E-commerce Sales Analysis

A Project Report

submitted in partial fulfillment of the requirements

of

Fundamentals with cloud computing and gen AI

by

KR. BHUVANESH

bhuvaneskr2003@gmail.com

1DFA1F1233D2DDA882B4B949989D9AFD (au912421114002)

912421114002

Under the Guidance of

P. Raja

Master trainer, Edunet foundation

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.

ABSTRACT of the Project

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

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:

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

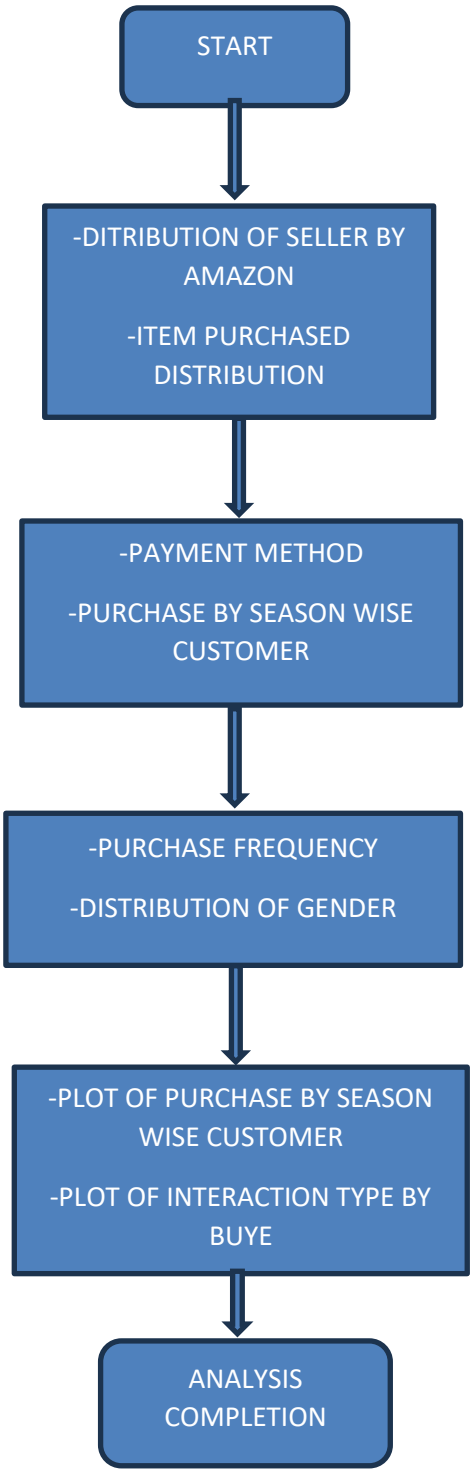
Existing methods often focus on large companies and offer broad insights. They may not provide detailed information, such as how gender affects purchasing or how preferences change seasonally. Additionally, smaller businesses may struggle to apply complex models.

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)

Project Requirements:

1. **Data Requirements**
    - Customer Data: Information about customers, including gender, demographics, and purchasing habits.
    - Sales Data: Detailed records of each sale, including item ID, category, quantity, purchase date, payment method, and season (or date for calculating season).
    - Seller Data: Information on each seller, including seller ID, if they are managed by Amazon or a third-party, and their product listings.
    - Interaction Data: Records of customer interactions, including actions such as browsing, adding items to the cart, and purchasing.
    - Product Data: Details of each product available in the e-commerce platform, like category, price, and availability.
2. **Analysis Requirements**
    - Descriptive Analytics: Provide a breakdown of key metrics such as the distribution of gender, payment methods, and seasonal purchase trends.
    - Behavioral Analytics: Track and categorize customer behavior, like interaction types, purchase frequency, and product preferences.

- Segmentation Analysis: Divide customers into segments (e.g., loyal customers, frequent buyers) for more targeted insights and marketing strategies.
- Trend Analysis: Identify patterns or trends in purchases based on seasonal, daily, or monthly patterns, helping to predict demand.
- Seller Analysis: Evaluate seller performance, especially differentiating between Amazon-managed sellers and third-party sellers.

3. **Visualization Requirements**
   - Create visualizations for each major entity to provide an intuitive understanding of the data. This includes:
     - Bar Charts for gender distribution, seller types, and payment methods.
     - Pie Charts to show proportional data, like the distribution of product categories or customer types.
     - Line Charts for trends over time (e.g., seasonal or monthly purchase trends).
     - Heatmaps to show frequency data or seasonal variations in customer purchases.
   - Ensure visualizations are clear and informative for a non-technical audience.
4. **Reporting Requirements**
   - Executive Summary: A high-level overview of findings, highlighting major trends and key insights.
   - Detailed Report: A comprehensive report breaking down each part of the analysis (e.g., gender distribution, payment preferences, seasonal trends).
   - Recommendations: Based on the analysis, provide actionable insights (e.g., recommended payment methods to promote, popular products to stock in certain seasons).
5. **Data Security and Compliance**
   - Privacy Compliance: Ensure all customer data is handled in compliance with data privacy regulations (e.g., GDPR, CCPA).

