Name - Atharv Khisti

Roll No - 573

DIV E

PRN - 202201040160

```
import pandas as pd
```

```
# Read the CSV file
data = pd.read_csv('grainsales.csv')
```

```
# 1. Identify 10 grains for the given dataset
grains = data['GrainName'].unique()[:10]
print("10 grains in the dataset:", grains)
```

```
10 grains in the dataset: ['Ragi' 'Bajra' 'Oats' 'Sattu ' 'Sooji' 'Brown rice ' 'Wheat' 'Corn']
```

```
# 2. Implement all 20 grains using Pandas methods
all_grains = data['GrainName'].unique()
print("All 20 grains in the dataset:", all_grains)
```

```
All 20 grains in the dataset: ['Ragi' 'Bajra' 'Oats' 'Sattu ' 'Sooji' 'Brown rice ' 'Wheat' 'Corn']
```

```
# 3. Which was the best month for sales? How much was earned that month?
monthly_sales = data.groupby('Months')['Sales'].sum()
best_month = monthly_sales.idxmax()
earning = monthly_sales.max()
print("Best month for sales:", best_month)
print("Earnings in the best month:", earning)
```

```
Best month for sales: JULY
Earnings in the best month: 16000000
```

```
# 4. Which product sold the most? Why do you think it did?
product_sales = data.groupby('GrainName')['Sales'].sum()
best_product = product_sales.idxmax()
print("Product sold the most:", best_product)
```

```
Product sold the most: Wheat
```

```
# 5. Which city sold the most products?
city_sales = data.groupby('City')['Sales'].sum()
best_city = city_sales.idxmax()
print("City that sold the most products:", best_city)
```

```
City that sold the most products: Asansole
```

```
# 6. What is the total sales revenue for each grain across all months?
total_sales_grain = data.groupby('GrainName')['Sales'].sum()
print("Total sales revenue for each grain:")
print(total_sales_grain)
```

```
Total sales revenue for each grain:
GrainName
Bajra          6000000
Brown rice    14000000
Corn          13500000
Oats           4000000
Ragi           5000000
Sattu          5000000
Sooji          9000000
Wheat         16000000
Name: Sales, dtype: int64
```

```
# 7. How does the sales revenue vary by state? Which state has the highest and lowest sales?
sales_by_state = data.groupby('State')['Sales'].sum()
highest_sales_state = sales_by_state.idxmax()
lowest_sales_state = sales_by_state.idxmin()
```

```
print("Sales revenue by state:")
print(sales_by_state)
print("State with the highest sales revenue:", highest_sales_state)
print("State with the lowest sales revenue:", lowest_sales_state)
```

```
Sales revenue by state:
State
Gujarat         5000000
Hariyana        4000000
Maharashtra     5000000
Panjab          6000000
Tamil Nadu      9000000
Telangana      14000000
UP             13500000
West Bengol    16000000
Name: Sales, dtype: int64
State with the highest sales revenue: West Bengol
State with the lowest sales revenue: Hariyana
```

```
# 8.What is the average sales revenue per month?
average_sales_month = data.groupby('Months')['Sales'].mean()
print("Average sales revenue per month:")
print(average_sales_month)
```

```
Average sales revenue per month:
Months
APRIL    2500000.0
AUG      4500000.0
FEB      1500000.0
JAN      1000000.0
JULY     4000000.0
JUNE     3500000.0
MARCH    2000000.0
MAY      3000000.0
Name: Sales, dtype: float64
```

```
# 9.Is there any correlation between the sales revenue and the year? If so, how strong is the correlation?
correlation = data['Sales'].corr(data['Year'])
print("Correlation between sales revenue and year:", correlation)
```

```
Correlation between sales revenue and year: nan
```

```
#10.Can you identify any seasonal trends in the sales of grains?
seasonal_trends = data.groupby('Months')['Sales'].sum().sort_values(ascending=False)
print("Seasonal trends in sales of grains:")
print(seasonal_trends)
```

```
Seasonal trends in sales of grains:
Months
JULY     16000000
JUNE     14000000
AUG      13500000
MAY       9000000
FEB       6000000
APRIL     5000000
JAN       5000000
MARCH     4000000
Name: Sales, dtype: int64
```

```
#11.Which month had the highest and lowest sales revenue for each grain?
highest_month_grain = data.groupby(['GrainName'])['Sales'].idxmax()
lowest_month_grain = data.groupby(['GrainName'])['Sales'].idxmin()

highest_month_sales = data.loc[highest_month_grain, ['GrainName', 'Months', 'Sales']]
lowest_month_sales = data.loc[lowest_month_grain, ['GrainName', 'Months', 'Sales']]

print("Month with the highest sales revenue for each grain:")
print(highest_month_sales)
print("Month with the lowest sales revenue for each grain:")
print(lowest_month_sales)
```

```
Month with the highest sales revenue for each grain:
      GrainName Months    Sales
1         Bajra    FEB  1500000
```

```
9   Brown rice    JUNE  3500000
11         Corn    AUG  4500000
6          Oats  MARCH  2000000
0          Ragi    JAN  1000000
7        Sattu   APRIL  2500000
8         Sooji    MAY  3000000
10        Wheat   JULY  4000000
Month with the lowest sales revenue for each grain:
      GrainName Months    Sales
1         Bajra    FEB  1500000
9   Brown rice    JUNE  3500000
11         Corn    AUG  4500000
6          Oats  MARCH  2000000
