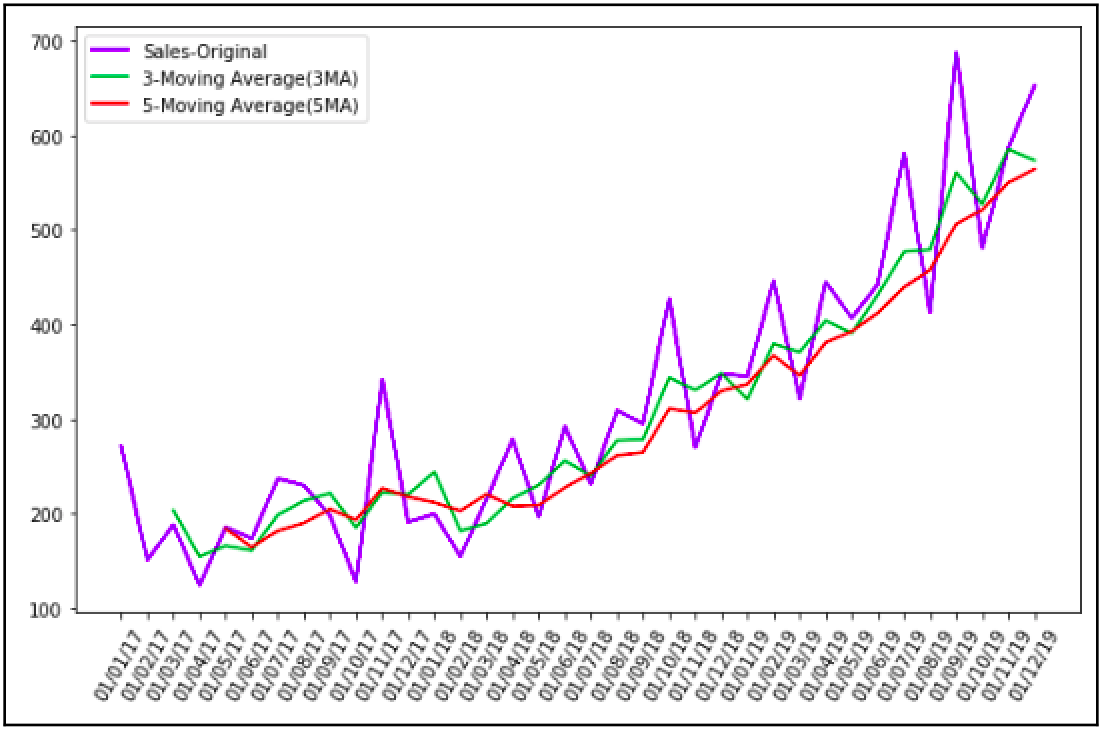
	name	age	income	gender	department	grade	performance_score	performance_std_scaler	performance_minmax_scaler	performance_robust_scaler
0	John Nash	23.0	25000.0	M	Sales	G1	619	0.035578	0.667453	-0.306306
1	Allen Smith	45.0	NaN	NaN	Operations	G3	723	0.528922	0.790094	0.443243
2	S Kumar	NaN	16000.0	F	Finance	G0	520	-0.434048	0.550708	-1.019820
3	Jack Morgan	32.0	35000.0	M	Finance	G2	674	0.296481	0.732311	0.090090
4	Ying Chin	45.0	65000.0	F	Sales	G3	556	-0.263275	0.593160	-0.760360

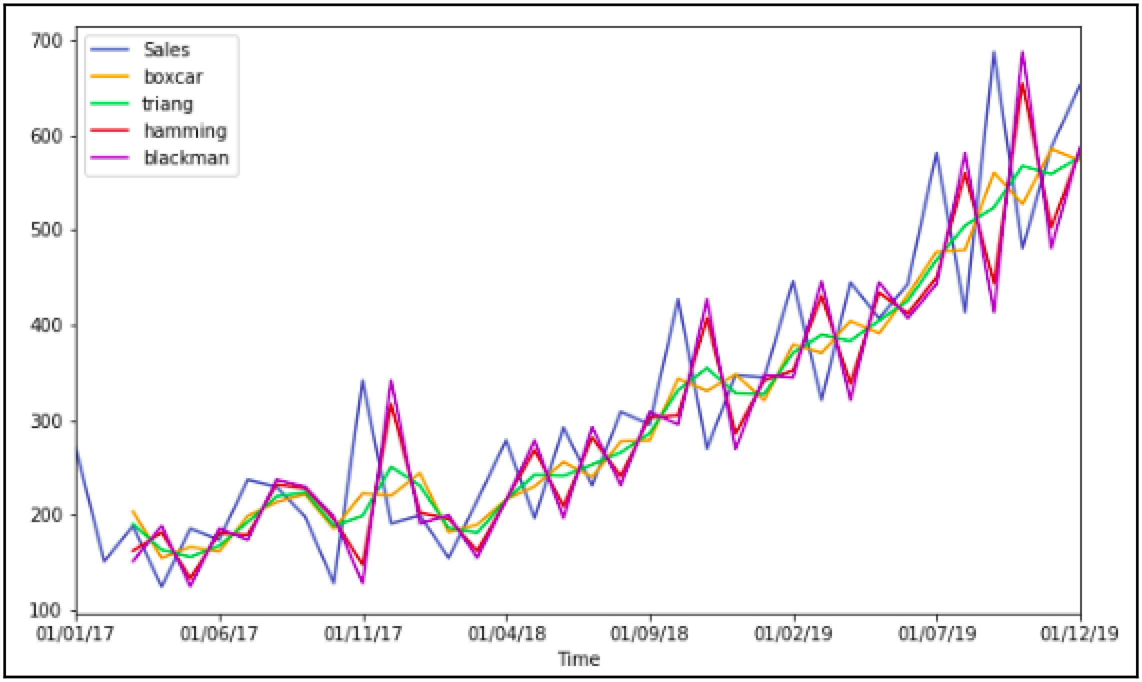
	name	age	income	gender	department	grade	performance_score	performance_grade
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723	A
1	S Kumar	NaN	16000.0	F	Finance	G0	520	B
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674	B
3	Ying Chin	45.0	65000.0	F	Sales	G3	556	B
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711	A

	name	age	income	gender	department	grade	performance_score	performance_grade	first_name	last_name
0	Allen Smith	45.0	NaN	NaN	Operations	G3	723	A	Allen	Smith
1	S Kumar	NaN	16000.0	F	Finance	G0	520	B	S	Kumar
2	Jack Morgan	32.0	35000.0	M	Finance	G2	674	B	Jack	Morgan
3	Ying Chin	45.0	65000.0	F	Sales	G3	556	B	Ying	Chin
4	Dheeraj Patel	30.0	42000.0	F	Operations	G2	711	A	Dheeraj	Patel

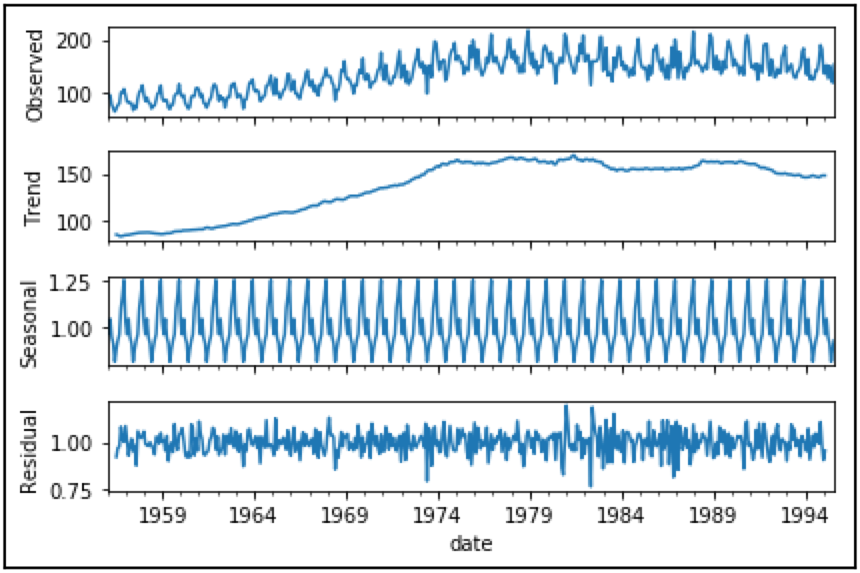
Chapter 8: Signal Processing and Time Series



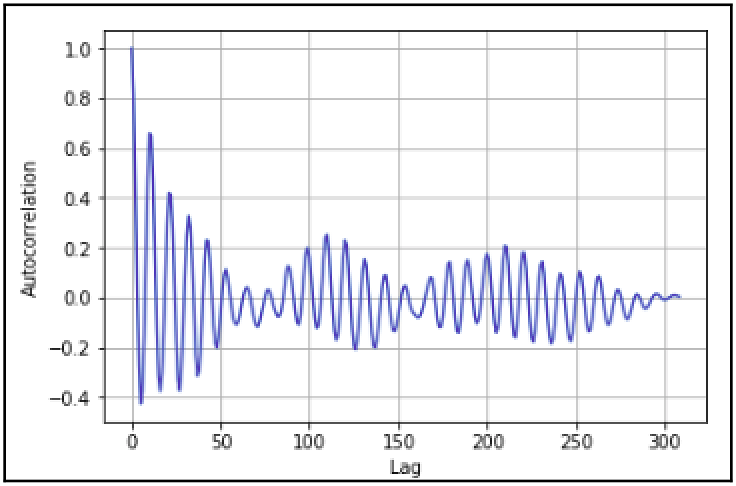

Sales-Original
3-Moving Average(3MA)
5-Moving Average(5MA)
700
600
500
400
300
200
100
01/01/17
01/02/17
01/03/17
01/04/17
01/05/17
01/06/17
01/07/17
01/08/17
01/09/17
01/10/17
01/11/17
01/12/17
01/01/18
01/02/18
01/03/18
01/04/18
01/05/18
01/06/18
01/07/18
01/08/18
01/09/18
01/10/18
01/11/18
01/12/18
01/01/19
01/02/19
01/03/19
01/04/19
01/05/19
01/06/19
01/07/19
01/08/19
01/09/19
01/10/19
01/11/19
01/12/19

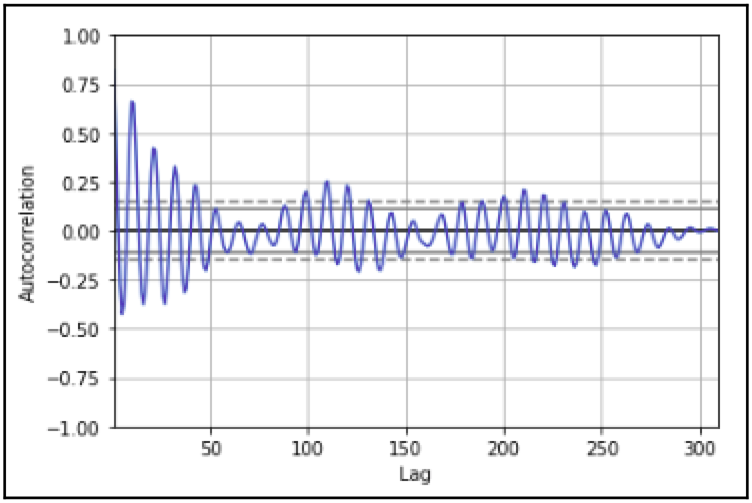


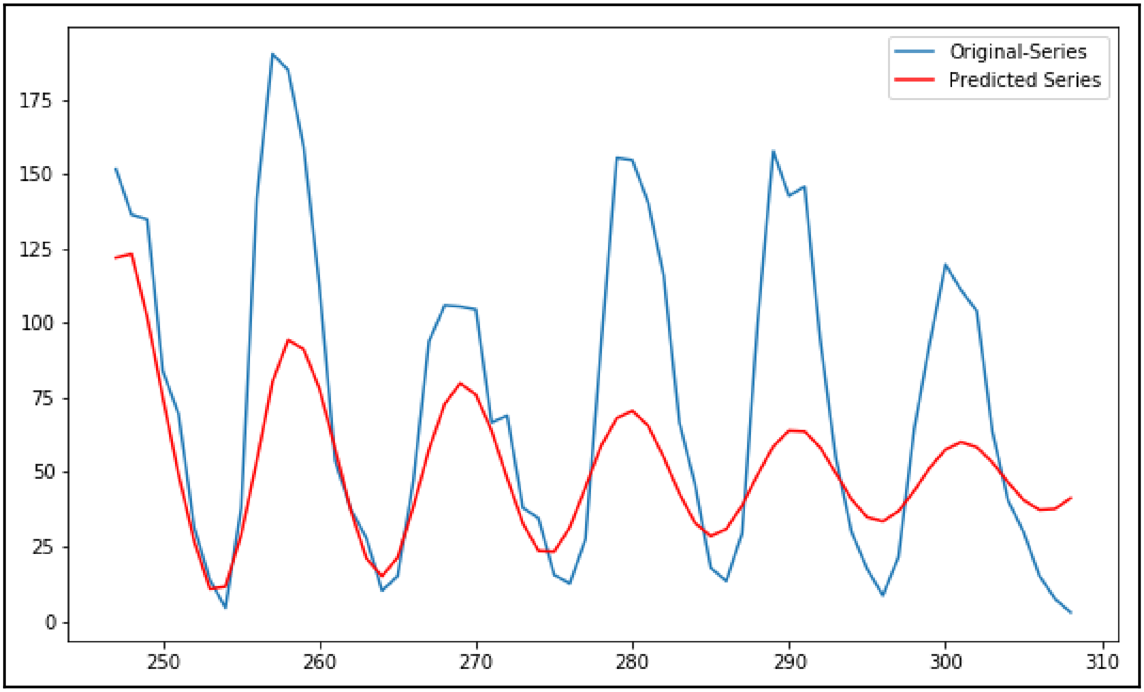
Sales
boxcar
triang
hamming
blackman
700
600
500
400
300
200
100
01/01/17
01/06/17
01/11/17
01/04/18
01/09/18
01/02/19
01/07/19
01/12/19
Time

