

# Exploratory Data Analysis of Online Sales in the USA

This report presents an exploratory data analysis (EDA) conducted on a dataset of online sales in the United States. The dataset, obtained from Kaggle, contains information about the sales of various products, merchandise, and electronics in different states. The aim of this analysis is to gain insights into customer purchasing behavior and patterns, which can help large retailers increase their profits.

## Dataset Overview:

The dataset is in CSV format and includes the following columns:

- order\_date: Date of the order
- status: Order status (e.g., received, complete)
- qty\_ordered: Quantity of items ordered
- price: Price of the item
- value: Total value of the order
- discount\_amount: Amount of discount applied
- total: Total amount paid by the customer
- category: Product category
- payment\_method: Payment method used
- bi\_st: Business intelligence status
- cust\_id: Customer ID
- year: Year of the order
- month: Month of the order
- First Name: Customer's first name
- Gender: Customer's gender
- age: Customer's age
- E Mail: Customer's email address
- Customer Since: Date the customer started shopping
- Phone No.: Customer's phone number

- Place Name: Place of residence
- County: County of residence
- City: City of residence
- State: State of residence
- Zip: ZIP code
- Region: Region of residence
- Discount\_Percent: Percentage of discount applied

## Data Preprocessing:

Before conducting the analysis, the dataset was preprocessed to remove unnecessary columns and handle missing values, if any. The following columns were identified as potentially unnecessary and removed from the analysis:

- order\_id
- item\_id
- sku
- ref\_num
- Prefix
- Middle Name
- Initial
- Last Name
- full\_name
- ssn
- User Name

## Key Performance Indicators (KPIs):

The following KPIs were identified to measure the performance of online sales:

1. **Total sales**
2. **Average order value**
3. **Sales by category**
4. **Customer acquisition rate**
5. **Customer retention rate**

## 6. Discount effectiveness

### Graphical Analysis:

To visualize the insights gained from the data, several graphs were created:

1. Sales trend over time: A line graph showcasing the total sales per month or year.
2. Sales by category: A bar chart or pie chart illustrating the proportion of sales for each product category.
3. Customer acquisition and retention: A line graph depicting the number of new and retained customers over time.
4. Age distribution of customers: A histogram or bar chart representing the age ranges of customers.
5. Sales by region: A choropleth map or bar chart displaying sales performance across different states or regions.
6. Payment method preference: A pie chart or bar chart showing the distribution of payment methods used by customers.
7. Discount impact: A scatter plot or line graph comparing sales with and without discounts.

### Insights and Conclusions:

The analysis of monthly sales data reveals a significant spike in sales during December 2020, with a total of \$5.7 million. This represents a remarkable 196.17% increase compared to the average monthly sales figure of \$1.9 million. The substantial increase in sales can be attributed to the holiday season, particularly Christmas and New Year celebrations, when consumer spending tends to be higher.

On the other hand, the lowest performing month was February, which experienced a 77.38% decrease from the average monthly sales. This decline in sales may be due to various factors such as post-holiday financial constraints, seasonal changes in consumer behavior, or external economic factors.

Mobiles and Tablets contribute to more than half of the total sales, followed by Appliances. This might suggest that greater investments in these categories can yield more profits in the future. Consider expanding the product range and offering competitive prices to capitalize on the demand in these categories.

The chart reveals notable spikes in new customer acquisition during the month of December. This seasonality can be attributed to the holiday shopping period, where the company likely experiences

increased sales and customer activity. Implement targeted marketing campaigns and promotions to attract more customers during these peak seasons.

The chart indicates a healthy balance between new and retained customers. While the company continues to acquire new customers, it also maintains a substantial base of retained customers. Focus on enhancing customer loyalty by providing personalized experiences, exclusive offers, and excellent customer service to encourage repeat purchases.

At the same time, the importance of retained customers suggests that the company should also invest in customer retention strategies, such as personalized marketing, loyalty programs, and exceptional customer service. Regularly engage with existing customers through email marketing, social media, and other channels to keep them informed about new products, offers, and promotions.

The majority of the customers are adults in their 50s, followed by the other age groups. Teenagers, Kids, and adults over 70 form the minority. To increase sales among these minority age groups:

- Develop products and marketing strategies tailored to the specific needs and preferences of teenagers and kids.
- Collaborate with influencers or celebrities popular among these age groups to promote your products.
- Create engaging and interactive content on social media platforms frequented by younger audiences.
- For adults over 70, focus on user-friendly products, clear instructions, and emphasize the benefits of your products for their lifestyle.

Texas is the best performing state as per the sales data, amounting as high as 15 million USD. Western US states, like Alaska, Nevada, Idaho, and the nearby states had sales at approximately 1 million USD each. To increase sales in these states:

- Conduct market research to understand the specific needs and preferences of customers in these regions.
- Collaborate with local businesses or influencers to increase brand visibility and trust.
- Offer region-specific promotions or discounts to incentivize purchases.
- Improve distribution channels and ensure easy availability of products in these states.

COD was the most preferred mode of payment, followed by Easypay. Financesettlement and cashatdoorstep were rarely used. Consider offering incentives or discounts for customers using

Easypay to encourage adoption of this payment method. Investigate the reasons behind the low usage of Financesettlement and cashatdoorstep, and make necessary improvements to make these options more attractive to customers.

There were not a lot of discounts. However, at around 15000 USD discounts, sales of around 30000 USD were attainable, but not beyond that. As the discount amounts increased to 20000 USD and 30000 USD, the sales didn't increase but rather got back closer to 20000 USD. This suggests that there might be a sweet spot for discount amounts. Experiment with different discount levels and analyze the impact on sales to find the optimal discount range that maximizes revenue without compromising profitability. Additionally, consider offering personalized discounts based on customer segments or purchase history to drive sales more effectively.

This report, along with the Python code used for the analysis, will be uploaded to GitHub as a PDF file. The code will demonstrate the data preprocessing steps, calculation of KPIs, and creation of the mentioned graphs using appropriate libraries such as Pandas, Matplotlib, and Seaborn.