Name :- Kapileshwar Ghatol.

Roll no :- 513

Div :- E

Batch :- E-1

Subject :- Essentials of data science.

Assignment No.4.

Practical no 4

```
In [1]:  ▶ import pandas as pd
           df = pd.read_csv("grainsales.csv")
           print(df)
```

```
        GrainName        State       City Months  Year    Sales
0            Ragi  Maharashtra     Nagpur    JAN  2023  1000000
1           Bajra       Panjab   Amritsar    FEB  2023  1500000
2            Ragi  Maharashtra     Nagpur    JAN  2023  1000000
3           Bajra       Panjab   Amritsar    FEB  2023  1500000
4            Ragi  Maharashtra     Nagpur    JAN  2023  1000000
5           Bajra       Panjab   Amritsar    FEB  2023  1500000
6            Oats     Hariyana   Gurugram  MARCH  2023  2000000
7           Sattu      Gujarat      Surat  APRIL  2023  2500000
8           Sooji   Tamil Nadu    Madurai    MAY  2023  3000000
9      Brown rice    Telangana  Hyderabad   JUNE  2023  3500000
10          Wheat  West Bengol   Asansole   JULY  2023  4000000
11           Corn           UP     Kanpur    AUG  2023  4500000
12           Ragi  Maharashtra     Nagpur    JAN  2023  1000000
13          Bajra       Panjab   Amritsar    FEB  2023  1500000
14           Oats     Hariyana   Gurugram  MARCH  2023  2000000
15          Sattu      Gujarat      Surat  APRIL  2023  2500000
16          Sooji   Tamil Nadu    Madurai    MAY  2023  3000000
17     Brown rice    Telangana  Hyderabad   JUNE  2023  3500000
18          Wheat  West Bengol   Asansole   JULY  2023  4000000
19           Corn           UP     Kanpur    AUG  2023  4500000
20          Sooji   Tamil Nadu    Madurai    MAY  2023  3000000
21     Brown rice    Telangana  Hyderabad   JUNE  2023  3500000
22          Wheat  West Bengol   Asansole   JULY  2023  4000000
23           Corn           UP     Kanpur    AUG  2023  4500000
```

```
In [2]:  ▶ print("Max Sales of grain:", df['Sales'].max())
```

```
Max Sales of grain: 4500000
```

```
In [3]:  ▶ print(df['Sales'].sum()/len(df['Sales']))
```

```
2685185.185185185
```

```
In [4]:  ▶ #print top 5 records
           print(df.head(5))
           #print bottom 5 records
           print(df.tail(5))
```

```
    GrainName        State       City Months  Year    Sales
0        Ragi  Maharashtra     Nagpur    JAN  2023  1000000
1       Bajra       Panjab   Amritsar    FEB  2023  1500000
2        Ragi  Maharashtra     Nagpur    JAN  2023  1000000
3       Bajra       Panjab   Amritsar    FEB  2023  1500000
4        Ragi  Maharashtra     Nagpur    JAN  2023  1000000
     GrainName        State       City Months  Year    Sales
22       Wheat  West Bengol   Asansole   JULY  2023  4000000
23        Corn           UP     Kanpur    AUG  2023  4500000
24        Ragi  Maharashtra     Nagpur    JAN  2023  1000000
25  Brown rice    Telangana  Hyderabad   JUNE  2023  3500000
26       Wheat  West Bengol   Asansole   JULY  2023  4000000
```

```
In [6]:  ▶ gdf2 = df.groupby('GrainName').count()

           print(gdf2['Sales'])
```

```
GrainName
Bajra         4
Brown rice    4
Corn          3
Oats          2
Ragi          5
Sattu         2
Sooji         3
Wheat         4
Name: Sales, dtype: int64
```

```
In [7]:  ▶ #Find Maximum grainmame for each sale
           print(df.groupby('GrainName').max())
```

```
                  State       City Months  Year    Sales
GrainName
Bajra            Panjab   Amritsar    FEB  2023  1500000
Brown rice    Telangana  Hyderabad   JUNE  2023  3500000
Corn                 UP     Kanpur    AUG  2023  4500000
Oats           Hariyana   Gurugram  MARCH  2023  2000000
Ragi        Maharashtra     Nagpur    JAN  2023  1000000
Sattu           Gujarat      Surat  APRIL  2023  2500000
Sooji        Tamil Nadu    Madurai    MAY  2023  3000000
Wheat       West Bengol   Asansole   JULY  2023  4000000
```

