EDS Assignment 4

Name – *Punit Kawadkar* **Div**. – *D2*

Roll No. – *431* **PRN No.** – *202201040138*

CSV FILE :

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

Code :

```python
import pandas as pd

# Read the CSV file into a DataFrame
df = pd.read_csv("/content/grainsales.csv")
```

```python
# Task 1: Which was the best month for sales? How much was earned that
month?
df['Sales'] = df['Sales'].astype(float) # Convert the 'Sales' column
to numeric type
best_month = df.groupby('Months')['Sales'].sum().idxmax()
revenue_earned = df.groupby('Months')['Sales'].sum().max()
print("The best month for sales was", best_month, "with a total revenue
of", revenue_earned)

# Task 2: Which product sold the most? Why do you think it did?
best_product = df.groupby('GrainName')['Sales'].sum().idxmax()
print("The product that sold the most was", best_product)

# Task 3: Which city sold the most products?
best_city = df.groupby('City')['Sales'].sum().idxmax()
print("The city that sold the most products was", best_city)

# Task 4: What products are most often sold together?

# w.r.t State

product_combinations = df.groupby('State')['GrainName'].apply(lambda x:
list(combinations(x, 2)))
product_counts =
pd.Series(product_combinations.explode()).value_counts().head()
print("Products most often sold together:")
print(product_counts)

# w.r.t City

product_combinations = df.groupby('City')['GrainName'].apply(lambda x:
list(combinations(x, 2)))
product_counts =
pd.Series(product_combinations.explode()).value_counts().head()
print("Products most often sold together:")
print(product_counts)

# w.r.t Months

product_combinations = df.groupby('Months')['GrainName'].apply(lambda
x: list(combinations(x, 2)))
product_counts =
pd.Series(product_combinations.explode()).value_counts().head()
print("Products most often sold together:")
print(product_counts)

# w.r.t Year
```

```python
product_combinations = df.groupby('Year')['GrainName'].apply(lambda x:
list(combinations(x, 2)))
product_counts =
pd.Series(product_combinations.explode()).value_counts().head()
print("Products most often sold together:")
print(product_counts)

# w.r.t Sales

product_combinations = df.groupby('Sales')['GrainName'].apply(lambda x:
list(combinations(x, 2)))
product_counts =
pd.Series(product_combinations.explode()).value_counts().head()
print("Products most often sold together:")
print(product_counts)
```

Output :

```
The best month for sales was JULY with a total revenue of 16000000.0
The product that sold the most was Wheat
The city that sold the most products was Asansole
Products most often sold together:
(Ragi, Ragi)                    10
(Bajra, Bajra)                   6
(Brown rice , Brown rice )       6
(Wheat, Wheat)                   6
(Sooji, Sooji)                   3
Name: GrainName, dtype: int64
Products most often sold together:
(Ragi, Ragi)                    10
(Bajra, Bajra)                   6
(Wheat, Wheat)                   6
(Brown rice , Brown rice )       6
(Corn, Corn)                     3
Name: GrainName, dtype: int64
Products most often sold together:
(Ragi, Ragi)                    10
(Bajra, Bajra)                   6
(Wheat, Wheat)                   6
(Brown rice , Brown rice )       6
(Corn, Corn)                     3
Name: GrainName, dtype: int64
Products most often sold together:
(Ragi, Brown rice )     16
(Ragi, Wheat)           16
(Bajra, Brown rice )    15
(Bajra, Wheat)          15
(Ragi, Sooji)           11
Name: GrainName, dtype: int64
Products most often sold together:
(Ragi, Ragi)                    10
```

```
(Bajra, Bajra)                 6
(Brown rice , Brown rice )     6
(Wheat, Wheat)                 6
(Sooji, Sooji)                 3
Name: GrainName, dtype: int64
```