CUSTOMER BEHAVIOUR ANALYSIS

Customer behaviour analysis looks at how customers make decisions, their product and buying preferences and how they respond to various initiatives, including marketing campaigns, product updates and offers. An analysis can include the psychological factors and social influences that lead to a decision.

Customer behaviour analysis is important because it provides valuable insights for businesses to tailor their offerings, enhance customer experiences and create effective marketing campaigns and product updates. Understanding customer decisions and preferences is essential for long-term success in the marketplace.

Objective:

To analyse customer purchasing behaviour and identify patterns that drive revenue, customer value, and churn risk, enabling data-driven business decisions.

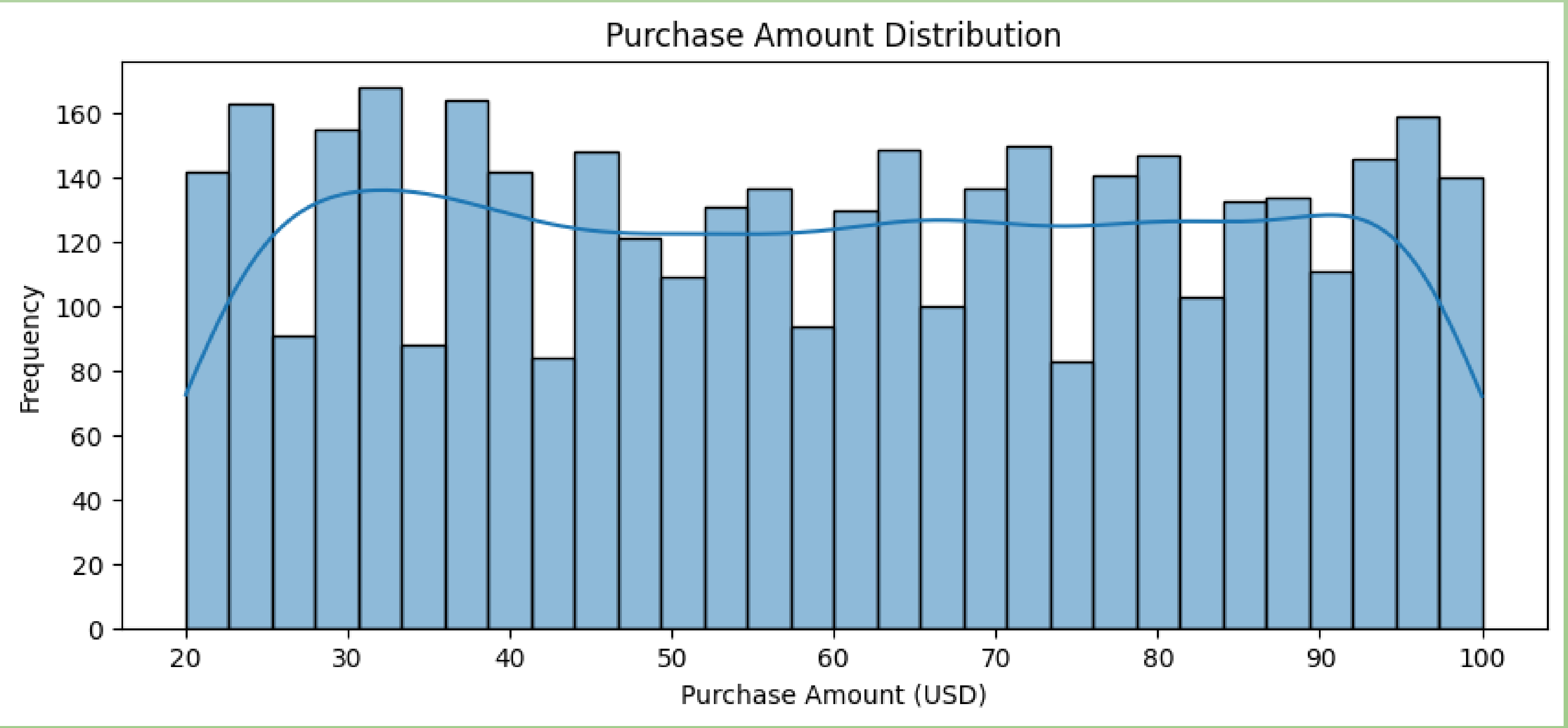
Approach:

