E-commerce Sales Analysis

A Project Report

submitted in partial fulfillment of the requirements

of

"AI/ML WITH CLOUD COMPUTING"

by

KR. BHUVANESH

bhuvaneskr2003@gmail.com

Under the Guidance of

Name of Guide (P.Raja, Mater Trainer)

ACKNOWLEDGEMENT

I would like to take this opportunity to sincerely thank everyone who helped me, directly or indirectly, in completing this project, *E-commerce Sales Analysis*.

First and foremost, I would like to express my gratitude to my supervisor, P. RAJA, for his continuous support and guidance throughout this project. His/her valuable insights and encouragement were essential in shaping this analysis and ensuring its successful completion. The confidence and trust placed in me were a constant source of motivation.

I also wish to thank class co-ordinator Dr.M.Anto alosius for providing the necessary resources and support for the analysis. It has been a valuable learning experience, and I am grateful for the opportunity to work on this project under P. RAJA'S mentorship.

E-commerce Sales Analysis is a project focused on analyzing the sales data of an e-commerce website to identify key business insights and inform decision-making.

Problem Statement: The objective is to analyze sales data to extract crucial information such as total sales, best-selling products, customer purchase patterns, and other vital metrics. The analysis helps the company better understand its performance, optimize marketing strategies, and improve inventory management.

Objectives:
1. Calculate total sales and revenue over a specified period.
2. Identify best-selling products based on units sold and revenue generated.
3. Analyze customer purchase patterns, including frequency of purchases and average order value.
4. Provide actionable insights for improving business operations and increasing sales.

Methodology:
- **Data Collection**: Sales data is gathered from the e-commerce platform, typically containing order details, product information, customer data, and transaction dates.
- **Data Cleaning**: The raw data is cleaned to handle missing or inconsistent values.
- **Data Analysis**: Key metrics are calculated, including total sales, sales per product category, customer segmentation, and purchase behavior trends. Various visualization techniques like bar charts and heatmaps are used to identify patterns and anomalies.
- **Tools**: Python, Excel, or data visualization tools like Tableau and Power BI may be used for the analysis.

Key Results:
- Total sales during the analysis period.
- Top-selling products contributing the most to revenue.
- Identification of high-value customers and their purchase patterns.
- Insights into seasonal or time-based purchase trends.

Conclusion: The analysis provides actionable insights into sales performance, enabling the company to optimize marketing efforts, adjust inventory, and target customer segments more effectively, ultimately driving revenue growth and improving customer satisfaction.

TABLE OF CONTENTS

LIST OF FIGURES

		Page No.
Figure 1	**Plot of Interaction Type by Buyer**	
Figure 2	**Distribution of seller By Amazon (YES/NO)**	
Figure 3	**Distribution of buyer's gender**	
Figure 4	**Purchase Category of Customer**	
Figure 5	**Item Purchased Distribution**	
Figure 6	**season vise customer**	
Figure 7	**Frequency of Purchases**	
Figure 8	**Payment Method**	
Figure 9	**Information of customer dataset**	

CHAPTER 1

Introduction

1.1 Problem Statement:

The project addresses the need to analyze sales data from an e-commerce website. As online shopping grows, businesses must understand their sales trends, best-selling products, and customer buying patterns. Without proper analysis, companies may miss opportunities, face inventory issues, or fail to improve their marketing strategies. This analysis is important because it helps businesses make better decisions based on data, improving overall performance.

1.2 Motivation:

This project was chosen because data is a key resource for e-commerce companies. With so much sales data available, businesses need insights to make informed decisions. The potential benefits include better stock management, targeted marketing, and understanding customer behavior. The project's impact could help the company increase sales, improve customer satisfaction, and stay competitive in the market.

1.3 Objective:

The main goals of this project are:
- To calculate the total sales over a certain period.
- To identify the best-selling products based on units sold and revenue.
- To study customer purchase patterns, such as how often they buy and how much they spend.
- To offer insights that will help the company improve marketing, manage stock, and engage customers better.

