

# Analyzing Sales Patterns in Online Retail

## 1. Introduction

- **Project Title:** Analyzing Sales Patterns in Online Retail
- **Objective:** The aim of this project is to explore sales patterns in an online retail dataset, uncover key trends, and provide actionable recommendations to optimize sales and marketing strategies.
- **Dataset:** The dataset used in this analysis is the Online Retail dataset, which includes transactions from a UK-based online retailer from 2010 to 2011.

## 2. Data Cleaning and Preparation

- **Initial Inspection:**
  - Loaded the dataset and inspected its structure and content.
  - Identified columns with missing values and the extent of those missing values.
- **Handling Missing Values:**
  - Removed rows with missing values.
- **Removing Duplicates:**
  - Identified and removed duplicate rows to ensure each record is unique.
- **Data Transformation:**
  - Converted the `InvoiceDate` column to datetime format.
  - Created a new column `TotalSales` by multiplying `Quantity` and `UnitPrice`.

## 3. Exploratory Data Analysis (EDA)

- **Summary Statistics:**
  - Calculated key metrics like mean, median, and standard deviation for numerical features.
- **Sales Trends Over Time:**
  - Visualized monthly sales trends to identify peak sales periods and seasonal patterns.
- **Product Performance:**
  - Analyzed sales by product category to identify top-selling products.
- **Customer Analysis:**
  - Explored customer purchase behavior to highlight high-value customers.
- **Regional Analysis:**
  - Visualized sales by region to identify areas with the highest sales.

## 4. Insights and Recommendations

- **Sales Trends:**
  - **Insight:** Peak sales are observed during the holiday season in December.
  - **Recommendation:** Increase inventory and marketing efforts during the holiday season to capitalize on peak sales periods.
- **Product Performance:**
  - **Insight:** Top-selling products include "White Hanging Heart T-Light Holder" and "Regency Cakestand 3 Tier."
  - **Recommendation:** Maintain adequate stock for top-selling products and consider promotional strategies for underperforming products like "Manual Hand Coffee Grinder."
- **Customer Analysis:**
  - **Insight:** High-value customers are primarily from the UK and Germany.
  - **Recommendation:** Implement loyalty programs and targeted marketing campaigns for high-value customers.
- **Regional Analysis:**
  - **Insight:** The UK accounts for the majority of sales, followed by Germany and France.
  - **Recommendation:** Focus on expanding market reach in regions with potential for growth, such as Spain and Portugal.

## 5. Conclusion

- **Summary:** This analysis provided insights into sales patterns, product performance, customer behavior, and regional sales trends. The recommendations aim to optimize inventory, target marketing efforts, and explore new market opportunities.
- **Next Steps:** Further analysis could include more detailed customer segmentation, analysis of marketing campaign effectiveness, and forecasting future sales trends.

```
import pandas as pd
df = pd.read_csv('Online_Retail.csv')
print(df.head())
df.info()
df.describe()
df.isnull().sum()
df_cleaned.head()
df_cleaned.duplicated().sum()
```

```
# Convert 'InvoiceDate' to datetime
df_cleaned['InvoiceDate'] = pd.to_datetime(df_cleaned['InvoiceDate'])
# Create a new column for total sales
```

```
df_cleaned['TotalSales'] = df_cleaned['Quantity'] *  
df_cleaned['UnitPrice']
```

```
# Distribution of TotalSales  
import matplotlib.pyplot as plt  
import seaborn as sns  
plt.figure(figsize=(10, 6))  
sns.histplot(df_cleaned['TotalSales'], bins=50, kde=True)  
plt.title('Distribution of Total Sales')  
plt.xlabel('Total Sales')  
plt.ylabel('Frequency')  
plt.show()  
  
# Sales Trends Over Time  
plt.figure(figsize=(12, 6))  
df_cleaned.set_index('InvoiceDate').resample('M')['TotalSales'].sum().plot()  
plt.title('Monthly Sales Trends')  
plt.xlabel('Month')  
plt.ylabel('Total Sales')  
plt.show()  
  
# Sales by Product  
plt.figure(figsize=(12, 8))  
top_products =  
df_cleaned.groupby('Description')['TotalSales'].sum().nlargest(10)  
sns.barplot(x=top_products.values, y=top_products.index)  
plt.title('Top 10 Products by Sales')  
plt.xlabel('Total Sales')  
plt.ylabel('Product')  
plt.show()  
  
# Customer Analysis  
plt.figure(figsize=(12, 8))  
top_customers =  
df_cleaned.groupby('CustomerID')['TotalSales'].sum().nlargest(10)  
sns.barplot(x=top_customers.values, y=top_customers.index)  
plt.title('Top 10 Customers by Sales')  
plt.xlabel('Total Sales')  
plt.ylabel('Customer ID')  
plt.show()
```

```
# Regional Analysis
plt.figure(figsize=(12, 8))
top_countries =
df_cleaned.groupby('Country')['TotalSales'].sum().nlargest(10)
sns.barplot(x=top_countries.values, y=top_countries.index)
plt.title('Top 10 Countries by Sales')
plt.xlabel('Total Sales')
plt.ylabel('Country')
plt.show()
```