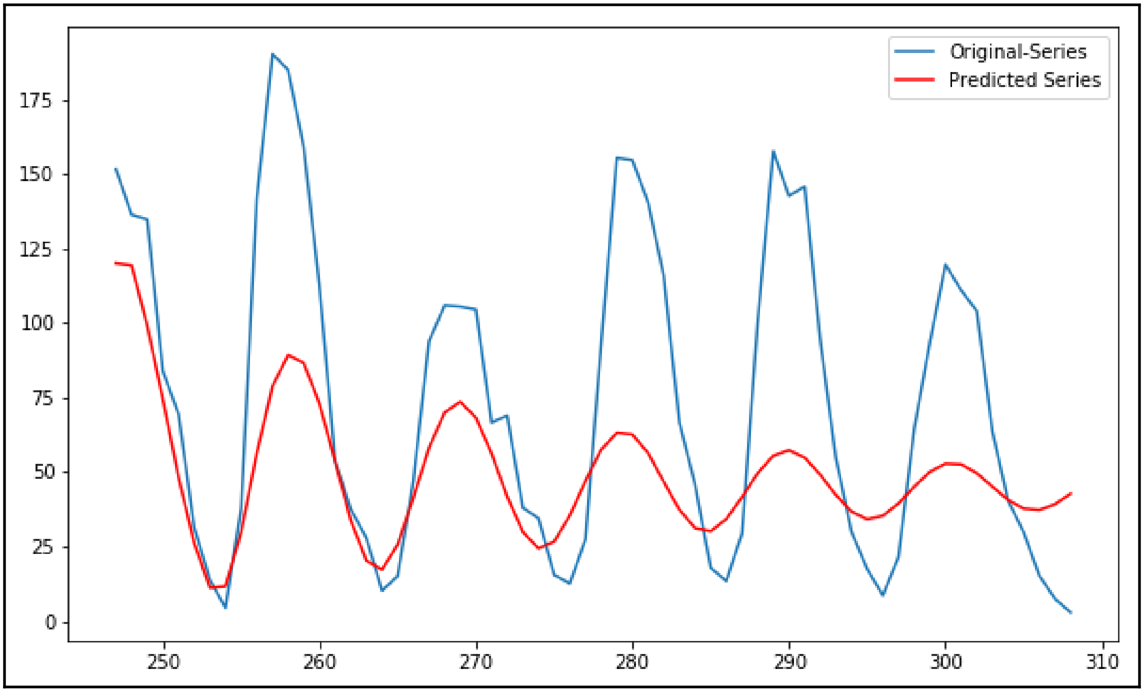


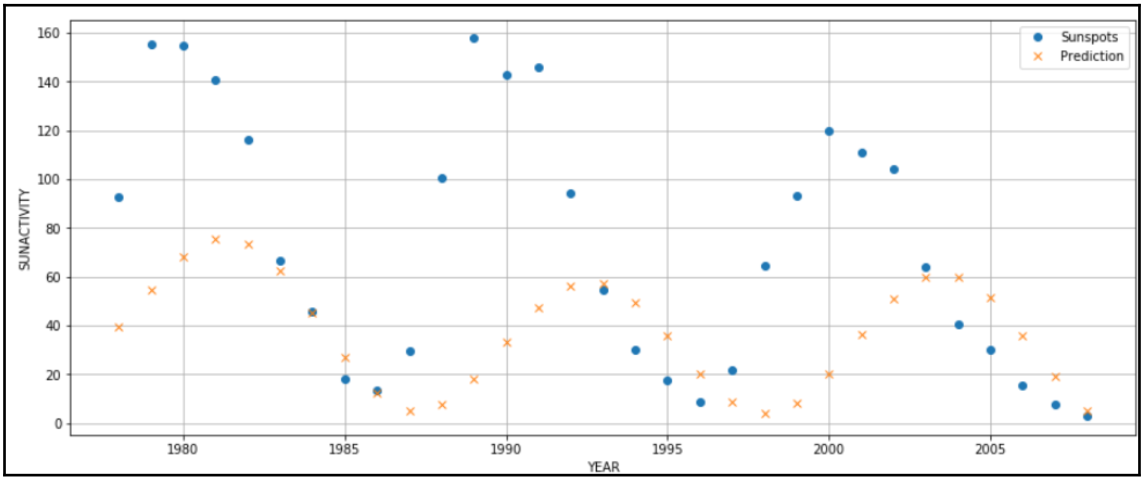
Observed
200
100
Trend
150
100
Seasonal
1.25
1.00
Residual
1.00
0.75
1959
1964
1969
1974
1979
1984
1989
1994
date

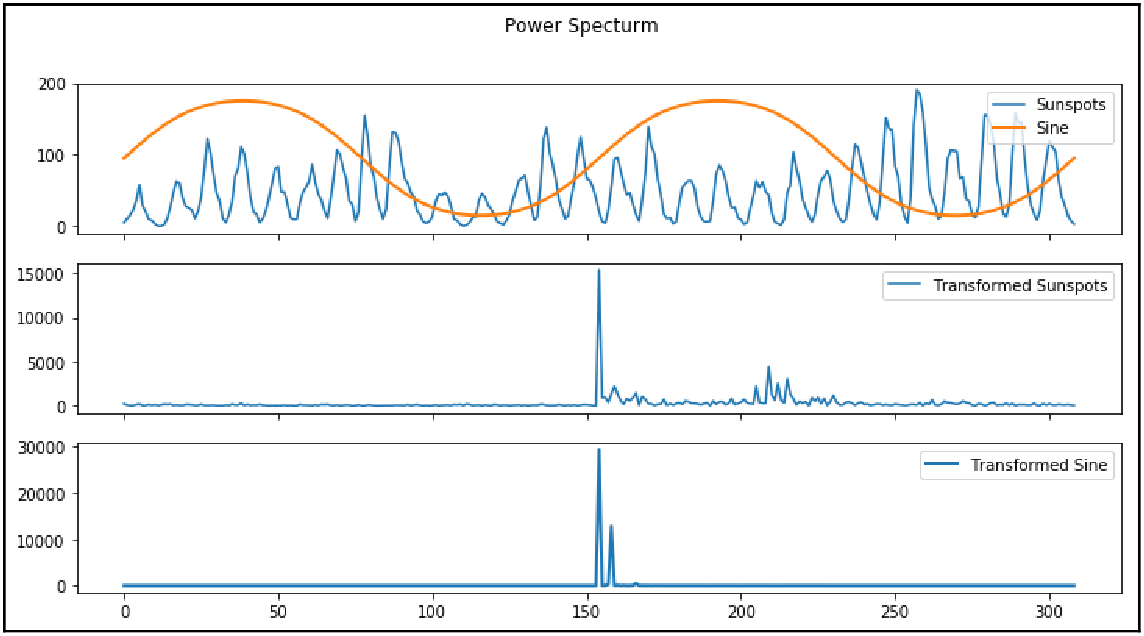

Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
50
100
150
200
250
300
Lag

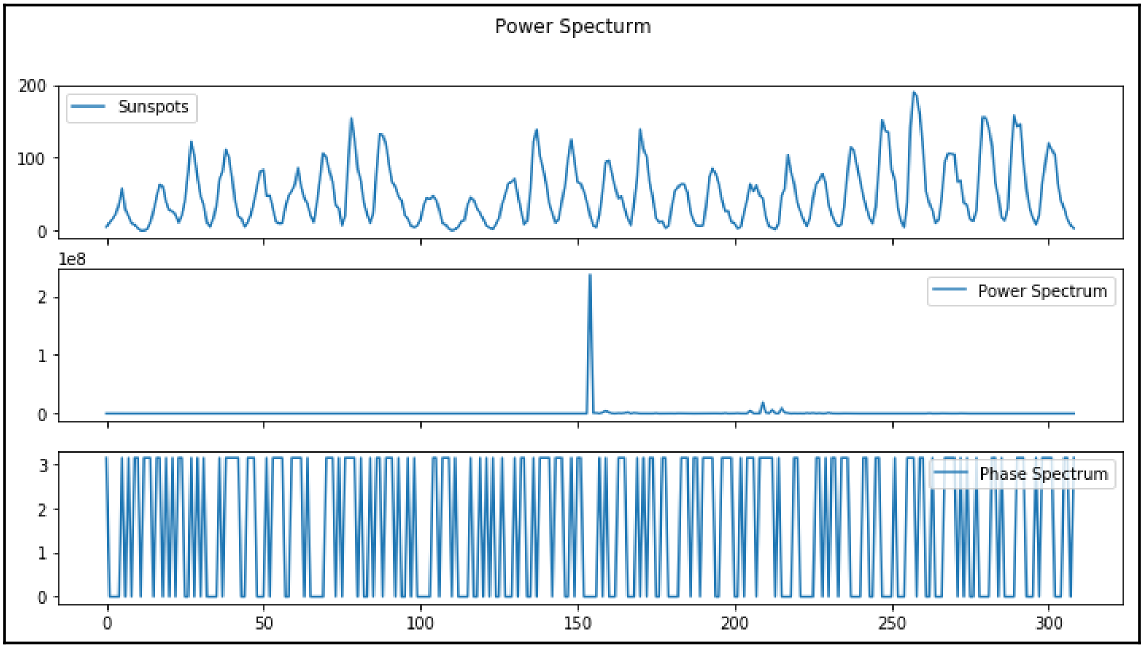

Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
50
100
150
200
250
300
Lag


Original-Series
Predicted Series
175
150
125
100
75
50
25
0
250
260
270
280
290
300
310

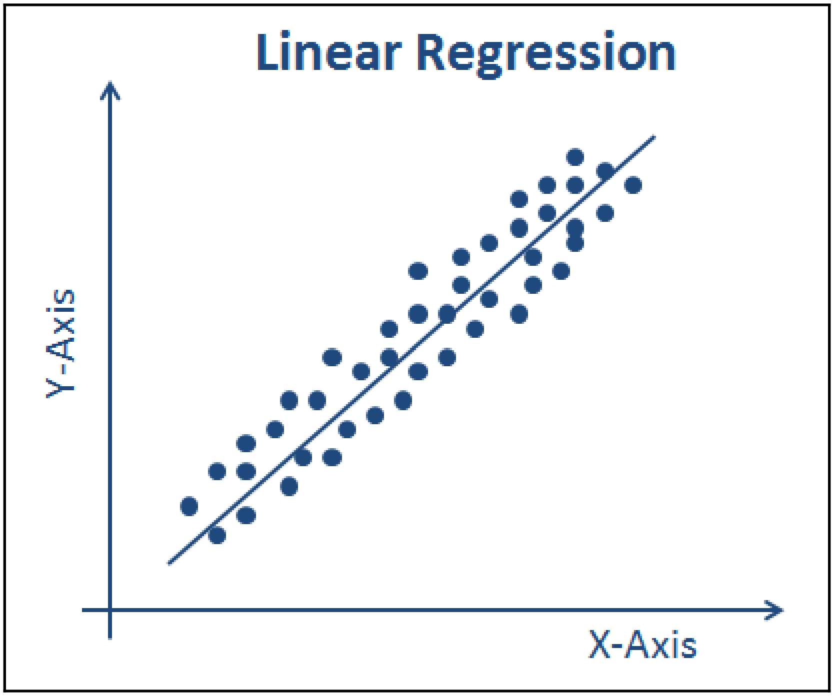

Original-Series
Predicted Series
175
150
125
100
75
50
25
0
250
260
270
280
290
300
310


Sunspots
Prediction
SUNACTIVITY
YEAR
160
140
120
100
80
60
40
20
0
1980
1985
1990
1995
2000
2005

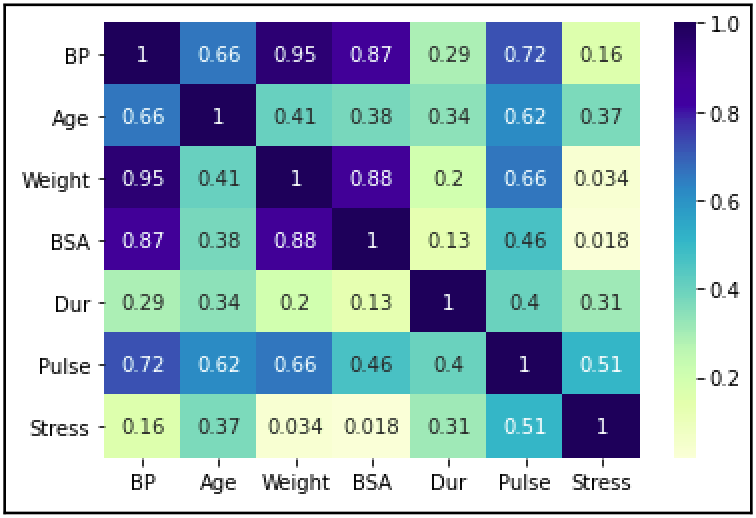

Power Specturm
Sunspots
Sine
Transformed Sunspots
Transformed Sine

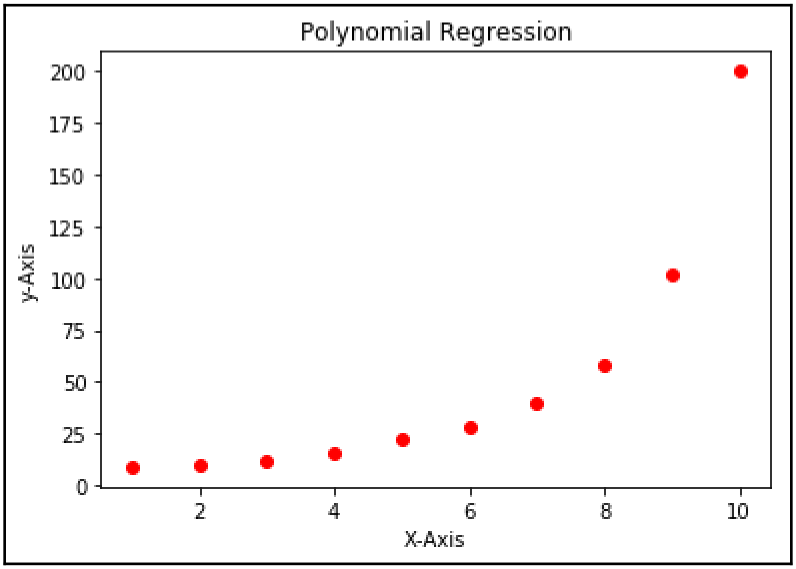

Power Specturm
Sunspots
Power Spectrum
Phase Spectrum

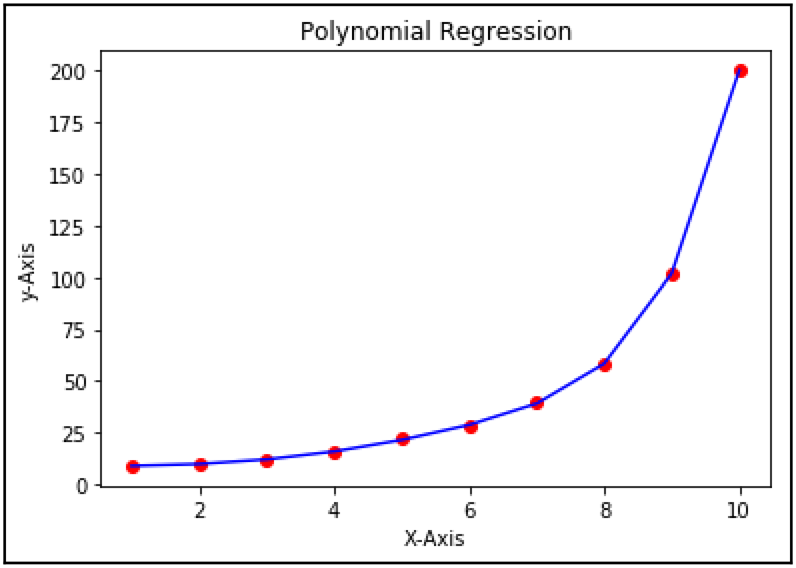
Chapter 9: Supervised Learning - Regression Analysis

