KIRANA STORE DEMAND FORECASTING

TEAM MEMBERS: Vanshika Dixit, Aadya Diwan, Yukti Vadehra, Khushi

Problem Statement

Objective:
To help kirana (local retail) store owners make smarter inventory decisions by predicting demand for products in advance.

Why it matters:

- Avoid stockouts & overstock
- Reduce waste
- Improve customer satisfaction

Dataset Overview

KEY FEATURES:

•Description: Product description.
•InvoiceDate: Date of transaction.
•Quantity: Number of units sold.
•CustomerID: Unique identifier for customers.
•UnitPrice: Price per unit.

OBSERVATIONS:

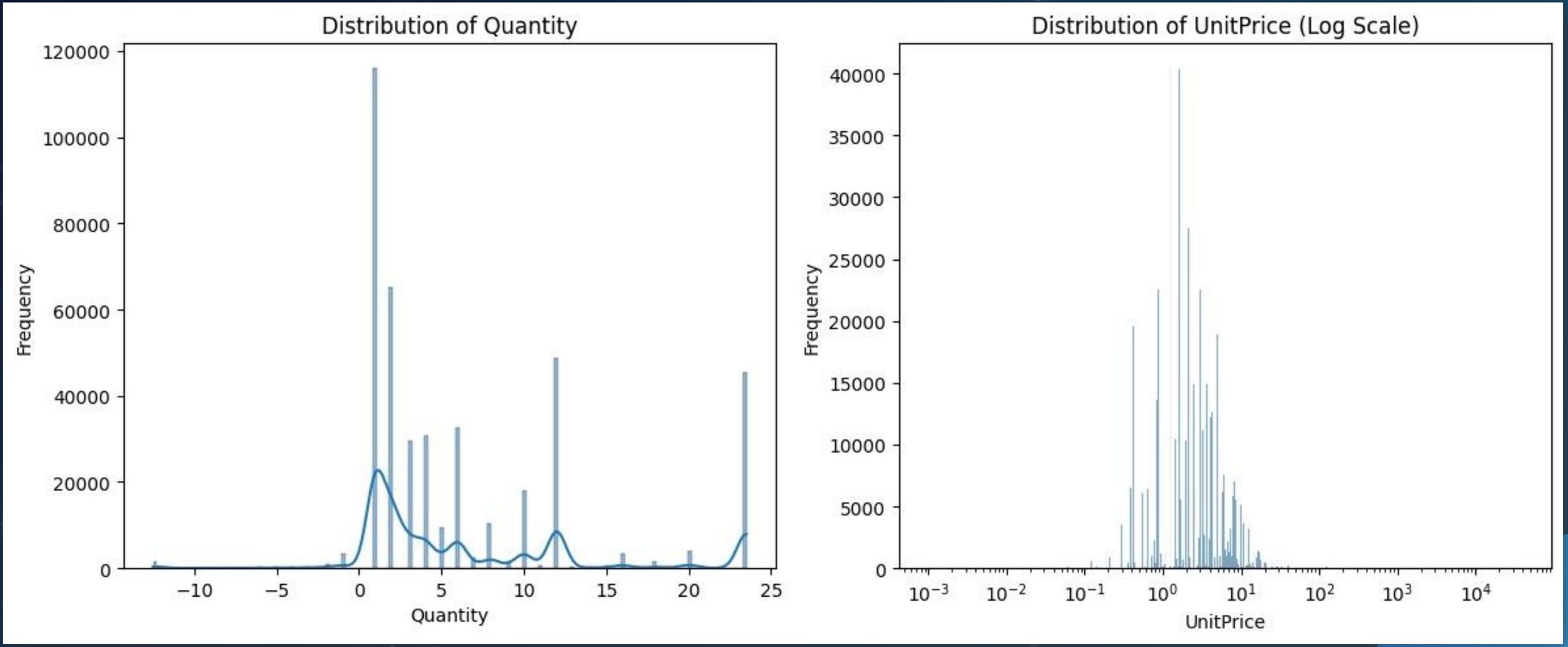
•Presence of missing values in CustomerID and Description.
•Duplicate entries detected.
•Outliers in Quantity values