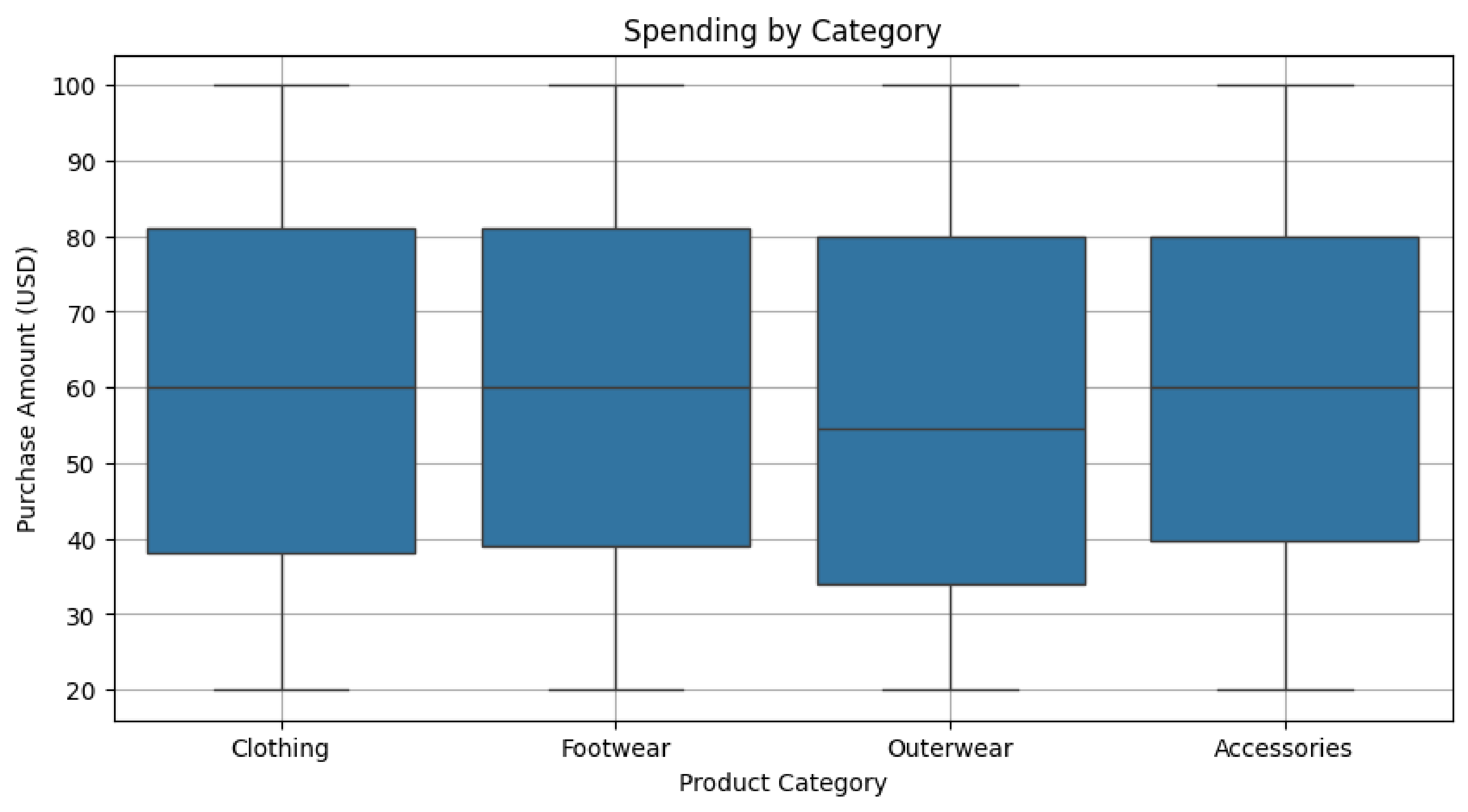
The analysis was conducted using Python with a structured analytical framework involving:

