MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

EDS Theory Activity 01

Name : Vedika Hemant Kulkarni

Roll No : CS8-43

PRN : 202401120013

Topic : Sales Dataset

Code for my dataset and problem statement :

vedika.py > ...

```python
import pandas as pd
import numpy as np

# Sample Sales Dataset
data = {
    'order_id': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
    'date': pd.to_datetime([
        '2021-01-15', '2021-01-18', '2021-02-02', '2021-02-20', '2021-03-05',
        '2021-03-18', '2021-04-01', '2021-04-15', '2021-05-01', '2021-05-18'
    ]),
    'customer': ['Alice', 'Bob', 'Charlie', 'Alice', 'Eve', 'Bob', 'Frank', 'Grace', 'Eve', 'Alice'],
    'region': ['West', 'East', 'East', 'North', 'West', 'South', 'North', 'East', 'West', 'North'],
    'product': ['Laptop', 'Printer', 'Phone', 'Monitor', 'Tablet', 'Laptop', 'Printer', 'Phone', 'Laptop', 'Tablet'],
    'category': ['Technology']*10,
    'sales': [1200, 350, 800, 400, 300, 1400, 375, 750, 1300, 310],
    'quantity': [1, 1, 2, 1, 2, 1, 1, 1, 1, 2],
    'discount': [0.1, 0.2, 0.15, 0.05, 0.0, 0.1, 0.2, 0.1, 0.05, 0.15],
    'profit': [200, 30, 100, 60, 50, 250, 25, 90, 210, 40]
}
df = pd.DataFrame(data)

# 1. Total Sales
print("1. Total Sales:", df['sales'].sum())

# 2. Average Profit
print("2. Average Profit:", df['profit'].mean())

# 3. Top 3 Customers by Sales
print("3. Top 3 Customers by Sales:\n", df.groupby('customer')['sales'].sum().sort_values(ascending=False).head(3))
```

vedika.py > ...

```python
# 4. Most Sold Product by Quantity
print("4. Most Sold Product:", df.groupby('product')['quantity'].sum().idxmax())

# 5. Monthly Sales
print("5. Monthly Sales Summary:\n", df.groupby(df['date'].dt.to_period('M'))['sales'].sum())

# 6. Region with Highest Sales
print("6. Region with Highest Sales:", df.groupby('region')['sales'].sum().idxmax())

# 7. Total Discount Given
print("7. Total Discount Given:", (df['sales'] * df['discount']).sum())

# 8. Product with Highest Total Profit
print("8. Product with Highest Profit:", df.groupby('product')['profit'].sum().idxmax())

# 9. Average Sales Per Order
print("9. Average Sales Per Order:", df['sales'].mean())

# 10. Total Sales and Profit per Region
print("10. Sales and Profit per Region:\n", df.groupby('region')[['sales', 'profit']].sum())

# 11. Orders with Discount > 10%
print("11. Orders with >10% Discount:\n", df[df['discount'] > 0.1])

# 12. Correlation between Discount and Profit
print("12. Correlation between Discount and Profit:", df['discount'].corr(df['profit']))

# 13. Average Quantity Sold per Product
print("13. Average Quantity Sold per Product:\n", df.groupby('product')['quantity'].mean())
```

vedika.py > ...

```python
# 14. Maximum Sales in a Single Order
print("14. Maximum Sales in a Single Order:", df['sales'].max())

# 15. Number of Unique Customers
print("15. Unique Customers Count:", df['customer'].nunique())

# 16. Orders Count per Region
print("16. Orders Count per Region:\n", df['region'].value_counts())

# 17. Days with Sales > 1000
print("17. Dates with Sales > 1000:\n", df[df['sales'] > 1000][['date', 'sales']])

# 18. Cumulative Sales Over Time
df_sorted = df.sort_values('date')
df_sorted['cumulative_sales'] = df_sorted['sales'].cumsum()
print("18. Cumulative Sales Over Time:\n", df_sorted[['date', 'cumulative_sales']])

# 19. Profit Margin per Order
df['profit_margin'] = df['profit'] / df['sales']
print("19. Profit Margin per Order:\n", df[['order_id', 'profit_margin']])

# 20. Month with Lowest Average Profit
print("20. Month with Lowest Avg Profit:", df.groupby(df['date'].dt.month)['profit'].mean().idxmin())

