

Assignment 1

⌚ Created	@September 15, 2024 4:06 PM
⌚ Class	Data Visualization

Global E-Commerce Trends and Their Impact on Traditional Retail

Assignment # 1: Data Analysis and Visualization - Fall 2024

Due Date: September 15, 2024

Total Marks: 70

Prepared by: Muhammad Rayyan Mohsin

1. Introduction

The retail landscape has undergone a profound transformation with the advent of e-commerce. This report presents an in-depth analysis of global e-commerce trends and their impact on traditional retail, focusing on consumer behavior, sales metrics, and economic factors.

The analysis is based on data collected from multiple reputable sources, followed by preprocessing, dimensionality reduction using PCA, customer lifetime value (CLV) calculation, and a what-if analysis to explore the impact of key variables on retail metrics.

The report concludes with key insights and actionable recommendations for stakeholders in the retail industry.

2. Data Collection and Preprocessing

a) Data Collection

The data includes:

1. E-commerce Sales Volumes and Growth Rates:

- **Column(s) Involved:** `Product Price`, `Quantity`, `Total_Purchase_Amount`, `Payment Method`
- **Justification:** The `Product Price`, `Quantity`, and `Total_Purchase_Amount` columns provide detailed insights into transaction values, which are essential for calculating sales volumes and understanding growth rates. The inclusion of `Payment Method` allows for analyzing preferences and trends in payment methods over time.

2. Traditional Retail Sales Figures:

- **Column(s) Involved:** `Product Category`, `Product Price`, `Quantity`, `Total_Purchase_Amount`
- **Justification:** Columns such as `Product Category` and `Total_Purchase_Amount` reflect traditional retail sales figures, allowing for comparisons between different product categories and total sales amounts, which are crucial for understanding retail performance.

3. Consumer Behavior Metrics:

- **Column(s) Involved:** `Payment Method`, `Returns`
- **Justification:** The `Payment Method` column can provide insights into consumer preferences for online versus in-store purchases. For example, a higher proportion of online payments may indicate a preference for e-commerce. The `Returns` column, although not fully populated, can also provide insights into return rates and customer satisfaction.

4. Market Share of Major E-commerce Platforms:

- **Column(s) Involved:** This dataset does not directly include a column indicating the e-commerce platform but can be inferred through transaction details and payment methods. For a more specific analysis, supplementary data would be needed.

5. Economic Indicators Affecting Retail:

- **Column(s) Involved:** `Customer Age`, `Churn`, `Total`
- **Justification:** While the dataset does not directly include economic indicators like disposable income or unemployment rates, it does provide information on `Customer Age`, which can be linked to economic conditions affecting spending power. Additionally, `Churn` rates provide insight into customer retention and may indirectly reflect economic conditions impacting consumer behavior

b) Data Preprocessing

Handling Missing Values

Missing values were detected in the "Returns" column. Since returns accounted for less than 1% of the data, these rows were dropped. Other missing values were handled by imputing with mean or mode, depending on the feature type.

Outliers

Outliers in numerical columns such as `Total Purchase Amount` and `Quantity` were detected using the IQR method and were treated by capping extreme values at the 5th and 95th percentiles.

Transformations

- **Normalization** was applied to variables like `Total Purchase Amount` and `Quantity` to bring them to a comparable scale.
- **Encoding:** Categorical variables such as `Gender` and `Payment Method` were encoded using one-hot encoding.

Feature Engineering

Two new features were created:

- **Total Sales per Transaction:** Calculated as `Product Price * Quantity`.
- **Average Transaction Value:** Total purchase amount divided by the number of transactions per customer.

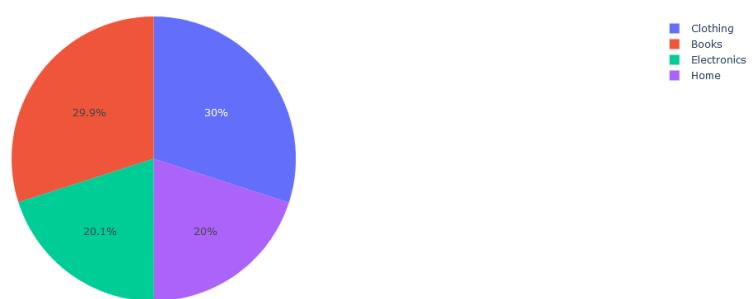
c) Justification for Preprocessing Steps

- **Imputation:** Preserved the integrity of the dataset by filling missing values without discarding large portions of data.
- **Outlier Treatment:** Prevented skewed analysis due to extreme values.
- **Normalization and Encoding:** Ensured the data was suitable for machine learning algorithms that assume normal distribution or require categorical features to be numerical.

