Assignment Number 1

Problem Statement:

Create your own dataset having 20 rows and 25 columns in a excel file and on created dataset:

1. Perform operations of loading data or doing feature engineering on data sets. Perform basic numerical computations like mean, median, range and creation of multi-dimensional arrays.
2. Perform data pre-processing tasks such as data cleaning, handling missing values, and data transformation.

In [25]:
```
#Importing pandas library and numpy library as pd and np resp so that we can use th
import pandas as pd
import numpy as np
file = pd.read_excel("hackathon_Excel.xlsx") #opened excel file i.e. dataframe usin
print(file)
```

	Registration ID	Participant Name	Email	Phone Number	\
0	RID001	Participant 1	participant1@gmail.com	91975463508	
1	RID002	Participant 2	participant2@gmail.com	96921638357	
2	RID003	Participant 3	participant3@gmail.com	98934404830	
3	RID004	Participant 4	participant4@gmail.com	93136397852	
4	RID005	Participant 5	participant5@gmail.com	93011267619	
5	RID006	Participant 6	participant6@gmail.com	97563969103	
6	RID007	Participant 7	participant7@gmail.com	93269634118	
7	RID008	Participant 8	participant8@gmail.com	91415821947	
8	RID009	Participant 9	participant9@gmail.com	95485571746	
9	RID010	Participant 10	participant10@gmail.com	92340810932	
10	RID011	Participant 11	participant11@gmail.com	97547603683	
11	RID012	Participant 12	participant12@gmail.com	91005001863	
12	RID013	Participant 13	participant13@gmail.com	98591183037	
13	RID014	Participant 14	participant14@gmail.com	93613189929	
14	RID015	Participant 15	participant15@gmail.com	96876230430	
15	RID016	Participant 16	participant16@gmail.com	95979420684	
16	RID017	Participant 17	participant17@gmail.com	94324094300	
17	RID018	Participant 18	participant18@gmail.com	98209935601	
18	RID019	Participant 19	participant19@gmail.com	96098708991	
19	RID020	Participant 20	participant20@gmail.com	92639968236	

	Team Name	Project Title	Technology Stack	College/University	\
0	Team 1	Project 1	Python, Flask	University 1	
1	Team 2	Project 2	JavaScript, React	University 2	
2	Team 3	Project 3	Java, Spring	University 3	
3	Team 4	Project 4	C++, Qt	University 4	
4	Team 5	Project 5	Python, Flask	University 5	
5	Team 6	Project 6	JavaScript, React	University 6	
6	Team 7	Project 7	Java, Spring	University 7	
7	Team 8	Project 8	C++, Qt	University 8	
8	Team 9	Project 9	Python, Flask	University 9	
9	Team 10	Project 10	JavaScript, React	University 10	
10	Team 11	Project 11	Java, Spring	University 11	
11	Team 12	Project 12	C++, Qt	University 12	
12	Team 13	Project 13	Python, Flask	University 13	
13	Team 14	Project 14	JavaScript, React	University 14	
14	Team 15	Project 15	Java, Spring	University 15	
15	Team 16	Project 16	C++, Qt	University 16	
16	Team 17	Project 17	Python, Flask	University 17	
17	Team 18	Project 18	JavaScript, React	University 18	
18	Team 19	Project 19	Java, Spring	University 19	
19	Team 20	Project 20	C++, Qt	University 20	