- Data cleaning and preparation
- Exploratory Data Analysis (EDA)
- Outlier detection using IQR
- Statistical testing (t-test, correlation analysis)
- Behavioural segmentation
- Predictive churn risk scoring
- Business-oriented interpretation of results

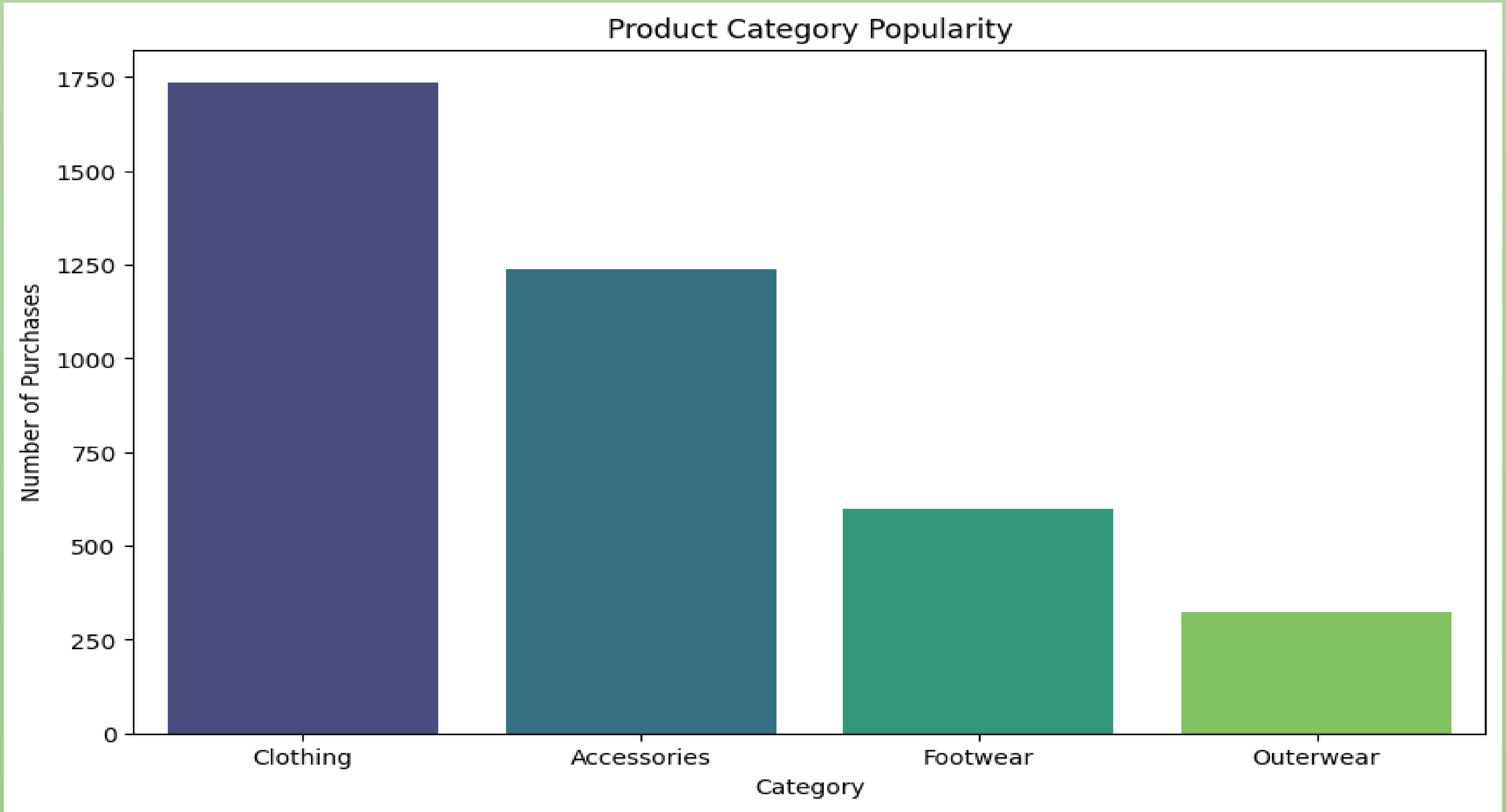
Key Analytical Focus Areas:

