import libraries ¶

In [2]:
```
import numpy as np
import pandas as pd
```

import dataset

In [3]:
```
data=pd.read_csv(r"C:\Users\user\Downloads\6_Salesworkload1.csv")
data
```

Out[3]:

	MonthYear	Time index	Country	StoreID	City	Dept_ID	Dept. Name	HoursOwn	HoursLea
0	10.2016	1.0	United Kingdom	88253.0	London (I)	1.0	Dry	3184.764	
1	10.2016	1.0	United Kingdom	88253.0	London (I)	2.0	Frozen	1582.941	
2	10.2016	1.0	United Kingdom	88253.0	London (I)	3.0	other	47.205	
3	10.2016	1.0	United Kingdom	88253.0	London (I)	4.0	Fish	1623.852	
4	10.2016	1.0	United Kingdom	88253.0	London (I)	5.0	Fruits & Vegetables	1759.173	
...	...	...	...	...	...	...	...	...	
7653	06.2017	9.0	Sweden	29650.0	Gothenburg	12.0	Checkout	6322.323	
7654	06.2017	9.0	Sweden	29650.0	Gothenburg	16.0	Customer Services	4270.479	
7655	06.2017	9.0	Sweden	29650.0	Gothenburg	11.0	Delivery	0	
7656	06.2017	9.0	Sweden	29650.0	Gothenburg	17.0	others	2224.929	
7657	06.2017	9.0	Sweden	29650.0	Gothenburg	18.0	all	39652.2	

7658 rows × 14 columns

mean

In [4]:
```
print(data.mean())
```

```
Time index      5.000000e+00
StoreID         6.199522e+04
Dept_ID         9.470588e+00
HoursLease      2.203608e+01
Sales units     1.076471e+06
Turnover        3.721393e+06
Customer                 NaN
dtype: float64
```

median

In [5]:
```
print(data.median())
```

```
Time index          5.0
StoreID         75400.5
Dept_ID             9.0
HoursLease          0.0
Sales units    293230.0
Turnover       931957.5
Customer            NaN
dtype: float64
```

mode

In [6]:
```
print(data.describe())
```

```
        Time index       StoreID      Dept_ID   HoursLease   Sales units  \
count  7650.000000   7650.000000  7650.000000  7650.000000  7.650000e+03
mean      5.000000  61995.220000     9.470588    22.036078  1.076471e+06
std       2.582158  29924.581631     5.337429   133.299513  1.728113e+06
min       1.000000  12227.000000     1.000000     0.000000  0.000000e+00
25%       3.000000  29650.000000     5.000000     0.000000  5.457125e+04
50%       5.000000  75400.500000     9.000000     0.000000  2.932300e+05
75%       7.000000  87703.000000    14.000000     0.000000  9.175075e+05
max       9.000000  98422.000000    18.000000  3984.000000  1.124296e+07

           Turnover  Customer
count  7.650000e+03       0.0
mean   3.721393e+06       NaN
std    6.003380e+06       NaN
min    0.000000e+00       NaN
25%    2.726798e+05       NaN
50%    9.319575e+05       NaN
75%    3.264432e+06       NaN
max    4.271739e+07       NaN
```

In [7]:
```
print(data.sum())
```

```
MonthYear      10.201610.201610.201610.201610.201610.201610.2...
Time index                                                38250.0
StoreID                                                474263433.0
Dept_ID                                                    72450.0
HoursLease                                                168576.0
Sales units                                            8235000965.0
Turnover                                              28468656015.0
Customer                                                     0.0
dtype: object
```

In [8]:
```
df = pd.DataFrame(data[["Sales units","Turnover"]])
df
```

Out[8]:

	Sales units	Turnover
0	398560.0	1226244.0
1	82725.0	387810.0
2	438400.0	654657.0
3	309425.0	499434.0
4	165515.0	329397.0
...	...	...
7653	3886530.0	14538825.0
7654	245.0	0.0
7655	0.0	0.0
7656	245.0	0.0
7657	3886530.0	15056214.0

