ACTIVITY 2

```
# TECHMART SALES DATA ANALYSIS CHALLENGE
# ========================================


# Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
    df = pd.read_csv("techmart_sales.csv")
    print("Dataset Loaded Successfully!")
except:
    print("Creating Sample Dataset...")
    np.random.seed(42)
 data = {
        "Order_ID": range(1001, 1101),
        "Order_Date": pd.date_range(start="2025-01-01", periods=100, freq='D'),
        "Region": np.random.choice(["East", "West", "North", "South"], 100),
        "Product_Category": np.random.choice(["Laptops", "Smartphones", "Accessories", "Home
Appliances"], 100),
        "Quantity": np.random.randint(1, 5, 100),
        "Unit_Price": np.random.randint(500, 50000, 100),
        "Discount": np.random.choice([0, 5, 10, 15], 100),
        "Customer_Age": np.random.randint(18, 60, 100)
    }
df = pd.DataFrame(data)


    # Create Sales column
```

```
    df["Sales"] = df["Quantity"] * df["Unit_Price"] * (1 - df["Discount"]/100)
    df["Profit"] = df["Sales"] * np.random.uniform(0.1, 0.3, 100)
    df.to_csv("techmart_sales.csv", index=False)
    print("Sample Dataset Created!")
print("\nChecking Missing Values:\n", df.isnull().sum())
df.fillna({
    "Quantity": df["Quantity"].median(),
    "Discount": 0
}, inplace=True)
df.drop_duplicates(inplace=True)
df["Order_Date"] = pd.to_datetime(df["Order_Date"])
df["Month"] = df["Order_Date"].dt.month_name()
df["Profit_Margin"] = (df["Profit"] / df["Sales"]) * 100
def age_group(age):
    if age <= 25:
        return "18-25"
    elif age <= 35:
        return "26-35"
    elif age <= 50:
        return "36-50"
    else:
        return "50+"
df["Age_Group"] = df["Customer_Age"].apply(age_group)
print("\n========== BUSINESS ANALYSIS ==========")
total_sales = df["Sales"].sum()
total_profit = df["Profit"].sum()
print("\nTotal Sales:", round(total_sales,2))
```

```
print("Total Profit:", round(total_profit,2))
# Sales by Region
region_sales = df.groupby("Region")["Sales"].sum()
print("\nSales by Region:\n", region_sales)
category_profit = df.groupby("Product_Category")["Profit"].sum()
print("\nProfit by Category:\n", category_profit)
monthly_sales = df.groupby("Month")["Sales"].sum()
print("\nMonthly Sales:\n", monthly_sales)
age_sales = df.groupby("Age_Group")["Sales"].sum()
print("\nSales by Age Group:\n", age_sales)
print("\n========== PIVOT TABLES ==========")
pivot_region_category = pd.pivot_table(
    df,
    values="Sales",
    index="Region",
    columns="Product_Category",
    aggfunc="sum"
)

print("\nRegion vs Category (Sales):\n", pivot_region_category)

pivot_month_category = pd.pivot_table(
    df,
    values="Profit",
    index="Month",
    columns="Product_Category",
    aggfunc="sum"
```

```
)
print("\nMonth vs Category (Profit):\n", pivot_month_category)
plt.figure()
region_sales.plot(kind="bar")
plt.title("Sales by Region")
plt.xlabel("Region")
plt.ylabel("Total Sales")
plt.show()
plt.figure()
category_profit.plot(kind="bar")
plt.title("Profit by Category")
plt.xlabel("Product Category")
plt.ylabel("Total Profit")
plt.show()
plt.figure()
monthly_sales.plot(kind="line", marker="o")
plt.title("Monthly Sales Trend")
plt.xlabel("Month")
plt.ylabel("Total Sales")
plt.xticks(rotation=45)
plt.show()
df.to_csv("techmart_sales_cleaned.csv", index=False)
print("\nCleaned dataset saved as techmart_sales_cleaned.csv"
print("\n========== KEY INSIGHTS ==========")
print("• The highest sales region is:", region_sales.idxmax())
print("• The most profitable category is:", category_profit.idxmax())
print("• The age group contributing most sales is:", age_sales.idxmax())
```

print("• Highest sales month is:", monthly_sales.idxmax())

print("\nAnalysis Completed Successfully!")