	Year of Study	Department	...	Participation Type	\
0	2nd Year	ME	...	Team	
1	3rd Year	ME	...	Team	
2	2nd Year	CE	...	Individual	
3	4th Year	CSE	...	Team	
4	3rd Year	ME	...	Individual	
5	2nd Year	CSE	...	Individual	
6	3rd Year	ME	...	Individual	
7	2nd Year	ECE	...	Team	
8	2nd Year	ME	...	Individual	
9	4th Year	ECE	...	Team	
10	4th Year	CSE	...	Individual	

```
11      3rd Year         CE  ...              Team
12      2nd Year         ME  ...        Individual
13      4th Year         ME  ...              Team
14      2nd Year        ECE  ...        Individual
15      1st Year        ECE  ...        Individual
16      2nd Year        ECE  ...              Team
17      3rd Year        CSE  ...              Team
18      1st Year         ME  ...        Individual
19      2nd Year         ME  ...        Individual

                     GitHub Profile                       LinkedIn Profile  \
0    https://github.com/participant1   https://linkedin.com/in/participant1
1    https://github.com/participant2   https://linkedin.com/in/participant2
2    https://github.com/participant3   https://linkedin.com/in/participant3
3    https://github.com/participant4   https://linkedin.com/in/participant4
4    https://github.com/participant5   https://linkedin.com/in/participant5
5    https://github.com/participant6   https://linkedin.com/in/participant6
6    https://github.com/participant7   https://linkedin.com/in/participant7
7    https://github.com/participant8   https://linkedin.com/in/participant8
8    https://github.com/participant9   https://linkedin.com/in/participant9
9   https://github.com/participant10  https://linkedin.com/in/participant10
10  https://github.com/participant11  https://linkedin.com/in/participant11
11  https://github.com/participant12  https://linkedin.com/in/participant12
12  https://github.com/participant13  https://linkedin.com/in/participant13
13  https://github.com/participant14  https://linkedin.com/in/participant14
14  https://github.com/participant15  https://linkedin.com/in/participant15
15  https://github.com/participant16  https://linkedin.com/in/participant16
16  https://github.com/participant17  https://linkedin.com/in/participant17
17  https://github.com/participant18  https://linkedin.com/in/participant18
18  https://github.com/participant19  https://linkedin.com/in/participant19
19  https://github.com/participant20  https://linkedin.com/in/participant20

                        Resume Link Preferred Language Dietary Preferences  \
0    https://example.com/resume1.pdf            Spanish               Vegan
1    https://example.com/resume2.pdf             French                 NaN
2    https://example.com/resume3.pdf             German               Vegan
3    https://example.com/resume4.pdf            English               Vegan
4    https://example.com/resume5.pdf            Spanish         Gluten-Free
5    https://example.com/resume6.pdf             German          Vegetarian
6    https://example.com/resume7.pdf             German         Gluten-Free
7    https://example.com/resume8.pdf             German                 NaN
8    https://example.com/resume9.pdf            Spanish                 NaN
9   https://example.com/resume10.pdf             French                 NaN
10  https://example.com/resume11.pdf            English                 NaN
11  https://example.com/resume12.pdf             French               Vegan
12  https://example.com/resume13.pdf            Spanish         Gluten-Free
13  https://example.com/resume14.pdf             French         Gluten-Free
14  https://example.com/resume15.pdf            English                 NaN
15  https://example.com/resume16.pdf            English         Gluten-Free
16  https://example.com/resume17.pdf            English               Vegan
17  https://example.com/resume18.pdf             German         Gluten-Free
18  https://example.com/resume19.pdf            Spanish          Vegetarian
19  https://example.com/resume20.pdf             French          Vegetarian

   Shirt Size Emergency Contact Name Emergency Contact Number  \
0          XL              Contact 1              97804948206
```

```
1          XL             Contact 2             94020355472
2           L             Contact 3             97001960593
3           S             Contact 4             99328258205
4           L             Contact 5             96783508421
5          XL             Contact 6             93405351786
6          XL             Contact 7             95723384555
7           L             Contact 8             98010467539
8           M             Contact 9             99651105048
9          XL            Contact 10             98022421191
10         XL            Contact 11             99950048028
11          M            Contact 12             96166275648
12          S            Contact 13             91481233124
13          M            Contact 14             97387434384
14          L            Contact 15             93301295426
15         XL            Contact 16             91329319672
16          S            Contact 17             99174257845
17          S            Contact 18             91487981811
18         XL            Contact 19             94175373517
19          L            Contact 20             91130348437

   Previous Participation
0                     Yes
1                      No
2                     Yes
3                      No
4                     Yes
5                     Yes
6                      No
7                     Yes
8                     Yes
9                     Yes
10                     No
11                    Yes
12                    Yes
13                     No
14                    Yes
15                     No
16                     No
17                     No
18                     No
19                    Yes

[20 rows x 25 columns]
```

Descriptions for each column in a hackathon registration database:

1. **Registration ID:** A unique identifier for each participant's registration.
2. **Participant Name:** The full name of the participant.
3. **Email:** The email address of the participant.
4. **Phone Number:** The contact phone number of the participant.
5. **Team Name:** The name of the participant's team.