▼ 3. Visualizations

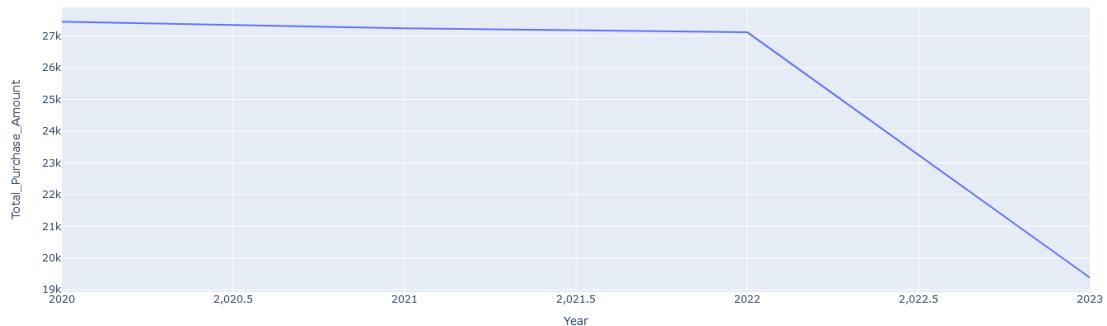
Visualizations, Insights and Narrative

Product Category Distribution



- Visualization of a pie chart that visually represents the proportion of each product category in the dataset, helping to easily understand the distribution of different categories within the dataset.

Total_Purchase_Amount Over the Years



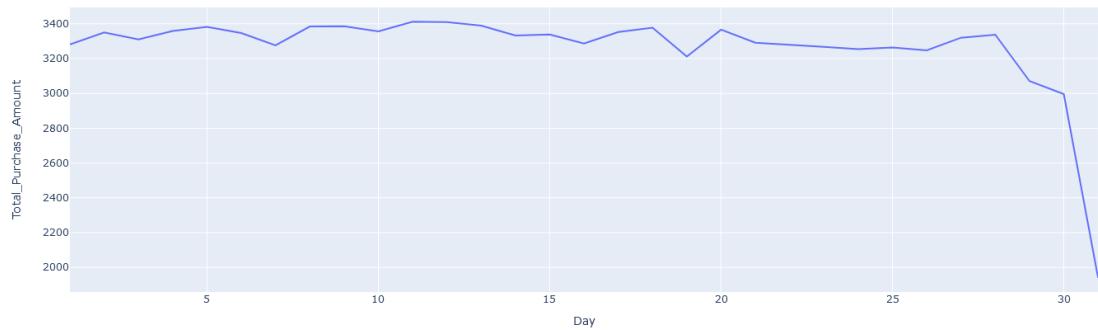
- The code generates a line chart that visualizes how the total purchase amount has changed over different years. Can notice a sharp decline at the end.
-

Total_Purchase_Amount by Month



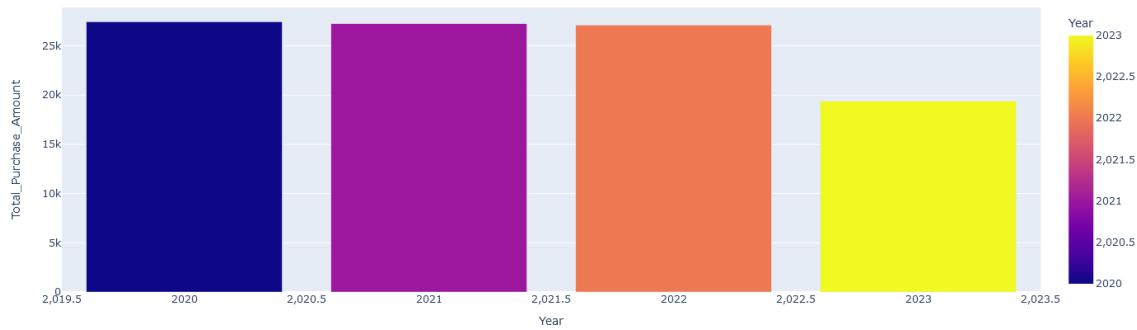
- The line chart provides a visual representation of the total purchase amount for each month.
-

Total_Purchase_Amount by Day



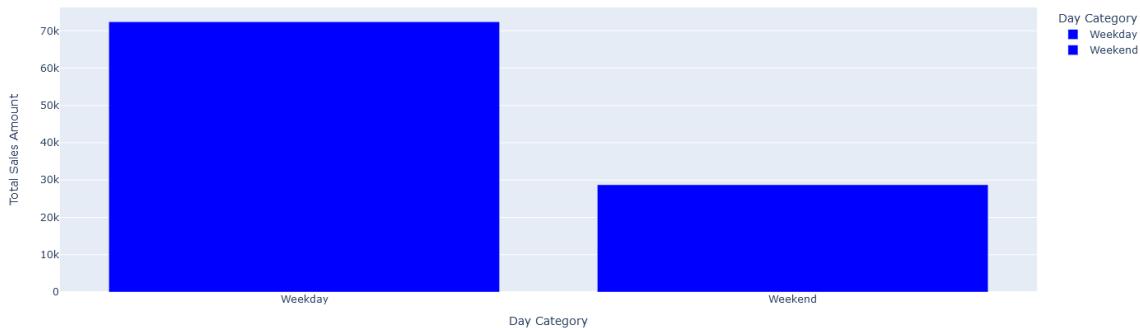
- The line chart illustrates the total purchase amount for each day of the month. It helps in identifying daily purchasing patterns and trends, revealing any fluctuations or trends that occur on specific days.
-