1.4 Scope of the Project:

The project focuses on analyzing sales data from the company's e-commerce website for a specific time period. It includes cleaning the data, calculating key sales figures, and spotting trends in customer behavior. The analysis is limited to past data and does not involve predicting future trends. While customer behavior is studied, detailed analysis of specific customer segments or external factors like competitor data is not covered in this project. The focus remains on improving internal business decisions based on sales data.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

E-commerce sales analysis has become important for businesses to understand customer behavior and improve decision-making. Past studies have focused on analyzing sales trends, customer demographics, and purchase behaviors to help businesses optimize their operations. Researchers like Kotler (2010) and Brynjolfsson et al. (2013) discuss how insights from data can guide marketing strategies and improve product offerings. Companies like Amazon have used these techniques to understand their customers better and boost sales.

2.2 Mention any existing models, techniques, or methodologies related to the problem.

Common techniques used in e-commerce sales analysis include:

- o Demographic Analysis: Understanding customer characteristics such as gender or location to target marketing efforts.
- o Product and Seller Analysis: Assessing how well products sell and how different sellers, such as Amazon, perform.
- o Purchase Behavior Analysis: Using tools like Market Basket Analysis to understand what customers buy and how often.
- o Payment Method Analysis: Analyzing customer preferences for payment options.
- o Seasonal Analysis: Studying how customer purchases change throughout the year to plan promotions and manage inventory.
- o Purchase Frequency: Analyzing how often customers buy to identify loyal customers and improve retention.

2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

Existing methods often focus on large companies and offer broad insights. They may not provide detailed information, such as how gender affects purchasing or how preferences change seasonally. Additionally, smaller businesses may struggle to apply complex models.

How This Project Will Address These Gaps: This project will provide detailed insights by analyzing gender distribution, seller performance (AMZN), product purchases, payment methods, customer interaction, seasonal trends, and purchase frequency. This approach will help smaller businesses understand their customers better and make data-driven decisions to improve sales.

CHAPTER 3
Proposed Methodology

3.1   Data Flow Diagram
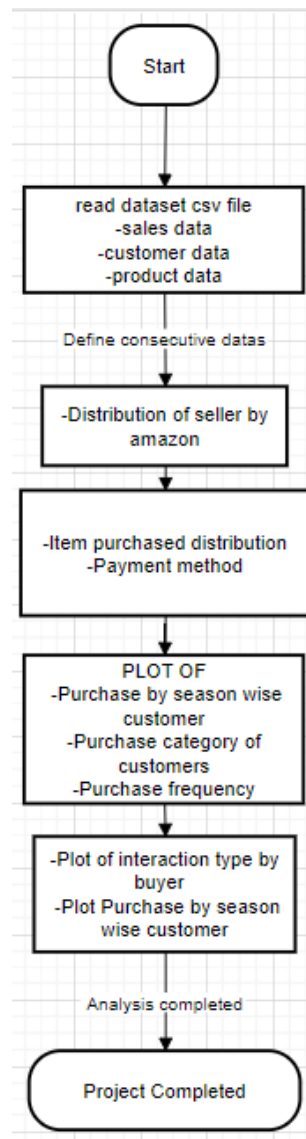
3.2   Advantages:

- Comprehensive Analysis This holistic approach helps identify trends and patterns that may be overlooked in single-factor analyses.
- Targeted Marketing Insights purchasing specific products helps create more effective advertising campaigns that resonate with the target audience.
- Seller Performance Evaluation
- Understanding Payment Preferences for increased customer satisfaction.
- Customer Interaction Insights help businesses enhance user experience and optimize product placement or recommendations.

- This foresight can improve sales during peak seasons and help manage stock levels effectively.
- By understanding how often customers buy, businesses can implement targeted loyalty programs and promotions to enhance customer engagement.
- Overall, this methodology provides actionable insights, businesses can make data-driven decisions that lead to improved sales and customer satisfaction.