0          Ragi    JAN  1000000
7        Sattu   APRIL  2500000
8         Sooji    MAY  3000000
10        Wheat   JULY  4000000
```

```
#12.What is the distribution of sales revenue among different cities?
sales_distribution_city = data.groupby('City')['Sales'].sum()
print("Distribution of sales revenue among different cities:")
print(sales_distribution_city)
```

```
Distribution of sales revenue among different cities:
City
Amritsar      6000000
Asansole     16000000
Gurugram      4000000
Hyderabad    14000000
Kanpur       13500000
Madurai       9000000
Nagpur        5000000
Surat         5000000
Name: Sales, dtype: int64
```

```
#13.Are there any outliers in the sales revenue? If so, which grains and months are affected?
Q1 = data['Sales'].quantile(0.25)
Q3 = data['Sales'].quantile(0.75)
IQR = Q3 - Q1

outliers = data[(data['Sales'] < Q1 - 1.5 * IQR) | (data['Sales'] > Q3 + 1.5 * IQR)]
print("Outliers in sales revenue:")
print(outliers)
```

```
Outliers in sales revenue:
Empty DataFrame
Columns: [GrainName, State, City, Months, Year, Sales]
Index: []
```

```
#14.Can you calculate the percentage contribution of each grain to the total sales revenue?
percentage_contribution = (data.groupby('GrainName')['Sales'].sum() / data['Sales'].sum()) * 100
print("Percentage contribution of each grain to total sales revenue:")
print(percentage_contribution)
```

```
Percentage contribution of each grain to total sales revenue:
GrainName
Bajra          8.275862
Brown rice    19.310345
Corn          18.620690
Oats           5.517241
Ragi           6.896552
Sattu          6.896552
Sooji         12.413793
Wheat         22.068966
Name: Sales, dtype: float64
```

```
#15.Is there any difference in sales revenue between different years?
yearly_sales_comparison = data.groupby('Year')['Sales'].sum()
print("Difference in sales revenue between different years:")
print(yearly_sales_comparison)
```

```
Difference in sales revenue between different years:
Year
2023    72500000
Name: Sales, dtype: int64
```

```
#16.What is the average sales revenue per grain across all cities?
average_sales_grain_city = data.groupby(['GrainName', 'City'])['Sales'].mean()
print("Average sales revenue per grain across all cities:")
print(average_sales_grain_city)
```

```
Average sales revenue per grain across all cities:
GrainName    City
Bajra        Amritsar     1500000.0
Brown rice   Hyderabad    3500000.0
Corn         Kanpur       4500000.0
Oats         Gurugram     2000000.0
Ragi         Nagpur       1000000.0
Sattu        Surat        2500000.0
Sooji        Madurai      3000000.0
Wheat        Asansole     4000000.0
Name: Sales, dtype: float64
```

```
#17.Which month had the highest sales revenue for each city?
highest_month_city = data.groupby(['City'])['Sales'].idxmax()
highest_month_sales_city = data.loc[highest_month_city, ['City', 'Months', 'Sales']]
print("Month with the highest sales revenue for each city:")
print(highest_month_sales_city)
```

```
Month with the highest sales revenue for each city:
         City Months    Sales
1    Amritsar    FEB  1500000
10   Asansole   JULY  4000000
6    Gurugram  MARCH  2000000
9   Hyderabad   JUNE  3500000
11     Kanpur    AUG  4500000
8     Madurai    MAY  3000000
0      Nagpur    JAN  1000000
7       Surat  APRIL  2500000
```

```
#18.Are there any seasonal differences in sales revenue between different states?
seasonal_sales_state = data.groupby(['State', 'Months'])['Sales'].sum()
print("Seasonal differences in sales revenue between different states:")
print(seasonal_sales_state)
```

```
Seasonal differences in sales revenue between different states:
State        Months
Gujarat      APRIL      5000000
Hariyana     MARCH      4000000
Maharashtra  JAN        5000000
Panjab       FEB        6000000
Tamil Nadu   MAY        9000000
Telangana    JUNE      14000000
UP           AUG       13500000
West Bengol  JULY      16000000
Name: Sales, dtype: int64
```

```
#19.What is the total sales revenue for each year?
total_sales_year = data.groupby('Year')['Sales'].sum()
print("Total sales revenue for each year:")
print(total_sales_year)
```

```
Total sales revenue for each year:
Year
2023    72500000
Name: Sales, dtype: int64
```

```
#20.Which grain has the highest sales revenue in each state?
highest_sales_grain_state = data.groupby(['State', 'GrainName'])['Sales'].sum().reset_index()
idx = highest_sales_grain_state.groupby('State')['Sales'].transform(max) == highest_sales_grain_state['Sales']
highest_sales_grain_state = highest_sales_grain_state[idx]
print("Grain with the highest sales revenue in each state:")
print(highest_sales_grain_state)
```

```
Grain with the highest sales revenue in each state:
         State    GrainName     Sales
0      Gujarat       Sattu    5000000
1     Hariyana         Oats   4000000
2  Maharashtra         Ragi   5000000
3       Panjab        Bajra   6000000
4   Tamil Nadu        Sooji   9000000
```

```
5    Telangana  Brown rice   14000000
6           UP         Corn  13500000
7  West Bengol        Wheat  16000000
```

0s completed at 1:27 AM