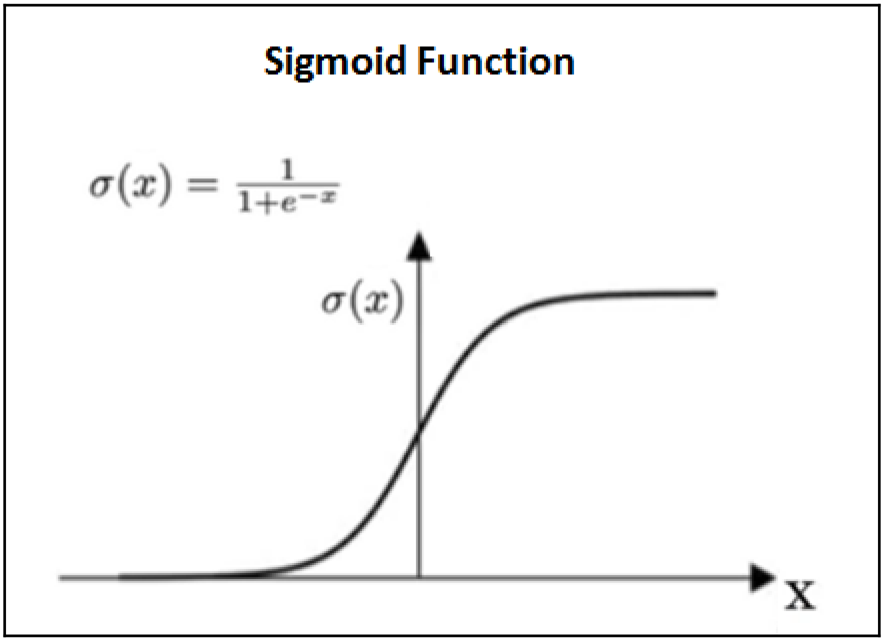


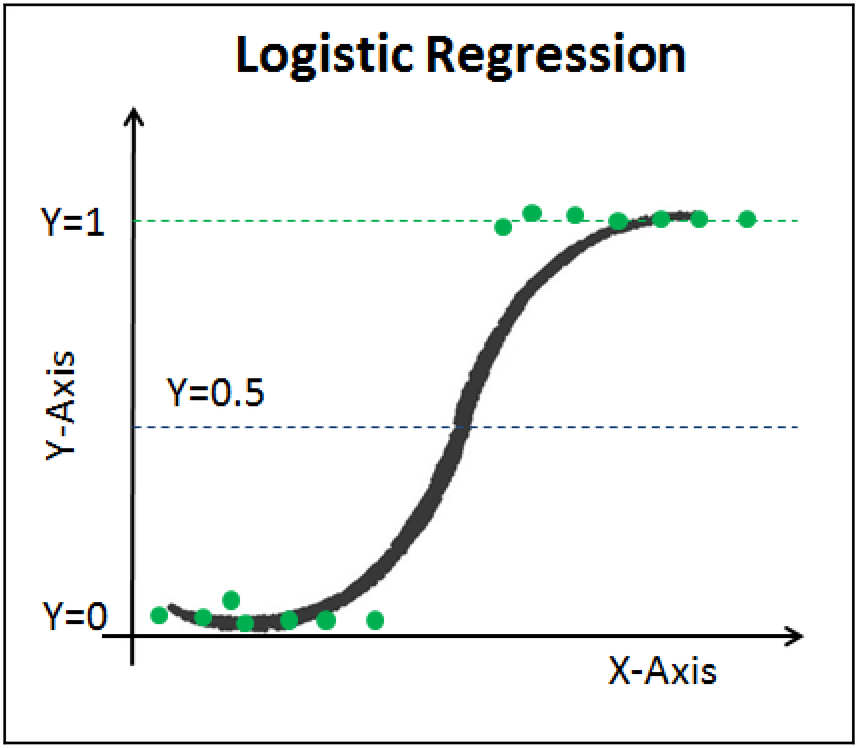
	BP	Age	Weight	BSA	Dur	Pulse	Stress
0	105	47	85.4	1.75	5.1	63	33
1	115	49	94.2	2.10	3.8	70	14
2	116	49	95.3	1.98	8.2	72	10
3	117	50	94.7	2.01	5.8	73	99
4	112	51	89.4	1.89	7.0	72	95


BP
Age
Weight
BSA
Dur
Pulse
Stress
1 0.66 0.95 0.87 0.29 0.72 0.16
0.66 1 0.41 0.38 0.34 0.62 0.37
0.95 0.41 1 0.88 0.2 0.66 0.034
0.87 0.38 0.88 1 0.13 0.46 0.018
0.29 0.34 0.2 0.13 1 0.4 0.31
0.72 0.62 0.66 0.46 0.4 1 0.51
0.16 0.37 0.034 0.018 0.31 0.51 1
1.0
0.8
0.6
0.4
0.2


Polynomial Regression
y-Axis
X-Axis
0
25
50
75
100
125
150
175
200
2
4
6
8
10

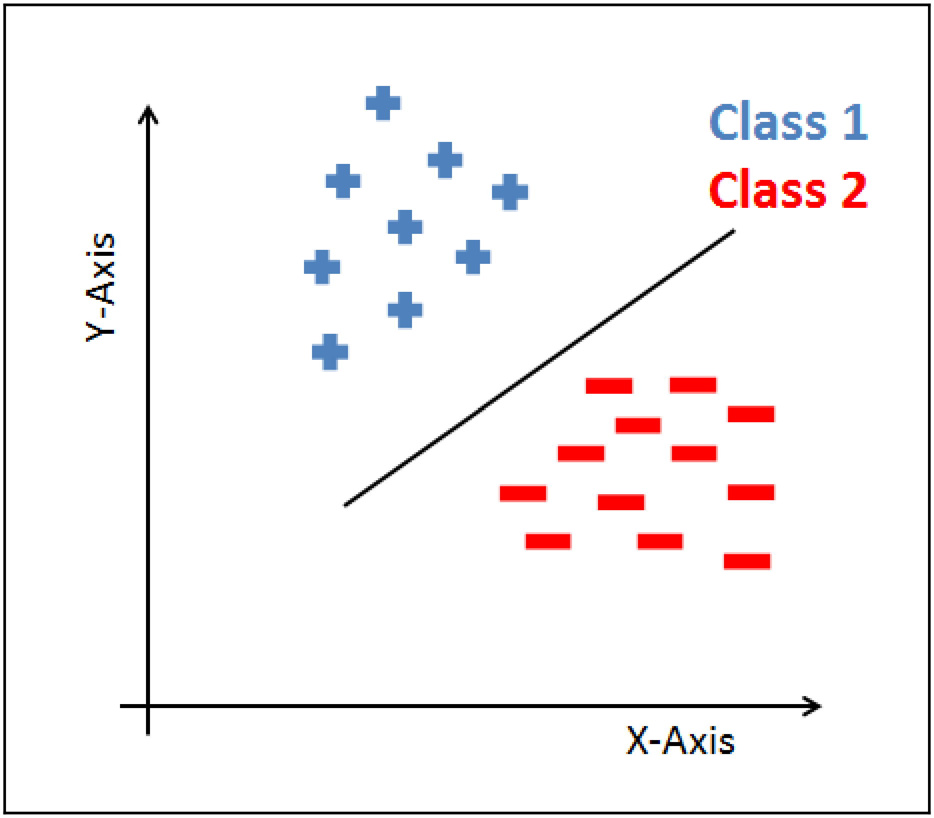

Polynomial Regression
200
175
150
125
100
75
50
25
0
y-Axis
2
4
6
8
10
X-Axis

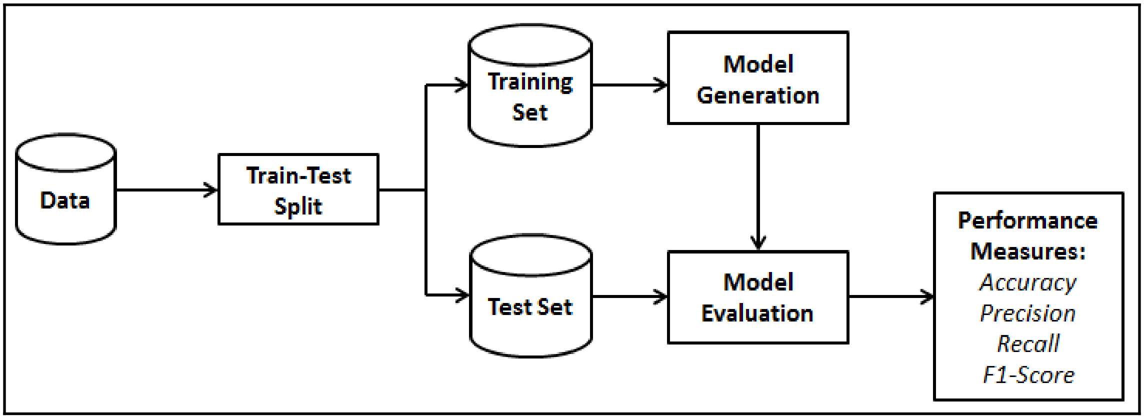

Sigmoid Function
$\sigma(x) = \frac{1}{1+e^{-x}}$
$\sigma(x)$
X



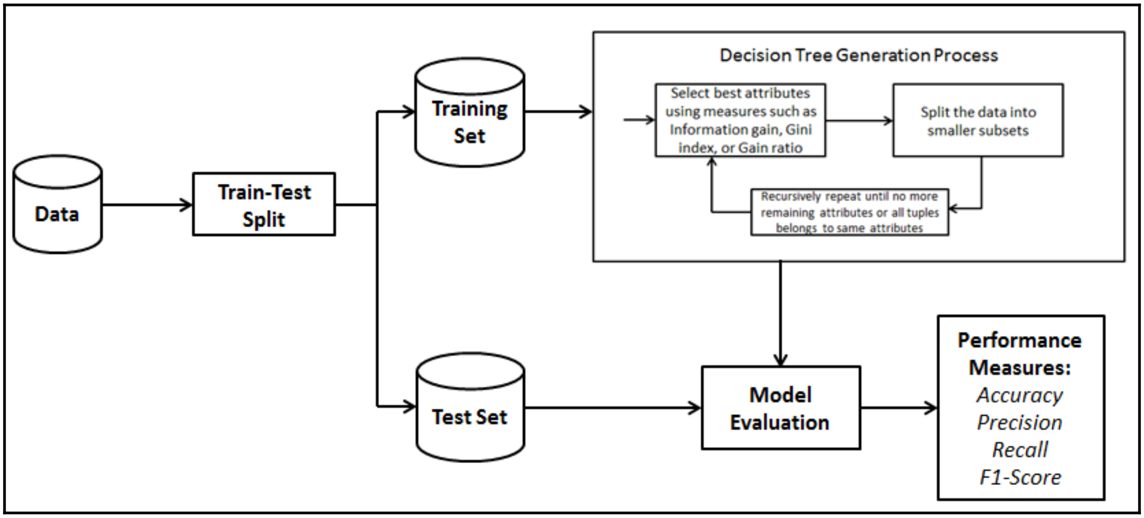
	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	label
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

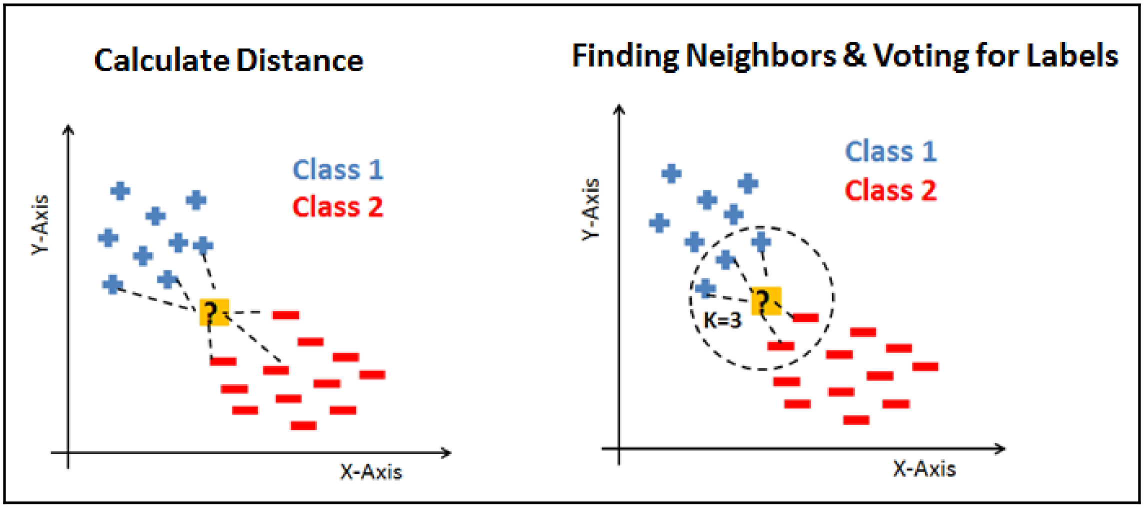
Chapter 10: Supervised Learning - Classification Techniques