6. **Project Title:** The title of the project submitted by the participant or team.
7. **Technology Stack:** The technologies used in the participant's project.
8. **College/University:** The name of the college or university the participant is attending.
9. **Year of Study:** The academic year the participant is currently in.
10. **Department:** The academic department the participant belongs to.
11. **City:** The city where the participant's college or university is located.
12. **State:** The state where the participant's college or university is located.
13. **Country:** The country where the participant's college or university is located.
14. **Date of Registration:** The date when the participant registered for the event.
15. **Time of Registration:** The time when the participant registered for the event.
16. **Participation Type:** Indicates whether the participant is participating individually or as part of a team.
17. **GitHub Profile:** The URL to the participant's GitHub profile.
18. **LinkedIn Profile:** The URL to the participant's LinkedIn profile.
19. **Resume Link:** The URL to the participant's resume.
20. **Preferred Language:** The language preferred by the participant for communication.
21. **Dietary Preferences:** Any dietary restrictions or preferences of the participant.
22. **Shirt Size:** The participant's shirt size.
23. **Emergency Contact Name:** The name of the participant's emergency contact.
24. **Emergency Contact Number:** The phone number of the participant's emergency contact.
25. **Previous Participation:** Indicates whether the participant has participated in the event before.

In [26]:
```
file.head() #.head() returns the first few rows (the "head" of the dataframe i.e. e
```

Out[26]:

	Registration ID	Participant Name	Email	Phone Number	Team Name	Project Title	Technolo Sta
0	RID001	Participant 1	participant1@gmail.com	91975463508	Team 1	Project 1	Pytho Fla
1	RID002	Participant 2	participant2@gmail.com	96921638357	Team 2	Project 2	JavaScri Rea
2	RID003	Participant 3	participant3@gmail.com	98934404830	Team 3	Project 3	Java, Spri
3	RID004	Participant 4	participant4@gmail.com	93136397852	Team 4	Project 4	C++,
4	RID005	Participant 5	participant5@gmail.com	93011267619	Team 5	Project 5	Pytho Fla

5 rows × 25 columns

In [27]:
```
file.tail() #.tail() method returns a specified number of last row
```

Out[27]:

	Registration ID	Participant Name	Email	Phone Number	Team Name	Project Title	Techno
15	RID016	Participant 16	participant16@gmail.com	95979420684	Team 16	Project 16	C+
16	RID017	Participant 17	participant17@gmail.com	94324094300	Team 17	Project 17	Py
17	RID018	Participant 18	participant18@gmail.com	98209935601	Team 18	Project 18	JavaS
18	RID019	Participant 19	participant19@gmail.com	96098708991	Team 19	Project 19	Java, S
19	RID020	Participant 20	participant20@gmail.com	92639968236	Team 20	Project 20	C+

5 rows × 25 columns

In [28]:
```
file.describe() #.describe() calculates a few summary statistics for each column
```

Out[28]:

	Phone Number	Emergency Contact Number
count	2.000000e+01	2.000000e+01
mean	9.494702e+10	9.576677e+10
std	2.561872e+09	3.005306e+09
min	9.100500e+10	9.113035e+10
25%	9.291844e+10	9.337934e+10
50%	9.490483e+10	9.647489e+10
75%	9.707813e+10	9.801346e+10
max	9.893440e+10	9.995005e+10