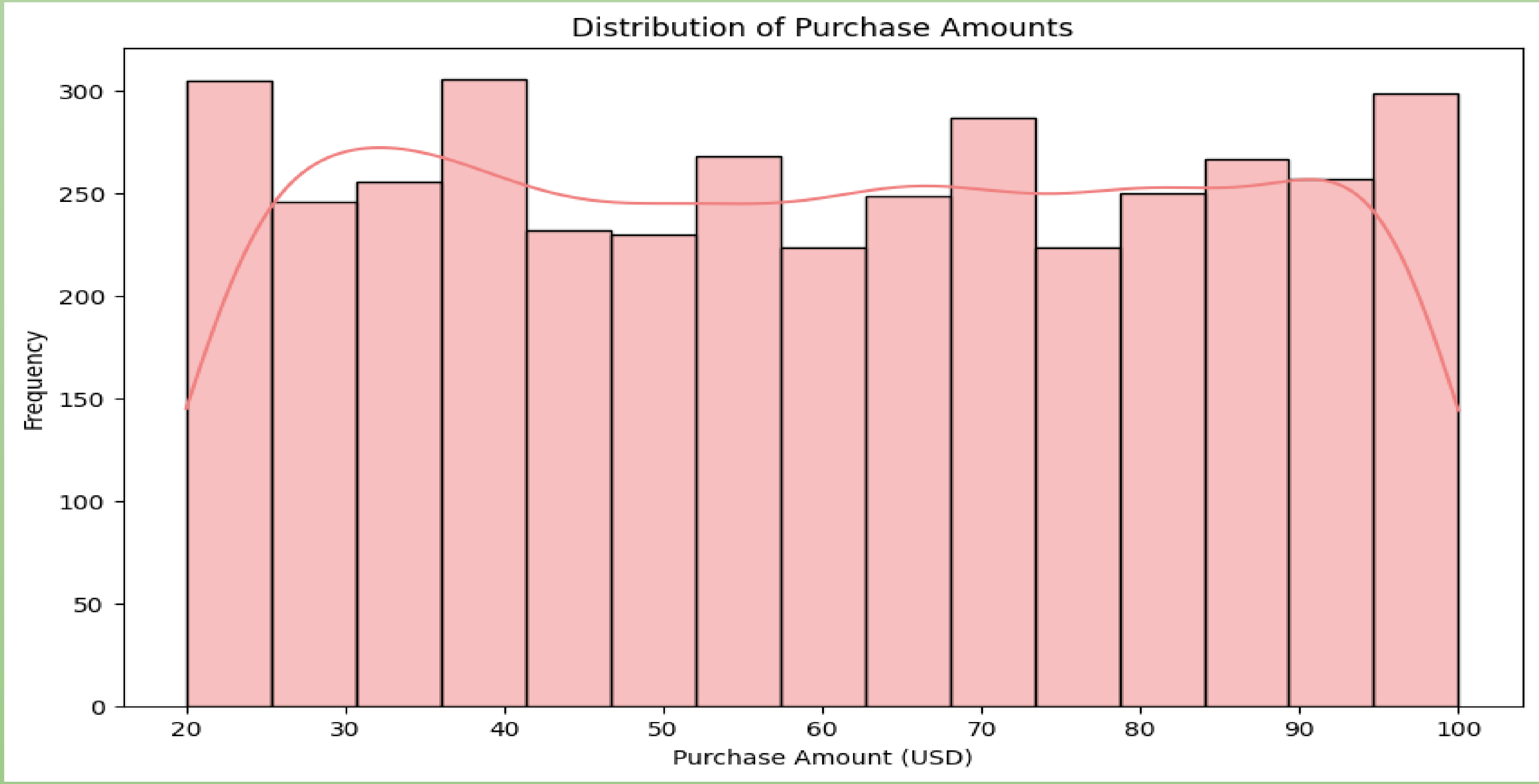
- Spending behaviour analysis
- Demographic influence on purchases
- Location-based purchasing trends
- Gender-based statistical comparison
- Purchase distribution and skewness
- High-value and risky customer detection
- Churn risk modelling
- Customer loyalty indicators (previous purchases)