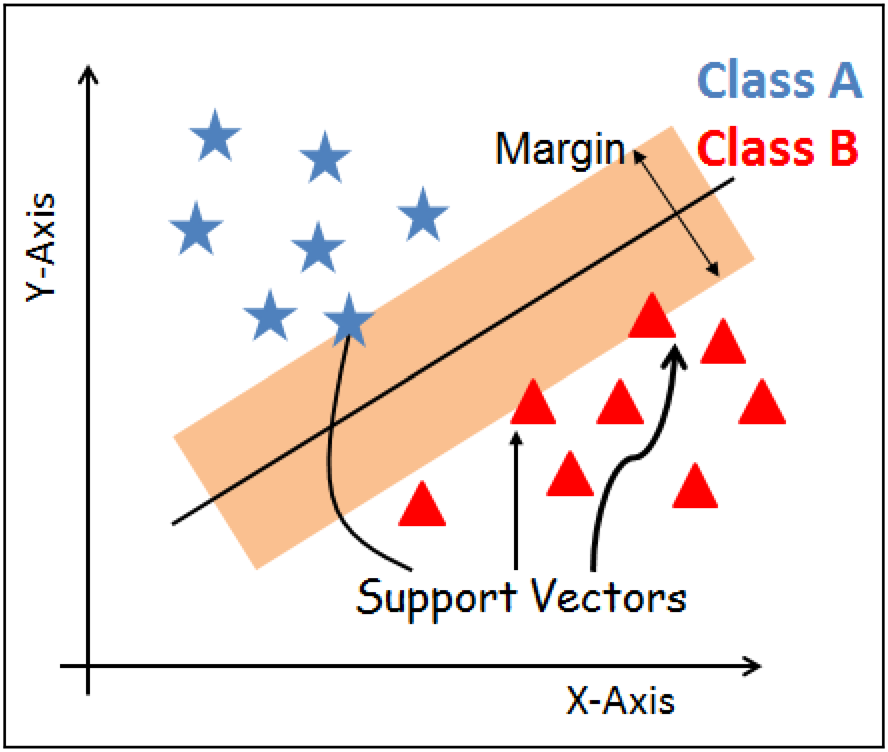




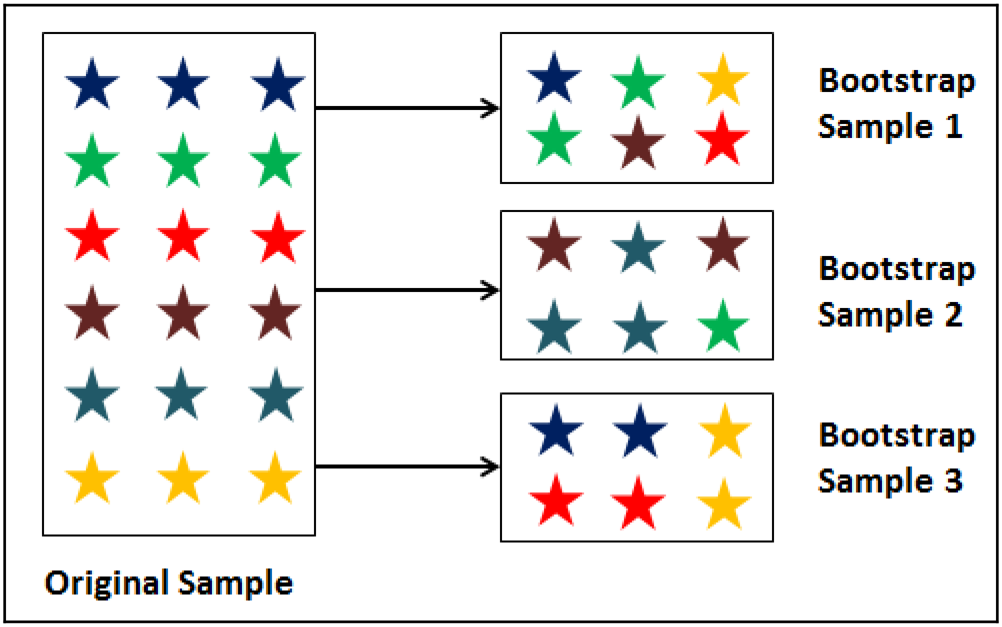
	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	label
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1




Calculate Distance
Class 1
Class 2
?
Y-Axis
X-Axis
Finding Neighbors & Voting for Labels
Class 1
Class 2
?
K=3
Y-Axis
X-Axis

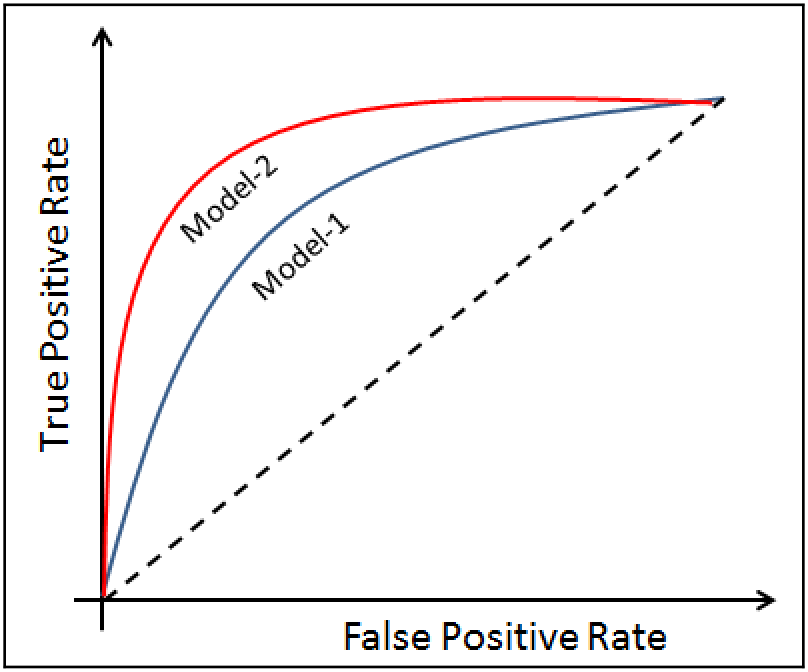

Class A
Class B
Margin
Y-Axis
Support Vectors
X-Axis

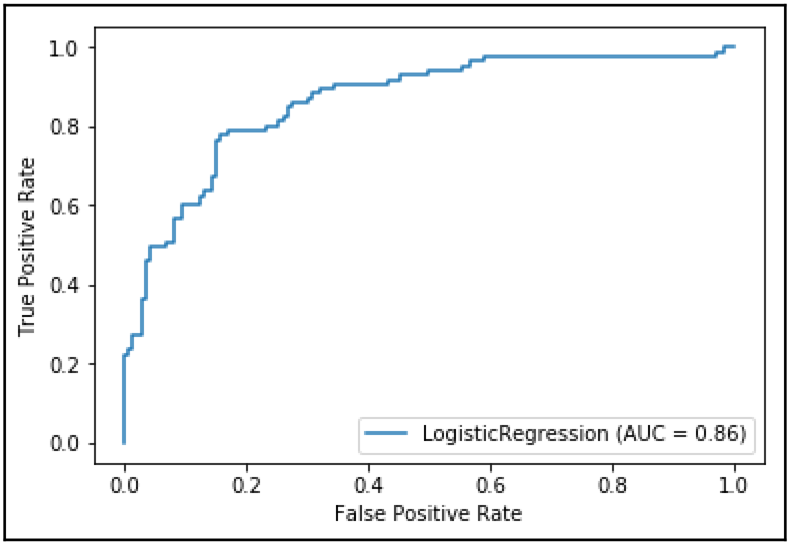
	Data				
	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5
Iteration-1	Test	Train	Train	Train	Train
Iteration-2	Train	Test	Train	Train	Train
Iteration-3	Train	Train	Test	Train	Train
Iteration-4	Train	Train	Train	Test	Train
Iteration-5	Train	Train	Train	Train	Test



	Predicted (YES)	Predicted (NO)	
Actual (YES)	TP= 500	FN=25	Recall $= \frac{TP}{TP+FN} = \frac{500}{500+25}$ = 0.9524
Actual (NO)	FP=50	TN=250	Specificity $= \frac{TN}{TN+FP} = \frac{250}{250+50}$ = 0.8333
	Precision $= \frac{TP}{TP+FP} = \frac{500}{500+50}$ = 0.9091	Negative Predictions $= \frac{TN}{TN+FN} = \frac{250}{250+25}$ = 0.9091	Total = 825

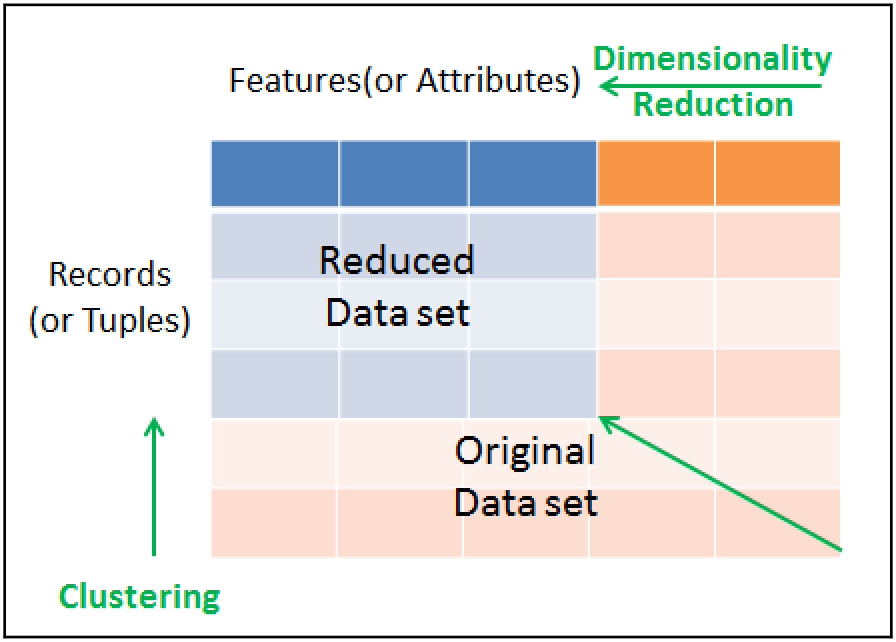
	precision	recall	f1-score	support
Yes(1)	0.79	0.90	0.84	146
No(0)	0.78	0.58	0.66	85
accuracy			0.78	231
macro avg	0.78	0.74	0.75	231
weighted avg	0.78	0.78	0.78	231

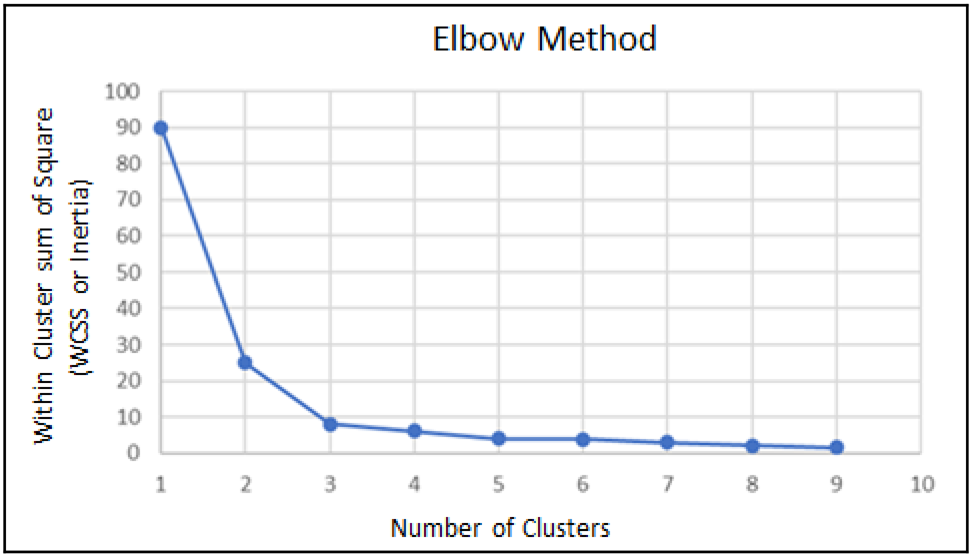

True Positive Rate
Model-2
Model-1
False Positive Rate

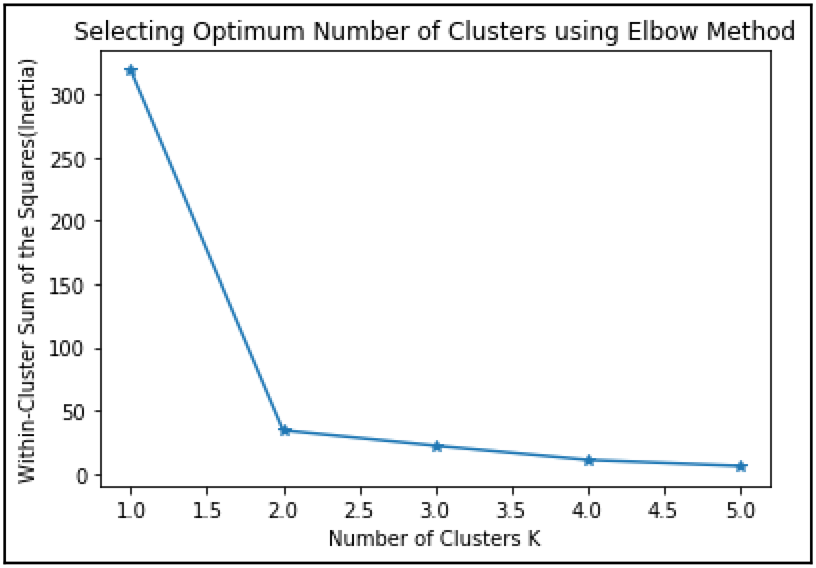

1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
LogisticRegression (AUC = 0.86)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate

$$\textbf{AUC} \begin{cases} 0.5 & \text{No Discrimination} \\ 0.6 - 0.7 & \text{Poor} \\ 0.7 - 0.8 & \text{Acceptable(fair)} \\ 0.8 - 0.9 & \text{Excellent(good)} \\ >0.9 & \text{Outstanding} \end{cases}$$

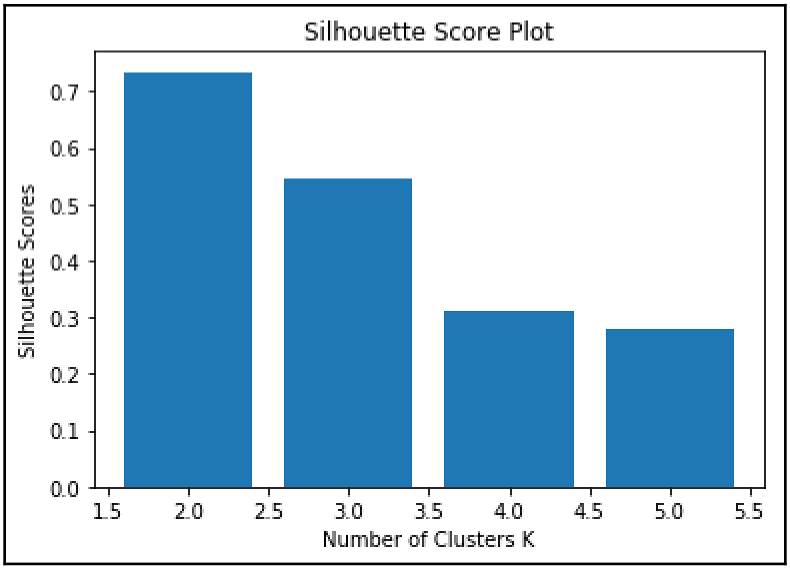
Chapter 11: Unsupervised Learning - PCA and Clustering