In [29]:
```
file.info() #.info() shows information on each of the columns such as the data type
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 25 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   Registration ID           20 non-null     object
 1   Participant Name          20 non-null     object
 2   Email                     20 non-null     object
 3   Phone Number              20 non-null     int64 
 4   Team Name                 20 non-null     object
 5   Project Title             20 non-null     object
 6   Technology Stack          20 non-null     object
 7   College/University        20 non-null     object
 8   Year of Study             20 non-null     object
 9   Department                20 non-null     object
 10  City                      20 non-null     object
 11  State                     20 non-null     object
 12  Country                   20 non-null     object
 13  Date of Registration      20 non-null     object
 14  Time of Registration      20 non-null     object
 15  Participation Type        20 non-null     object
 16  GitHub Profile            20 non-null     object
 17  LinkedIn Profile          20 non-null     object
 18  Resume Link               20 non-null     object
 19  Preferred Language        20 non-null     object
 20  Dietary Preferences       14 non-null     object
 21  Shirt Size                20 non-null     object
 22  Emergency Contact Name    20 non-null     object
 23  Emergency Contact Number  20 non-null     int64 
 24  Previous Participation    20 non-null     object
dtypes: int64(2), object(23)
memory usage: 4.0+ KB
```

In [30]:
```
file.isnull().sum() #this will return the number of missing values in dataset
```

Out[30]:
```
Registration ID            0
Participant Name           0
Email                      0
Phone Number               0
Team Name                  0
Project Title              0
Technology Stack           0
College/University         0
Year of Study              0
Department                 0
City                       0
State                      0
Country                    0
Date of Registration       0
Time of Registration       0
Participation Type         0
GitHub Profile             0
LinkedIn Profile           0
Resume Link                0
Preferred Language         0
Dietary Preferences        6
Shirt Size                 0
Emergency Contact Name     0
Emergency Contact Number   0
Previous Participation     0
dtype: int64
```

In [31]:
```
#removing duplicates
cleaned_data = file.drop_duplicates()
print(cleaned_data)
```

	Registration ID	Participant Name	Email	Phone Number	\
0	RID001	Participant 1	participant1@gmail.com	91975463508	
1	RID002	Participant 2	participant2@gmail.com	96921638357	
2	RID003	Participant 3	participant3@gmail.com	98934404830	
3	RID004	Participant 4	participant4@gmail.com	93136397852	
4	RID005	Participant 5	participant5@gmail.com	93011267619	
5	RID006	Participant 6	participant6@gmail.com	97563969103	
6	RID007	Participant 7	participant7@gmail.com	93269634118	
7	RID008	Participant 8	participant8@gmail.com	91415821947	
8	RID009	Participant 9	participant9@gmail.com	95485571746	
9	RID010	Participant 10	participant10@gmail.com	92340810932	
10	RID011	Participant 11	participant11@gmail.com	97547603683	
11	RID012	Participant 12	participant12@gmail.com	91005001863	
12	RID013	Participant 13	participant13@gmail.com	98591183037	
13	RID014	Participant 14	participant14@gmail.com	93613189929	
14	RID015	Participant 15	participant15@gmail.com	96876230430	
15	RID016	Participant 16	participant16@gmail.com	95979420684	
16	RID017	Participant 17	participant17@gmail.com	94324094300	
17	RID018	Participant 18	participant18@gmail.com	98209935601	
18	RID019	Participant 19	participant19@gmail.com	96098708991	
19	RID020	Participant 20	participant20@gmail.com	92639968236	

	Team Name	Project Title	Technology Stack	College/University	\
0	Team 1	Project 1	Python, Flask	University 1	
1	Team 2	Project 2	JavaScript, React	University 2	
2	Team 3	Project 3	Java, Spring	University 3	
3	Team 4	Project 4	C++, Qt	University 4	
4	Team 5	Project 5	Python, Flask	University 5	
5	Team 6	Project 6	JavaScript, React	University 6	
6	Team 7	Project 7	Java, Spring	University 7	
7	Team 8	Project 8	C++, Qt	University 8	
8	Team 9	Project 9	Python, Flask	University 9	
9	Team 10	Project 10	JavaScript, React	University 10	
10	Team 11	Project 11	Java, Spring	University 11	
11	Team 12	Project 12	C++, Qt	University 12	
12	Team 13	Project 13	Python, Flask	University 13	
13	Team 14	Project 14	JavaScript, React	University 14	
14	Team 15	Project 15	Java, Spring	University 15	
15	Team 16	Project 16	C++, Qt	University 16	
16	Team 17	Project 17	Python, Flask	University 17	
17	Team 18	Project 18	JavaScript, React	University 18	
18	Team 19	Project 19	Java, Spring	University 19	
19	Team 20	Project 20	C++, Qt	University 20	