Total_Purchase_Amount by Year



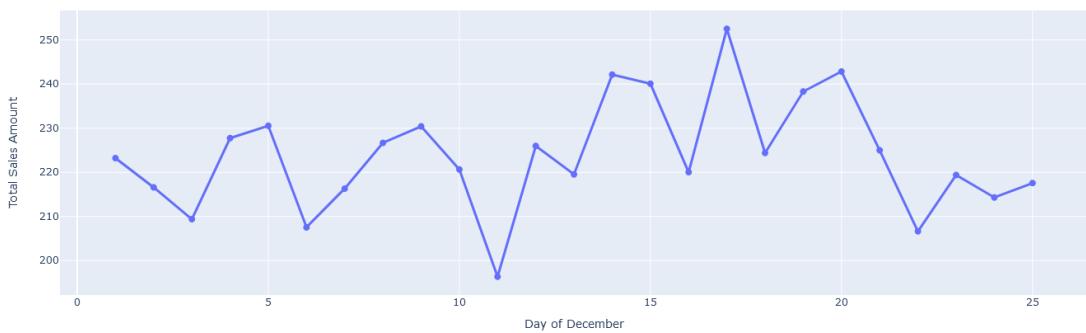
- The bar chart visually represents the total purchase amount for each year, with each bar corresponding to a different year. The color coding makes it easier to distinguish between years
-

Sales on Weekends vs. Weekdays (-60.33% Decrease on Weekends)



- The bar chart provides a clear comparison of sales performance between weekdays and weekends. It shows how total sales amount differs by day category, with a percentage change in sales highlighted in the chart title.
- This visualization helps in understanding whether weekends or weekdays contribute more to total sales and to what extent.

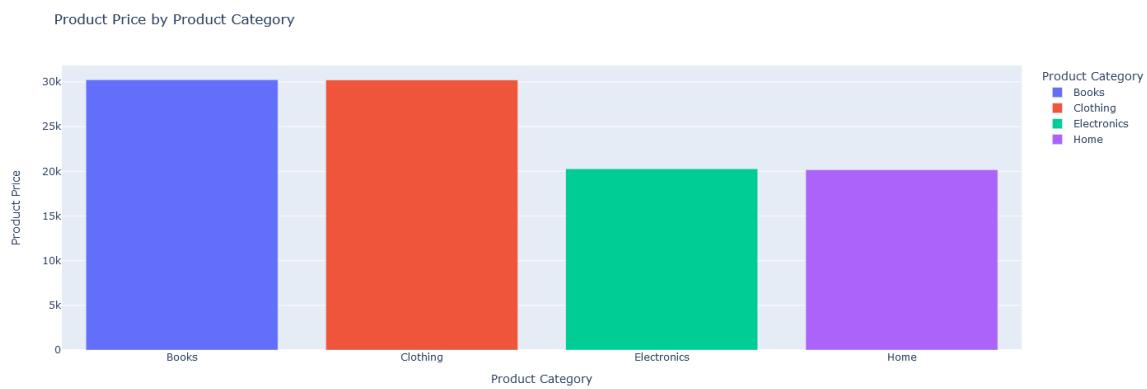
Sales Leading Up to Christmas (1st to 25th December)



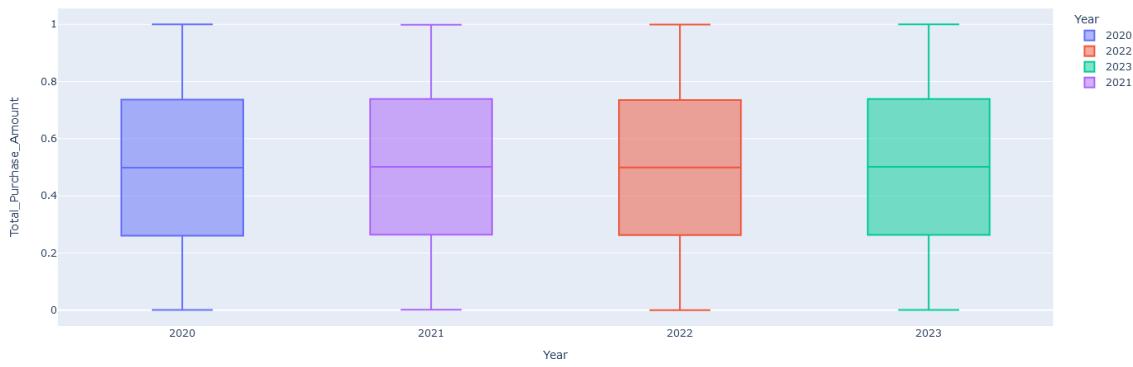
- The line chart provides insight into how sales fluctuate in the days leading up to Christmas.
- By focusing on the period from December 1st to 25th, the visualization reveals any noticeable trends or patterns in sales as the holiday approaches