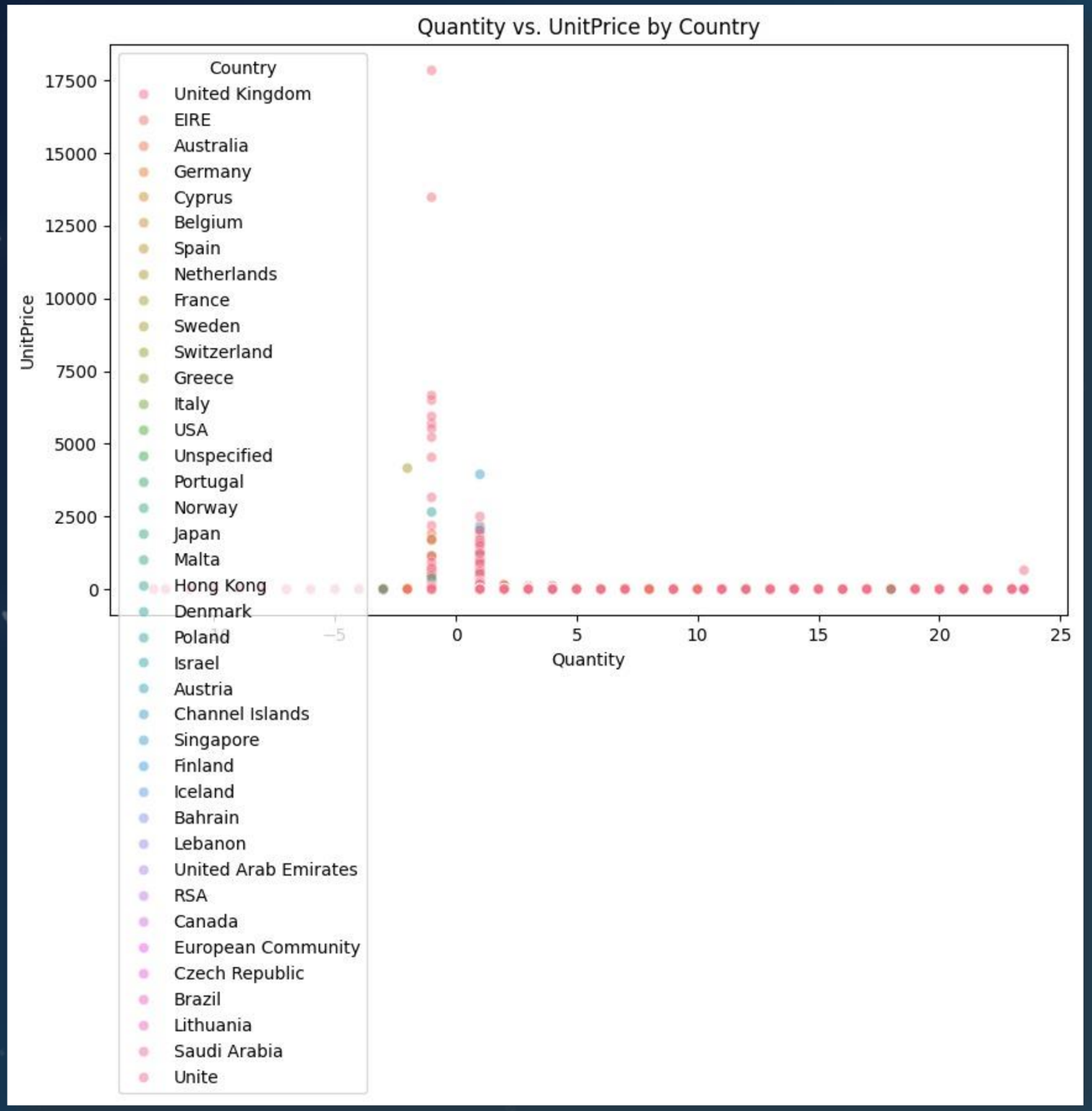
DATA CLEANING STEPS

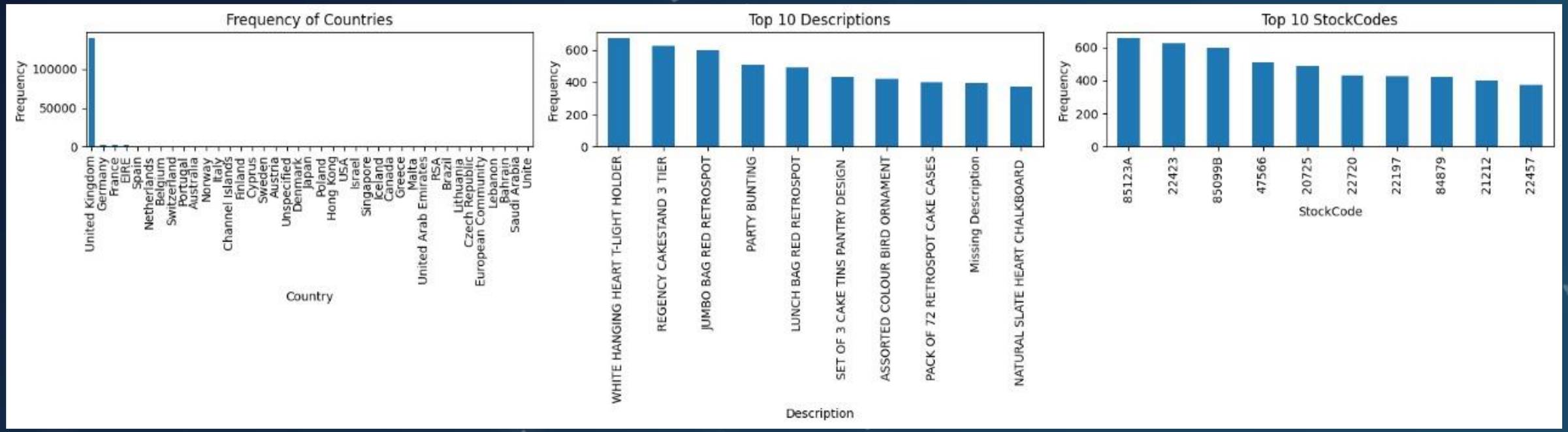
•**Handling Missing Values:**
•missing CustomerID with the mode (most frequent value).
•Replaced missing Description entries with "Missing Description".
•Data Type Conversion:
•Converted InvoiceDate to datetime format.
•Cast CustomerID to integer type.
•**Outlier Treatment**:
•Applied Interquartile Range (IQR) method to detect outliers in Quantity.
•Clipped outliers to the IQR boundaries to mitigate their impact.
•**Duplicate Removal**:
•Identified and removed duplicate rows to ensure data integrity.