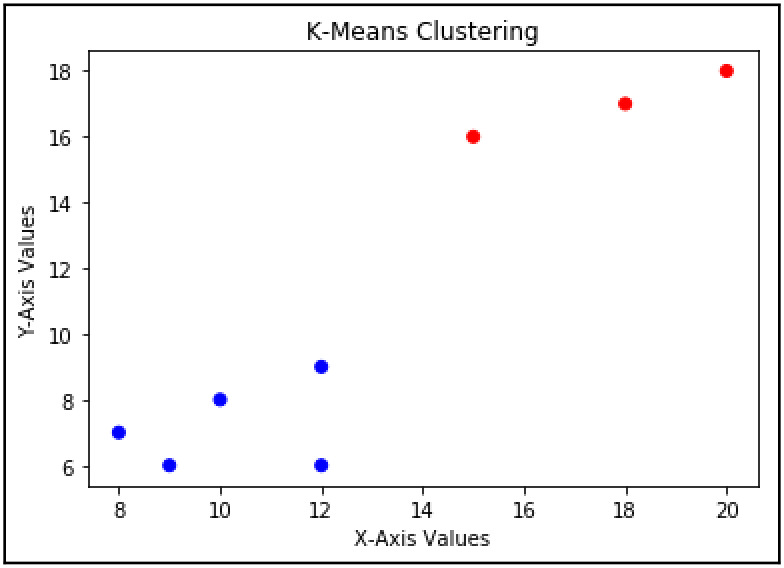


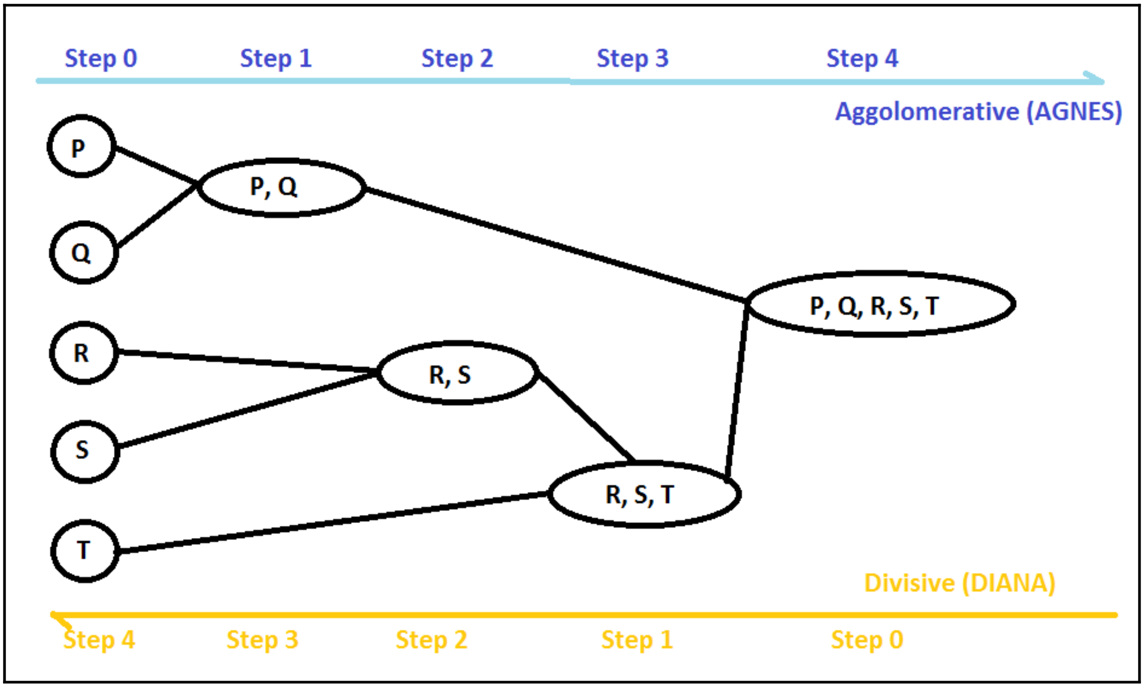


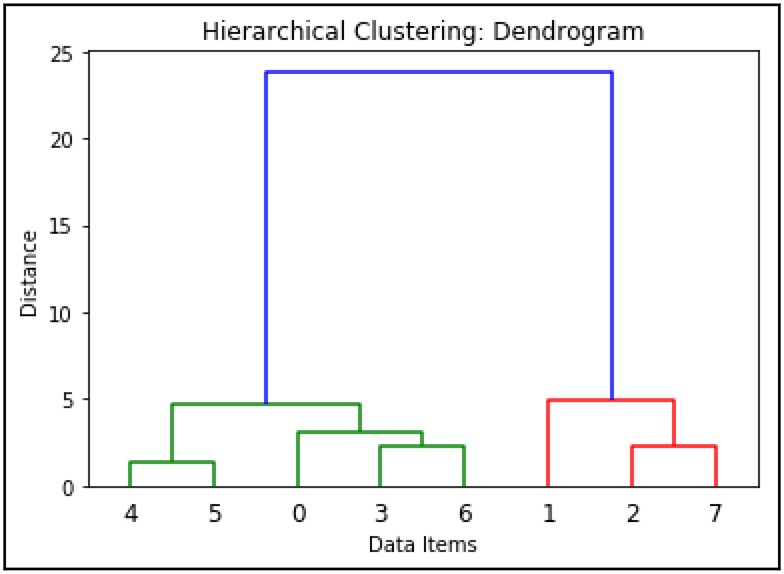
Selecting Optimum Number of Clusters using Elbow Method
Within-Cluster Sum of the Squares(Inertia)
Number of Clusters K

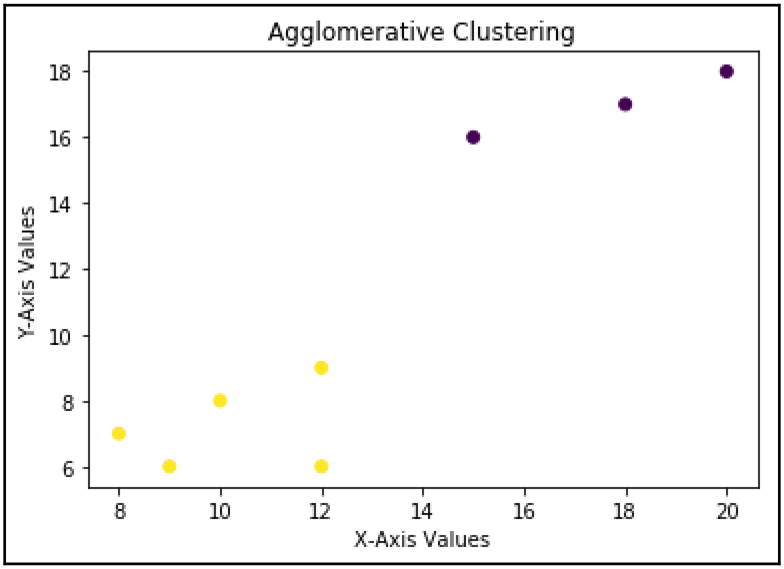


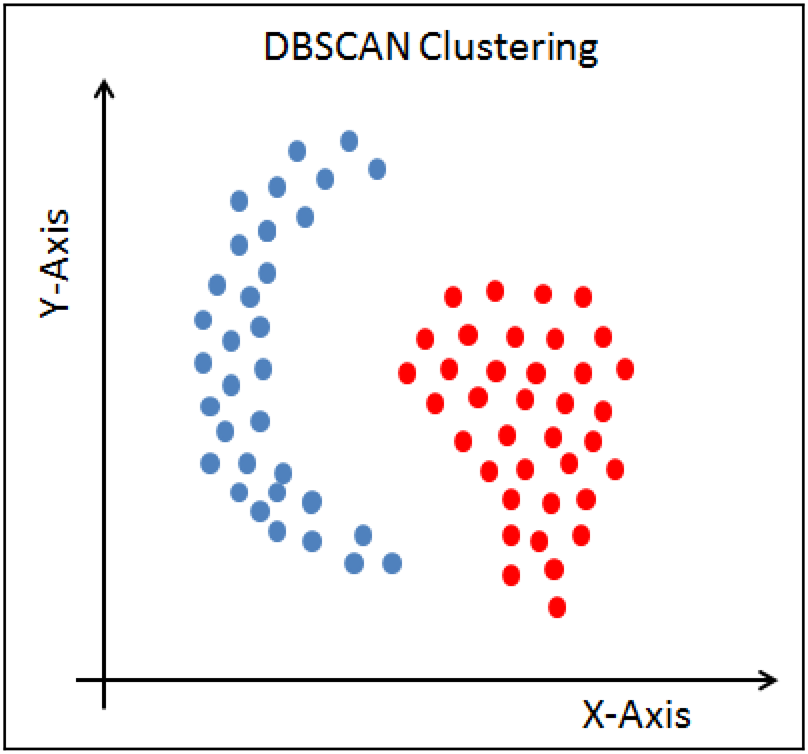
Silhouette Score Plot
Silhouette Scores
Number of Clusters K

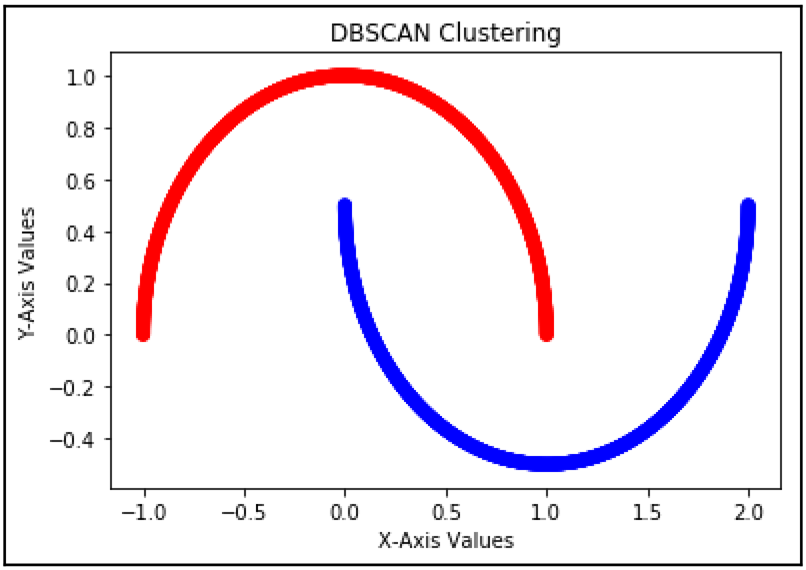

K-Means Clustering
Y-Axis Values
X-Axis Values
18
16
14
12
10
8
6
8
10
12
14
16
18
20


Step 0
Step 1
Step 2
Step 3
Step 4
Aggolomerative (AGNES)
P
Q
R
S
T
P, Q
R, S
R, S, T
P, Q, R, S, T
Divisive (DIANA)
Step 4
Step 3
Step 2
Step 1
Step 0

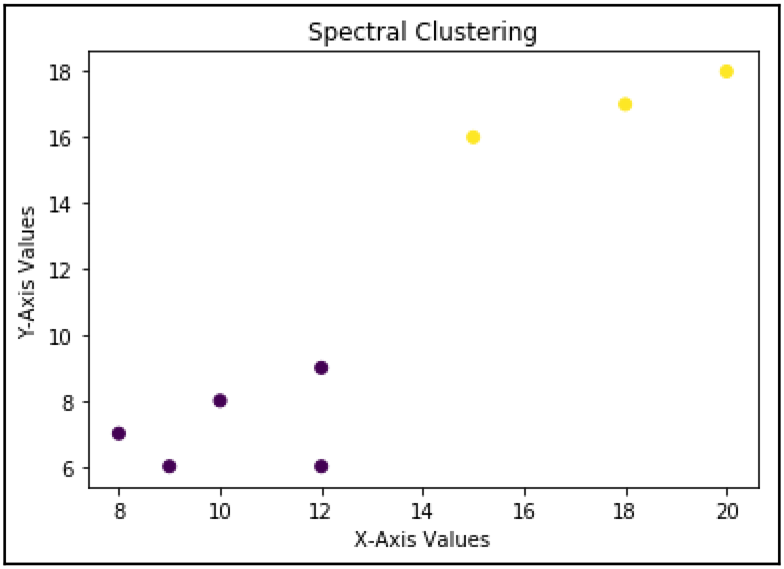

Hierarchical Clustering: Dendrogram
Distance
25
20
15
10
5
0
4
5
0
3
6
1
2
7
Data Items


Agglomerative Clustering
Y-Axis Values
18
16
14
12
10
8
6
8
10
12
14
16
18
20
X-Axis Values

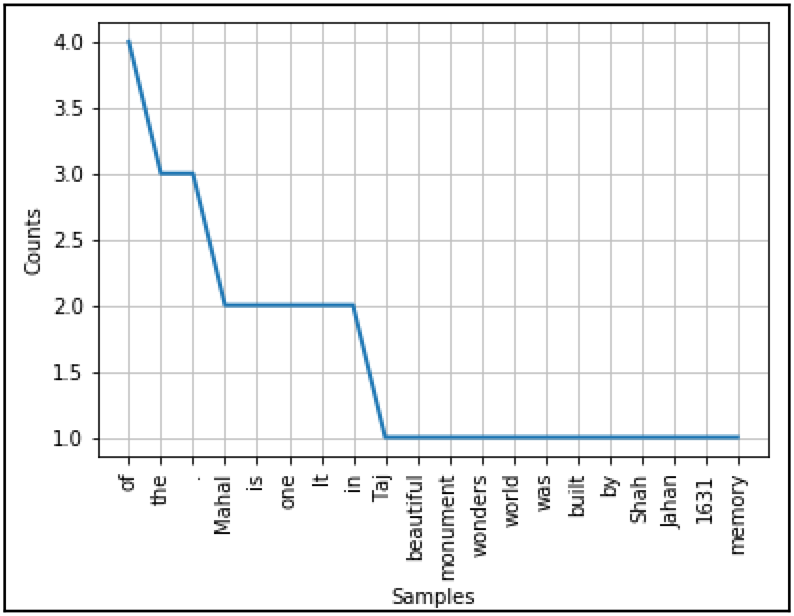

DBSCAN Clustering
Y-Axis
X-Axis

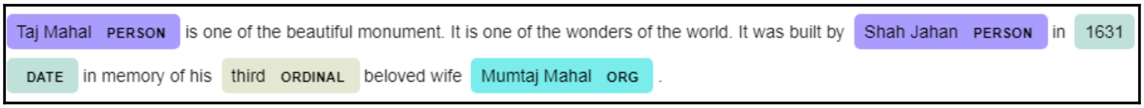

DBSCAN Clustering
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
Y-Axis Values
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
X-Axis Values

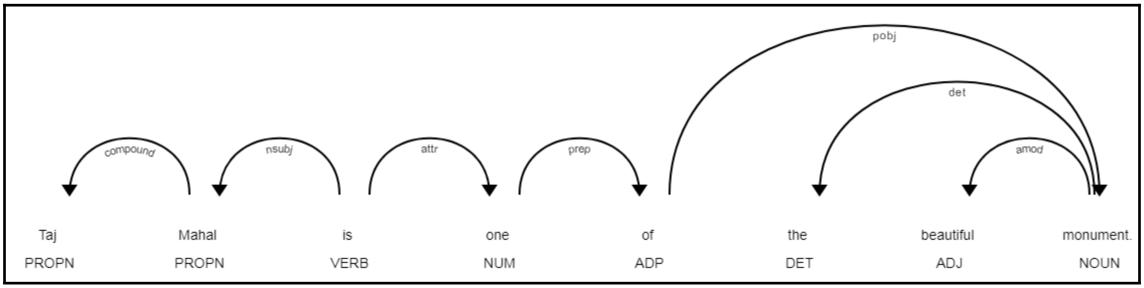
$$K(x_i, x_j) = e^{\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right)}$$