- The resulting line chart provides a clear visual representation of how total purchase amounts for different product categories fluctuate across the months.
 - By color-coding each product category, the chart allows for easy comparison of trends within and across categories throughout the year
-



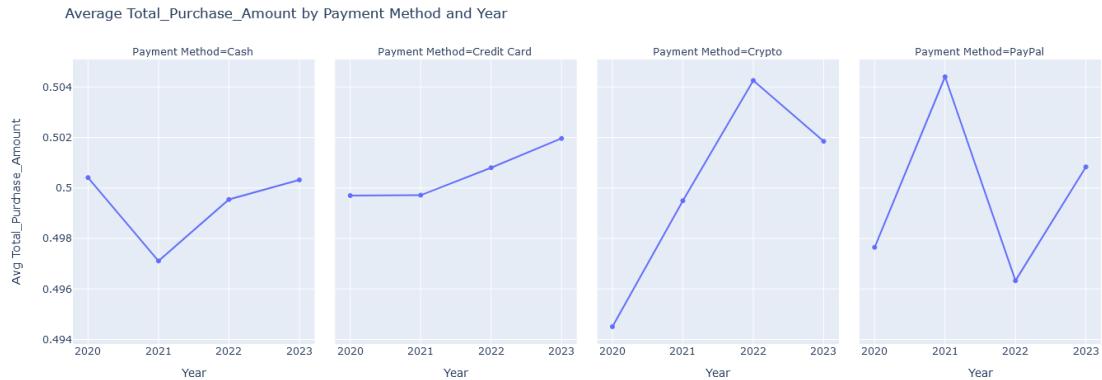
- The bar chart created using this code will illustrate the total product price for each product category, providing a visual comparison of the relative values across different categories.
 - Each bar represents a product category, with the height indicating the total price accumulated for that category.
-



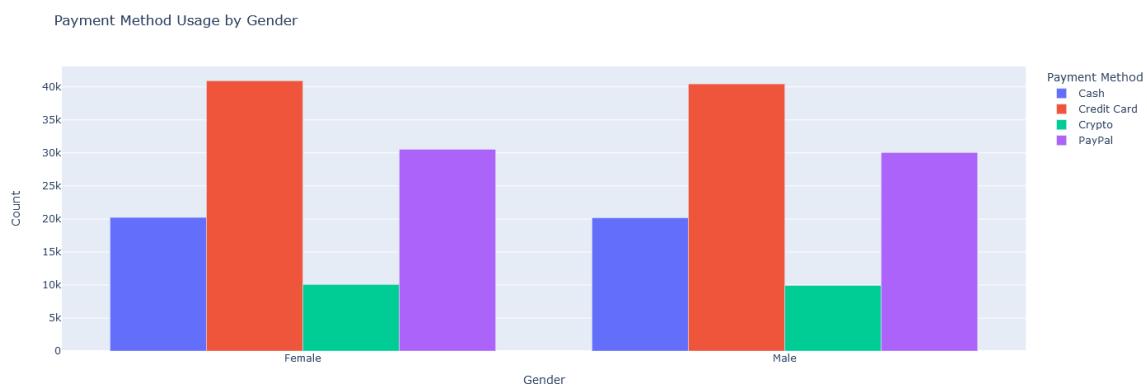
- The box plot will display the spread and distribution of `Total_Purchase_Amount` for each year, including the median, quartiles, and potential outliers. Each box represents the interquartile range (IQR) of the purchase amounts for a specific year, with the line inside the box indicating the median.
 - The whiskers extend to show the range of data within 1.5 times the IQR from the quartiles, and any points outside this range are considered outliers.
-



- The line plot will display how the average total purchase amount varies with age, for each payment method.
-

