

Code:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler

sns.set_theme(style="whitegrid")

df = pd.read_csv('retail_sales_final.csv')

print("Dataset Preview:")
print(df.head())

#-----Data Preprocessing-----
#-----

# 1. Handle Missing Values
print("\nMissing values per column:")
print(df.isnull().sum())

# Fill missing numerical values with median
numerical_columns = df.select_dtypes(include=['float64', 'int64']).columns
df[numerical_columns] =
df[numerical_columns].fillna(df[numerical_columns].median())

# Fill missing categorical values with mode
categorical_columns = df.select_dtypes(include=['object']).columns
df[categorical_columns] =
df[categorical_columns].fillna(df[categorical_columns].mode().iloc[0])

# 2. Handle Outliers using IQR
for column in numerical_columns:
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    df[column] = np.where(df[column] < lower_bound, lower_bound,
df[column])
    df[column] = np.where(df[column] > upper_bound, upper_bound,
df[column])

# 3. Normalize Numerical Data
scaler = MinMaxScaler()
df[numerical_columns] = scaler.fit_transform(df[numerical_columns])
```

```

#-----Exploratory Data Analysis-----
-----

# 1. Top-Selling Product Categories
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='Product Category', order=df['Product
Category'].value_counts().index)
plt.title("Top Selling Product Categories")
plt.xticks(rotation=45)
plt.show()

# 2. Monthly Sales Trends
df['Purchase Date'] = pd.to_datetime(df['Purchase Date'])
monthly_sales = df.groupby(df['Purchase
Date'].dt.month)['Revenue'].sum().reset_index()
monthly_sales.columns = ['Month', 'Revenue']

plt.figure(figsize=(10, 6))
sns.lineplot(data=monthly_sales, x='Month', y='Revenue', marker='o')
plt.title("Monthly Sales Trends")
plt.xticks(range(1, 13))
plt.show()

# 3. Revenue by Region
plt.figure(figsize=(8, 6))
sns.barplot(data=df, x='Region', y='Revenue', estimator=np.sum)
plt.title("Revenue by Region")
plt.show()

# 4. Profit Analysis by Holiday Season
plt.figure(figsize=(8, 6))
sns.barplot(data=df, x='Holiday Season', y='Profit', estimator=np.mean,
palette="viridis")
plt.title("Average Profit by Holiday Season")
plt.show()

# 5. Revenue by Product Name
top_products = df.groupby('Product
Name')['Revenue'].sum().nlargest(10).reset_index()

plt.figure(figsize=(12, 6))
sns.barplot(data=top_products, x='Revenue', y='Product Name',
palette="Blues_r")
plt.title("Top 10 Products by Revenue")
plt.show()

