I

In [1]:
```python
import numpy as np
import pandas as pd
```

In [2]:
```python
df=pd.read_csv(r"C:\Users\Admin\Downloads\2015 - 2015.csv")
df
```

Out[2]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Fre
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.6
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.6
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.6
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.6
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.6
...	...	...	...	...	...	...	...	...	
153	Rwanda	Sub-Saharan Africa	154	3.465	0.03464	0.22208	0.77370	0.42864	0.5
154	Benin	Sub-Saharan Africa	155	3.340	0.03656	0.28665	0.35386	0.31910	0.4
155	Syria	Middle East and Northern Africa	156	3.006	0.05015	0.66320	0.47489	0.72193	0.1
156	Burundi	Sub-Saharan Africa	157	2.905	0.08658	0.01530	0.41587	0.22396	0.1
157	Togo	Sub-Saharan Africa	158	2.839	0.06727	0.20868	0.13995	0.28443	0.3

158 rows × 12 columns

In [3]:
```python
df1=df[['Family','Freedom']]
df1
```

Out[3]:

	Family	Freedom
0	1.34951	0.66557
1	1.40223	0.62877
2	1.36058	0.64938
3	1.33095	0.66973
4	1.32261	0.63297
...	...	...
153	0.77370	0.59201
154	0.35386	0.48450
155	0.47489	0.15684
156	0.41587	0.11850
157	0.13995	0.36453

158 rows × 2 columns

Find mean, median, mode and describe

In [4]:
```python
print(df1.mean())
```

```
Family     0.991046
Freedom    0.428615
dtype: float64
```

In [5]:
```python
print(df1.median())
```

```
Family     1.029510
Freedom    0.435515
dtype: float64
```

```
In [6]: print(df1.mode())
```

```
        Family  Freedom
0      0.00000  0.00000
1      0.13995  0.07699
2      0.30285  0.09245
3      0.35386  0.10081
4      0.38174  0.10384
..         ...      ...
153    1.34043  0.65821
154    1.34951  0.65980
155    1.36058  0.66246
156    1.36948  0.66557
157    1.40223  0.66973

[158 rows x 2 columns]
```

```
In [7]: print(df1.describe())
```

```
            Family      Freedom
count   158.000000   158.000000
mean      0.991046     0.428615
std       0.272369     0.150693
min       0.000000     0.000000
25%       0.856823     0.328330
50%       1.029510     0.435515
75%       1.214405     0.549092
max       1.402230     0.669730
```

Find sum(), cumsum(), count, min and max values

```
In [8]: print(df1.sum())
```

```
Family     156.58526
Freedom     67.72116
dtype: float64
```

```
In [9]: print(df1.cumsum())
```

```
         Family    Freedom
0       1.34951    0.66557
1       2.75174    1.29434
2       4.11232    1.94372
3       5.44327    2.61345
4       6.76588    3.24642
..          ...        ...
153   155.20069   66.59679
154   155.55455   67.08129
155   156.02944   67.23813
156   156.44531   67.35663
157   156.58526   67.72116

[158 rows x 2 columns]
```

```
In [10]: print(df1.min())
```

```
Family     0.0
Freedom    0.0
dtype: float64
```

```
In [11]: print(df1.max())
```

```
Family     1.40223
Freedom    0.66973
dtype: float64
```

Find covariance and correlation (spearman and pearsons)

```
In [12]: print(df1.cov())
```

```
           Family    Freedom
Family   0.074185   0.018122
Freedom  0.018122   0.022708
```

II

```
In [15]: import numpy as np
         import pandas as pd
```

In [16]:
```python
df2=pd.read_csv(r"C:\Users\Admin\Downloads\fiat500_VehicleSelection_Dataset - f
df2
```

Out[16]:

	ID	model	engine_power	age_in_days	km	previous_owners	lat	l
0	1.0	lounge	51.0	882.0	25000.0	1.0	44.907242	8.6115598
1	2.0	pop	51.0	1186.0	32500.0	1.0	45.666359	12.2418889
2	3.0	sport	74.0	4658.0	142228.0	1.0	45.503300	11.417
3	4.0	lounge	51.0	2739.0	160000.0	1.0	40.633171	17.6346092
4	5.0	pop	73.0	3074.0	106880.0	1.0	41.903221	12.4956502
...	...	...	...	...	...	...	...	
1544	NaN	NaN	NaN	NaN	NaN	NaN	NaN	leng
1545	NaN	NaN	NaN	NaN	NaN	NaN	NaN	cond
1546	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Null valu
1547	NaN	NaN	NaN	NaN	NaN	NaN	NaN	fi
1548	NaN	NaN	NaN	NaN	NaN	NaN	NaN	sear

1549 rows × 11 columns

In [17]:
```python
df3=df2[['age_in_days','price']]
df3
```

Out[17]:

	age_in_days	price
0	882.0	8900
1	1186.0	8800
2	4658.0	4200
3	2739.0	6000
4	3074.0	5700
...	...	...
1544	NaN	5
1545	NaN	lonprice
1546	NaN	NO
1547	NaN	1
1548	NaN	1

1549 rows × 2 columns

Find mean, median, mode and describe¶

In [18]: `print(df3.mean())`

```
age_in_days    1650.980494
dtype: float64
```

In [19]: `print(df3.median())`

```
age_in_days    1035.0
dtype: float64
```

In [20]: `print(df3.mode())`

```
   age_in_days  price
0        366.0  10500
1        790.0    NaN
```

In [21]: `print(df3.describe())`

```
       age_in_days
count  1538.000000
mean   1650.980494
std    1289.522278
min     366.000000
25%     670.000000
50%    1035.000000
75%    2616.000000
max    4658.000000
```

Find sum(), cumsum(), count, min and max values

In [22]: `print(df3.sum())`

```
age_in_days                                            2.53921e+06
price          89008800420060005700790010750919056006000089501...
dtype: object
```

In [23]: `print(df3.cumsum())`

```
      age_in_days                                                          price
0            882.0                                                           8900
1           2068.0                                                       89008800
2           6726.0                                                   890088004200
3           9465.0                                               8900880042006000
4          12539.0                                           89008800420060005700
...             ...                                                            ...
1544           NaN   89008800420060005700790010750919056006000089501...
1545           NaN   89008800420060005700790010750919056006000089501...
1546           NaN   89008800420060005700790010750919056006000089501...
1547           NaN   89008800420060005700790010750919056006000089501...
1548           NaN   89008800420060005700790010750919056006000089501...

[1549 rows x 2 columns]
```

In [24]: `print(df3.min())`

```
age_in_days     366.0
price             1.0
dtype: float64
```

In [25]: `print(df3.max())`

```
age_in_days        4658
price          lonprice
dtype: object
```

Find covariance and correlation (spearman and pearsons)

In [26]: `print(df3.cov())`

```
              age_in_days
age_in_days   1.662868e+06
```

In [27]:
```
from scipy.stats import pearsonr
print(pearsonr(d1,d2))
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-27-4232186c37ad> in <module>
      1 from scipy.stats import pearsonr
----> 2 print(pearsonr(d1,d2))

NameError: name 'd1' is not defined
```

```
In [28]:  from scipy.stats import spearmanr
          print(spearmanr(d1,d2))
```

```
          ---------------------------------------------------------------------------
          NameError                                 Traceback (most recent call last)
          <ipython-input-28-c093abd64074> in <module>
                1 from scipy.stats import spearmanr
          ----> 2 print(spearmanr(d1,d2))

          NameError: name 'd1' is not defined
```

|||

```
In [29]:  import numpy as np
          import pandas as pd
```

```
In [30]:  df4=pd.read_csv(r"C:\Users\Admin\Downloads\3_Fitness-1 - 3_Fitness-1.csv")
          df4
```

Out[30]:

	Row Labels	Sum of Jan	Sum of Feb	Sum of Mar	Sum of Total Sales
0	A	5.62%	7.73%	6.16%	75
1	B	4.21%	17.27%	19.21%	160
2	C	9.83%	11.60%	5.17%	101
3	D	2.81%	21.91%	7.88%	127
4	E	25.28%	10.57%	11.82%	179
5	F	8.15%	16.24%	18.47%	167
6	G	18.54%	8.76%	17.49%	171
7	H	25.56%	5.93%	13.79%	170
8	Grand Total	100.00%	100.00%	100.00%	1150

In [31]:
```python
df5=df4[['Sum of Total Sales']]
df5
```

Out[31]:

	Sum of Total Sales
0	75
1	160
2	101
3	127
4	179
5	167
6	171
7	170
8	1150

Find mean, median, mode and describe

In [32]:
```python
print(df5.mean())
```

```
Sum of Total Sales    255.555556
dtype: float64
```

In [33]:
```python
print(df5.median())
```

```
Sum of Total Sales    167.0
dtype: float64
```

In [34]:
```python
print(df5.mode())
```

```
   Sum of Total Sales
0                  75
1                 101
2                 127
3                 160
4                 167
5                 170
6                 171
7                 179
8                1150
```

Find sum(), cumsum(), count, min and max values

```
In [35]:  print(df5.sum())
```

```
Sum of Total Sales    2300
dtype: int64
```

```
In [36]:  print(df5.cumsum())
```

```
      Sum of Total Sales
0                     75
1                    235
2                    336
3                    463
4                    642
5                    809
6                    980
7                   1150
8                   2300
```

```
In [37]:  print(df5.min())
```

```
Sum of Total Sales    75
dtype: int64
```

```
In [38]:  print(df5.max())
```

```
Sum of Total Sales    1150
dtype: int64
```

Find covariance and correlation (spearman and pearsons)

```
In [39]:  print(df5.cov())
```

```
                    Sum of Total Sales
Sum of Total Sales         113793.527778
```

IV

```
In [42]:  import numpy as np
          import pandas as pd
```

In [43]: df6=pd.read_csv(r"C:\Users\Admin\Downloads\8_BreastCancerPrediction - 8_BreastC
df6

Out[43]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_
0	842302	M	17.99	10.38	122.80	1001.0	0.
1	842517	M	20.57	17.77	132.90	1326.0	0.0
2	84300903	M	19.69	21.25	130.00	1203.0	0.
3	84348301	M	11.42	20.38	77.58	386.1	0.
4	84358402	M	20.29	14.34	135.10	1297.0	0.
...	...	...	...	...	...	...	
564	926424	M	21.56	22.39	142.00	1479.0	0.
565	926682	M	20.13	28.25	131.20	1261.0	0.0
566	926954	M	16.60	28.08	108.30	858.1	0.0
567	927241	M	20.60	29.33	140.10	1265.0	0.
568	92751	B	7.76	24.54	47.92	181.0	0.0

569 rows × 32 columns

In [44]: df7=df6[['radius_mean','area_mean']]
df7

Out[44]:

	radius_mean	area_mean
0	17.99	1001.0
1	20.57	1326.0
2	19.69	1203.0
3	11.42	386.1
4	20.29	1297.0
...	...	...
564	21.56	1479.0
565	20.13	1261.0
566	16.60	858.1
567	20.60	1265.0
568	7.76	181.0

569 rows × 2 columns

Find mean, median, mode and describe

In [45]: `print(df7.mean())`

```
radius_mean      14.127292
area_mean       654.889104
dtype: float64
```

In [46]: `print(df7.median())`

```
radius_mean      13.37
area_mean       551.10
dtype: float64
```

In [47]: `print(df7.mode())`

```
   radius_mean  area_mean
0        12.34      512.2
```

Find sum(), cumsum(), count, min and max values

In [48]: `print(df7.sum())`

```
radius_mean       8038.429
area_mean       372631.900
dtype: float64
```