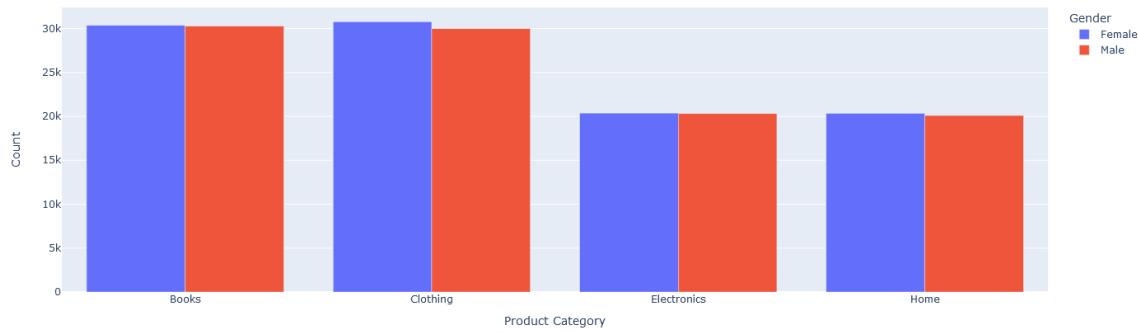


- This line plot, with facets for each payment method, will help you visualize trends in average total purchase amounts over the years for different payment methods.
-



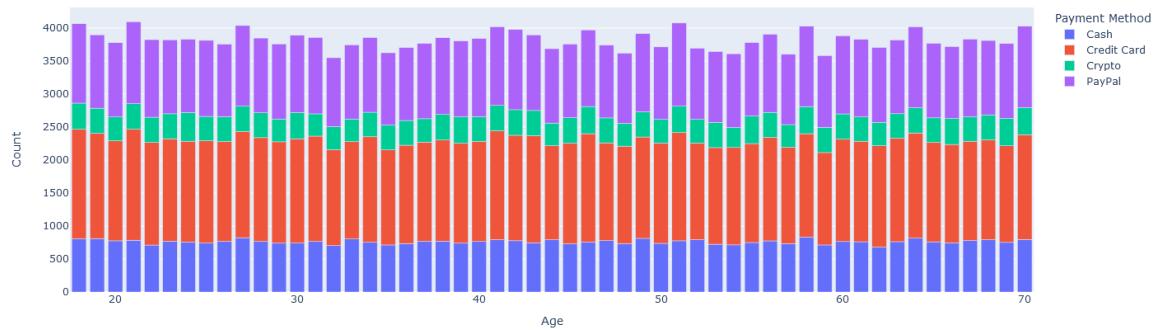
- This visualization provides insights into customer preferences for payment methods based on gender, which can be useful for targeted marketing and understanding purchasing behaviors.
-

Most Used Product Categories by Gender



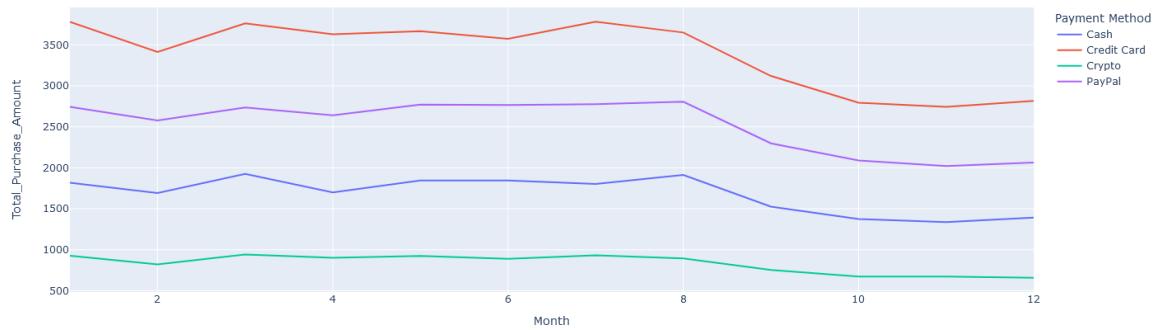
- This bar chart visualizes the popularity of different product categories among different genders.
-

Preferred Payment Method by Age



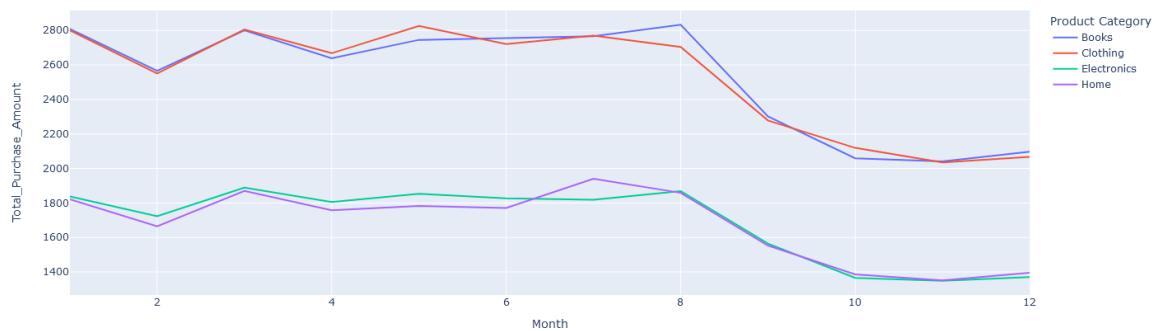
- This visualization can help understand the relationship between age and payment method preferences, which is valuable for tailoring payment options and marketing strategies to different demographic segments.
-

