**PIYUSH KUMAR MISHRA**

**230957212**

**ROLL N0:70**

**WEEK 9:**

**Exercise 1: Sales Performance Analysis of XYZ Company Data Overview:   
The sales data of XYZ Company contains the following attributes for each transaction:   
• Date: The date of the sale.   
• Product Name: The name of the product sold.   
• Units Sold: The number of units sold in the transaction.   
• Revenue: The total revenue generated by the sale.   
• Region: The geographical region where the sale occurred.   
• Discount Offered (%): The percentage of discount offered on the sale.   
• Salesperson: The name of the salesperson responsible for the sale.   
Insert At least 20 data and create the dataframe.  
 1. What are the top 3 sales transactions with the highest revenue?   
2. How many units of each product were sold?   
3. What is the total revenue after applying discounts?   
4. Which sales transaction had the highest discount offered, and how much revenue did it generate after applying the discount?   
5. Which salesperson generated the highest total revenue?   
6. What is the average discount offered by each salesperson?   
7. How much revenue was generated in each region?   
8. In which region did Alice generate the highest sales?   
9. Which product generated the highest revenue per unit sold?   
10. How many transactions were rated as "High" performance?   
11. Which salesperson sold the most units in the North region without offering any discount?   
12. What is the average revenue per unit sold in each region for each product?   
13. Which salesperson has the highest average revenue after discounts, and how does it compare between regions?   
14. What is the cumulative total revenue over time for each salesperson?   
15. For each salesperson, rank the transactions by revenue, and find the top 2 transactions for each.   
16. How has the total revenue generated by each product changed over time? Show cumulative revenue for each product per day.   
17. Analyze how discounts affect revenue. For each product, what is the average revenue generated with a discount compared to without a discount?   
18. What is the weighted average discount offered by each salesperson, weighted by the revenue they generated?   
19. What percentage of the total revenue does each region contribute? Compare it to the total revenue per region.**

import pandas as pd

import numpy as np

np.random.seed(0)

data = {

"Date": pd.date\_range(start="2023-01-01", periods=20, freq='D'),

"Product Name": np.random.choice(['Product A', 'Product B', 'Product C'], 20),

"Units Sold": np.random.randint(1, 10, size=20),

"Revenue": np.random.uniform(100, 1000, size=20).round(2),

"Region": np.random.choice(['North', 'South', 'East', 'West'], 20),

"Discount Offered (%)": np.random.randint(0, 30, size=20),

"Salesperson": np.random.choice(['Alice', 'Bob', 'Charlie', 'David'], 20)

}

sales\_data = pd.DataFrame(data)

sales\_data['Revenue After Discount'] = sales\_data['Revenue'] \* (1 - sales\_data['Discount Offered (%)'] / 100)

top\_revenue\_transactions = sales\_data.nlargest(3, 'Revenue')

units\_sold\_per\_product = sales\_data.groupby('Product Name')['Units Sold'].sum()

total\_revenue\_after\_discount = sales\_data['Revenue After Discount'].sum()

max\_discount\_transaction = sales\_data.loc[sales\_data['Discount Offered (%)'].idxmax()]

highest\_revenue\_salesperson = sales\_data.groupby('Salesperson')['Revenue'].sum().idxmax()

average\_discount\_per\_salesperson = sales\_data.groupby('Salesperson')['Discount Offered (%)'].mean()

revenue\_per\_region = sales\_data.groupby('Region')['Revenue'].sum()

alice\_sales = sales\_data[sales\_data['Salesperson'] == 'Alice']

region\_highest\_alice = alice\_sales.groupby('Region')['Revenue'].sum().idxmax()

sales\_data['Revenue Per Unit'] = sales\_data['Revenue'] / sales\_data['Units Sold'].replace(0, np.nan)

highest\_revenue\_per\_unit\_product = sales\_data.loc[sales\_data['Revenue Per Unit'].idxmax()]

high\_performance\_transactions\_count = (sales\_data['Revenue'] > 500).sum()

north\_no\_discount\_sales = sales\_data[(sales\_data['Region'] == 'North') & (sales\_data['Discount Offered (%)'] == 0)]

top\_north\_salesperson\_no\_discount = north\_no\_discount\_sales.groupby('Salesperson')['Units Sold'].sum().idxmax()

average\_revenue\_per\_unit\_region\_product = sales\_data.groupby(['Region', 'Product Name']).apply(

lambda x: (x['Revenue'] / x['Units Sold'].replace(0, np.nan)).mean()

