

Exercise 1: Pandas and Numpy

```
In [1]: #Importing the required libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import matplotlib.pyplot as plt
```

Matrix Multiplication

To create a numpy matrix X of dimensions n x m, where n = 100 and m = 20. Initialize Matrix X with random values. Then to create a numpy vector Y of dimensions m x 1. Initialize the vector Y with values from a normal distribution using $\mu = 2$ and $\sigma = 0.01$

```
In [2]: X=np.random.randint(10,size =(100,20)) # Initializing X with random values

#Initializing vector Y with the mean and std given
Y=np.random.normal(2,0.01,(20,1))
print("Y vector=",Y)
```

```
Y vector= [[2.01000581]
 [2.00726404]
 [1.98657501]
 [2.01951683]
 [1.99636613]
 [1.98793939]
 [1.98743152]
 [1.99813393]
 [2.00009748]
 [2.00239709]
 [2.00766363]
 [1.98906483]
 [1.97472742]
 [1.97709336]
 [1.99021826]
 [1.98266885]
 [1.99235989]
 [2.0068948 ]
 [1.98881392]
 [2.00478144]]
```

```
In [3]: print("X matrix=",X)
print("XnShape of X":, X.shape)
```

```
X matrix= [[7 1 8 ... 0 1 9]
 [2 4 9 ... 4 2 8]
 [6 2 ... 9 7 0]
 ...
 [7 1 2 ... 3 9 1]
 [2 4 5 ... 7 7 0]
 [1 8 2 ... 7 6 1]]
```

Shape of X: (100, 20)

To iteratively multiply (element-wise) each row of the matrix X with vector Y and sum the result of each iteration in another vector C.

```
In [4]: C= np.zeros(100) #Initialize vector C
```

```
In [5]: for i in range(100):                #for each row in matrix
sum=0
    for j in range(20):                #for each column in the row
sum += X[i][j]* Y[j]                # multiplying the required elements and adding them to the previous value to get sum
C[i] = sum                            # storing the sum of the element wise multiplication of a row
```

```
In [6]: print("Value of Vector C: ", C)
print("XnShape of Vector C: ",C.shape)
```

```
Value of Vector C: [169.53826628 179.47048188 179.87307917 197.79865232 193.3427728
169.59957395 143.65645206 191.57583815 197.55025403 163.61980952
193.89712075 183.48840336 157.62613832 187.65068285 235.5222964
225.32873432 129.72596375 197.38872024 167.68677748 163.3308359
159.58569034 161.61832615 205.78520145 219.70539728 231.54898159
163.52403256 181.43324894 151.49153916 189.71152758 187.50560027
197.65753738 163.30770068 211.54388739 177.31304166 201.76864024
171.10098336 171.41487113 141.63395496 203.90382745 195.51118964
169.57177795 191.48877752 201.56327364 193.69237433 177.69313089
187.43336237 199.61501659 161.56966324 175.40113267 201.56893834
187.47860964 147.67260725 197.57866356 187.44870576 227.60922086
151.62432979 177.68565965 177.4577411 163.60066097 159.37342826
191.31559879 177.66330284 177.49977323 141.59172136 189.30276042
161.84783869 161.44213589 221.63728464 139.64235813 167.59589802
207.5844341 181.57412204 179.50236655 187.57287835 205.511776
195.68177793 177.76134528 181.52397667 191.59175957 185.7443218
163.33796465 197.63073902 205.61437728 171.65100274 141.63215605
197.7230497 155.7030299 195.63512215 153.71877722 169.60863844
229.69191531 157.565221 177.6542577 165.58663796 117.70904857
173.86317469 197.58873911 161.61126047 151.59724585 143.7130126 ]
```

Shape of vector C: (100,)

To find mean and standard deviation of the new vector c. Here I am using the numpy functions for the same.

```
In [7]: mean = np.mean(C)                #Finds average of all the values in the vector
print("Mean of vector C:",mean)
```