Creating Sample Dataset...

Sample Dataset Created!

Checking Missing Values:

Order_ID 0

Order_Date 0

Region 0

Product_Category 0

Quantity 0

Unit_Price 0

Discount 0

Customer_Age 0

Sales 0

Profit 0

dtype: int64

========== BUSINESS ANALYSIS ==========

Total Sales: 5853655.6

Total Profit: 1266227.1

Sales by Region:

Region

East 1157231.10

North 1355072.50

South 2031559.05

West 1309792.95

Name: Sales, dtype: float64

Profit by Category:

Product_Category

Accessories 420082.963470

Home Appliances 309784.089748

Laptops 290981.803233

Smartphones 245378.247963

Name: Profit, dtype: float64

Monthly Sales:

Month

April 613450.85

February 1595030.50

January 1958755.50

March 1686418.75

Name: Sales, dtype: float64

Sales by Age Group:

Age_Group

18-25 1051804.55

26-35 1021125.05

36-50 2565597.75

50+ 1215128.25

Name: Sales, dtype: float64

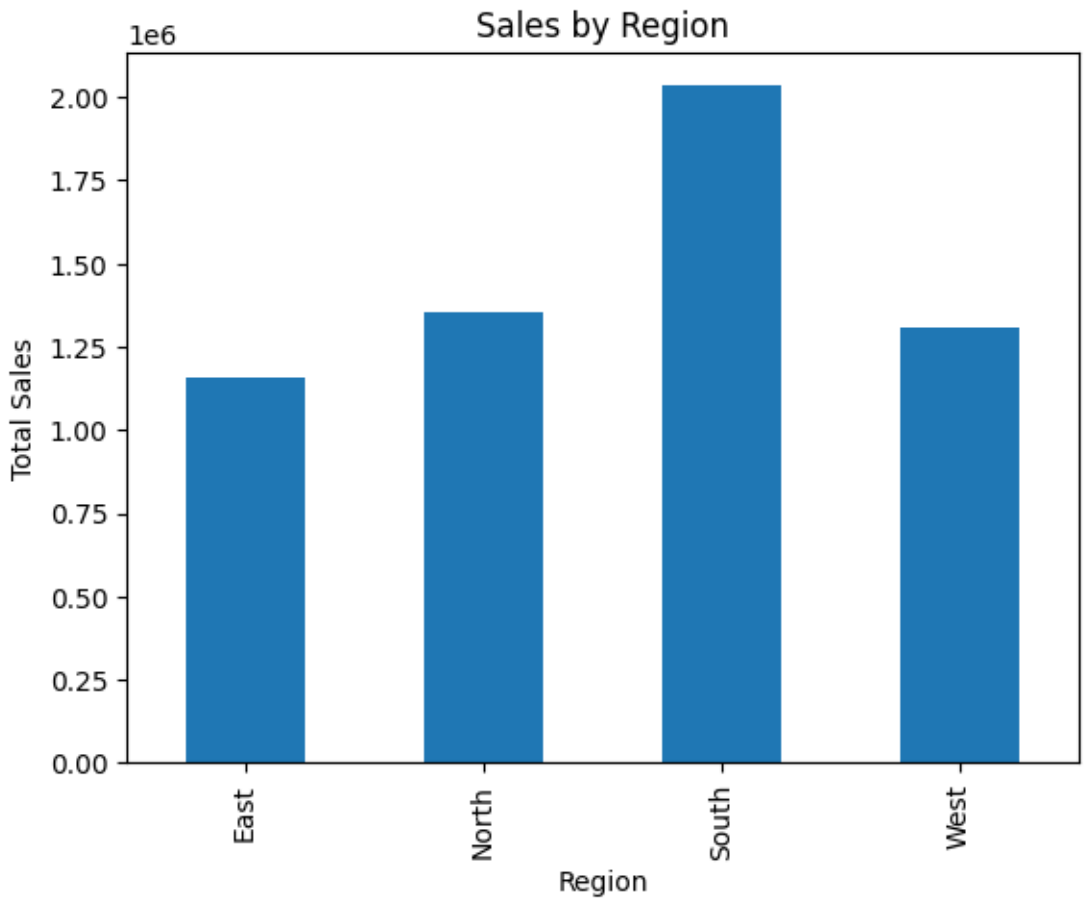
========== PIVOT TABLES ==========

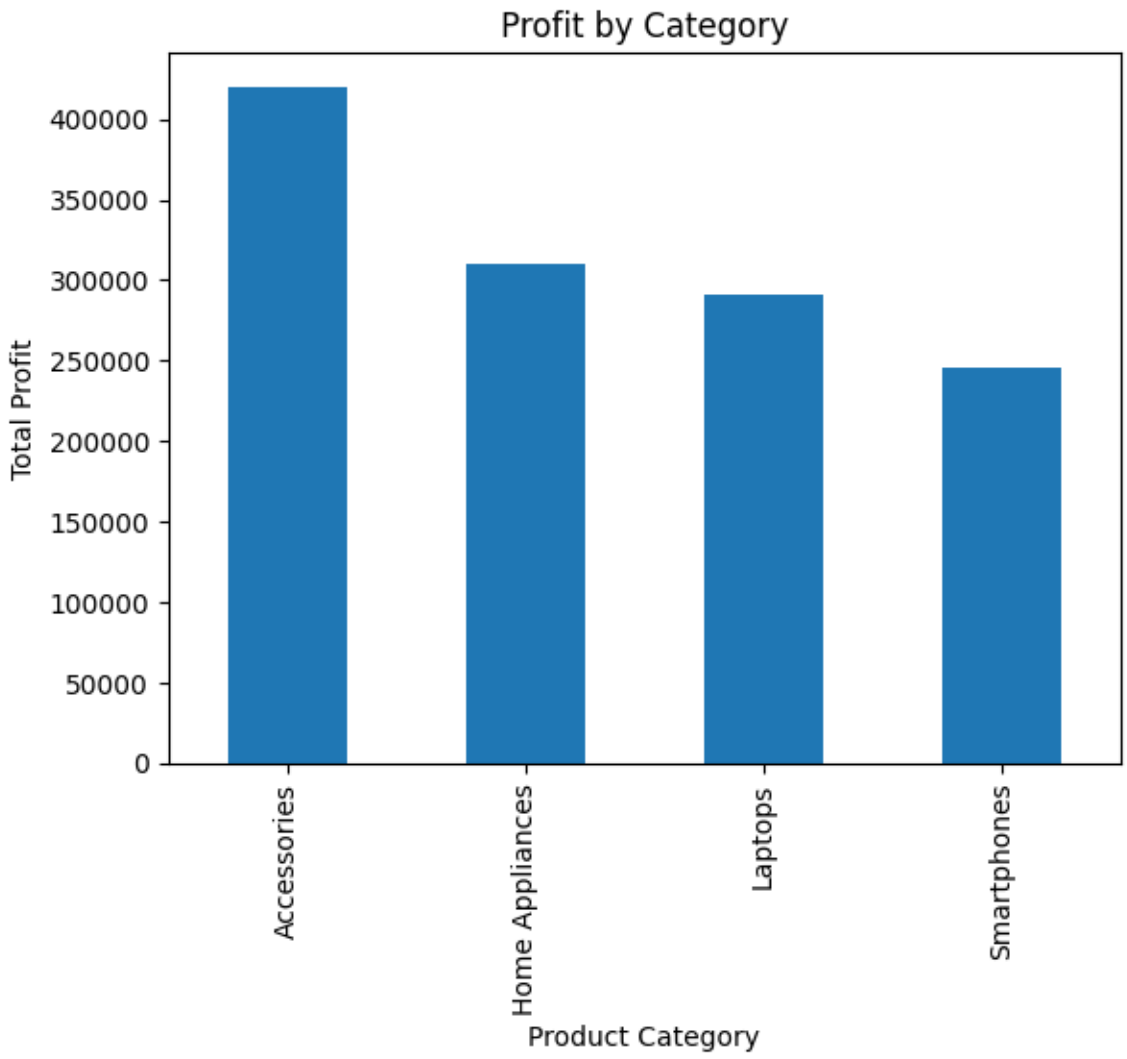
Region vs Category (Sales):