)

average\_revenue\_after\_discount = sales\_data.groupby('Salesperson')['Revenue After Discount'].mean()

highest\_avg\_revenue\_salesperson = average\_revenue\_after\_discount.idxmax()

sales\_data['Cumulative Revenue'] = sales\_data.groupby('Salesperson')['Revenue'].cumsum()

top\_transactions\_per\_salesperson = sales\_data.sort\_values('Revenue', ascending=False).groupby('Salesperson').head(2)

cumulative\_revenue\_product\_over\_time = sales\_data.groupby(['Date', 'Product Name'])['Revenue'].sum().groupby(level=1).cumsum()

avg\_revenue\_with\_discount = sales\_data[sales\_data['Discount Offered (%)'] > 0].groupby('Product Name')['Revenue After Discount'].mean()

avg\_revenue\_without\_discount = sales\_data[sales\_data['Discount Offered (%)'] == 0].groupby('Product Name')['Revenue'].mean()

discount\_impact\_analysis = pd.DataFrame({

'Avg Revenue With Discount': avg\_revenue\_with\_discount,

'Avg Revenue Without Discount': avg\_revenue\_without\_discount

}).fillna(0)

weighted\_average\_discount = (sales\_data['Discount Offered (%)'] \* sales\_data['Revenue']).groupby(sales\_data['Salesperson']).sum() / sales\_data.groupby('Salesperson')['Revenue'].sum()

total\_revenue = sales\_data['Revenue'].sum()

percentage\_revenue\_by\_region = (revenue\_per\_region / total\_revenue) \* 100

comparison\_total\_revenue\_per\_region = pd.DataFrame({

'Total Revenue': revenue\_per\_region,

'Percentage of Total Revenue': percentage\_revenue\_by\_region

})

print("Top 3 Sales Transactions by Revenue:\n", top\_revenue\_transactions)

print("\nUnits Sold for Each Product:\n", units\_sold\_per\_product)

print("\nTotal Revenue After Discounts:", total\_revenue\_after\_discount)

print("\nTransaction with Highest Discount Offered:\n", max\_discount\_transaction)

print("\nSalesperson with Highest Total Revenue:", highest\_revenue\_salesperson)

print("\nAverage Discount Offered by Each Salesperson:\n", average\_discount\_per\_salesperson)

print("\nRevenue Generated in Each Region:\n", revenue\_per\_region)

print("\nRegion where Alice Generated Highest Sales:", region\_highest\_alice)

print("\nProduct with Highest Revenue Per Unit Sold:\n", highest\_revenue\_per\_unit\_product[['Product Name', 'Revenue Per Unit']])

print("\nNumber of High Performance Transactions:", high\_performance\_transactions\_count)

print("\nSalesperson in North Region with Most Units Sold (No Discount):", top\_north\_salesperson\_no\_discount)

print("\nAverage Revenue Per Unit Sold in Each Region for Each Product:\n", average\_revenue\_per\_unit\_region\_product)

print("\nSalesperson with Highest Average Revenue After Discounts:", highest\_avg\_revenue\_salesperson)

print("\nCumulative Total Revenue Over Time for Each Salesperson:\n", sales\_data[['Date', 'Salesperson', 'Cumulative Revenue']])

print("\nTop 2 Transactions for Each Salesperson:\n", top\_transactions\_per\_salesperson)

print("\nCumulative Revenue Over Time for Each Product:\n", cumulative\_revenue\_product\_over\_time)

print("\nAverage Revenue With vs Without Discount:\n", discount\_impact\_analysis)

print("\nWeighted Average Discount Offered by Each Salesperson:\n", weighted\_average\_discount)

print("\nPercentage of Total Revenue by Region:\n", comparison\_total\_revenue\_per\_region)

**OUTPUT:**

Top 3 Sales Transactions by Revenue:

Date Product Name Units Sold Revenue Region Discount Offered (%) \

8 2023-01-09 Product A 4 949.37 West 9

1 2023-01-02 Product B 5 796.81 East 15

12 2023-01-13 Product B 4 727.87 North 0

Salesperson Revenue After Discount

8 Alice 863.9267

1 Charlie 677.2885

12 Charlie 727.8700

Units Sold for Each Product:

Product Name

Product A 30

Product B 30

Product C 25

Name: Units Sold, dtype: int32

Total Revenue After Discounts: 9376.795900000003

Transaction with Highest Discount Offered:

Date 2023-01-19 00:00:00

Product Name Product B

Units Sold 5

Revenue 383.89

Region West

Discount Offered (%) 29

Salesperson Charlie

Revenue After Discount 272.5619

Name: 18, dtype: object

Salesperson with Highest Total Revenue: Charlie

Average Discount Offered by Each Salesperson:

Salesperson

Alice 15.000000

Bob 15.500000

Charlie 9.375000

David 11.666667

Name: Discount Offered (%), dtype: float64

Revenue Generated in Each Region:

Region

East 2583.59

North 4283.50

South 338.10

West 3312.99

Name: Revenue, dtype: float64

Region where Alice Generated Highest Sales: West

Product with Highest Revenue Per Unit Sold:

Product Name Product A

Revenue Per Unit 703.57

Name: 15, dtype: object

Number of High Performance Transactions: 11

Salesperson in North Region with Most Units Sold (No Discount): Charlie

Average Revenue Per Unit Sold in Each Region for Each Product:

Region Product Name

East Product A 703.570000

Product B 106.389750

Product C 109.311667

North Product A 285.939444

Product B 142.408333

Product C 83.131250

South Product A 37.566667

West Product A 224.561250

Product B 277.679333

Product C 218.413333

dtype: float64

Salesperson with Highest Average Revenue After Discounts: Alice

Cumulative Total Revenue Over Time for Each Salesperson:

Date Salesperson Cumulative Revenue

0 2023-01-01 Charlie 338.10

1 2023-01-02 Charlie 1134.91

2 2023-01-03 Bob 510.54

3 2023-01-04 Alice 611.59

4 2023-01-05 David 116.91

5 2023-01-06 Bob 1166.41

6 2023-01-07 Bob 1817.30

7 2023-01-08 David 772.15

8 2023-01-09 Alice 1560.96

9 2023-01-10 Charlie 1848.55

10 2023-01-11 Charlie 2272.11

11 2023-01-12 David 1265.48

12 2023-01-13 Charlie 2999.98

13 2023-01-14 David 1419.68

14 2023-01-15 David 2119.77

15 2023-01-16 David 2823.34

16 2023-01-17 Charlie 3289.32

17 2023-01-18 Bob 2033.33

18 2023-01-19 Charlie 3673.21

19 2023-01-20 Charlie 4100.55

Top 2 Transactions for Each Salesperson:

Date Product Name Units Sold Revenue Region Discount Offered (%) \

8 2023-01-09 Product A 4 949.37 West 9

1 2023-01-02 Product B 5 796.81 East 15

12 2023-01-13 Product B 4 727.87 North 0

15 2023-01-16 Product A 1 703.57 East 4

14 2023-01-15 Product C 8 700.09 North 18

5 2023-01-06 Product C 6 655.87 East 16

6 2023-01-07 Product A 1 650.89 North 17

3 2023-01-04 Product B 1 611.59 West 21

Salesperson Revenue After Discount Revenue Per Unit Cumulative Revenue

8 Alice 863.9267 237.342500 1560.96

1 Charlie 677.2885 159.362000 1134.91

12 Charlie 727.8700 181.967500 2999.98

15 David 675.4272 703.570000 2823.34

14 David 574.0738 87.511250 2119.77

5 Bob 550.9308 109.311667 1166.41

6 Bob 540.2387 650.890000 1817.30

3 Alice 483.1561 611.590000 611.59

Cumulative Revenue Over Time for Each Product:

Date Product Name

2023-01-01 Product A 338.10

2023-01-02 Product B 796.81

2023-01-03 Product A 848.64

2023-01-04 Product B 1408.40

2023-01-05 Product B 1525.31

2023-01-06 Product C 655.87

2023-01-07 Product A 1499.53

2023-01-08 Product C 1311.11

2023-01-09 Product A 2448.90

2023-01-10 Product A 3162.54

2023-01-11 Product A 3586.10

2023-01-12 Product C 1804.44

2023-01-13 Product B 2253.18

2023-01-14 Product C 1958.64

2023-01-15 Product C 2658.73

2023-01-16 Product A 4289.67

2023-01-17 Product B 2542.52

2023-01-18 Product B 2758.55

2023-01-19 Product B 3142.44

2023-01-20 Product B 3569.78

Name: Revenue, dtype: float64

Average Revenue With vs Without Discount:

Avg Revenue With Discount Avg Revenue Without Discount

Product Name

Product A 579.388367 423.56

Product B 343.557657 727.87

Product C 468.826420 0.00

Weighted Average Discount Offered by Each Salesperson:

Salesperson

Alice 13.701645

Bob 15.566853

Charlie 8.502001

David 9.292278

dtype: float64

Percentage of Total Revenue by Region:

Total Revenue Percentage of Total Revenue

Region

East 2583.59 24.563090

North 4283.50 40.724726

South 338.10 3.214434

West 3312.99 31.497750