In [49]: `print(df7.cumsum())`

```
     radius_mean  area_mean
0         17.990     1001.0
1         38.560     2327.0
2         58.250     3530.0
3         69.670     3916.1
4         89.960     5213.1
..           ...        ...
564     7973.339   369066.8
565     7993.469   370327.8
566     8010.069   371185.9
567     8030.669   372450.9
568     8038.429   372631.9

[569 rows x 2 columns]
```

In [50]: `print(df7.min())`

```
radius_mean       6.981
area_mean       143.500
dtype: float64
```

In [51]: `print(df7.max())`

```
radius_mean       28.11
area_mean       2501.00
dtype: float64
```

Find covariance and correlation (spearman and pearsons)

In [52]: `print(df7.cov())`

```
             radius_mean       area_mean
radius_mean    12.418920     1224.483409
area_mean    1224.483409   123843.554318
```

V

In [55]:
```python
import numpy as np
import pandas as pd
```

```
In [64]: df8=pd.read_csv(r"C:\Users\Admin\Downloads\9_bottle.csv")
         df8
```

```
C:\Users\Admin\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3
071: DtypeWarning: Columns (47,73) have mixed types.Specify dtype option on i
mport or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Out[64]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sa
0	1	1	054.0 056.0	19-4903CR-HY-060-0930-05400560-0000A-3	0	10.500	33.4400	NaN	25.64900	Na
1	1	2	054.0 056.0	19-4903CR-HY-060-0930-05400560-0008A-3	8	10.460	33.4400	NaN	25.65600	Na
2	1	3	054.0 056.0	19-4903CR-HY-060-0930-05400560-0010A-7	10	10.460	33.4370	NaN	25.65400	Na
3	1	4	054.0 056.0	19-4903CR-HY-060-0930-05400560-0019A-3	19	10.450	33.4200	NaN	25.64300	Na
4	1	5	054.0 056.0	19-4903CR-HY-060-0930-05400560-0020A-7	20	10.450	33.4210	NaN	25.64300	Na
...	...	...	...	...	...	...	...	...	...	.
864858	34404	864859	093.4 026.4	20-1611SR-MX-310-2239-09340264-0000A-7	0	18.744	33.4083	5.805	23.87055	108.7
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805	23.87072	108.7
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796	23.88911	108.4
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816	24.01426	107.7

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sa
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297	105.6

864863 rows × 74 columns

In [65]:
```python
df9=df8[['Cst_Cnt','R_PRES']]
df9
```

Out[65]:

	Cst_Cnt	R_PRES
0	1	0
1	1	8
2	1	10
3	1	19
4	1	20
...	...	...
864858	34404	0
864859	34404	2
864860	34404	5
864861	34404	10
864862	34404	15

864863 rows × 2 columns

Find mean, median, mode and describe

In [58]:
```python
print(df9.mean())
```

```
Cst_Cnt    17138.790958
R_PRES       228.395694
dtype: float64
```

In [59]:
```python
print(df9.median())
```

```
Cst_Cnt    16848.0
R_PRES       126.0
dtype: float64
```

In [60]: `print(df9.mode())`

```
     Cst_Cnt   R_PRES
0     14467       10
```

Find sum(), cumsum(), count, min and max values

In [61]: `print(df9.sum())`

```
Cst_Cnt     14822706164
R_PRES        197530985
dtype: int64
```

In [62]: `print(df9.cumsum())`

```
              Cst_Cnt        R_PRES
0                   1             0
1                   2             8
2                   3            18
3                   4            37
4                   5            57
...               ...           ...
864858    14822568548     197530953
864859    14822602952     197530955
864860    14822637356     197530960
864861    14822671760     197530970
864862    14822706164     197530985

[864863 rows x 2 columns]
```

In [63]: `print(df9.min())`

```
Cst_Cnt     1
R_PRES      0
dtype: int64
```

In [66]: `print(df9.max())`

```
Cst_Cnt     34404
R_PRES       5458
dtype: int64
```

Find covariance and correlation (spearman and pearsons)

```
In [67]: print(df9.cov())
```

```
                  Cst_Cnt          R_PRES
Cst_Cnt  1.048771e+08 -529861.015146
R_PRES  -5.298610e+05   102052.603292
```