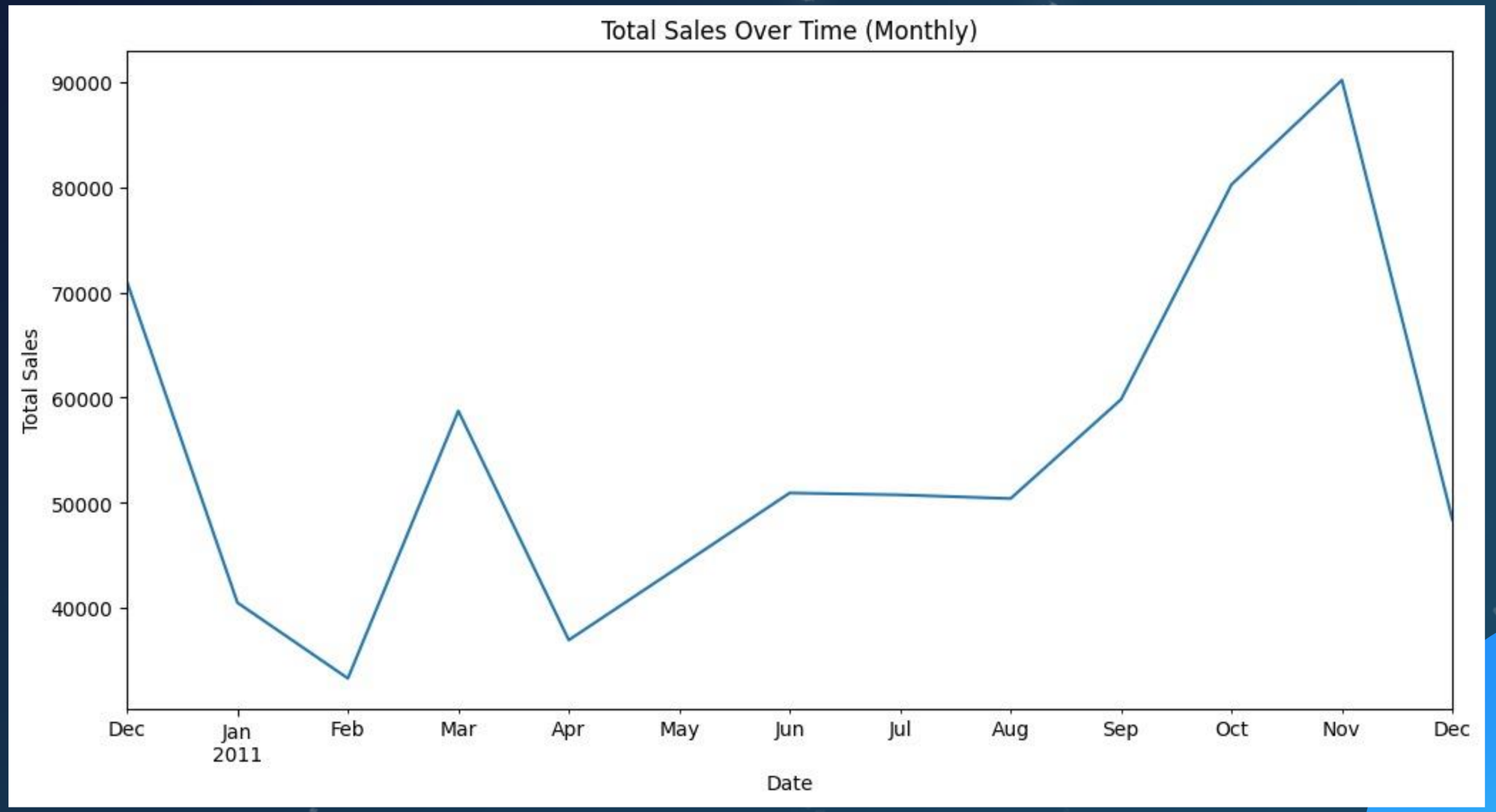
DATA VISUALIZATION

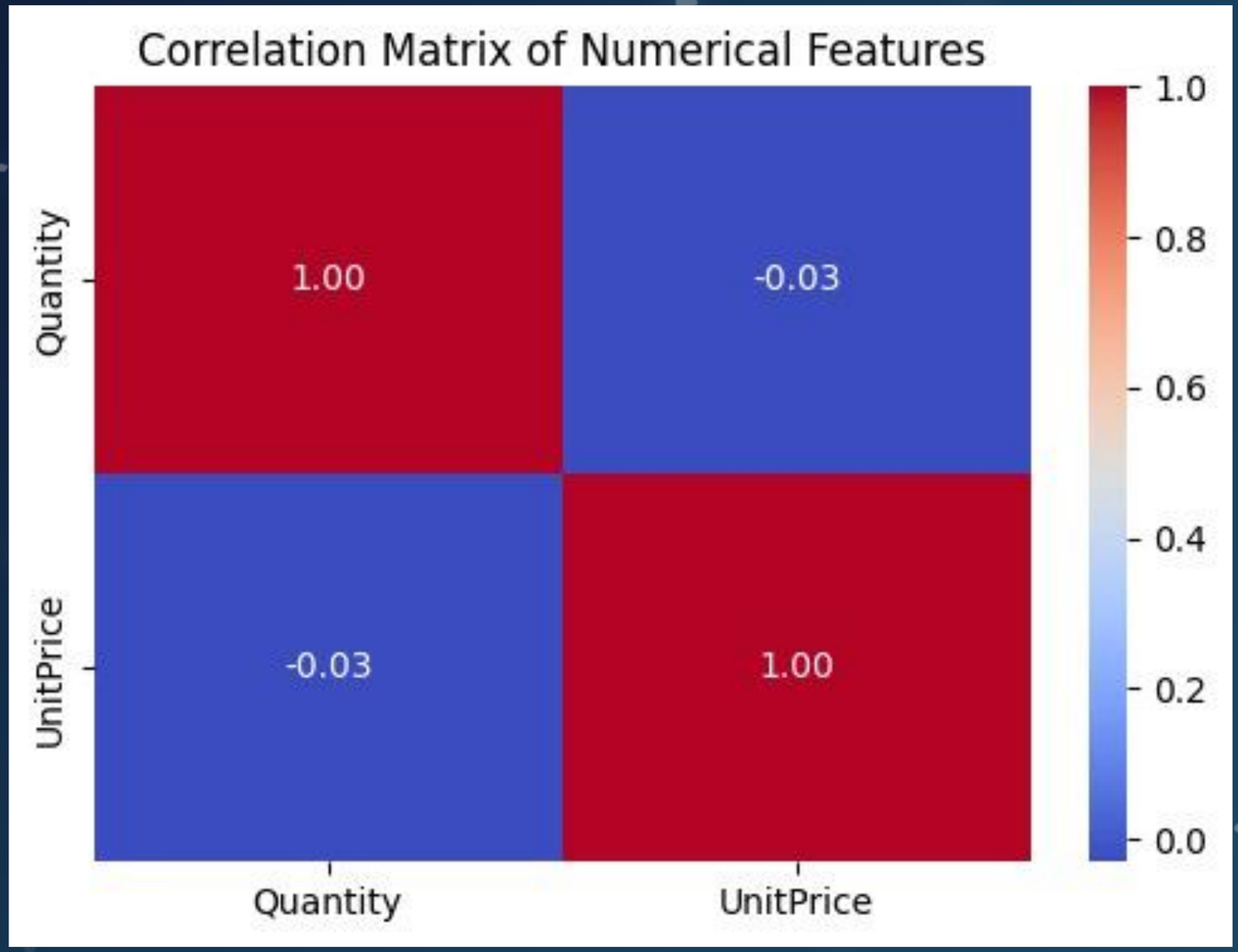
•Sales Trends:
•Line charts depicting sales over time.
•Product Performance:
•Bar charts showcasing top-selling products.
•Customer Behavior:
•Histograms illustrating purchase frequency.
•Outlier Detection:
•Boxplots used to visualize distribution and identify outliers in quantity.