	Year of Study	Department	...	Participation Type	\
0	2nd Year	ME	...	Team	
1	3rd Year	ME	...	Team	
2	2nd Year	CE	...	Individual	
3	4th Year	CSE	...	Team	
4	3rd Year	ME	...	Individual	
5	2nd Year	CSE	...	Individual	
6	3rd Year	ME	...	Individual	
7	2nd Year	ECE	...	Team	
8	2nd Year	ME	...	Individual	
9	4th Year	ECE	...	Team	
10	4th Year	CSE	...	Individual	

```
11      3rd Year         CE  ...              Team
12      2nd Year         ME  ...        Individual
13      4th Year         ME  ...              Team
14      2nd Year        ECE  ...        Individual
15      1st Year        ECE  ...        Individual
16      2nd Year        ECE  ...              Team
17      3rd Year        CSE  ...              Team
18      1st Year         ME  ...        Individual
19      2nd Year         ME  ...        Individual

                     GitHub Profile                       LinkedIn Profile  \
0    https://github.com/participant1   https://linkedin.com/in/participant1
1    https://github.com/participant2   https://linkedin.com/in/participant2
2    https://github.com/participant3   https://linkedin.com/in/participant3
3    https://github.com/participant4   https://linkedin.com/in/participant4
4    https://github.com/participant5   https://linkedin.com/in/participant5
5    https://github.com/participant6   https://linkedin.com/in/participant6
6    https://github.com/participant7   https://linkedin.com/in/participant7
7    https://github.com/participant8   https://linkedin.com/in/participant8
8    https://github.com/participant9   https://linkedin.com/in/participant9
9   https://github.com/participant10  https://linkedin.com/in/participant10
10  https://github.com/participant11  https://linkedin.com/in/participant11
11  https://github.com/participant12  https://linkedin.com/in/participant12
12  https://github.com/participant13  https://linkedin.com/in/participant13
13  https://github.com/participant14  https://linkedin.com/in/participant14
14  https://github.com/participant15  https://linkedin.com/in/participant15
15  https://github.com/participant16  https://linkedin.com/in/participant16
16  https://github.com/participant17  https://linkedin.com/in/participant17
17  https://github.com/participant18  https://linkedin.com/in/participant18
18  https://github.com/participant19  https://linkedin.com/in/participant19
19  https://github.com/participant20  https://linkedin.com/in/participant20

                        Resume Link Preferred Language Dietary Preferences  \
0    https://example.com/resume1.pdf            Spanish               Vegan
1    https://example.com/resume2.pdf             French                 NaN
2    https://example.com/resume3.pdf             German               Vegan
3    https://example.com/resume4.pdf            English               Vegan
4    https://example.com/resume5.pdf            Spanish         Gluten-Free
5    https://example.com/resume6.pdf             German          Vegetarian
6    https://example.com/resume7.pdf             German         Gluten-Free
7    https://example.com/resume8.pdf             German                 NaN
8    https://example.com/resume9.pdf            Spanish                 NaN
9   https://example.com/resume10.pdf             French                 NaN
10  https://example.com/resume11.pdf            English                 NaN
11  https://example.com/resume12.pdf             French               Vegan
12  https://example.com/resume13.pdf            Spanish         Gluten-Free
13  https://example.com/resume14.pdf             French         Gluten-Free
14  https://example.com/resume15.pdf            English                 NaN
15  https://example.com/resume16.pdf            English         Gluten-Free
16  https://example.com/resume17.pdf            English               Vegan
17  https://example.com/resume18.pdf             German         Gluten-Free
18  https://example.com/resume19.pdf            Spanish          Vegetarian
19  https://example.com/resume20.pdf             French          Vegetarian

   Shirt Size Emergency Contact Name Emergency Contact Number  \
0          XL              Contact 1              97804948206
```

```
1          XL             Contact 2             94020355472
2           L             Contact 3             97001960593
3           S             Contact 4             99328258205
4           L             Contact 5             96783508421
5          XL             Contact 6             93405351786
6          XL             Contact 7             95723384555
7           L             Contact 8             98010467539
8           M             Contact 9             99651105048
9          XL            Contact 10             98022421191
10         XL            Contact 11             99950048028
11          M            Contact 12             96166275648
12          S            Contact 13             91481233124
13          M            Contact 14             97387434384
14          L            Contact 15             93301295426
15         XL            Contact 16             91329319672
16          S            Contact 17             99174257845
17          S            Contact 18             91487981811
18         XL            Contact 19             94175373517
19          L            Contact 20             91130348437

   Previous Participation
0                     Yes
1                      No
2                     Yes
3                      No
4                     Yes
5                     Yes
6                      No
7                     Yes
8                     Yes
9                     Yes
10                     No
11                    Yes
12                    Yes
13                     No
14                    Yes
15                     No
16                     No
17                     No
18                     No
19                    Yes

[20 rows x 25 columns]
```