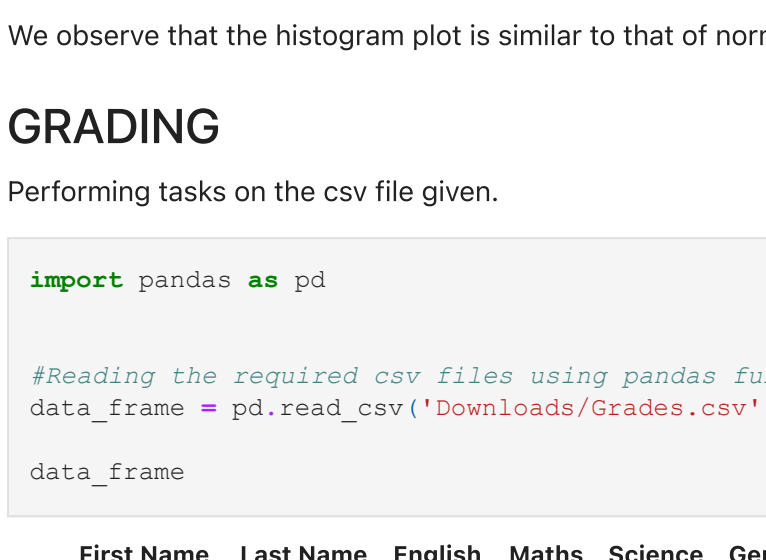
Mean of vector C: 179.44284313045046

```
In [8]: std = np.std(C)                 # Finds the variation of the values in the vector from its mean value
print("Standard Deviation of vector C:",std)
```

Standard Deviation of vector C: 23.355834124633724

Plotting the histogram of vector C using 5 bins

```
In [9]: plt.hist(C, edgecolor="red", bins=5)
plt.title("Histogram of vector C")
plt.xlabel("The range of values in vector")
plt.ylabel("The number of values ")
plt.show()
```



We observe that the histogram plot is similar to that of normal distribution.

GRADING

Performing tasks on the csv file given.

```
In [10]: import pandas as pd

#Reading the required csv files using pandas functionality
data_frame = pd.read_csv('Downloads/Grades.csv')
data_frame
```

	First Name	Last Name	English	Maths	Science	German	Sports	Final Grade
0	Robyn	Hobgood	60.95	24.77	20.60	69.32	8.36	184.00
1	Eddy	Sweargin	100.00	12.99	100.00	52.24	100.00	365.23
2	Leoma	Bridgman	83.37	100.00	78.69	100.00	19.50	381.56
3	Arnetta	Pearl	87.75	100.00	86.93	87.90	41.73	404.31
4	Maryland	Colby	100.00	100.00	100.00	18.87	88.72	407.59
5	Sherron	Sherron	92.06	55.91	93.93	-56.74	77.71	262.87
6	Glendora	Christopher	78.26	100.00	25.60	100.00	100.00	403.86
7	Darlana	Gunn	100.00	64.53	100.00	23.21	79.01	366.75
8	Aldo	Armas	100.00	83.49	100.00	100.00	92.32	475.81
9	Tiny	Jack	94.35	33.09	82.57	31.13	100.00	341.14
10	Carlton	Elms	100.00	36.52	5.54	33.82	12.07	187.95
11	Lauretta	Herbert	50.73	-0.10	67.76	100.00	55.98	274.37
12	Almeta	Dimond	80.37	100.00	69.02	100.00	79.62	429.01
13	Phoebe	Schill	100.00	70.37	100.00	47.00	77.37	394.74
14	Krystyna	Paris	18.75	73.80	87.00	59.30	100.00	338.85
15	Miyoko	Laffoon	100.00	100.00	100.00	34.98	94.55	429.53
16	Rebecca	Duck	70.79	97.81	52.25	19.76	-13.93	226.68
17	Elwanda	Hyland	45.69	74.86	43.10	45.00	76.72	285.37
18	Gretchen	Kerrick	68.70	75.87	-14.87	57.32	84.28	271.30
19	Winnifred	Colonna	83.79	100.00	85.66	3.94	100.00	373.39
20	Gidget	Casseus	100.00	100.00	100.00	99.52	62.59	462.11
21	Elaina	Mcdougal	100.00	100.00	80.29	100.00	-36.81	343.48
22	Shoshana	Goldberger	55.10	100.00	100.00	100.00	100.00	455.10
23	Argentina	Nelson	100.00	100.00	40.44	100.00	100.00	440.44
24	Lyle	Millsaps	100.00	71.88	100.00	-17.33	100.00	354.55
25	Janay	Julius	41.70	100.00	55.90	75.53	100.00	373.13
26	Deborah	Heyden	0.84	18.61	50.14	83.58	65.68	218.85
27	Thelma	Romberger	72.41	61.00	100.00	76.38	51.37	361.16
28	Armanda	Hendley	35.40	66.92	69.67	71.08	-2.34	240.73
29	Raymon	Myerson	49.83	27.36	61.90	72.97	13.11	225.17