3.3 Requirement Specification

3.3.1 Hardware Requirements:

1. **Processor (CPU)**:(Minimum: Dual-core processor, Recommended: Quad-core or higher for faster data processing and model training.)
2. **Memory (RAM)**:(Minimum: 8 GB, Recommended: 16 GB or more for handling larger datasets and running complex models)
3. **Storage**:(Minimum: 10 GB of available disk space, Recommended: 50 GB or more for storing datasets, models, and outputs)
4. **Graphics Processing Unit (GPU)**:(Recommended: NVIDIA GPU with CUDA support for accelerated model training, especially for deep learning tasks. Google Colab provides access to GPUs and TPUs)
5. **Network**:(Reliable internet connection to access Google Colab and datasets stored online (e.g., Google Drive, Kaggle))

3.3.2 Software Requirements:

1. **Operating System**:(Google Colab runs in a cloud-based environment, so no specific operating system is required on the user's local machine. However, a web browser (Chrome, Firefox, etc.) is needed)
2. **Programming Language**:(Python 3.x: Ensure the project is compatible with Python 3, which is the default in Google Colab)
3. **Libraries and Frameworks**:
     - **Data Manipulation**:(Pandas, NumPy)
     - **Data Visualization**:(Matplotlib,Seaborn)
     - **Machine Learning**:(Scikit-learn,TensorFlow or PyTorch)
4. **Development Environment**:(Google Colab, which provides an interactive Jupyter notebook environment with built-in support for many libraries.)
5. **Version Control (Optional)**:(Git for tracking changes and collaborating on code (can be integrated with GitHub))
6. **Data Storage Solutions**:(Google Drive for storing and accessing datasets easily within Google Colab)

CHAPTER 4

# Implementation and Result

4.1 E-commerce Sales Analysis

Ploting of Interaction Type by Buyer

Fig.1 Plot of Interaction Type by Buyer

Fig.2 Distribution of seller By Amazon (YES/NO)

Fig.3 Distribution of buyer's gender

```python
import matplotlib.pyplot as plt
import seaborn as sns
df['Interaction type'].value_counts().plot(kind='pie',autopct='%1.1f%%',labels=['like','view','purchease'],explode=[0,0,0
plt.ylabel(' ')
plt.title('Distribution of Type of intraction')
plt.show()
```
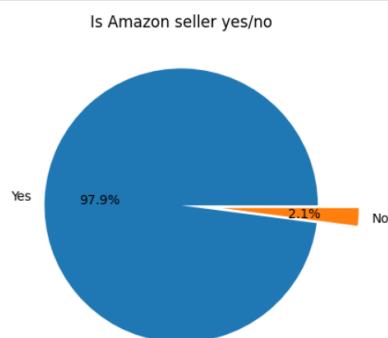
Distribution of Type of intraction
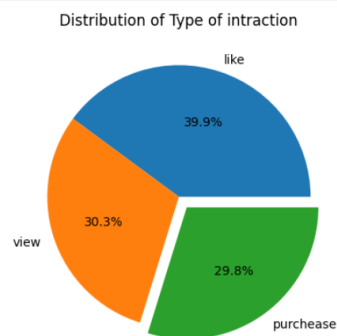
Distribution of seller By Amazon(YES/NO)

```python
df1['Is Amazon Seller'].value_counts().plot(kind='pie',autopct='%1.1f%%',labels=['Yes','No'],explode=[0,0.3])
plt.ylabel(' ')
plt.title('Is Amazon seller yes/no')
plt.show()
```

Is Amazon seller yes/no

Distribution of gender

```python
df3['Gender'].value_counts().plot(kind='bar')
plt.title('Buyer distribution by gender')
plt.tight_layout()
plt.show()
```