- Data Anonymization: Mask or anonymize personal identifiers in customer data to protect privacy, especially for data used in public or shared reports.

6. **Performance Requirements**

   - Data Processing Speed: Ensure data processing is efficient enough to handle large datasets within reasonable time frames.
   - Data Accuracy: Accuracy in data cleaning, analysis, and reporting is essential to avoid misleading insights and decisions.
   - Scalability: The analysis pipeline should be adaptable to growing datasets, which may increase over time as sales or customer bases grow.

7. **Documentation and Maintenance**

   - Documentation: Include clear documentation of data sources, preprocessing steps, and analysis methods, enabling future users or analysts to understand and replicate the process.
   - Version Control: Use version control to track changes in datasets, code, and reports, which helps manage updates and improvements over time.

These requirements will help you structure the project comprehensively, ensuring all necessary data, analysis, and reporting elements are addressed while maintaining data accuracy, security, and performance. Let me know if you'd like further details on any specific requirement!

CHAPTER 4

Implementation and Result

4.1 E-commerce Sales Analysis

1. Data Loading and Preprocessing
    - Begin by loading your e-commerce dataset and examining its structure to understand the types of data you're working with, such as dates, categories, numerical values, and text fields.
    - Perform any necessary data cleaning, like handling missing values, converting data types, or correcting inconsistent values.

Analysis Steps:

1. Distribution of Gender

- Determine the proportion of male and female customers within your data.
- This analysis provides insight into the demographics of your customer base, which is valuable for targeted marketing.

2. Distribution of Sellers by Amazon

- Identify the distribution of different sellers who list their products on Amazon, which could include sellers directly managed by Amazon and third-party sellers.
- This analysis shows the variety and competitiveness of sellers, helping to understand if there are certain seller segments dominating the platform.

3. Item Purchased Distribution

- Analyze the types of items most frequently purchased by customers. This could involve looking at categories of products or specific items.
- High-frequency items indicate customer preferences, which can inform inventory management, promotional campaigns, and personalization.

4. Payment Method Distribution

- Examine the preferred payment methods of customers (e.g., credit cards, PayPal, cash on delivery).
- This information is useful for understanding customer payment preferences, potentially guiding partnerships with payment providers and improving customer experience.

5. Plot of Interaction Type by Buyer

- Assess how customers interact with the platform, such as browsing, adding items to the cart, or making purchases.
- This analysis can help you understand customer behavior on the platform and highlight areas where user experience improvements could encourage more conversions.

6. Purchase by Season-Wise Customer

- Use seasonal analysis to examine if there are specific times of the year when purchases spike, like during holiday seasons or summer sales.
- Identifying seasonal trends allows for strategic stocking, staffing, and marketing efforts, maximizing sales during peak seasons.

7. Purchase Category of Customers

- Classify customers based on their purchasing behavior, such as occasional shoppers, loyal customers, or bulk buyers.
- This segmentation can help in crafting personalized marketing strategies and loyalty programs, enhancing customer retention.

8. Purchase Frequency

- Measure how often customers make purchases, looking for trends like repeat customers or high-frequency buyers.
- Understanding purchase frequency can highlight loyal customers or those who may benefit from loyalty incentives, increasing engagement and brand loyalty.

Results Interpretation:

- **Demographics (Gender, Seller Distribution):**
  Helps understand customer and seller demographics, supporting personalized marketing strategies and identifying successful seller profiles.
- **Product and Payment Preferences:**
  Provides insight into what customers like to buy and how they prefer to pay, guiding inventory planning and payment method optimizations.
- **Seasonal Trends:**
  Offers a seasonal purchase overview, allowing preparations for high-demand periods with targeted promotions and inventory management.

- **Customer Segmentation and Purchase Frequency:**
  Identifies high-value customer segments and frequent buyers, which can enhance the effectiveness of loyalty programs and promotional strategies.

This structured approach will give a comprehensive view of e-commerce data, highlighting trends that can drive customer engagement, sales, and overall strategic decision-making.

Interpreting these results together provides a full picture of customer demographics, product preferences, seasonal trends, and engagement patterns. This comprehensive view can guide strategies in marketing, inventory management, and customer relationship building, all aimed at enhancing customer satisfaction, loyalty, and overall sales performance on the platform.