In [32]:
```
file.shape #.shape shows the number of rows and columns of the dataframe/excel
```

Out[32]: (20, 25)

In [33]:
```
file["New Column"] = range(1, 21) #using range funcn to add new column which will h
print(file["New Column"])
```

```
0      1
1      2
2      3
3      4
4      5
5      6
6      7
7      8
8      9
9     10
10    11
11    12
12    13
13    14
14    15
15    16
16    17
17    18
18    19
19    20
Name: New Column, dtype: int64
```

In [34]:
```
d = pd.DataFrame(file) #The `DataFrame` function in pandas is used to create a 2-di
```

In [35]:
```
nums = d["New Column"] #Here we have accessed the phone number column of our datase
print(nums)
```

```
0      1
1      2
2      3
3      4
4      5
5      6
6      7
7      8
8      9
9     10
10    11
11    12
12    13
13    14
14    15
15    16
16    17
17    18
18    19
19    20
Name: New Column, dtype: int64
```

In [36]:
```
mean = nums.mean()
print(mean)
```

```
10.5
```

In [37]:
```
mode = nums.mode()
print(mode)
```

```
0      1
1      2
2      3
3      4
4      5
5      6
6      7
7      8
8      9
9     10
10    11
11    12
12    13
13    14
14    15
15    16
16    17
17    18
18    19
19    20
Name: New Column, dtype: int64
```

In [38]:
```
median = nums.median()
print(median)
```

```
10.5
```

In [39]:
```
maximum_value = nums.max()
print(maximum_value)
```

```
20
```

In [40]:
```
minimum_value = nums.min()
print(minimum_value)
```

```
1
```

In [41]:
```
range_a  = maximum_value - minimum_value
print(range_a)
```

```
19
```

In [42]:
```
#transforming catergorical data to numerical/binary
from sklearn.preprocessing import StandardScaler, LabelEncoder
data =  d['Participant Name']
label_encoder = LabelEncoder()
encoded_data = label_encoder.fit_transform(data)
print(encoded_data)
```

```
[ 0 11 13 14 15 16 17 18 19  1  2  3  4  5  6  7  8  9 10 12]
```

In [43]:
```
array_2d = np.array([[1,4,5], [4,5,6]])
print(array_2d)
```

```
[[1 4 5]
 [4 5 6]]
```

In [44]:
```
array_3d = np.array([[[1, 2, 3, 4], [5, 6, 7, 8], [9, 10, 11, 12]],
                     [[13, 14, 15, 16], [17, 18, 19, 20], [21, 22, 23, 24]]])
```

```
print(array_3d)
```

```
[[[ 1  2  3  4]
  [ 5  6  7  8]
  [ 9 10 11 12]]

 [[13 14 15 16]
  [17 18 19 20]
  [21 22 23 24]]]
```

In [45]:
```
zeros_array = np.zeros((2, 3))
print(zeros_array)
```

```
[[0. 0. 0.]
 [0. 0. 0.]]
```

In [46]:
```
ones_array = np.ones((2,2))
print(ones_array)
```

```
[[1. 1.]
 [1. 1.]]
```