Buyer distribution by gender
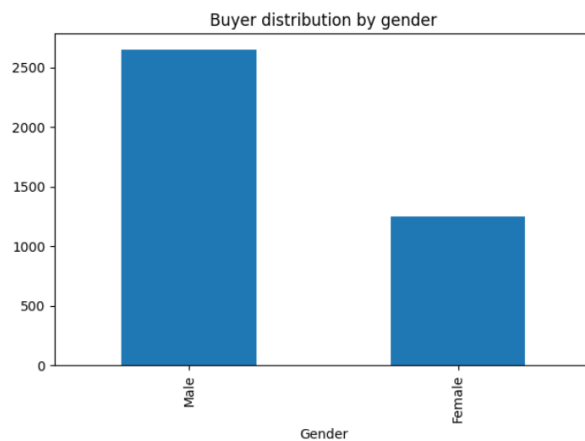
4.2 Customer purchase

Purchase Category of Customer

```
In [23]: df3['Category'].value_counts().plot(kind='bar')
         plt.title('Category')
         plt.tight_layout()
         plt.show()
```
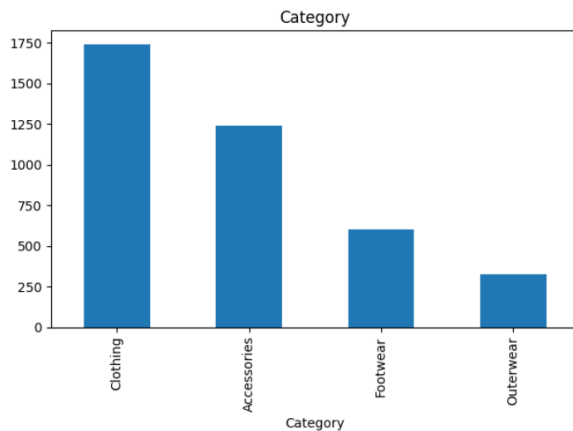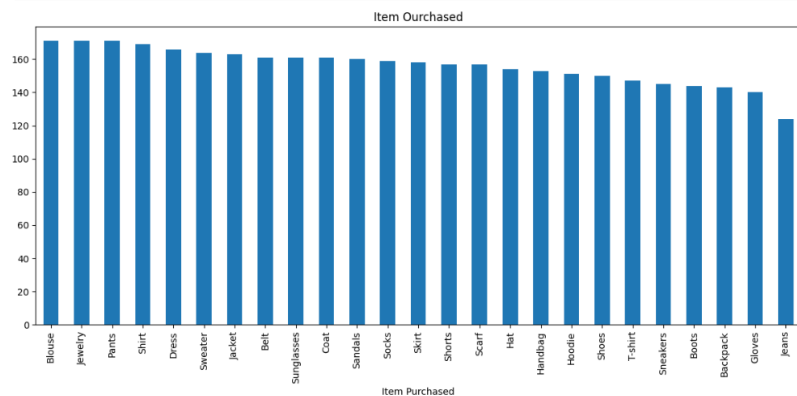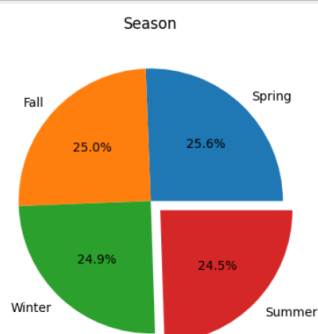

Item Purchased Distribution

```
In [24]: plt.figure(figsize=(12,6))
         df3['Item Purchased'].value_counts().plot(kind='bar')
         plt.title('Item Ourchased')
         plt.tight_layout()
         plt.show()
```
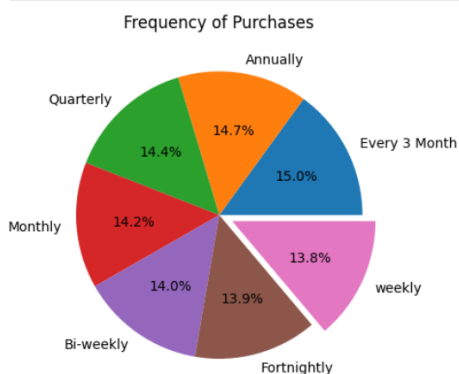

season vise customer

```
In [25]: df3['Season'].value_counts().plot(kind='pie',autopct='%1.1f%%',labels=['Spring','Fall','Winter','Summer'],explode=[0,0,0,
         plt.ylabel(' ')
         plt.title('Season')
         plt.show()
```