# --- Recommendations ---
print("\n--- Recommendations ---")

```

```

print("1. Focus on top-selling product categories and high-revenue products
for marketing campaigns.")
print("2. Offer promotions and discounts during the holiday season to boost
sales further.")
print("3. Optimize inventory management in high-performing regions to avoid
stockouts.")
print("4. Increase discounts on less popular products to clear inventory
and attract more customers.")
print("5. Leverage monthly sales trends to plan seasonal promotions and
optimize staff scheduling.")

```

Output:

Dataset Preview:

| | Product ID | Product Name | Product Category | Price | Cost | Revenue | ... |
|-----------------|---------------|--------------|------------------|----------------|---------|---------|-----|
| Customer Gender | Purchase Date | Discount (%) | Region | Holiday Season | Profit | | |
| 0 | 2665 | Noodles | Groceries | 314.11 | 289.41 | 1655.39 | ... |
| Male | 2024-07-12 | 20 | North | No | 1365.98 | | |
| 1 | 1777 | Charger | Electronics | 240.26 | 120.63 | 1350.52 | ... |
| Female | 2023-12-24 | 20 | Central | Yes | 1229.89 | | |
| 2 | 8453 | Gloves | Clothing | 815.66 | 352.99 | 335.80 | ... |
| Female | 2024-04-02 | 20 | Central | No | -17.19 | | |
| 3 | 4542 | Laptop | Electronics | 218.76 | 815.09 | 1629.63 | ... |
| Female | 2024-09-21 | 15 | South | No | 814.54 | | |
| 4 | 4483 | Desk | Furniture | 117.76 | 360.08 | 1915.94 | ... |
| Male | 2023-12-27 | 10 | North | Yes | 1555.86 | | |