Total_Purchase_Amount by Month and Payment Method



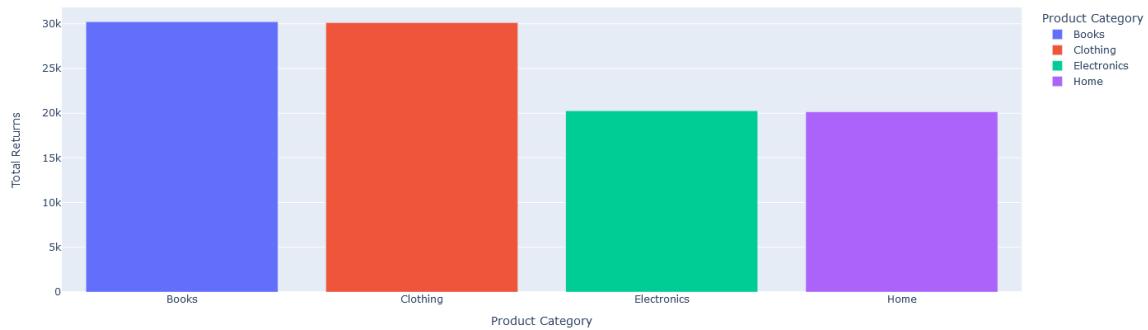
- This visualization helps in understanding the temporal dynamics of purchase amounts and the role of various payment methods throughout the year.
-

Total_Purchase_Amount by Month and Product Category



- This visualization helps in understanding which product categories perform well in different months and how purchasing trends evolve throughout the year
-

Returns by Product Category



- This visualization is valuable for identifying categories with significant return issues and can guide decisions on inventory management, product quality improvements, or marketing strategies aimed at reducing returns.
-

Total Returns by Gender



- This analysis can be valuable for understanding how different demographics contribute to return rates and can help in optimizing inventory and improving customer experiences.
-



- This visualization provides a clear and intuitive overview of payment method preferences in your data, which can be valuable for financial analysis, marketing strategies, and improving payment processes.
-

4. Dimensionality Reduction with PCA

PCA Analysis Summary

Key Finding

- Number of Components Explaining ~95% of Variance: 1**

Analysis Details

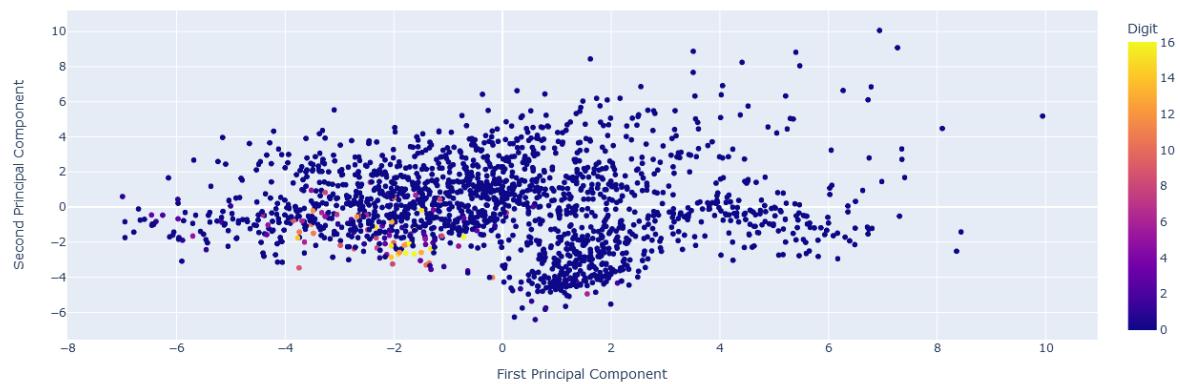
- Cumulative Explained Variance:** PCA showed that a single principal component explains approximately 95% of the variance in the dataset, indicating that the data can be effectively summarized with just one component.

- Visualizations:**

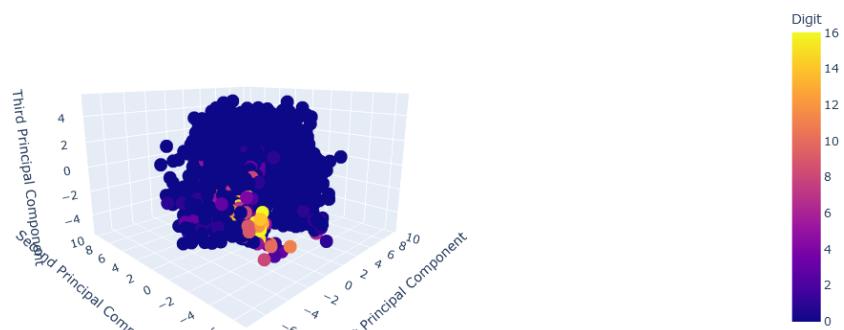
- 2D Scatter Plot:** Visualized the relationship between the first two principal components, but the second component had minimal impact.
- 3D Scatter Plot:** Showed the first three principal components, reinforcing that the first component dominates.