```

vedika.py > ...

```python
# Problem Statements and Results
print("1. Total Sales:", df['sales'].sum())
print("2. Average Profit:", df['profit'].mean())
print("3. Top 3 Customers by Sales:\n", df.groupby('customer')['sales'].sum().sort_values(ascending=False).head(3))
print("4. Most Sold Product by Quantity:", df.groupby('product')['quantity'].sum().idxmax())
print("5. Monthly Sales Summary:\n", df.groupby(df['date'].dt.to_period('M'))['sales'].sum())
print("6. Region with Highest Sales:", df.groupby('region')['sales'].sum().idxmax())
print("7. Total Discount Given:", (df['sales'] * df['discount']).sum())
print("8. Product with Highest Profit:", df.groupby('product')['profit'].sum().idxmax())
print("9. Average Sales Per Order:", df['sales'].mean())
print("10. Sales and Profit per Region:\n", df.groupby('region')[['sales', 'profit']].sum())
print("11. Orders with >10% Discount:\n", df[df['discount'] > 0.1])
print("12. Correlation between Discount and Profit:", df['discount'].corr(df['profit']))
print("13. Average Quantity Sold per Product:\n", df.groupby('product')['quantity'].mean())
print("14. Maximum Sales in a Single Order:", df['sales'].max())
print("15. Unique Customers Count:", df['customer'].nunique())
print("16. Orders Count per Region:\n", df['region'].value_counts())
print("17. Dates with Sales > 1000:\n", df[df['sales'] > 1000][['date', 'sales']])
print("18. Cumulative Sales Over Time:\n", df_sorted[['date', 'cumulative_sales']])
print("19. Profit Margin per Order:\n", df[['order_id', 'profit_margin']])
print("20. Month with Lowest Avg Profit:", df.groupby(df['date'].dt.month)['profit'].mean().idxmin())
```

Problem Statement and their output :

1. Total Sales

```
1. Total Sales: 7185
```

2. Average Profit

```
2. Average Profit: 105.5
```

3. Top 3 Customers by Sales

```
3. Top 3 Customers by Sales:
 customer
Alice    1910
Bob      1750
Eve      1600
Name: sales, dtype: int64
```

4. Most Sold Product by Quantity

```
4. Most Sold Product: Tablet
```

5. Monthly Sales

```
5. Monthly Sales Summary:
 date
2021-01    1550
2021-02    1200
2021-03    1700
2021-04    1125
2021-05    1610
Freq: M, Name: sales, dtype: int64
```

6. Region with Highest Sales

```
6. Region with Highest Sales: West
```

7. Total Discount Given

```
7. Total Discount Given: 731.5
```

8. Product with Highest Total Profit

```
8. Product with Highest Profit: Laptop
```

9. Average Sales Per Order

```
9. Average Sales Per Order: 718.5
```

10. Total Sales and Profit per Region

```
10. Sales and Profit per Region:
         sales  profit
region
East      1900     220
North     1085     125
South     1400     250
West      2800     460
```

11. Orders with Discount > 10%

```
11. Orders with >10% Discount:
   order_id       date customer region product    category sales quantity discount profit profit_margin
1         2 2021-01-18      Bob   East Printer  Technology   350        1     0.20     30      0.085714
2         3 2021-02-02  Charlie   East   Phone  Technology   800        2     0.15    100      0.125000
6         7 2021-04-01    Frank  North Printer  Technology   375        1     0.20     25      0.066667
9        10 2021-05-18    Alice  North  Tablet  Technology   310        2     0.15     40      0.129032
```

12. Correlation between Discount and Profit

```
12. Correlation between Discount and Profit: -0.33488870023461303
```

13. Average Quantity Sold per Product

```
13. Average Quantity Sold per Product:
 product
Laptop     1.0
Monitor    1.0
Phone      1.5
Printer    1.0
Tablet     2.0
Name: quantity, dtype: float64
```

14. Maximum Sales in a Single Order

```
14. Maximum Sales in a Single Order: 1400
```

15. Number of Unique Customers

```
16. Orders Count per Region:
 region
West      3
East      3
North     3
South     1
Name: count, dtype: int64
```

16. Orders Count per Region


```
16. Orders Count per Region:
 region
West      3
East      3
North     3
South     1
Name: count, dtype: int64
```


17. Days with Sales > 1000

```
17. Dates with Sales > 1000:
         date  sales
0  2021-01-15   1200
5  2021-03-18   1400
8  2021-05-01   1300
```

18. Cumulative Sales Over Time

```
18. Cumulative Sales Over Time:
        date  cumulative_sales
0 2021-01-15              1200
1 2021-01-18              1550
2 2021-02-02              2350
3 2021-02-20              2750
4 2021-03-05              3050
5 2021-03-18              4450
6 2021-04-01              4825
7 2021-04-15              5575
8 2021-05-01              6875
9 2021-05-18              7185
```

19. Profit Margin per Order

```
19. Profit Margin per Order:
   order_id  profit_margin
0         1       0.166667
1         2       0.085714
2         3       0.125000
3         4       0.150000
4         5       0.166667
5         6       0.178571
6         7       0.066667
7         8       0.120000
8         9       0.161538
9        10       0.129032
```

20. Month with Lowest Average Profit