Purchase Amount Distribution
Frequency
0
20
40
60
80
100
120
140
160
20
30
40
50
60
70
80
90
100
Purchase Amount (USD)

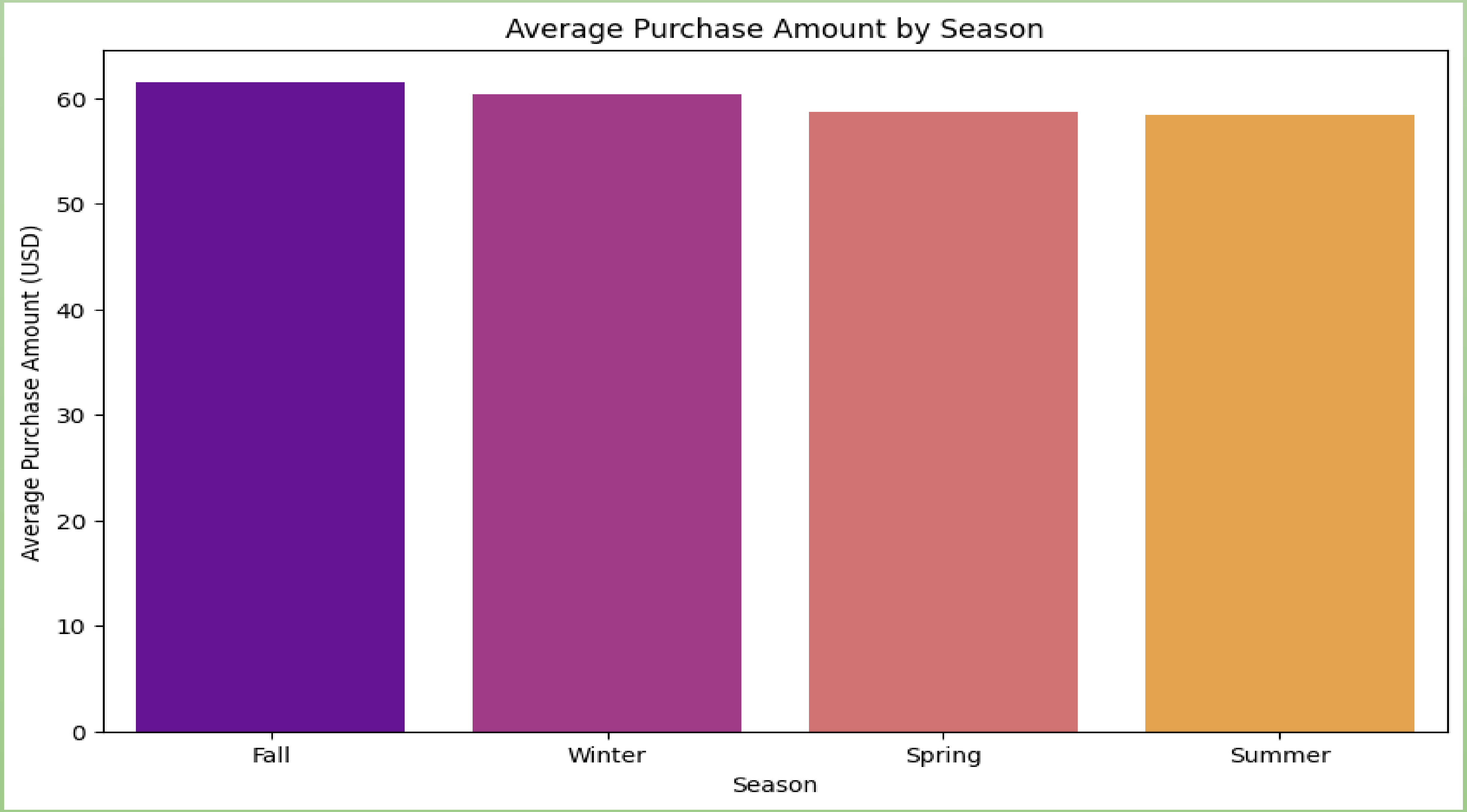