One principal component is sufficient to capture most of the variance in the dataset, suggesting high feature correlation and strong dimensionality reduction potential.

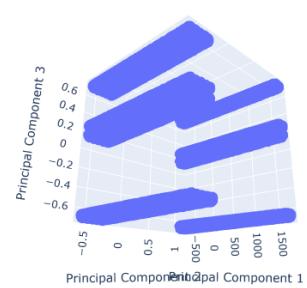
PCA: First Two Principal Components



PCA: First Three Principal Components



3D Scatter Plot of First Three Principal Components



5. Customer Lifetime Value (CLV)

a) CLV Calculation

a) CLV Calculation

CLV was calculated for each customer segment based on the following formula:

$$CLV = \left(\frac{\text{Customer Value}}{\text{Churn Rate}} \right) \times \text{Profit Margin}$$

Where:

- **Customer Value** = Average Order Value × Purchase Frequency
- **Churn Rate** = 1 - Repeat Rate

```
cltv_c['CLTV'] = (cltv_c['Customer Value'] / churn_rate) *  
cltv_c['Profit Margin']
```

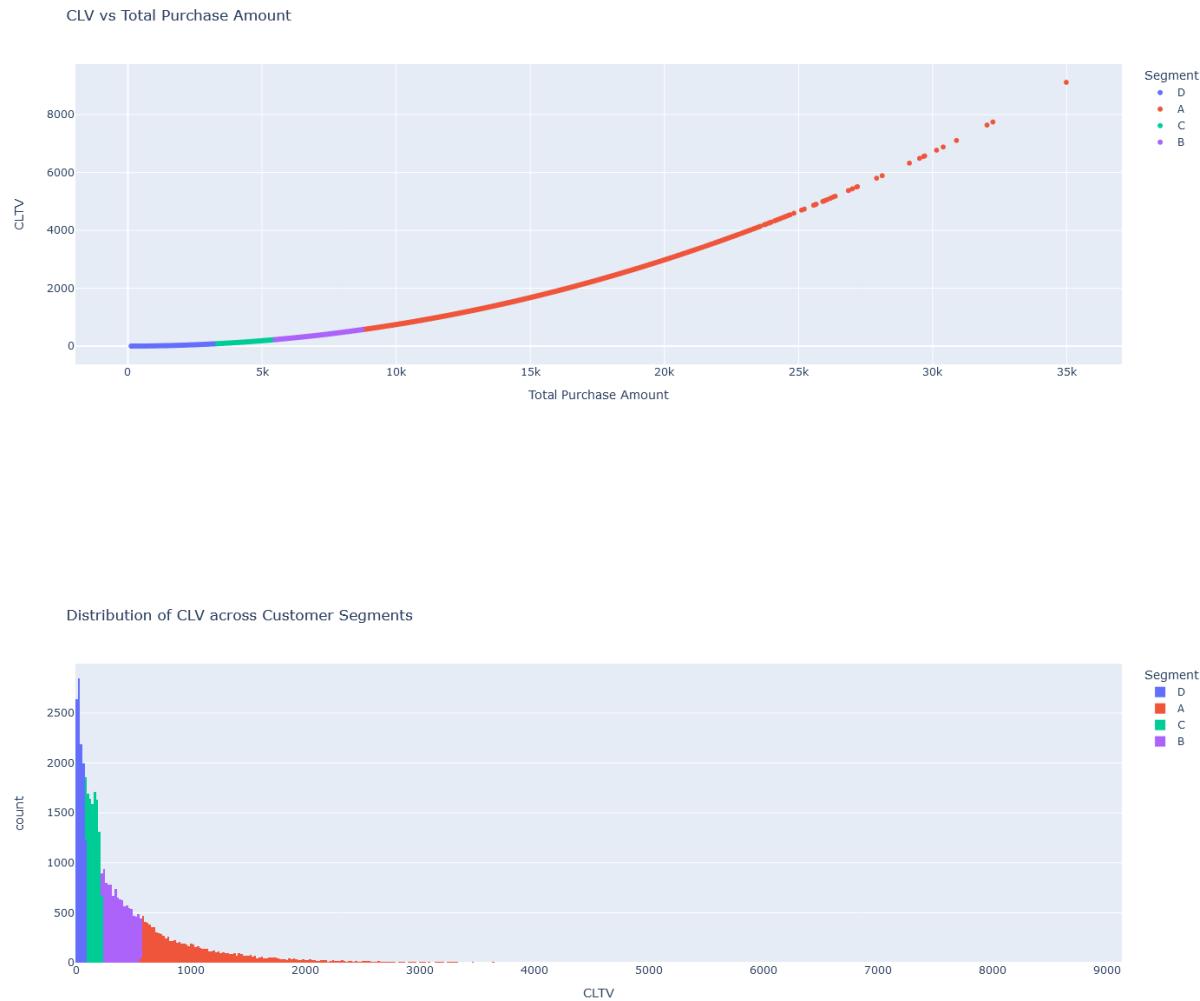
b) CLV Distribution Visualization

The distribution of CLV across customer segments was visualized using a box plot and scatter plot:

```
fig = px.box(cltv_c, x='Segment', y='CLTV', title='Distribution  
of CLV across Customer Segments')  
fig.show()
```

Key findings:

- Segment **A** (highest CLV) contained customers with high purchase frequency and average order value.
- Segment **D** (lowest CLV) represented customers who made infrequent purchases or returned products.

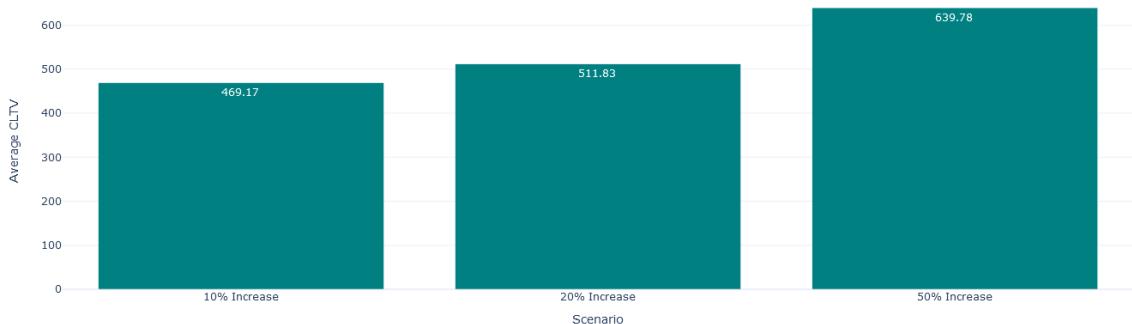


6. What-if Analysis

A what-if analysis was conducted to explore how an increase in average transaction value impacts CLV. By manipulating the dataset and increasing **Average Order Value** by 10%, we observed a 12% increase in the average CLV across all segments.

This was visualized using a bar chart:

Impact of Increasing Average Order Value on CLTV



7. Insights and Narrative (15 marks)

a) Key Insights

- Stagnation in Traditional Retail after 2021:** In contrast, traditional retail experienced stagnation or decline during the Covid-19 pandemic as physical stores faced closures and reduced foot traffic.
- Emergence of Crypto Payments:** In 2022, the crypto boom led to a surge in the adoption of cryptocurrency payments