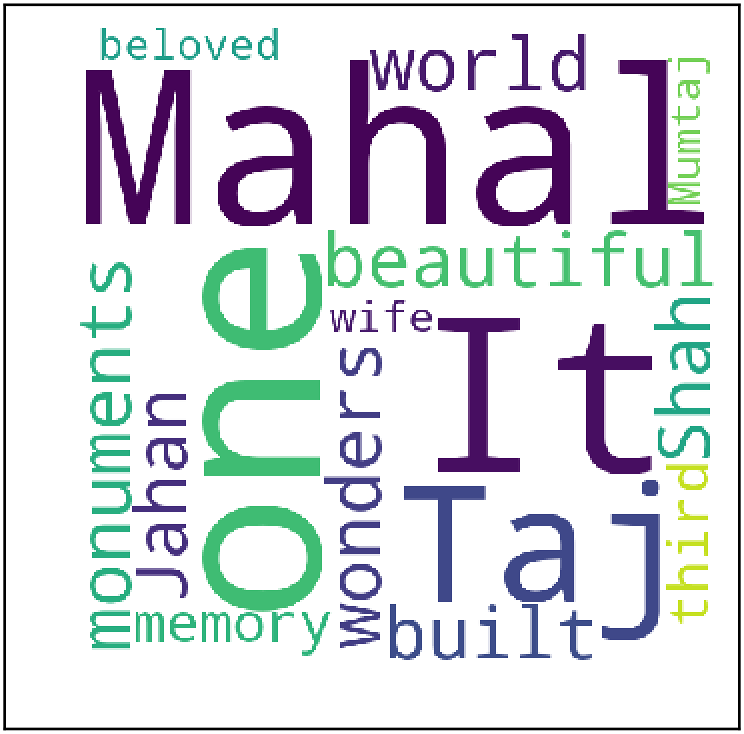


	pregnant	glucose	bp	skin	insulin	bmi	pedigree	age	label
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

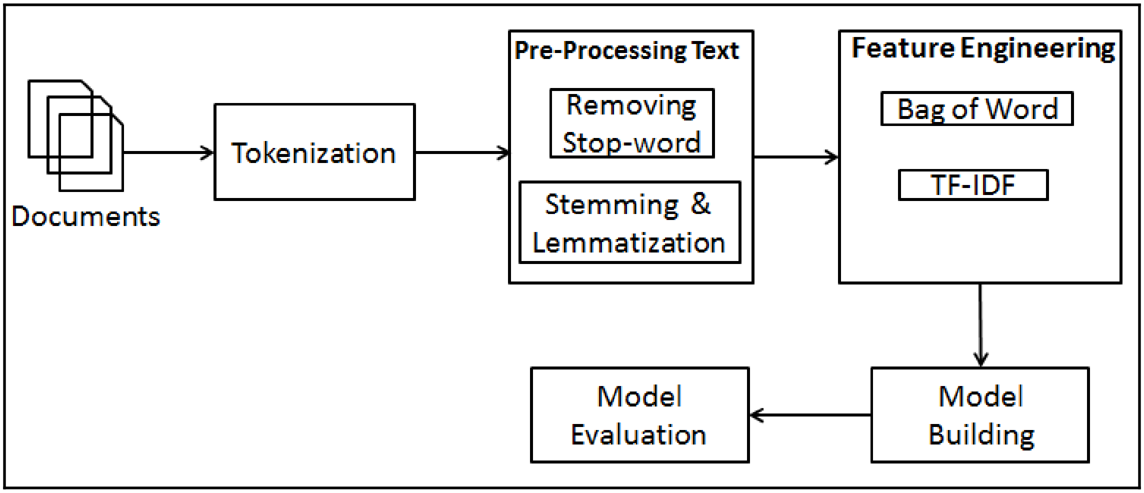

Counts
4.0
3.5
3.0
2.5
2.0
1.5
1.0
of
the
.
Mahal
is
one
It
in
Taj
beautiful
monument
wonders
world
was
built
by
Shah
Jahan
1631
memory
Samples


Taj Mahal PERSON
is one of the beautiful monument. It is one of the wonders of the world. It was built by
Shah Jahan PERSON
in
1631 DATE
in memory of his
third ORDINAL
beloved wife
Mumtaj Mahal ORG
.

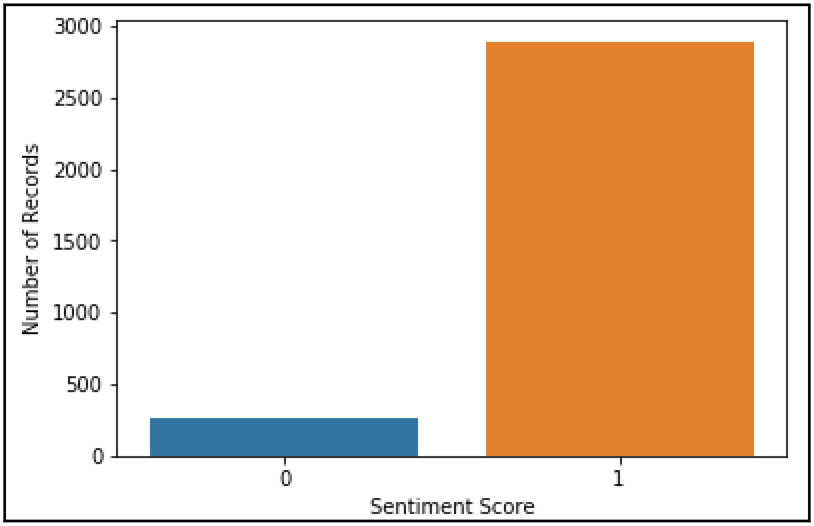

compound
nsubj
attr
prep
pobj
det
amod
Taj
PROPN
Mahal
PROPN
is
VERB
one
NUM
of
ADP
the
DET
beautiful
ADJ
monument.
NOUN



$$IDF(Word) = log_2\left[\frac{Number\ of\ Documents}{Number\ of\ Documents\ that\ Contains\ the\ Word}\right]$$

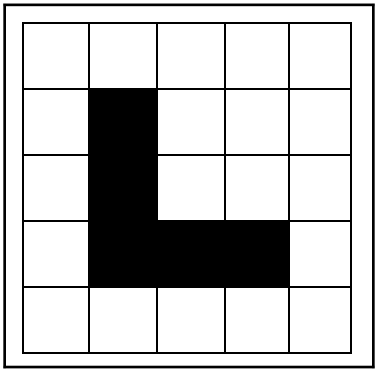


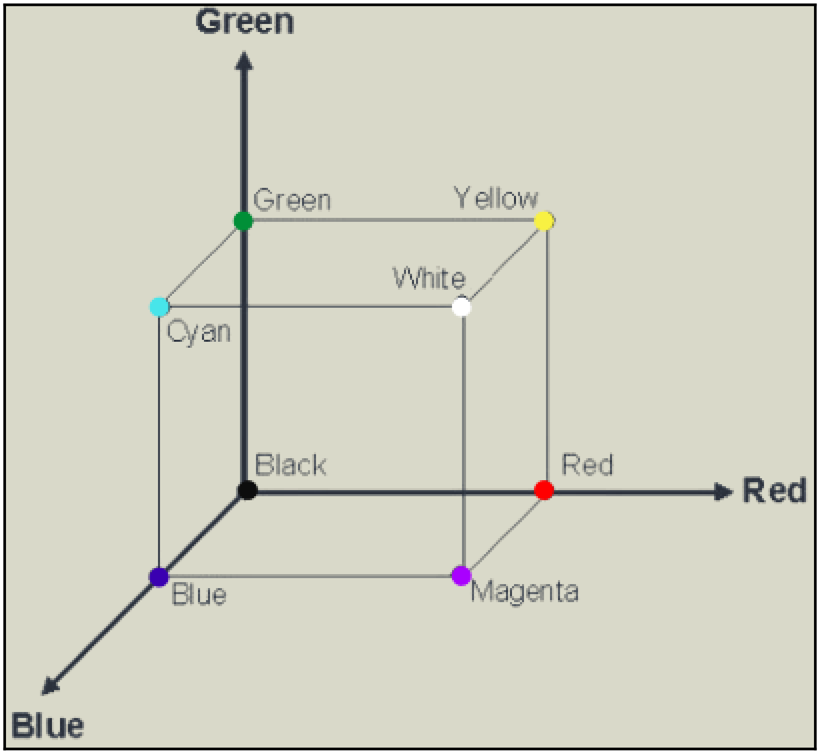
	rating	date	variation	verified_reviews	feedback
0	5	31-Jul-18	Charcoal Fabric	Love my Echo!	1
1	5	31-Jul-18	Charcoal Fabric	Loved it!	1
2	4	31-Jul-18	Walnut Finish	Sometimes while playing a game, you can answer...	1
3	5	31-Jul-18	Charcoal Fabric	I have had a lot of fun with this thing. My 4 ...	1
4	5	31-Jul-18	Charcoal Fabric	Music	1

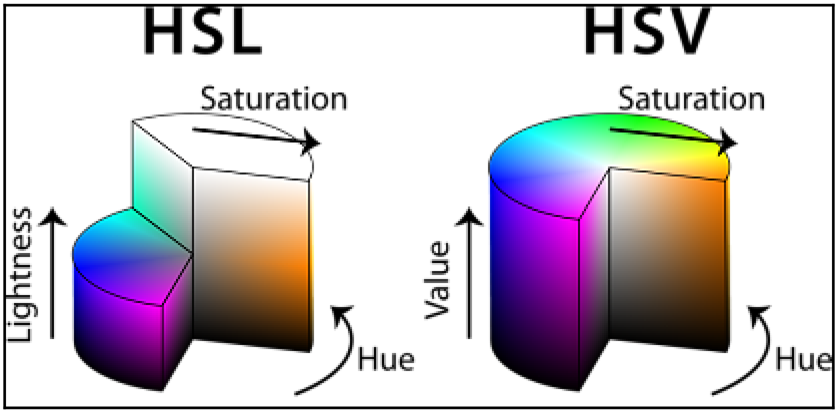


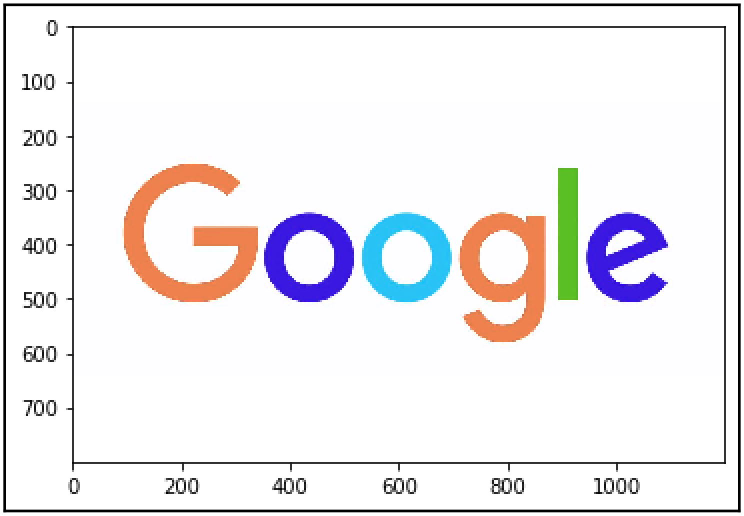
	rating	date	variation	verified_reviews	feedback
0	5	31-Jul-18	Charcoal Fabric	Love my Echo!	1
1	5	31-Jul-18	Charcoal Fabric	Loved it!	1
2	4	31-Jul-18	Walnut Finish	Sometimes while playing a game, you can answer...	1
3	5	31-Jul-18	Charcoal Fabric	I have had a lot of fun with this thing. My 4 ...	1
4	5	31-Jul-18	Charcoal Fabric	Music	1