[5 rows x 14 columns]

Missing values per column:

| | |
|------------------|---|
| Product ID | 0 |
| Product Name | 0 |
| Product Category | 0 |
| Price | 0 |
| Cost | 0 |
| Revenue | 0 |
| Sales Volume | 0 |
| Customer Age | 0 |
| Customer Gender | 0 |
| Purchase Date | 0 |
| Discount (%) | 0 |
| Region | 0 |
| Holiday Season | 0 |
| Profit | 0 |

dtype: int64

--- Recommendations ---

1. Focus on top-selling product categories and high-revenue products for marketing campaigns.

--- Recommendations ---

1. Focus on top-selling product categories and high-revenue products for marketing campaigns.
1. Focus on top-selling product categories and high-revenue products for marketing campaigns.
2. Offer promotions and discounts during the holiday season to boost sales further.
3. Optimize inventory management in high-performing regions to avoid stockouts.
4. Increase discounts on less popular products to clear inventory and attract more customers.
3. Optimize inventory management in high-performing regions to avoid stockouts.
4. Increase discounts on less popular products to clear inventory and attract more customers.
4. Increase discounts on less popular products to clear inventory and attract more customers.
5. Leverage monthly sales trends to plan seasonal promotions and optimize staff scheduling.

PS C:\Users\Parv Saini\Desktop\AIML_PROJECT>

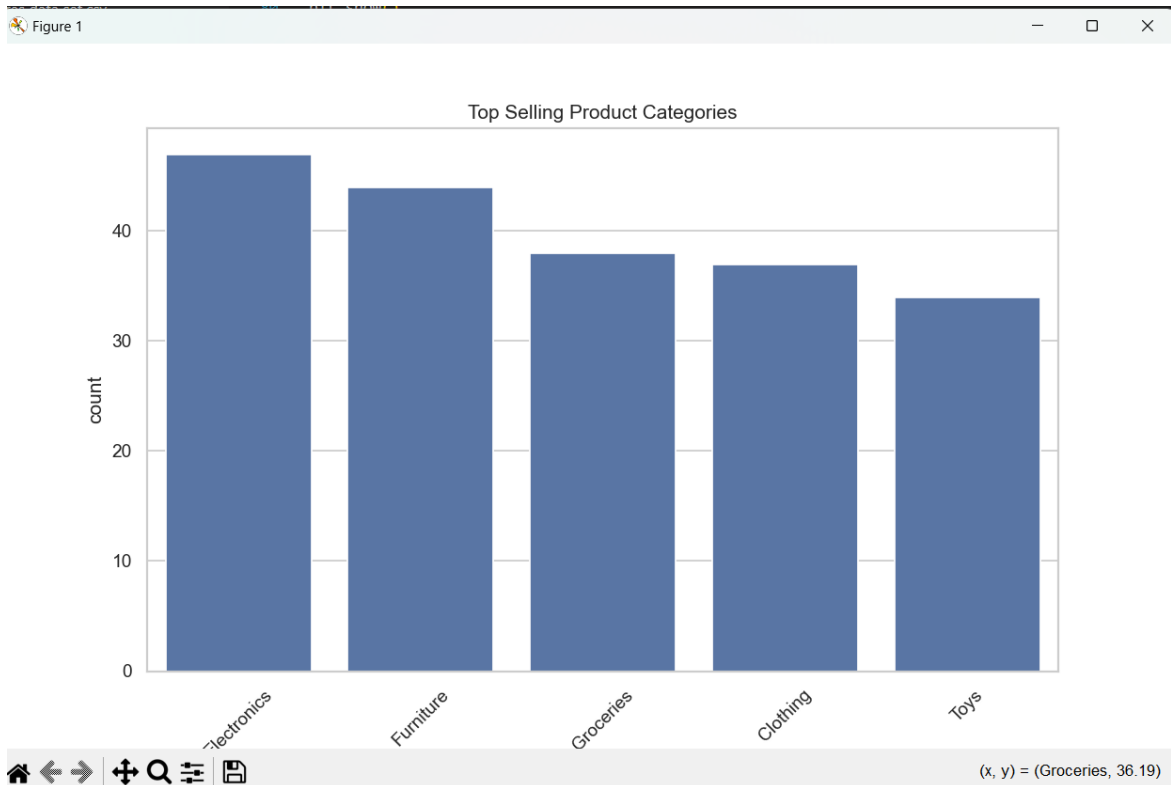


Figure 1

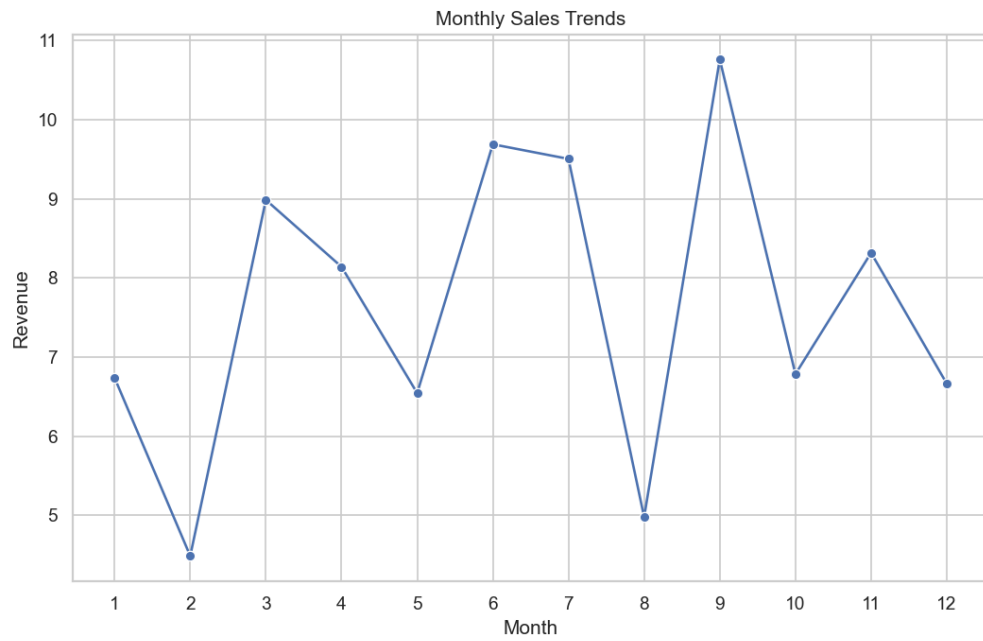


Figure 1

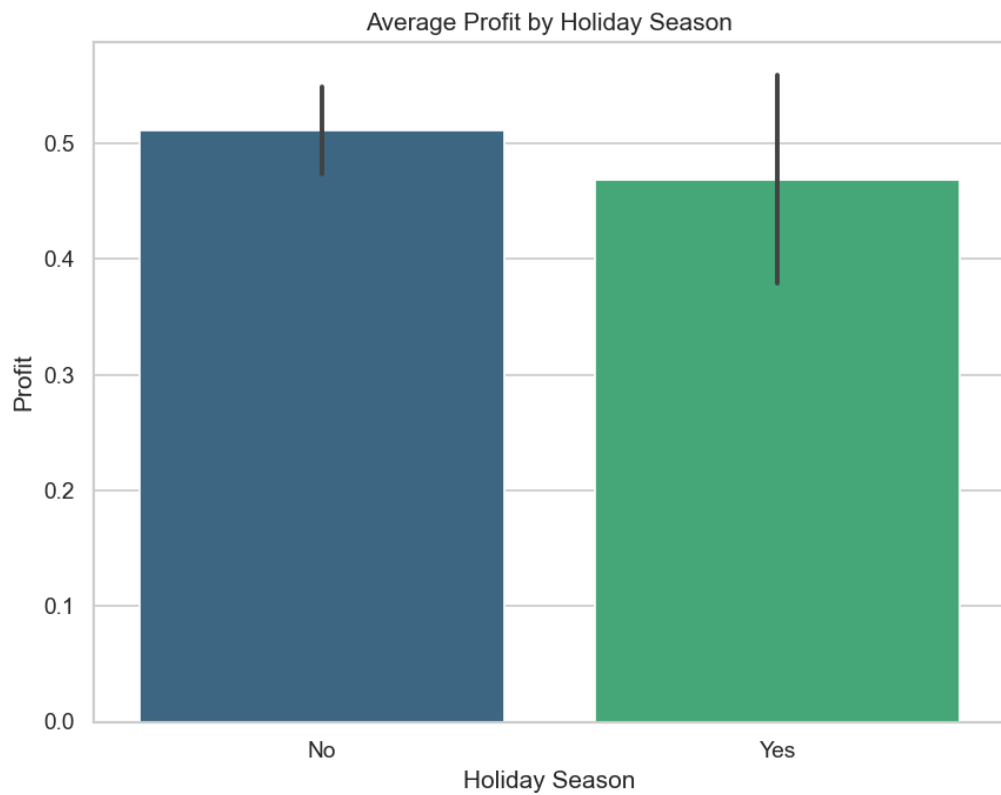


Figure 1