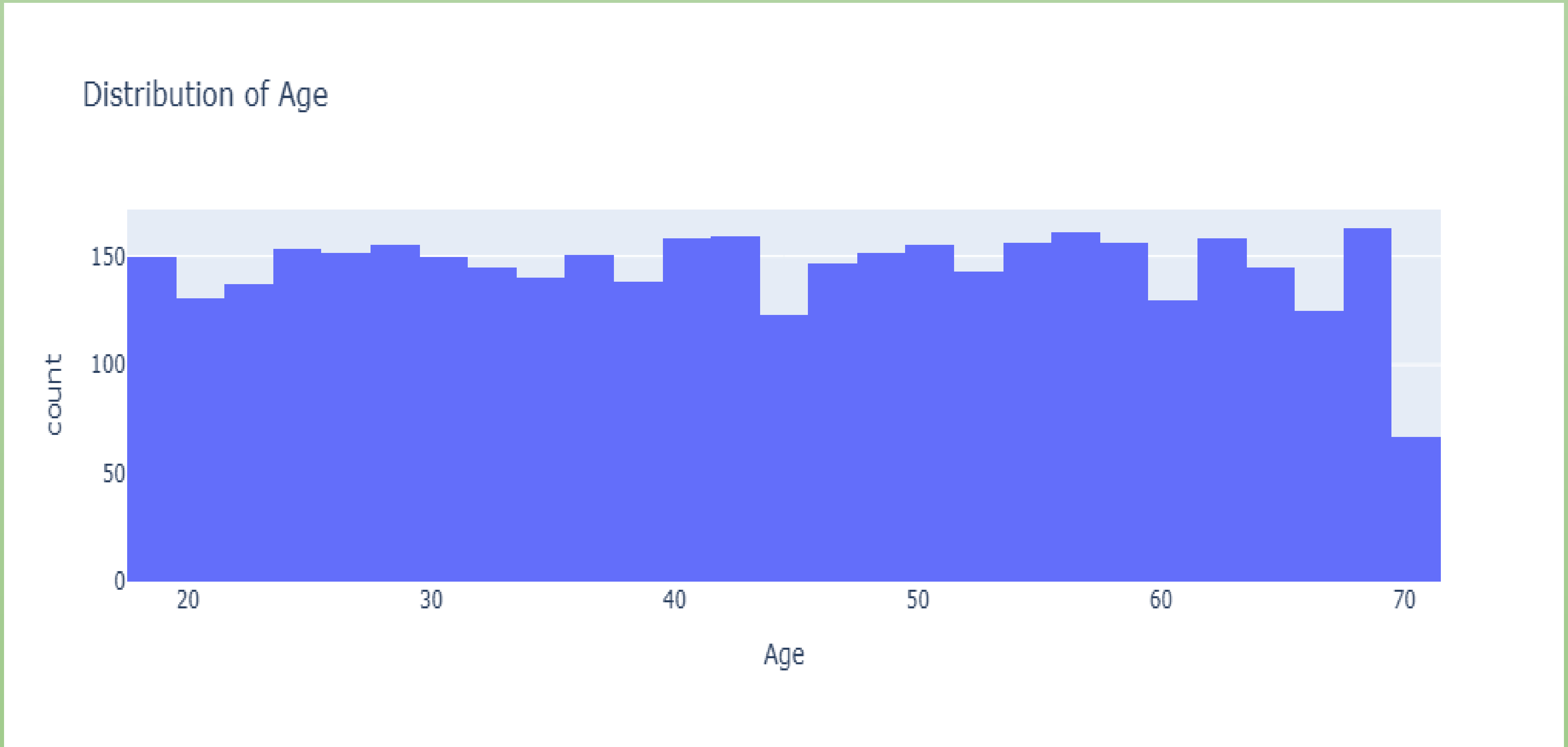
Spending by Category
Purchase Amount (USD)
100
90
80
70
60
50
40
30
20
Clothing
Footwear
Outerwear
Accessories
Product Category