Practical no 4

```
In [6]:  ▶ gdf2 = df.groupby('GrainName').count()

            print(gdf2['Sales'])

            GrainName
            Bajra        4
            Brown rice   4
            Corn         3
            Oats         2
            Ragi         5
            Sattu        2
            Sooji        3
            Wheat        4
            Name: Sales, dtype: int64
```

```
In [7]:  ▶ #Find Maximum grainmame for each sale
            print(df.groupby('GrainName').max())

                              State       City Months  Year    Sales
            GrainName
            Bajra            Panjab   Amritsar    FEB  2023  1500000
            Brown rice    Telangana  Hyderabad   JUNE  2023  3500000
            Corn                 UP     Kanpur    AUG  2023  4500000
            Oats           Hariyana   Gurugram  MARCH  2023  2000000
            Ragi        Maharashtra     Nagpur    JAN  2023  1000000
            Sattu           Gujarat      Surat  APRIL  2023  2500000
            Sooji        Tamil Nadu    Madurai    MAY  2023  3000000
            Wheat       West Bengol   Asansole   JULY  2023  4000000
```

```
In [9]:  ▶ #Find Minimum Sales for each grainname
            print(df.groupby('Sales').min())

                        GrainName        State       City Months  Year
            Sales
            1000000          Ragi  Maharashtra     Nagpur    JAN  2023
            1500000         Bajra       Panjab   Amritsar    FEB  2023
            2000000          Oats     Hariyana   Gurugram  MARCH  2023
            2500000         Sattu      Gujarat      Surat  APRIL  2023
            3000000         Sooji   Tamil Nadu    Madurai    MAY  2023
            3500000    Brown rice    Telangana  Hyderabad   JUNE  2023
            4000000         Wheat  West Bengol   Asansole   JULY  2023
            4500000          Corn           UP     Kanpur    AUG  2023
```

```
In [12]:  ▶ df['Sales']

Out[12]:  0     1000000
          1     1500000
          2     1000000
          3     1500000
          4     1000000
          5     1500000
          6     2000000
          7     2500000
          8     3000000
          9     3500000
          10    4000000
          11    4500000
          12    1000000
          13    1500000
          14    2000000
          15    2500000
          16    3000000
          17    3500000
          18    4000000
          19    4500000
          20    3000000
          21    3500000
          22    4000000
          23    4500000
          24    1000000
          25    3500000
          26    4000000
```

Practical no 4

In [13]: ▶ df[['City','State']]

Out[13]:

	City	State
0	Nagpur	Maharashtra
1	Amritsar	Panjab
2	Nagpur	Maharashtra
3	Amritsar	Panjab
4	Nagpur	Maharashtra
5	Amritsar	Panjab
6	Gurugram	Hariyana
7	Surat	Gujarat
8	Madurai	Tamil Nadu
9	Hyderabad	Telangana
10	Asansole	West Bengol
11	Kanpur	UP
12	Nagpur	Maharashtra
13	Amritsar	Panjab
14	Gurugram	Hariyana
15	Surat	Gujarat

In [15]: ▶ df[10:20]

Out[15]:

