Nikhil Sharma

Customer Shopping Data

17 July 2023

Online Sales in US Data Analysis Report

1. Introduction

This report presents findings from the analysis of the "Online Sales in USA" dataset. The dataset, uploaded by YT Gangster on Kaggle.com, includes sales data for various products, merchandise, and electronics in different states of the USA. As a significant portion of the population with internet access transitions to online shopping, large retailers are actively seeking ways to increase their profits. Sales analysis serves as a vital technique for retailers to understand customer purchasing behavior and patterns, enabling them to enhance their sales strategies. In this analysis, we will perform exploratory data analysis (EDA) on the dataset and predict sales for the next 7 days from the last date of the training dataset. The insights gained from this analysis will provide valuable information for retailers aiming to optimize their sales and understand customer preferences in the online market.

2. Data Analysis

The analysis used Python, including the NumPy, Pandas, Matplotlib, Seaborn libraries. The Jupyter Notebook code was used for the analysis, which explored and understood the dataset. The key findings are as follows:

2.1 Data Shape and Columns

The dataset consists of 286,392 rows and 36 columns. The columns are: 'order_id', 'order_date', 'status', 'item_id', 'sku', 'qty_ordered', 'price', 'value', 'discount_amount', 'total', 'category', 'payment_method', 'bi_st', 'cust_id', 'year', 'month', 'ref_num', 'Name Prefix', 'First Name', 'Middle Initial', 'Last Name', 'Gender', 'age', 'full_name', 'E Mail', 'Customer Since', 'SSN', 'Phone No. ', 'Place Name', 'County', 'City', 'State', 'Zip', 'Region', 'User Name', 'Discount_Percent'.

2.2 Data Types and Cleaning

The data types of the columns were checked and aligned with their respective attributes. The "Order Date" and "shipping date (DateOrders)" columns were converted to datetime for further analysis.

3. Findings

3.1 Online Shopping Trends and Customer Behavior

• Customer Behavior: The customer with ID 107853, named **Adolfo H Hord**, has the highest average purchase value of **$41,868**. The customer with ID 109038 has made the highest total purchases in terms of quantity (361 items) and value ($1,900,290).

• Product Analysis: The product with SKU MATSAM59DB75ADB2F80 has the highest total sales amounting to **$4,972,349** and has been sold 3,775 times.

• **Order Status:** There were **286,392** total orders, of which **141,699** were either refunded or canceled. This implies that almost 50% of the orders were not successful.

• **Sales by Region:** The **South** region has the highest total revenue of $89,647,849.

3.2 Revenue Analysis

• **Total and Average Revenue by Payment Method:** The payment method generating the highest total revenue is Easypay with $54,661,648. The payment method generating the highest average revenue per transaction is bankalfalah with $2,075.

• **Total and Average Revenue by Order Status:** The order status generating the highest total revenue is "canceled" with $132,642,666. The order status generating the highest average revenue per transaction is "holded" with $4,447.

• **City: Noxpater** has the highest total revenue at **$1,909,971**, ahead of Dennis and Goodland.

• **State: Texas** has the highest total revenue of **$15,467,455**, with California and New York next.

• **Order Status: Canceled orders** accounted for a total of **$132,642,666** from the revenue, more than the completed and received ones. Holden, processing, and pending have the highest share of average revenue.

• **Month: December 2020** has the most revenue at **$57,666,990**, significantly ahead of April 2021 and March 2021.

• **Age**: **39** year old customers brought in the most revenue, followed by 31, and 52 year old.

• **Gender:** Males brought in most revenue, slightly ahead of Females.

• **Category: Mobiles & Tablets** was the most popular category with a total revenue of **$130,111,993**. Appliances and Entertainment were next.

4. Conclusion

The analysis revealed key insights into customer behavior, product popularity, order status, sales distribution by region, payment methods, and revenue generation. The customer with the highest average purchase value is Adolfo H Hord, while the customer with the highest total purchases is customer ID 109038. The product with SKU MATSAM59DB75ADB2F80 is the top-

selling product. There is a significant number of refunded or canceled orders, nearly half of all total orders. The South region has the highest total sales, and Easypay is the most revenue-generating payment method. However, the order status generating the highest average revenue per transaction is "holded." The cities and states generating the most revenue are Noxpater and Texas, respectively. Despite a high number of canceled orders, they account for the most substantial total revenue. Revenue generation peaked in December 2020, with 39-year-old customers contributing the most. Males slightly outpaced females in generating revenue, and the Mobiles & Tablets category was the most popular.

5. Recommendations

- Based on the analysis, the following recommendations are suggested:

- Investigate the reasons for the high number of refunded or canceled orders and implement measures to reduce them.

- Enhance customer loyalty programs, focusing on high-value customers such as Adolfo H Hord and customer ID 109038.

- Optimize inventory and supply chain management for popular products, like the one with SKU MATSAM59DB75ADB2F80.

- Investigate the reasons why the "holded" order status generates the highest average revenue per transaction and apply insights to other order statuses.

- Focus marketing and sales strategies on areas with high sales, such as the South region, Noxpater, and Texas.

- Review the payment methods and provide incentives for using Easypay, which generates the highest total revenue.

- Develop targeted marketing campaigns for the customer demographics contributing most to the revenue, namely 39-year-old males.

- Consider expanding the Mobiles & Tablets category or offering related products, given its popularity.

6. Acknowledgments

I would like to express our gratitude to YT Gangster for providing the "Online Sales in US" dataset on Kaggle.

7. References

The analysis used the following libraries and tools:

• NumPy: A library for numerical computing in Python.

• Pandas: A data manipulation and analysis library.

• Matplotlib: A plotting library.

• Seaborn: A data visualization library.

• Jupyter Notebook: An interactive coding environment.

For a detailed understanding of the analysis process, including the Jupyter Notebook code used, please refer to the original analysis documentation.

Best regards,

Nikhil Sharma