Frequency of Purchases

```
In [26]: df3['Frequency of Purchases'].value_counts().plot(kind='pie',autopct='%1.1f%%',labels=['Every 3 Month','Annually','Quarte
         plt.ylabel(' ')
         plt.title('Frequency of Purchases')
         plt.show()
```


Payment Method

```
In [27]: df3['Payment Method'].value_counts().plot(kind='bar')
         plt.title('Payment Method')
         plt.tight_layout()
         plt.show()
```

Fig.7 Frequency of Purchases
Fig.8 Payment Method
Fig.9 Information of customer dataset
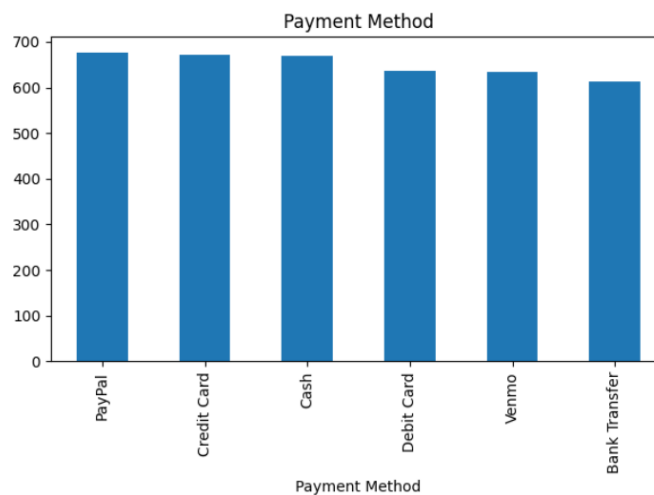
Customer Dataset

Import Customer dataset

```
In [20]: df3=pd.read_csv('/content/customer_details.csv')
         df3.head()
```

Out[20]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping

Information of dataset

CHAPTER 5

Discussion and Conclusion

5.1   Key Findings:

This analysis could help you make informed decisions in customer segmentation, marketing strategies, and operational improvements.

5.2   Git Hub Link of the Project:

Repository link: https://github.com/BHUVANESH-KR/AI_ML_WITH_CC.git

5.3   Video Recording of Project

Demonstration: https://drive.google.com/file/d/14J-WDWaNZIFWr4_muLwUU0Rifh1tUx1B/view?usp=drive_link

5.4   Limitations:

- The data may not fully represent all customer groups, which could lead to biased results.
- Seasonal trends are based on a short time period, possibly missing long-term patterns.
- Insights from Amazon sellers may not apply to other platforms.
- The focus on gender limits a deeper understanding of other customer demographics.
- Not all payment options are included, potentially missing newer trends.
- Customer preferences and market changes over time are not accounted for.
- Simplified interaction data may overlook more complex customer behaviors.

5.5   Future Work:

To improve the model, consider expanding the dataset to include more diverse customer demographics (age, income, location) for a broader analysis. Include data from other e-commerce platforms to enhance generalizability, applying machine learning techniques for predicting customer behavior could enhance decision-making and accuracy in future analyses.

5.6   Conclusion:

The project highlights key areas for enhancing user experience, particularly through checkout and product interaction improvements. Though there are some limitations, such as incomplete data and platform-specific focus, the analysis sets the stage for future improvements and more precise strategies in e-commerce operations and customer engagement.

REFERENCES

[1] **BigCommerce_ECommerce company: Ecommerce Analytics: How to Leverage the Power of Data for Your Business.**
https://www.bigcommerce.com/articles/ecommerce/ecommerce-analytics/

Appendices (if applicable)

Include any additional information such as code snippets, data tables, extended results, or other supplementary materials.