

Name :- Ankit S. Pandey.

Roll no :- ETI-88

Batch :- E-15

PRN :- 202401070149

Page No.	
Date	

Topic :- Dataset - Sales Dataset
Solution using Numpy and pandas.

1) Total revenue generated.

total revenue = sales['Revenue'].sum()

2) Average discount given.

average discount = sales['Discount'].mean()

3) Product with highest number of sales.
top-product = sales.groupby('product')
['quantity'].sum().idxmax()

4) Total quantity sold per region

quantity-per-region = sales.groupby('region')
['quantity'].sum()

5) Day with highest number of orders

sales['Date'] = pd.to_datetime(sales['Date'])
day-highest-order = sales['Date'].value_counts()
idxmax()

6) Top 5 customers by purchase value

top-customers = sales.groupby('customer_id')
['Revenue'].sum().sort_values().head(5)

7) Category generating most.

top-category = sales.groupby('category')
['Revenue'].sum().idxmax()

8) Month with highest revenue.

sales['month'] = sales['Date'].dt.month
month-highest-revenue = sales.groupby('month')
['Revenue'].sum().idxmax()

9) Average order value

→ $\text{Average-order-value} = \text{sales}['\text{Revenue}'].sum() / \text{sales}['\text{Order ID}'].nunique()$

10) Percentages of sales offering discount.

$\text{discounted-sales-percentage} = (\text{sales}[\text{sales}['Discount'] > 0].shape[0] / \text{sales}.shape[0]) * 100$

11) Region with highest average order value.

$\text{avgorder-value-region} = \text{sales}.groupby('Region')['Revenue'].mean().idxmax()$

12) Trend of monthly sales.

$\text{monthly-sales} = \text{trend_plot}(\text{kind} = 'line')$

13) Number of unique product sold.

$\text{unique-product-sold} = \text{sales}['product'].nunique()$

14) Customers with more than 10 purchases.

→ $\text{frequency-customers} = \text{sales}['CustomerID'].value_counts().country$
 $\text{frequency-customers} = \text{frequency-customer}[\text{frequency-customer} > 10]$

15) Correlation between quantity and revenue

→ $\text{Correlation} = \text{sales}['Quantity'].corr(\text{sales}['Revenue'])$

16] Product never sold.
product_list = pd.read_csv('product_list.csv')
sold_products = sales['product'].unique()

17] Max, min, mean revenue per order.
→ revenue_per_order = sales.groupby('Order ID')['Revenue'].sum()
max_revenue = revenue_per_order.max()
min_revenue = revenue_per_order.min()
mean_revenue = revenue_per_order.mean()

18] Most common discount rate.
→ Common_discount = sales['Discount'].mode()[0]

19] fastest growing product category by month

→ category_growth = sales.groupby(['month', 'category'])['Revenue'].sum().unstack().pct_change().mean()
fastest_growing_category = category_growth()

20] Average quantity sold per order for each product.

→ avg_quantity_per_order = sales.groupby('product')['Quantity'].mean()