Product Category Popularity
1750
1500
1250
1000
750
500
250
0
Number of Purchases
Clothing
Accessories
Footwear
Outerwear
Category



Distribution of Purchase Amounts
300
250
200
150
100
50
0
Frequency
20
30
40
50
60
70
80
90
100
Purchase Amount (USD)


Average Purchase Amount by Season
Average Purchase Amount (USD)
0
10
20
30
40
50
60
Fall
Winter
Spring
Summer
Season


Distribution of Age
count
150
100
50
0
20
30
40
50
60
70
Age


Previous Purchases by Gender
Previous Purchases
0
1
2
3
Female
Male
Gender


Previous Purchases by Location
Previous Purchases
0
10
20
30
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
Location

Gender-Based Purchase Comparison (Hypothesis Testing)

Testing if gender affects spending.

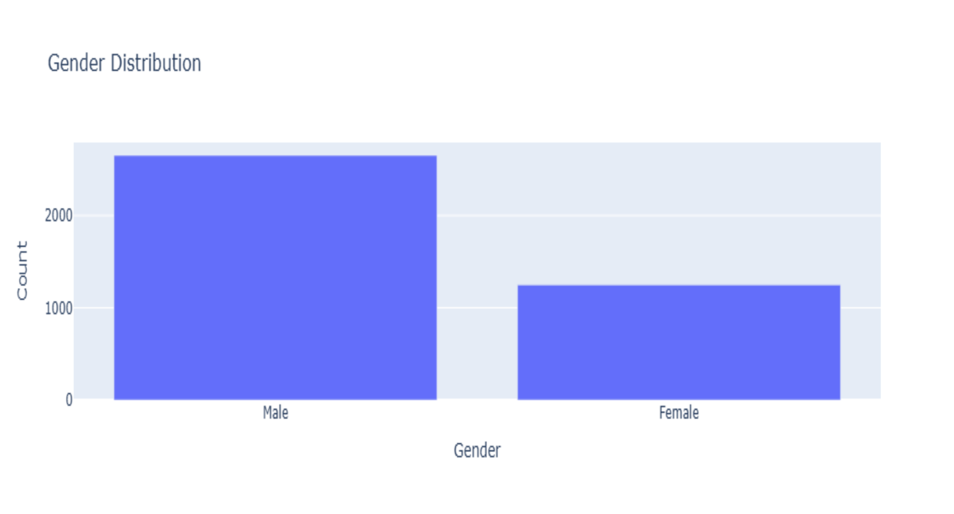
[42]:
```
from scipy import stats
import scipy.stats as stats
```

[52]:
```
#Group Data
male = df[df['Gender'] == 'Male']['Purchase Amount (USD)']
female = df[df['Gender'] == 'Female']['Purchase Amount (USD)']
```

[53]:
```
#Independent t-test
_, p_value = stats.ttest_ind(a=male, b=female,equal_var=False)
```

[54]:
```
if p_value < 0.05:
    print("That means gender affecting spending.")
else:
    print("That means gender doesn't affect spending.")
```

```
That means gender doesn't affect spending.
```



Correlation Analysis

Measure relationship between age & spending.

[91]:
```
df[['Age', 'Purchase Amount (USD)']].corr()
```

[91]:

	Age	Purchase Amount (USD)
Age	1.000000	-0.010424
Purchase Amount (USD)	-0.010424	1.000000

[92]:
```
sns.scatterplot(x='Age', y='Purchase Amount (USD)', data=df)
plt.title("Age vs Purchase Amount")
plt.show()
```


```
The plot shows NO CORRELATION between AGE & PURCHASE AMOUNT. The value -0.010124 is very close to zero, indicating that there is virtually no linear
relationship between AGE & PURCHASE AMOUNT.
```

Business Insights (MOST IMPORTANT)

Estimate true population mean.