Chapter 13: Analyzing Image Data

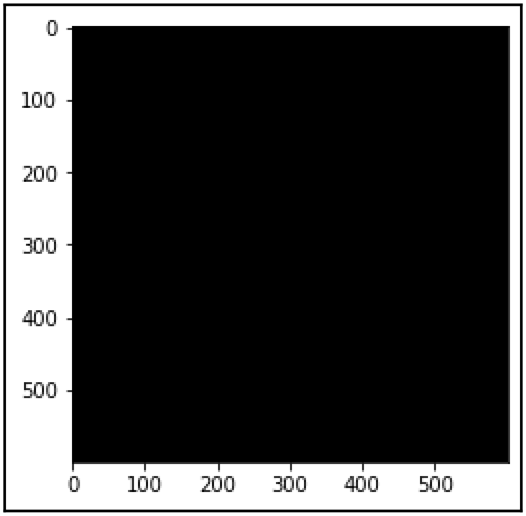


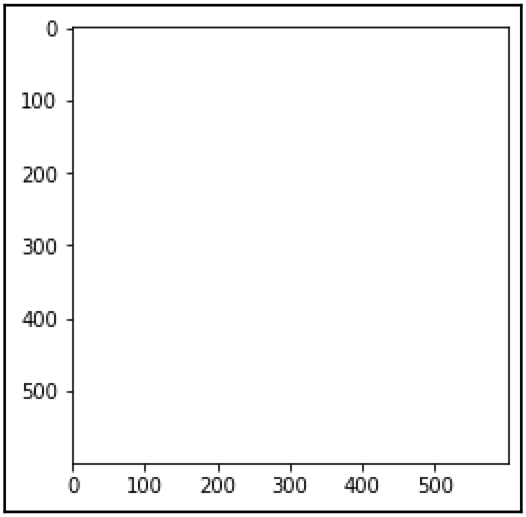

Green
Green
Yellow
White
Cyan
Black
Red
Red
Blue
Magenta
Blue

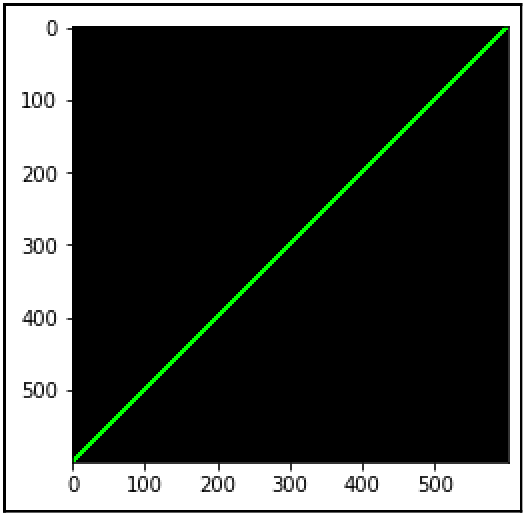

HSL
Saturation
Lightness
Hue
HSV
Saturation
Value
Hue

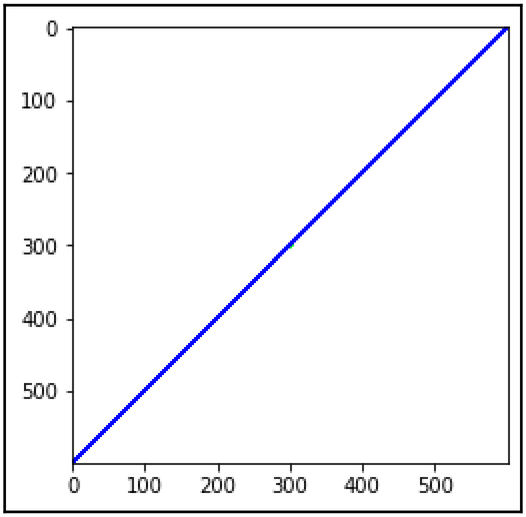

0
100
200
300
400
500
600
700
0
200
400
600
800
1000
Google

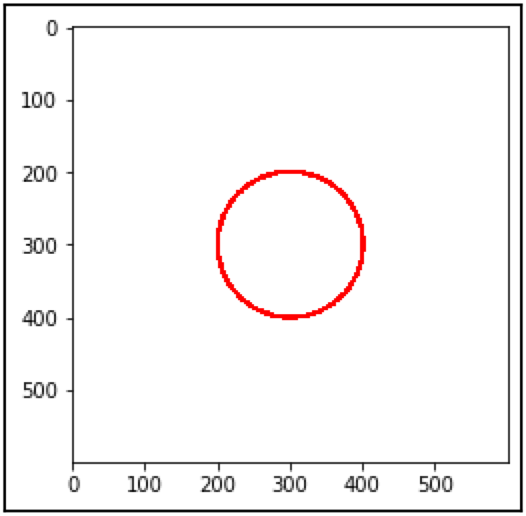

0
100
200
300
400
500
600
700
0
200
400
600
800
1000
Google

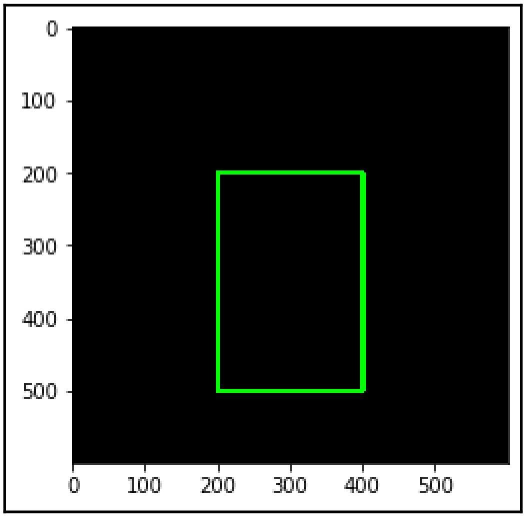

0
100
200
300
400
500
0
100
200
300
400
500

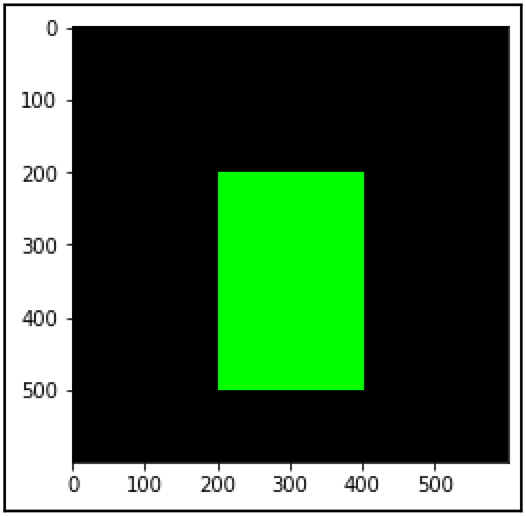

0
100
200
300
400
500
0
100
200
300
400
500

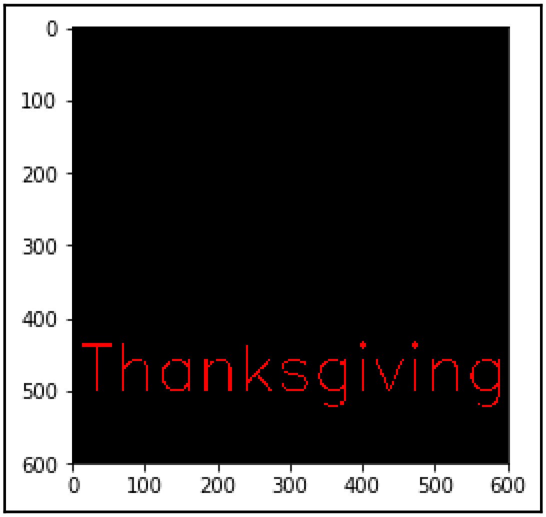

0
100
200
300
400
500
0
100
200
300
400
500

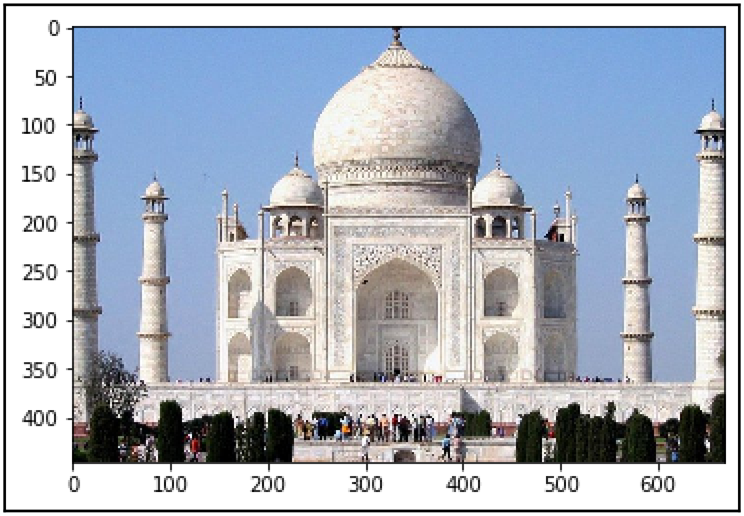

0
100
200
300
400
500
0
100
200
300
400
500

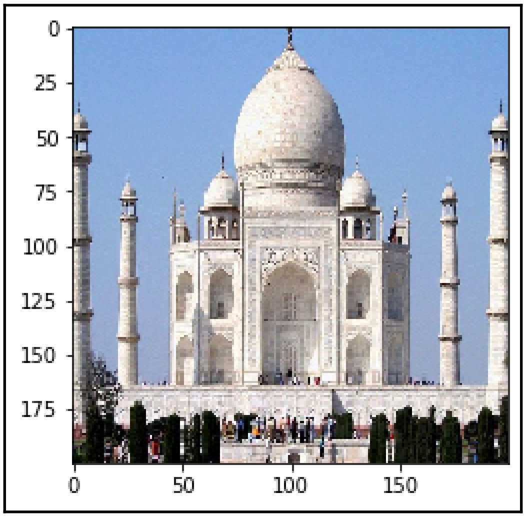

0
100
200
300
400
500
0
100
200
300
400
500

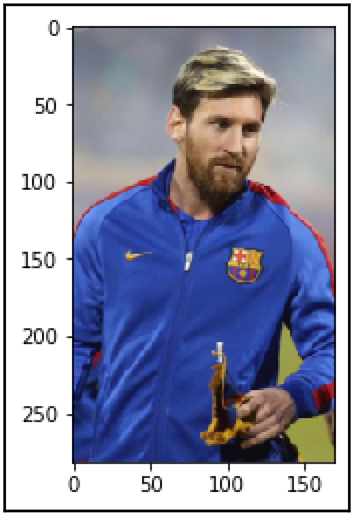

0
100
200
300
400
500
0
100
200
300
400
500

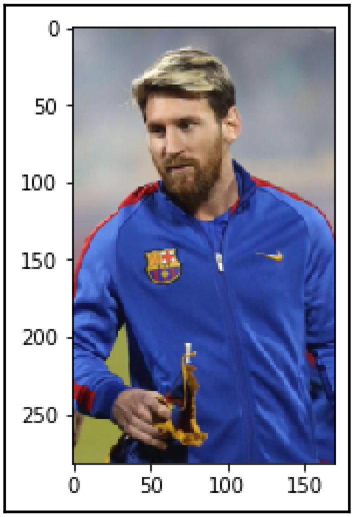

0
100
200
300
400
500
0
100
200
300
400
500

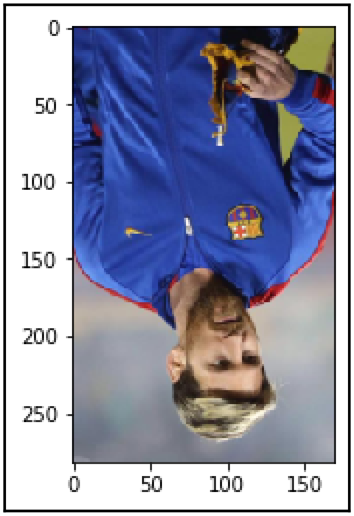

Thanksgiving
0
100
200
300
400
500
600
0
100
200
300
400
500
600

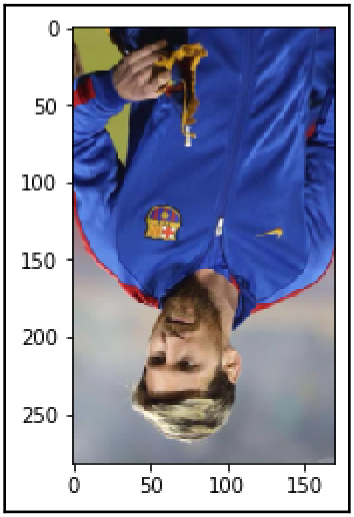

0
50
100
150
200
250
300
350
400
0
100
200
300
400
500
600


0
25
50
75
100
125
150
175
0
50
100
150


0
50
100
150
200
250
0
50
100
150

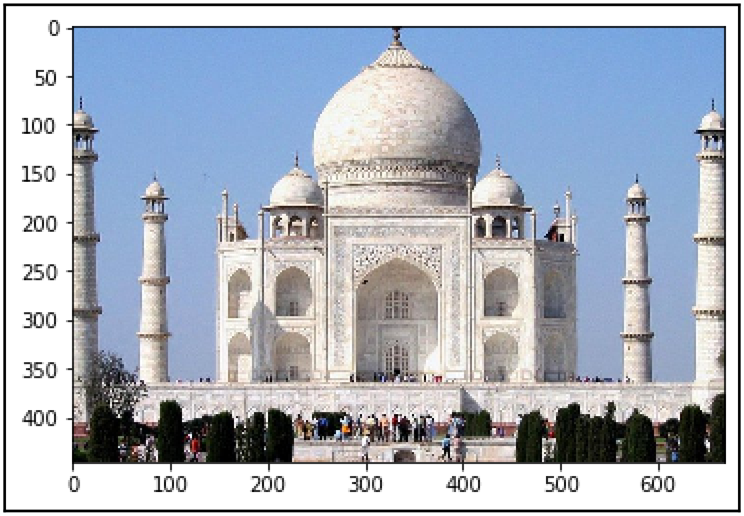

0
50
100
150
200
250
0
50
100
150

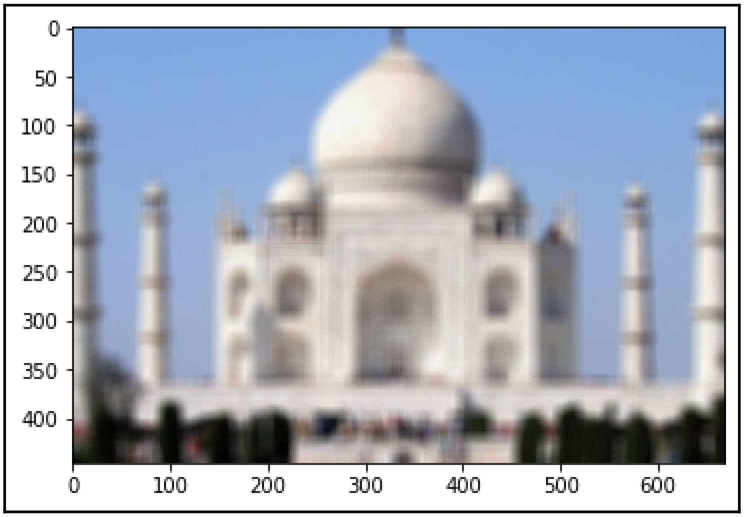

0
50
100
150
200
250
0
50
100
150

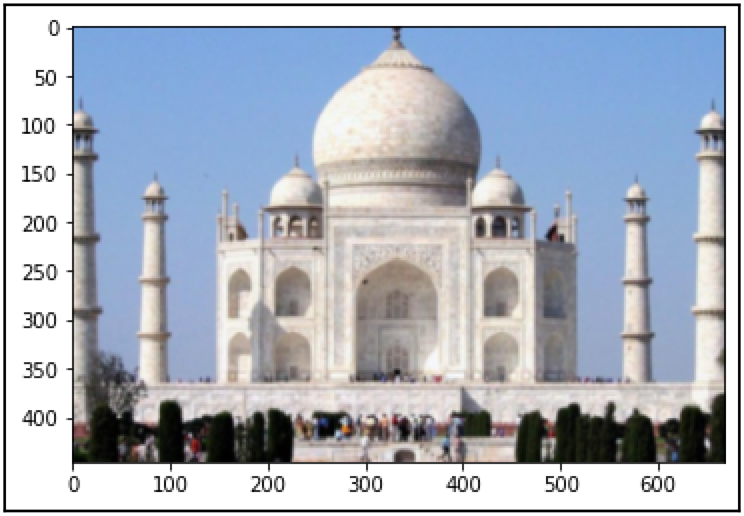

0
50
100
150
200
250
0
50
100
150

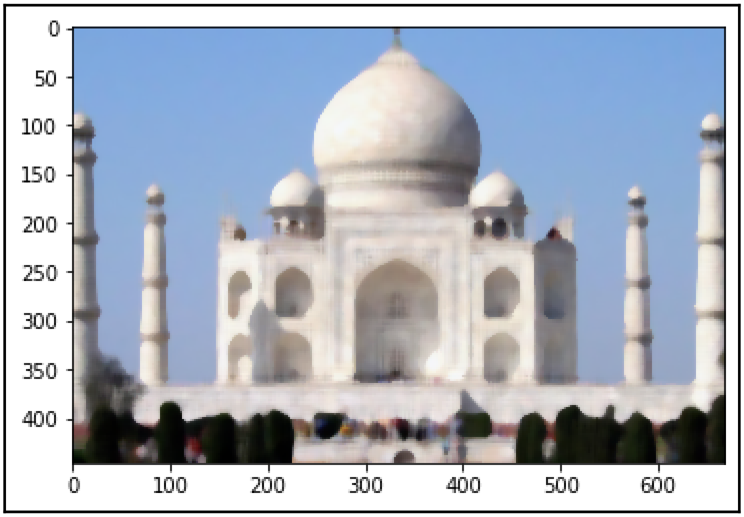

0
25
50
75
100
125
150
175
0
50
100
150
200
250

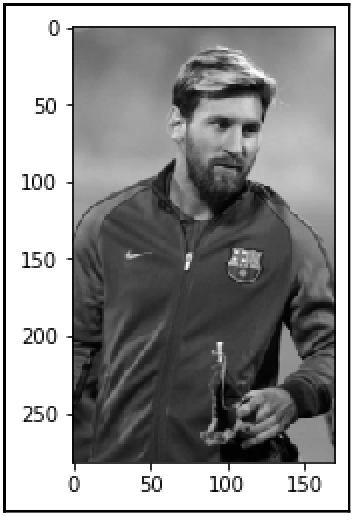

0
25
50
75
100
125
150
175
0
50
100
150
200
250

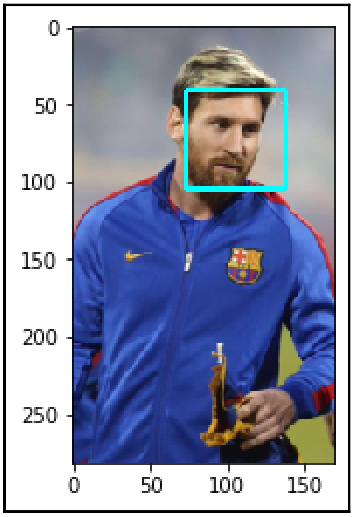

0
50
100
150
200
250
300
350
400
0
100
200
300
400
500
600

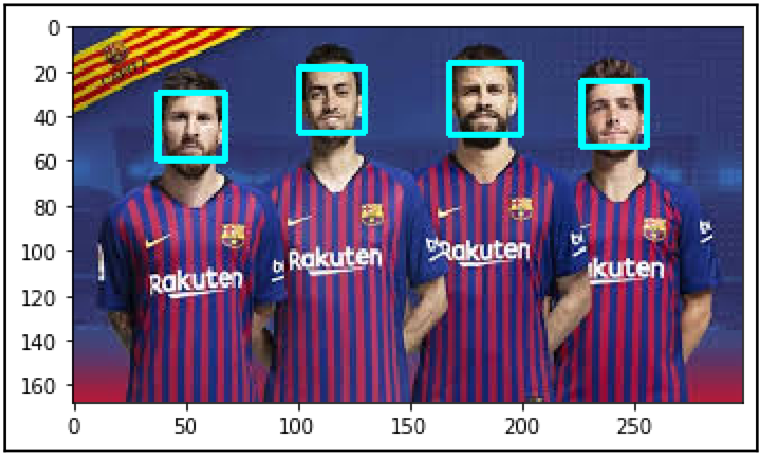

0
50
100
150
200
250
300
350
400
0
100
200
300
400
500
600


0
50
100
150
200
250
300
350
400
0
100
200
300
400
500
600

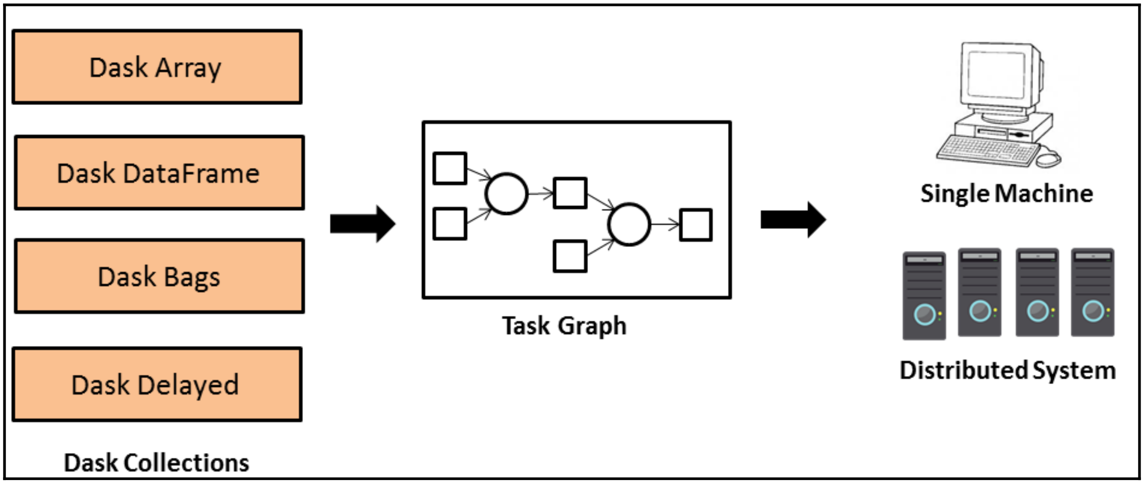

0
50
100
150
200
250
300
350
400
0
100
200
300
400
500
600

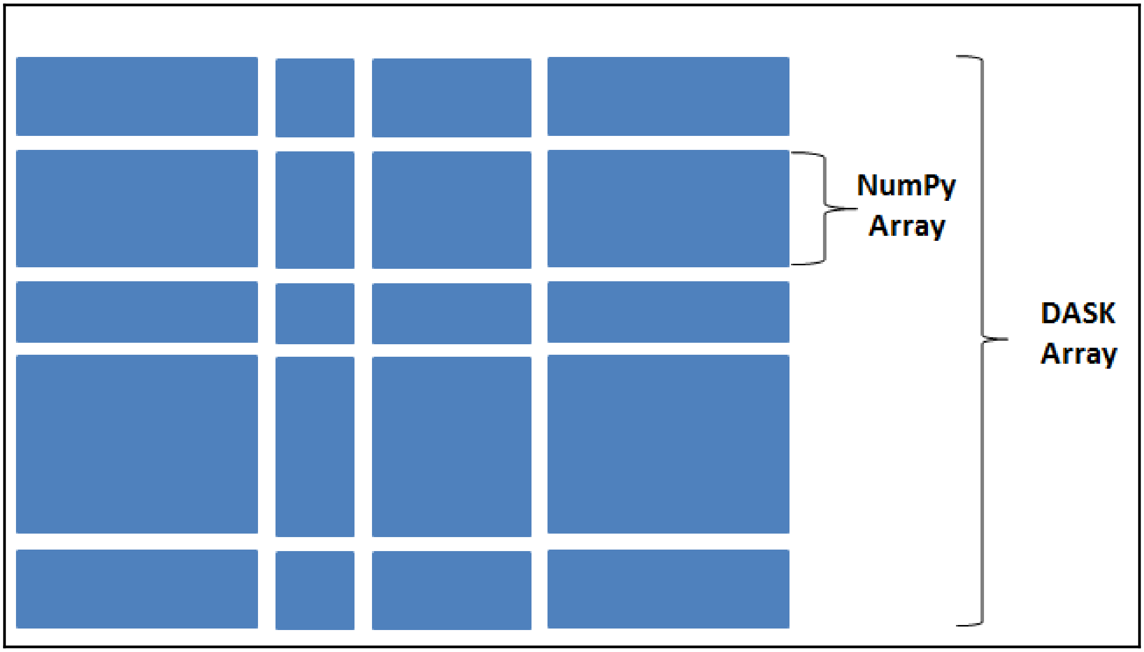