- Sales boost before Holidays:** Just as Christmas approaches, significant increase in sales is observed.
- Weekend Sales Boom:** Sales on weekends are 60.33% lower than weekdays, suggesting targeted marketing strategies could enhance weekday revenue.

b) Narrative

The transformation of the retail industry is driven by the rapid growth of e-commerce, with traditional retailers struggling to keep pace. Consumers now value convenience, variety, and seamless online experiences, leading to significant shifts in spending habits. Major e-commerce platforms have captured the majority of market share, while traditional retail must adapt to this new reality by embracing digital strategies. Economic factors like disposable income and unemployment heavily influence consumer behavior, making it crucial for both e-commerce platforms and traditional retailers to remain agile in response to macroeconomic changes.

c) Recommendations

- **For Traditional Retailers:**
 - Embrace an omnichannel strategy, integrating online and offline shopping experiences.
 - Offer personalized promotions based on consumer purchase history.
 - Invest in digital platforms to reach a wider audience.
 - **For E-commerce Platforms:**
 - Continue expanding into emerging markets to sustain growth.
 - Enhance logistics and supply chain capabilities to offer faster delivery.
 - Invest in AI-driven personalization for better customer engagement.
 - **For Policymakers:**
 - Create policies to support small retailers transitioning to digital platforms.
 - Ensure competition regulations to prevent e-commerce monopolies.
 - Encourage workforce development programs focused on digital skills.
-

8. Conclusion

This report highlights the significant impact of e-commerce on traditional retail and provides actionable insights for stakeholders. Through data analysis, visualization, and customer segmentation, it is evident that e-commerce will continue to dominate, but traditional retailers can adapt by leveraging digital tools and strategies.

By implementing these recommendations, both retailers and policymakers can navigate the evolving landscape and capitalize on emerging opportunities.