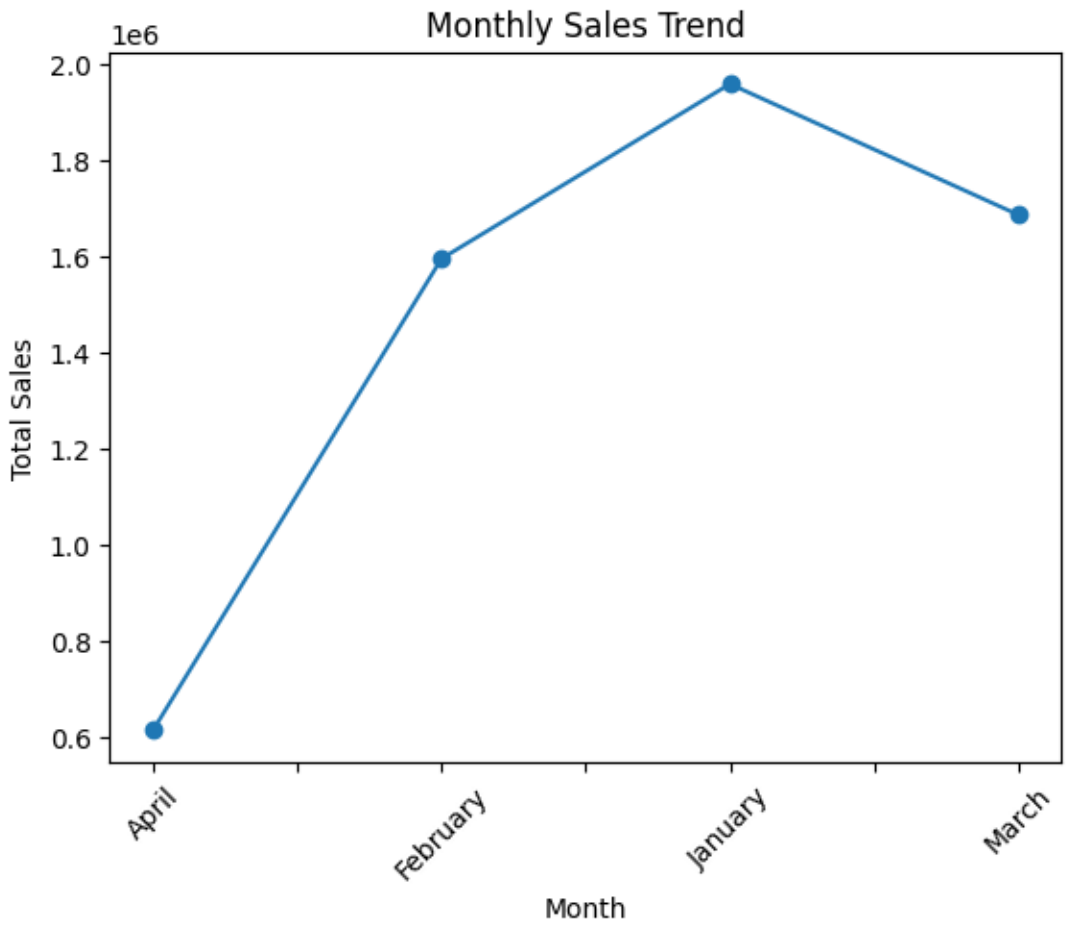
Product_Category	Accessories	Home Appliances	Laptops	Smartphones
Region				
East	380931.20	321602.40	248013.6	206683.9
North	229831.65	133674.65	631069.9	360496.3
South	959930.35	564481.90	241555.8	265591.0
West	204444.20	473787.75	338000.3	293560.7

Month vs Category (Profit):

Product_Category	Accessories	Home Appliances	Laptops	Smartphones
Month				
April	70923.154633	28247.517989	7074.604889	32337.215188
February	105171.055522	58386.500386	141127.618573	20151.710276
January	100512.360762	118001.723356	52060.434592	146011.700096
March	143476.392553	105148.348018	90719.145180	46877.622404


Sales by Region
1e6
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Total Sales
East
North
South
West
Region


Profit by Category
Total Profit
400000
350000
300000
250000
200000
150000
100000
50000
0
Accessories
Home Appliances
Laptops
Smartphones
Product Category



Cleaned dataset saved as techmart_sales_cleaned.csv

========== KEY INSIGHTS ==========

• The highest sales region is: South

• The most profitable category is: Accessories

• The age group contributing most sales is: 36-50

• Highest sales month is: January

Analysis Completed Successfully!