0
50
100
150
200
250
0
50
100
150

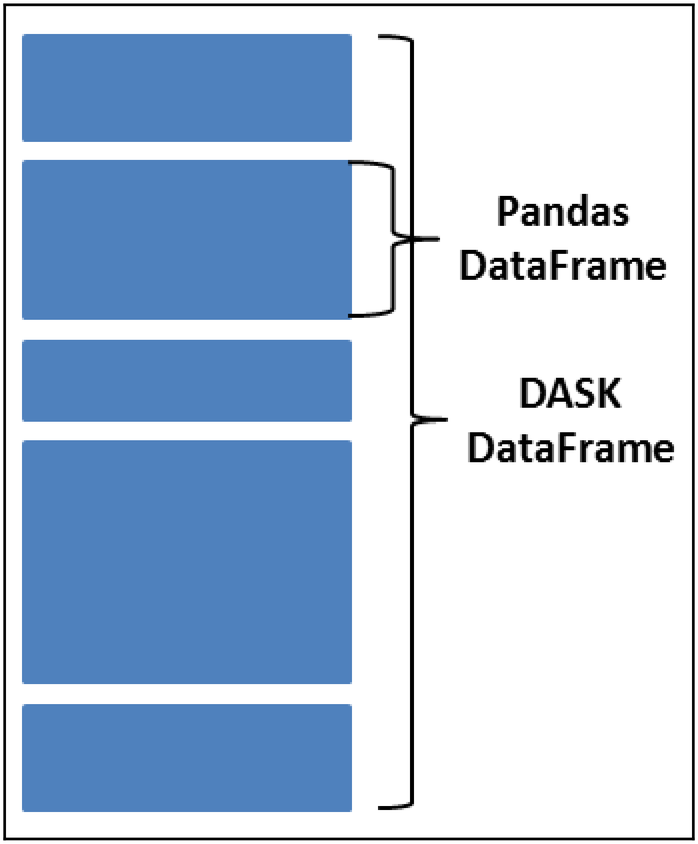

0
50
100
150
200
250
0
50
100
150


0
20
40
60
80
100
120
140
160
0
50
100
150
200
250
Rakuten
Rakuten
Rakuten
kuten

Chapter 14: Parallel Computing Using Dask





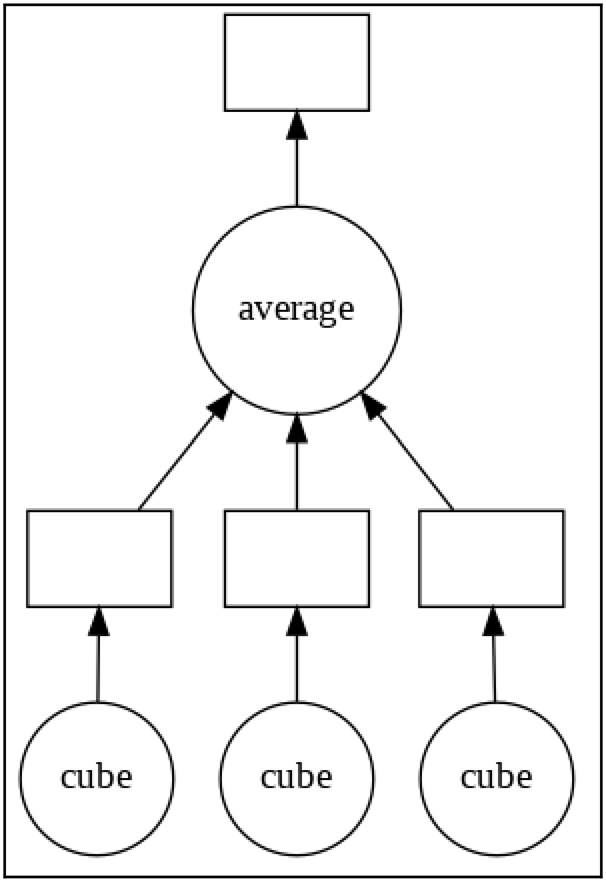


	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left
0	0.38	0.53	2	157	3	0	1
1	0.80	0.86	5	262	6	0	1
2	0.11	0.88	7	272	4	0	1
3	0.72	0.87	5	223	5	0	1
4	0.37	0.52	2	159	3	0	1

number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Departments	salary
2	157	3	0	1	0	sales	low
5	223	5	0	1	0	sales	low
2	159	3	0	1	0	sales	low
2	153	3	0	1	0	sales	low
6	247	4	0	1	0	sales	low

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident
left						
0	0.666810	0.715473	3.786664	199.060203	3.380032	0.175009
1	0.440098	0.718113	3.855503	207.419210	3.876505	0.047326

	item_name	price
0	Egg	5
1	Bread	20
0	Milk	54



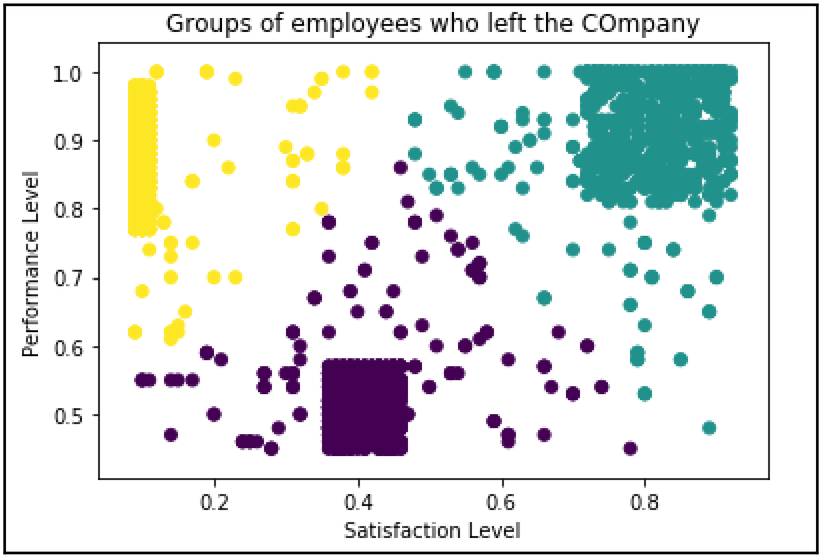
	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Departments
0	0.38	0.53	2	157	3	0	1	0	sales
1	0.80	0.86	5	262	6	0	1	0	sales
2	0.11	0.88	7	272	4	0	1	0	sales
3	0.72	0.87	5	223	5	0	1	0	sales
4	0.37	0.52	2	159	3	0	1	0	sales

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Departments
0	0.38	0.53	2	157	3	0	1	0	sales
1	0.80	0.86	5	262	6	0	1	0	sales
2	0.11	0.88	7	272	4	0	1	0	sales
3	0.72	0.87	5	223	5	0	1	0	sales
4	0.37	0.52	2	159	3	0	1	0	sales

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Departments
0	0.38	0.53	2	157	3	0	1	0	sales
1	0.80	0.86	5	262	6	0	1	0	sales
2	0.11	0.88	7	272	4	0	1	0	sales
3	0.72	0.87	5	223	5	0	1	0	sales
4	0.37	0.52	2	159	3	0	1	0	sales

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Departments
0	0.38	0.53	2	157	3	0	1	0	sales
1	0.80	0.86	5	262	6	0	1	0	sales
2	0.11	0.88	7	272	4	0	1	0	sales
3	0.72	0.87	5	223	5	0	1	0	sales
4	0.37	0.52	2	159	3	0	1	0	sales

	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Departments
0	0.38	0.53	2	157	3	0	1	0	sales
1	0.80	0.86	5	262	6	0	1	0	sales
2	0.11	0.88	7	272	4	0	1	0	sales
3	0.72	0.87	5	223	5	0	1	0	sales
4	0.37	0.52	2	159	3	0	1	0	sales


Groups of employees who left the COmpany
Performance Level
1.0
0.9
0.8
0.7
0.6
0.5
0.2
0.4
0.6
0.8
Satisfaction Level