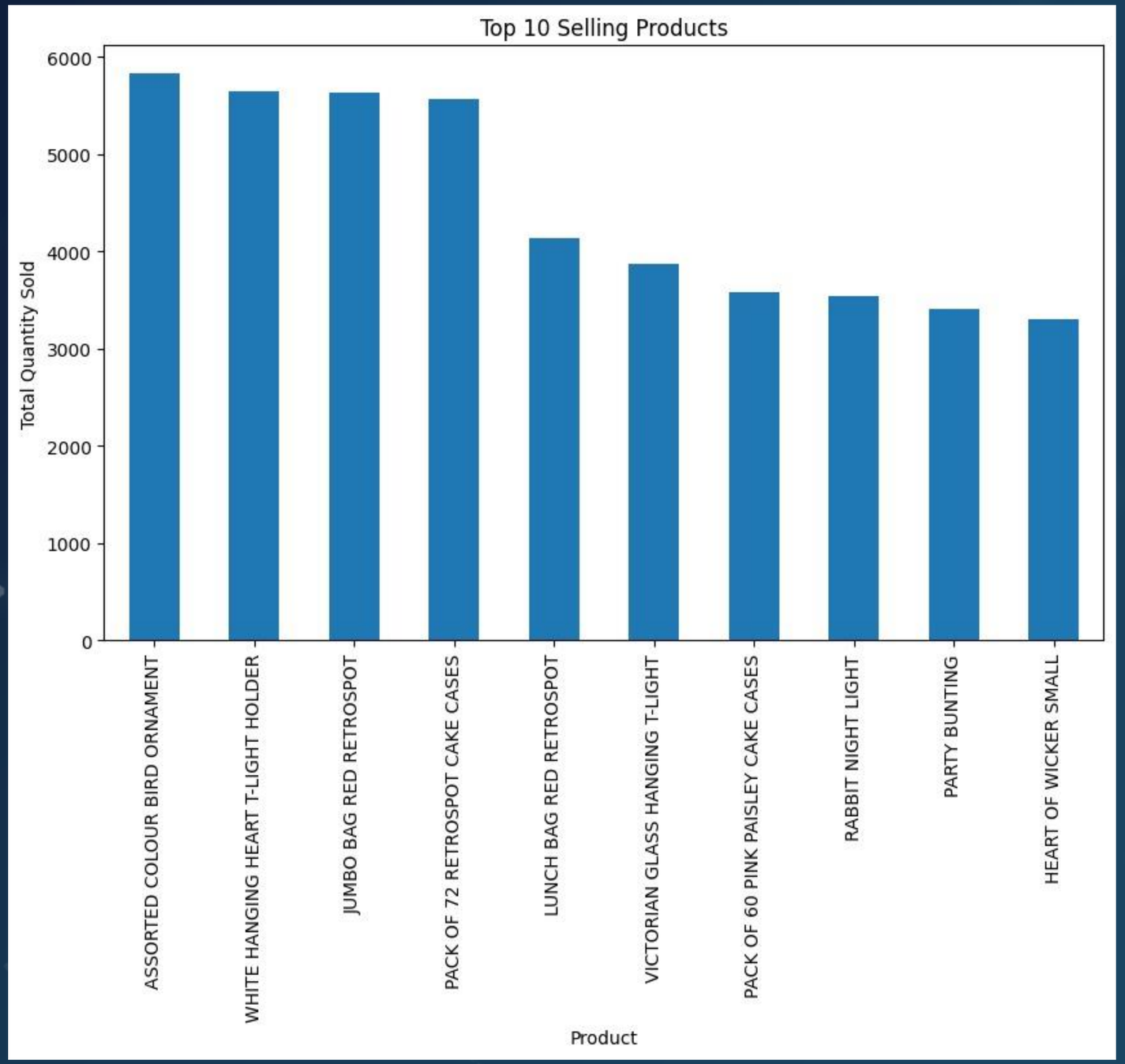

Distribution of Quantity
Frequency
Quantity
120000
100000
80000
60000
40000
20000
0
−10
−5
0
5
10
15
20
25
Distribution of UnitPrice (Log Scale)
Frequency
UnitPrice
40000
35000
30000
25000
20000
15000
10000
5000
0
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}


Quantity vs. UnitPrice by Country
Country
United Kingdom
EIRE
Australia
Germany
Cyprus
Belgium
Spain
Netherlands
France
Sweden
Switzerland
Greece
Italy
USA
Unspecified
Portugal
Norway
Japan
Malta
Hong Kong
Denmark
Poland
Israel
Austria
Channel Islands
Singapore
Finland
Iceland
Bahrain
Lebanon
United Arab Emirates
RSA
Canada
European Community
Czech Republic
Brazil
Lithuania
Saudi Arabia
Unite
UnitPrice
17500
15000
12500
10000
7500
5000
2500
0
−5
0
5
10
15
20
25
Quantity


Frequency of Countries
Frequency
100000
50000
0
United Kingdom
Germany
France
EIRE
Spain
Netherlands
Belgium
Switzerland
Portugal
Australia
Norway
Italy
Channel Islands
Finland
Cyprus
Sweden
Austria
Unspecified
Denmark
Japan
Poland
Hong Kong
USA
Israel
Singapore
Iceland
Canada
Greece
Malta
United Arab Emirates
RSA
Brazil
Lithuania
Czech Republic
European Community
Lebanon
Bahrain
Saudi Arabia
Unite
Country
Top 10 Descriptions
Frequency
600
400
200
0
WHITE HANGING HEART T-LIGHT HOLDER
REGENCY CAKESTAND 3 TIER
JUMBO BAG RED RETROSPOT
PARTY BUNTING