7658 rows × 2 columns

In [9]:
```
print(df.mode())
```

```
   Sales units  Turnover
0          0.0       0.0
```

In [10]:
```
print(df.mean())
```

```
Sales units    1.076471e+06
Turnover       3.721393e+06
dtype: float64
```

In [11]:
```
print(df.median())
```

```
Sales units    293230.0
Turnover       931957.5
dtype: float64
```

In [12]:
```
print(df.describe())
```

```
        Sales units      Turnover
count  7.650000e+03  7.650000e+03
mean   1.076471e+06  3.721393e+06
std    1.728113e+06  6.003380e+06
min    0.000000e+00  0.000000e+00
25%    5.457125e+04  2.726798e+05
50%    2.932300e+05  9.319575e+05
75%    9.175075e+05  3.264432e+06
max    1.124296e+07  4.271739e+07
```

In [13]:
```
print(df.sum())
```

```
Sales units    8.235001e+09
Turnover       2.846866e+10
dtype: float64
```

In [20]:
```
print(df.cumsum())
```

```
       Sales units      Turnover
0     3.985600e+05  1.226244e+06
1     4.812850e+05  1.614054e+06
2     9.196850e+05  2.268711e+06
3     1.229110e+06  2.768145e+06
4     1.394625e+06  3.097542e+06
...            ...           ...
7653  8.231114e+09  2.845360e+10
7654  8.231114e+09  2.845360e+10
7655  8.231114e+09  2.845360e+10
7656  8.231114e+09  2.845360e+10
7657  8.235001e+09  2.846866e+10

[7658 rows x 2 columns]
```

In [14]:
```
print(df.min())
```

```
Sales units    0.0
Turnover       0.0
dtype: float64
```

In [15]:
```
print(df.max())
```

```
Sales units    11242955.0
Turnover       42717390.0
dtype: float64
```

In [16]:
```
print(df.count())
```

```
Sales units    7650
Turnover       7650
dtype: int64
```

In [17]:
```
from numpy import cov
```

In [18]:
```
print(cov(df))
```

```
[[ 3.42530402e+11  1.26256987e+11  8.94962294e+10 ...  0.00000000e+00
  -1.01391290e+08  4.62248437e+12]
 [ 1.26256987e+11  4.65384286e+10  3.29883834e+10 ...  0.00000000e+00
  -3.73729125e+07  1.70385152e+12]
 [ 8.94962294e+10  3.29883834e+10  2.33835450e+10 ...  0.00000000e+00
  -2.64914825e+07  1.20776118e+12]
 ...
 [ 0.00000000e+00  0.00000000e+00  0.00000000e+00 ...  0.00000000e+00
   0.00000000e+00  0.00000000e+00]
 [-1.01391290e+08 -3.73729125e+07 -2.64914825e+07 ...  0.00000000e+00
   3.00125000e+04 -1.36828629e+09]
 [ 4.62248437e+12  1.70385152e+12  1.20776118e+12 ...  0.00000000e+00
  -1.36828629e+09  6.23809203e+13]]
```

In [19]:
```
from scipy.stats import pearsonr
```

In [21]:
```
df1 = df["Sales units"][0:100]
df2 = df["Turnover"][0:100]
df1
df2
```

Out[21]:
```
0      1226244.0
1       387810.0
2       654657.0
3       499434.0
4       329397.0
         ...
95     2319717.0
96           0.0
97    12161196.0
98           0.0
99     2204589.0
Name: Turnover, Length: 100, dtype: float64
```

In [22]:
```
print(pearsonr(df1,df2))
```

```
(0.9861714333836227, 2.5825285818653907e-78)
```

In [25]:
```
from scipy.stats import spearmanr
```

In [26]:
```
print(spearmanr(df1,df2))
```

```
SpearmanrResult(correlation=0.9127255342617159, pvalue=7.054492671709463e-40)
```

In [ ]:
```

```