

Decoding Online Retail: A Data-Driven Deep Dive

Done By
L.Smitha Rai



Project description

The project provides a dataset focused on retail transactional data. As the main data analyst, you will analyze sales and customer experience and create insightful dashboards to inform business stakeholders about potential problems and propose structural business improvements.

Objective

In this project, my primary focus is on addressing key challenges related to changing consumer behaviour within the retail transaction data. To achieve this goal efficiently, the project has been divided into four main objectives:

1. Sales Analysis

- The first objective analyzing which products are bestsellers and which ones have low sales.

- By determining this insights, we can identify the organization's identify top-selling products or services, and detect underperforming areas to focus on improvement.

2. Customer Analysis

- The objective of customer analysis is to understand customer satisfaction and their engagement in purchases.
- Through this analysis, I aim to offer valuable insights into customer satisfaction for each product or purchase, allowing us to shift our focus and enhance the products accordingly.

3. Demographic Analysis

- The objective of demographic analysis is to understand the characteristics and behaviors of customer base .
- This analysis offers insights into customer preferences, allowing us to deliver personalized recommendations for each customer group.

Data Overview

- The data analysis is performed using Python in Jupyter Notebook to gain insights about the data. This involves various data preprocessing, cleaning, and exploratory data analysis tasks to prepare the data for further analysis.

Notebook link : [link here](#)

- The link will lead you to a comprehensive notebook or script that outlines the data analysis procedures. A Detailed step by step process from data wrangling to exploratory data analysis is done.
- The dataset provides single data table After examining the data fields, I noticed that the dataset generally represents the following key information.

1. Customer : General information about customers which includes identifiers and demographics.

2. Order : Information about the order including date of order and quantity ordered, order value.

3. Product: Specific information about the ordered item including product name, product category, product department.

4. Feedback : Information about feedback on each purchase and ratings on each product.

Data Cleaning

1.Drop unnecessary columns

(Email, Phone, Address, Zipcode, Product_Type, Month, State, City)

2.Fix the datatype of the columns

3.Check for missing value

4.Check for duplicate value

Feature Metric

1.Created date time frame

Key Insights:

1. Sales Analysis

Top selling product categories:

- Electronics and grocery products are the top revenue contributors, accounting for 24% and 22% respectively, while other categories make up 18%.
- The top-selling product categories have received the highest rating of 4, indicating that customers are satisfied with these top sellers.
- The grocery category is the most preferred among individuals aged 32-35.

Top selling product Brands:

- Pepsi is the top seller, representing 24% of the total revenue, while BlueStar is the least, contributing just 2%.
- People aged 34, within the 30 to 40 middle-age bracket, are consuming more Pepsi.

Sales Over Time:

- Sales were high in April and August 2023, followed by a decline from September 2023 to February 2024.
- Sales of the top selling Pepsi brand experienced a steep decline from April 2023 to February 2024.

2.Customer Analysis

Customer Type Breakdown:

- Regular customers generate the highest revenue, accounting for approximately 49%, while new customers contribute 30% and premium customers account for 21% of the revenue.

Customer Satisfaction levels:

- The customer feedback data reveals that a significant portion of purchases are met with positive experiences, as evidenced by the 33.36% of customers rating their purchase as "Excellent" and 31.54% rating it as "Good."

- The average customer rating is around 3.16.
- Ratings were high in October 2023 but predominantly declined from November onward.
- Young adults aged 18 to 30, along with individuals aged 45 to 47, are the most active customers.

3.Demographic Analysis

Income breakdown:

- Customers with a decent income, representing the middle class, contribute the most to revenue, accounting for 43%. This is followed by low-income customers at 32%, and high-income customers at 25%.

Payment breakdown:

- Credit cards lead as the preferred payment method, followed closely by debit cards and cash.

Suggestions :

- Due to data limitations, the root cause of underperforming product categories and declining sales remains unclear. However, analyzing inventory and logistics data could help identify the underlying issues.
- Since credit cards are favored by most of the customer base, offering EMI options on high-priced products and providing cashback incentives could enhance customer satisfaction.
- For regular customers, implementing loyalty programs could further increase engagement and retention.
- Although electronic brands currently generate lower revenue, they receive positive customer feedback. Therefore, maintaining adequate stock levels, ensuring product quality, expanding the product range, and optimizing logistics could significantly boost revenue.