LUNCH BAG RED RETROSPOT
SET OF 3 CAKE TINS PANTRY DESIGN
ASSORTED COLOUR BIRD ORNAMENT
PACK OF 72 RETROSPOT CAKE CASES
Missing Description
NATURAL SLATE HEART CHALKBOARD
Description
Top 10 StockCodes
Frequency
600
400
200
0
85123A
22423
85099B
47566
20725
22720
22197
84879
21212
22457
StockCode


Total Sales Over Time (Monthly)
90000
80000
70000
60000
50000
40000
Total Sales
Dec
Jan
2011
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Date


Correlation Matrix of Numerical Features
Quantity
UnitPrice
1.00
-0.03
-0.03
1.00
Quantity
UnitPrice
1.0
0.8
0.6
0.4
0.2
0.0


Top 10 Selling Products
Total Quantity Sold
6000
5000
4000
3000
2000
1000
0
ASSORTED COLOUR BIRD ORNAMENT
WHITE HANGING HEART T-LIGHT HOLDER
JUMBO BAG RED RETROSPOT
PACK OF 72 RETROSPOT CAKE CASES
LUNCH BAG RED RETROSPOT
VICTORIAN GLASS HANGING T-LIGHT
PACK OF 60 PINK PAISLEY CAKE CASES
RABBIT NIGHT LIGHT
PARTY BUNTING
HEART OF WICKER SMALL
Product

SUMMARY

Correlation Analysis:
A heatmap visualizes the correlation between numerical features.Hypothesis Testing: A t-test compares average unit prices between the UK and other countries.Customer Segmentation: RFM analysis segments customers based on recency, frequency, and monetary value.Time Series Analysis: A line plot shows the monthly sales trend.Product Analysis: A bar chart displays the top 10 selling products.