[93]:
```
mean = df['Purchase Amount (USD)'].mean()
std = df['Purchase Amount (USD)'].std()
n = len(df)

confidence = 0.95
z = stats.norm.ppf(1 - (1 - confidence) / 2)

lower = mean - z * (std / np.sqrt(n))
upper = mean + z * (std / np.sqrt(n))

(lower, upper)
```

[93]:
```
(np.float64(59.021003799812476), np.float64(60.50771414890548))
```

We are 95% confident that the true average purchase amount of all customers in the population lies between $59.02 and $60.51.
This range was calculated using:

- Sample mean
- Sample standard deviation
- Sample size
- Z-score for 95% confidence

Predicted churn rate of the customers

[24]:
```python
def churn_score(freq, subscription):
    score = 0

    if freq in ['annually', 'every 3 months']:
        score += 2
    elif freq in ['quarterly']:
        score += 1

    if subscription == 'No':
        score += 2

    return score

df['churn_score'] = df.apply(
    lambda x: churn_score(
        x['Frequency of Purchases'].lower(),
        x['Subscription Status']
    ), axis=1
)

# Churn threshold
df['churn_risk'] = df['churn_score'].apply(lambda x: 1 if x >= 3 else 0)

churn_rate = (df['churn_risk'].sum() / len(df)) * 100
print(f"Predicted Churn Rate: {churn_rate:.2f}%")
```

```
Predicted Churn Rate: 32.44%
```

🔍 Key Insights

- **Spending Patterns:** Purchase amounts show variability and skewness, indicating the presence of high-value customers and revenue concentration among specific segments.
- **Outlier Detection:** Identification of high-spending customers using IQR highlights potential premium segments and VIP customer profiles.
- **Gender Analysis:** Statistical testing (independent t-test) indicates whether gender significantly impacts spending behaviour, supporting evidence-based segmentation rather than assumptions.
- **Age vs Spending Relationship:** Correlation and scatter analysis reveal the relationship between customer age and purchase value, supporting targeted demographic strategies.
- **Geographical Insights:** Location-based analysis of previous purchases identifies regions with higher customer engagement and repeat purchase behaviour.
- **Customer Loyalty:** Previous purchase patterns reveal strong indicators of retention potential and customer lifetime value (CLV).
- **Churn Risk Modelling:** A churn scoring model based on behavioural variables (frequency, subscription, engagement) classifies customers into churn-risk categories for proactive retention strategies.
- **Statistical Confidence:** Population mean estimation using confidence intervals provides reliable business forecasting for average customer spending.

📈 Business Impact

This analysis enables the organisation to:

- Identify **high-value customer segments** for premium targeting
- Detect **churn-prone customers** for retention campaigns
- Optimise **marketing segmentation strategies**
- Improve **customer lifetime value (CLV)**
- Support **data-driven decision making**
- Enhance **revenue forecasting accuracy**
- Strengthen **customer engagement strategies**
- Enable **risk-based customer prioritisation**

🎯 Strategic Recommendations

Focus retention strategies on **high churn-risk segments** identified by churn scoring.
Build loyalty programs for **high-value outlier customers** to maximise CLV.
Use behavioural segmentation instead of demographic assumptions for marketing.
Prioritise **high-performing locations** for regional campaign investment.
Apply predictive scoring models for **early churn detection**.
Design **personalised engagement strategies** using previous purchase patterns.

🛠 Tools & Technologies

- Python
- Pandas, NumPy
- Matplotlib, Seaborn
- Plotly
- SciPy (Statistical Testing)
- Jupyter Notebook

🧠 Business Value Statement

This project transforms raw customer data into actionable business intelligence by identifying revenue drivers, customer risk factors, and growth opportunities, enabling strategic decision-making, targeted marketing, and long-term customer value optimisation.

THANK YOU.

Author – Soumi Mukherjee

LinkedIn - www.linkedin.com/in/soumimukherjeeofficial