**Step1: Data Loading and Cleaning**

import pandas as pd

# Load the dataset

retail\_sales\_data = pd.read\_csv('path\_to\_your\_dataset.csv')

# Convert 'Date' column to datetime format

retail\_sales\_data['Date'] = pd.to\_datetime(retail\_sales\_data['Date'], format='%d-%m-%Y')

# Check for missing or duplicate values

missing\_values = retail\_sales\_data.isnull().sum()

duplicates = retail\_sales\_data.duplicated().sum()

# Display updated information after conversion and check for issues

retail\_sales\_data\_info = retail\_sales\_data.info()

missing\_values, duplicates, retail\_sales\_data\_info

**Step 2: Descriptive Statistics**

Next, we calculate the basic statistics for the numeric columns.

# Descriptive statistics for numeric columns

descriptive\_stats = retail\_sales\_data[['Age', 'Quantity', 'Price per Unit', 'Total Amount']].describe()

descriptive\_stats

**Step 3: Time Series Analysis**

We analyze sales trends over time by grouping the data by Date and calculating the total sales per day.

# Group sales by date and calculate the total amount

daily\_sales = retail\_sales\_data.groupby('Date')['Total Amount'].sum()

# Visualize the sales trend over time

import matplotlib.pyplot as plt

daily\_sales.plot(kind='line', title='Daily Sales Trend', ylabel='Total Sales Amount', xlabel='Date')

plt.show()

**Step 4: Customer and Product Analysis**

Analyze customer demographics and product performance.

# Analyze customer demographics

customer\_age\_gender = retail\_sales\_data.groupby(['Age', 'Gender']).size().unstack(fill\_value=0)

# Analyze product performance

product\_performance = retail\_sales\_data.groupby('Product Category')['Total Amount'].sum()

customer\_age\_gender, product\_performance

**Step 5: Visualization**

Create visualizations to summarize the insights.

import seaborn as sns

# Bar chart for product performance

plt.figure(figsize=(10, 6))

sns.barplot(x=product\_performance.index, y=product\_performance.values, palette="viridis")

plt.title('Product Performance')

plt.xlabel('Product Category')

plt.ylabel('Total Sales Amount')

plt.xticks(rotation=45)

plt.show()

# Heatmap for customer demographics

plt.figure(figsize=(10, 6))

sns.heatmap(customer\_age\_gender, annot=True, cmap="YlGnBu", cbar=False)

plt.title('Customer Demographics (Age and Gender)')

plt.xlabel('Gender')

plt.ylabel('Age')

plt.show()

**Step 6: Recommendations**

Based on the analysis, derive actionable recommendations for the retail business.

**Recommendations:**

1. **Targeted Marketing**: Focus on age groups and genders that have higher purchasing behavior.
2. **Promote Best-Selling Products**: Invest in marketing and stock management for high-performing product categories.
3. **Seasonal Campaigns**: Utilize time series analysis to identify peak sales periods and plan seasonal campaigns accordingly.

And there you have it! We've walked through a comprehensive data analysis pipeline for retail sales data. Happy analyzing! 🚀