- Compute the sum for all subjects for each student.

```
In [11]: #Finding sum along the subject columns
data_frame['sum'] = data_frame[['English','Science','German','Sports',' Maths']].sum(axis=1)
```

```
In [12]: data_frame
```

	First Name	Last Name	English	Maths	Science	German	Sports	Final Grade	sum
0	Robyn	Hobgood	60.95	24.77	20.60	69.32	8.36	184.00	184.00
1	Eddy	Sweargin	100.00	12.99	100.00	52.24	100.00	365.23	365.23
2	Leoma	Bridgman	83.37	100.00	78.69	100.00	19.50	381.56	381.56
3	Arnetta	Pearl	87.75	100.00	86.93	87.90	41.73	404.31	404.31
4	Maryland	Colby	100.00	100.00	100.00	18.87	88.72	407.59	407.59
5	Sherron	Sherron	92.06	55.91	93.93	-56.74	77.71	262.87	262.87
6	Glendora	Christopher	78.26	100.00	25.60	100.00	100.00	403.86	403.86
7	Darlana	Gunn	100.00	64.53	100.00	23.21	79.01	366.75	366.75
8	Aldo	Armas	100.00	83.49	100.00	100.00	92.32	475.81	475.81
9	Tiny	Jack	94.35	33.09	82.57	31.13	100.00	341.14	341.14
10	Carlton	Elms	100.00	36.52	5.54	33.82	12.07	187.95	187.95
11	Lauretta	Herbert	50.73	-0.10	67.76	100.00	55.98	274.37	274.37
12	Almeta	Dimond	80.37	100.00	69.02	100.00	79.62	429.01	429.01
13	Phoebe	Schill	100.00	70.37	100.00	47.00	77.37	394.74	394.74
14	Krystyna	Paris	18.75	73.80	87.00	59.30	100.00	338.85	338.85
15	Miyoko	Laffoon	100.00	100.00	100.00	34.98	94.55	429.53	429.53
16	Rebecca	Duck	70.79	97.81	52.25	19.76	-13.93	226.68	226.68
17	Elwanda	Hyland	45.69	74.86	43.10	45.00	76.72	285.37	285.37
18	Gretchen	Kerrick	68.70	75.87	-14.87	57.32	84.28	271.30	271.30
19	Winnifred	Colonna	83.79	100.00	85.66	3.94	100.00	373.39	373.39
20	Gidget	Casseus	100.00	100.00	100.00	99.52	62.59	462.11	462.11
21	Elaina	Mcdougal	100.00	100.00	80.29	100.00	-36.81	343.48	343.48
22	Shoshana	Goldberger	55.10	100.00	100.00	100.00	100.00	455.10	455.10
23	Argentina	Nelson	100.00	100.00	40.44	100.00	100.00	440.44	440.44
24	Lyle	Millsaps	100.00	71.88	100.00	-17.33	100.00	354.55	354.55
25	Janay	Julius	41.70	100.00	55.90	75.53	100.00	373.13	373.13
26	Deborah	Heyden	0.84	18.61	50.14	83.58	65.68	218.85	218.85
27	Thelma	Romberger	72.41	61.00	100.00	76.38	51.37	361.16	361.16
28	Armanda	Hendley	35.40	66.92	69.67	71.08	-2.34	240.73	240.73
29	Raymon	Myerson	49.83	27.36	61.90	72.97	13.11	225.17	225.17

