import pandas as pd import numpy as np

import matplotlib.pyplot as plt import seaborn as sns

from scipy import stats

*# Load dataset*

df = pd.read\_csv(r"C:\Users\maya0\OneDrive\Desktop\New folder\fashion\_boutique\_sampl

*# Display first rows*

df.head()

*# Shape of dataset*

print("Shape:", df.shape)

*# Data types* print("\nData Types:") print(df.dtypes)

*# Missing values* print("\nMissing Values:") print(df.isnull().sum())

*# Duplicate rows*

print("\nDuplicate Rows:", df.duplicated().sum())

*# Descriptive statistics for numeric columns*

df.describe()

*# Most common product categories*

df[Vproduct\_categoryV].value\_counts()

*# Correlation between numeric columns*

df.select\_dtypes(include=VnumberV).corr()

Q1 = df[VpriceV].quantile(0.25) Q3 = df[VpriceV].quantile(0.75)

IQR = Q3 - Q1

outliers = df[(df[VpriceV] < Q1 - 1.5\*IQR) | (df[VpriceV] > Q3 + 1.5\*IQR)] outliers

*# Histogram of price* plt.figure(figsize=(6,4)) sns.histplot(df[VpriceV], kde=True) plt.title("Price Distribution") plt.show()

*# Barplot of categories*

plt.figure(figsize=(6,4))

sns.countplot(data=df, x=Vproduct\_categoryV, order=df[Vproduct\_categoryV].value\_coun plt.title("Product Category Counts")

plt.show()

# Conclusions & Insights

1. **Dataset Size & Structure**
   * The dataset contains **10 rows** and **9 columns**, with no missing values or dupli- cates detected.
   * Key features include: purchase\_date, product\_category, price, quan- tity, and region.

# Data Quality

* + All fields are complete, with no null values.
  + Data types are consistent: numeric columns (price, quantity) and categorical columns (product\_category, payment\_method, etc.) are correctly formatted.

# Key Patterns

* + The **most common product categories** are **Dresses** and **Footwear** (3 purchases each).
  + The **average price** across products is **2,840**, with most prices between **2,050** and

# 3,150.

1. **Statistical Insights**
   * **Median price** is 2,600, indicating a slight right skew in pricing.
   * **Correlation analysis** shows a **moderate negative relationship** (-0.44) between price and quantity, meaning higher-priced items tend to be bought in smaller quantities.

# Anomalies & Outliers

* + One potential outlier in the price column was detected — a product priced at 5,000, significantly higher than others.
  + This could represent a premium product or a special edition.

# Business Recommendations

* + Continue offering a balanced product mix, but consider bundling premium items with complementary lower-priced products to increase sales volume.
  + For top-selling categories like **Dresses** and **Footwear**, run targeted promotions to maintain demand.
  + Monitor high-priced items to determine if they should remain niche or be ad- justed to improve volume sales.