	GrainName	State	City	Months	Year	Sales
10	Wheat	West Bengol	Asansole	JULY	2023	4000000
11	Corn	UP	Kanpur	AUG	2023	4500000
12	Ragi	Maharashtra	Nagpur	JAN	2023	1000000
13	Bajra	Panjab	Amritsar	FEB	2023	1500000
14	Oats	Hariyana	Gurugram	MARCH	2023	2000000
15	Sattu	Gujarat	Surat	APRIL	2023	2500000
16	Sooji	Tamil Nadu	Madurai	MAY	2023	3000000
17	Brown rice	Telangana	Hyderabad	JUNE	2023	3500000
18	Wheat	West Bengol	Asansole	JULY	2023	4000000
19	Corn	UP	Kanpur	AUG	2023	4500000

Practical no 4

```
In [17]: ▶| df.iloc[0]
```

```
Out[17]: GrainName          Ragi
         State       Maharashtra
         City             Nagpur
         Months              JAN
         Year               2023
         Sales           1000000
         Name: 0, dtype: object
```

```
In [18]: ▶| df.iloc[1]
```

```
Out[18]: GrainName         Bajra
         State            Panjab
         City           Amritsar
         Months              FEB
         Year               2023
         Sales           1500000
         Name: 1, dtype: object
```

```
In [19]: ▶| df.iloc[-1]
```

```
Out[19]: GrainName           Wheat
         State        West Bengol
         City            Asansole
         Months              JULY
         Year                2023
         Sales            4000000
```

Practical no 4

```
In [29]: ▶ flights[flights.isnull().any(axis=1)].head()
```

Out[29]:

	GrainName	State	City	Months	Year	Sales
27	Wheat	NaN	NaN	NaN	NaN	NaN
28	Wheat	NaN	NaN	NaN	NaN	NaN

```
In [31]: ▶ df.dropna(axis=1,how='all')
```

Out[31]:

	GrainName	State	City	Months	Year	Sales
0	Ragi	Maharashtra	Nagpur	JAN	2023	1000000
1	Bajra	Panjab	Amritsar	FEB	2023	1500000
2	Ragi	Maharashtra	Nagpur	JAN	2023	1000000
3	Bajra	Panjab	Amritsar	FEB	2023	1500000
4	Ragi	Maharashtra	Nagpur	JAN	2023	1000000
5	Bajra	Panjab	Amritsar	FEB	2023	1500000
6	Oats	Hariyana	Gurugram	MARCH	2023	2000000
7	Sattu	Gujarat	Surat	APRIL	2023	2500000
8	Sooji	Tamil Nadu	Madurai	MAY	2023	3000000
9	Brown rice	Telangana	Hyderabad	JUNE	2023	3500000

```
In [21]: ▶ df_sorted = df.sort_values(by = 'Year')
           df_sorted.head()
```

Out[21]:

	GrainName	State	City	Months	Year	Sales
0	Ragi	Maharashtra	Nagpur	JAN	2023	1000000
24	Ragi	Maharashtra	Nagpur	JAN	2023	1000000
23	Corn	UP	Kanpur	AUG	2023	4500000
22	Wheat	West Bengol	Asansole	JULY	2023	4000000
21	Brown rice	Telangana	Hyderabad	JUNE	2023	3500000

```
In [22]: ▶ df_sorted = df.sort_values(by = 'State')
           df_sorted.head()
```

Out[22]:

	GrainName	State	City	Months	Year	Sales
15	Sattu	Gujarat	Surat	APRIL	2023	2500000
7	Sattu	Gujarat	Surat	APRIL	2023	2500000
14	Oats	Hariyana	Gurugram	MARCH	2023	2000000
6	Oats	Hariyana	Gurugram	MARCH	2023	2000000
0	Ragi	Maharashtra	Nagpur	JAN	2023	1000000

Practical no 4

```
In [35]:  ▶  plt.bar(df['GrainName'],df['Sales'])
              plt.title("Bar Chart")
              plt.show()
```


Bar Chart

1e6

(y-axis: 0, 1, 2, 3, 4)

x-axis labels: Ragi, Bajra, Oats, Sattu, Sooji, Brown rice, Wheat, Corn