We observe that the column Final grade already contains the sum of all the subjects.

```
In [13]: data_frame.columns
```

```
Out[13]: Index(['First Name', 'Last Name', 'English', ' Maths', 'Science', 'German',
'Sports', 'Final Grade', 'sum'],
      dtype='object')
```

- Compute the average of the point for each student(Total= 500) - 5 subjects are there

```
In [14]: #Finding average of the subjects and assigning it to new column
data_frame['average'] = data_frame[['sum']]/5
```

```
In [15]: data_frame
```

	First Name	Last Name	English	Maths	Science	German	Sports	Final Grade	sum	average
0	Robyn	Hobgood	60.95	24.77	20.60	69.32	8.36	184.00	184.00	36.800
1	Eddy	Sweargin	100.00	12.99	100.00	52.24	100.00	365.23	365.23	73.046
2	Leoma	Bridgman	83.37	100.00	78.69	100.00	19.50	381.56	381.56	76.312
3	Arnetta	Pearl	87.75	100.00	86.93	87.90	41.73	404.31	404.31	80.862
4	Maryland	Colby	100.00	100.00	100.00	18.87	88.72	407.59	407.59	81.518
5	Sherron	Sherron	92.06	55.91	93.93	-56.74	77.71	262.87	262.87	52.574
6	Glendora	Christopher	78.26	100.00	25.60	100.00	100.00	403.86	403.86	80.772
7	Darlana	Gunn	100.00	64.53	100.00	23.21	79.01	366.75	366.75	73.350
8	Aldo	Armas	100.00	83.49	100.00	100.00	92.32	475.81	475.81	95.162
9	Tiny	Jack	94.35	33.09	82.57	31.13	100.00	341.14	341.14	68.228
10	Carlton	Elms	100.00	36.52	5.54	33.82	12.07	187.95	187.95	37.590
11	Lauretta	Herbert	50.73	-0.10	67.76	100.00	55.98	274.37	274.37	54.874
12	Almeta	Dimond	80.37	100.00	69.02	100.00	79.62	429.01	429.01	85.802
13	Phoebe	Schill	100.00	70.37	100.00	47.00	77.37	394.74	394.74	78.948
14	Krystyna	Paris	18.75	73.80	87.00	59.30	100.00	338.85	338.85	67.770
15	Miyoko	Laffoon	100.00	100.00	100.00	34.98	94.55	429.53	429.53	85.906
16	Rebecca	Duck	70.79	97.81	52.25	19.76	-13.93	226.68	226.68	45.336
17	Elwanda	Hyland	45.69	74.86	43.10	45.00	76.72	285.37	285.37	57.074
18	Gretchen	Kerrick	68.70	75.87	-14.87	57.32	84.28	271.30	271.30	54.260
19	Winnifred	Colonna	83.79	100.00	85.66	3.94	100.00	373.39	373.39	74.678
20	Gidget	Casseus	100.00	100.00	100.00	99.52	62.59	462.11	462.11	92.422
21	Elaina	Mcdougal	100.00	100.00	80.29	100.00	-36.81	343.48	343.48	68.696
22	Shoshana	Goldberger	55.10	100.00	100.00	100.00	100.00	455.10	455.10	91.020
23	Argentina	Nelson	100.00	100.00	40.44	100.00	100.00	440.44	440.44	88.088
24	Lyle	Millsaps	100.00	71.88	100.00	-17.33	100.00	354.55	354.55	70.910
25	Janay	Julius	41.70	100.00	55.90	75.53	100.00	373.13	373.13	