RESULT PICTURES:

Fig.1 Plot of Interaction Type by Buyer

Ploting of Interaction Type by Buyer

In [10]:
```
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()
```

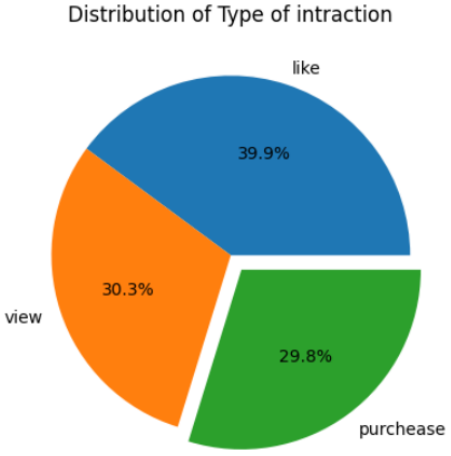


Fig.2 Distribution of seller By Amazon (YES/NO)

Distribution of seller By Amazon(YES/NO)

In [19]:
```
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()
```

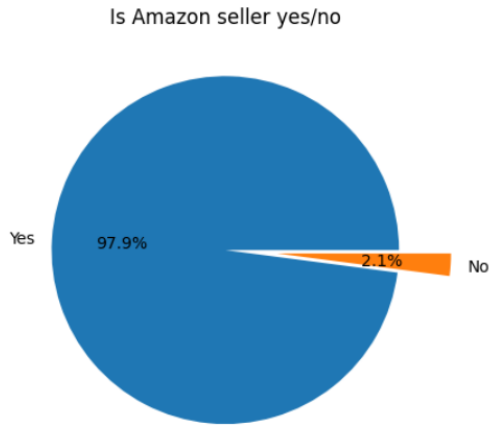


Fig.3 Distribution of buyer's gender

Distribution of gender

In [22]:
```
df3['Gender'].value_counts().plot(kind='bar')
plt.title('Buyer distribution by gender')
plt.tight_layout()
plt.show()
```

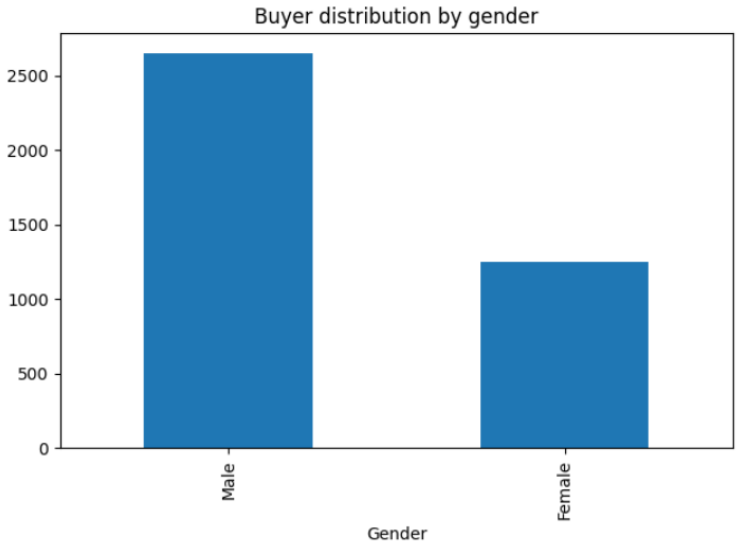


Fig.4 Purchase Category of Customer

Purchase Category of Customer

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

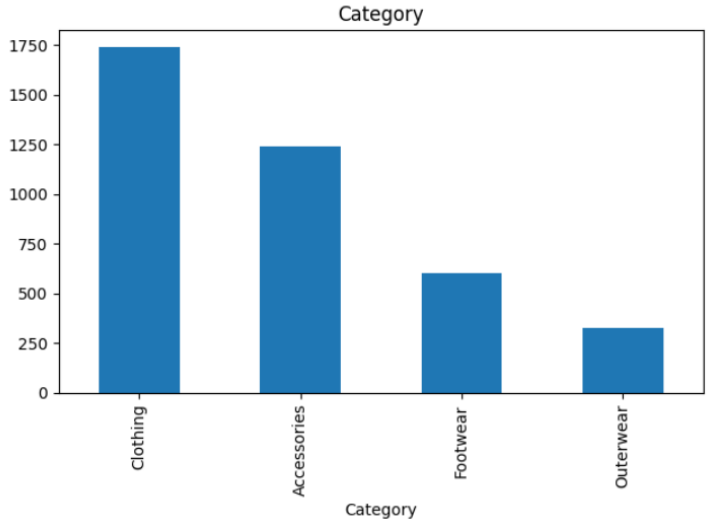


Fig.5 Item Purchased Distribution

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()
```

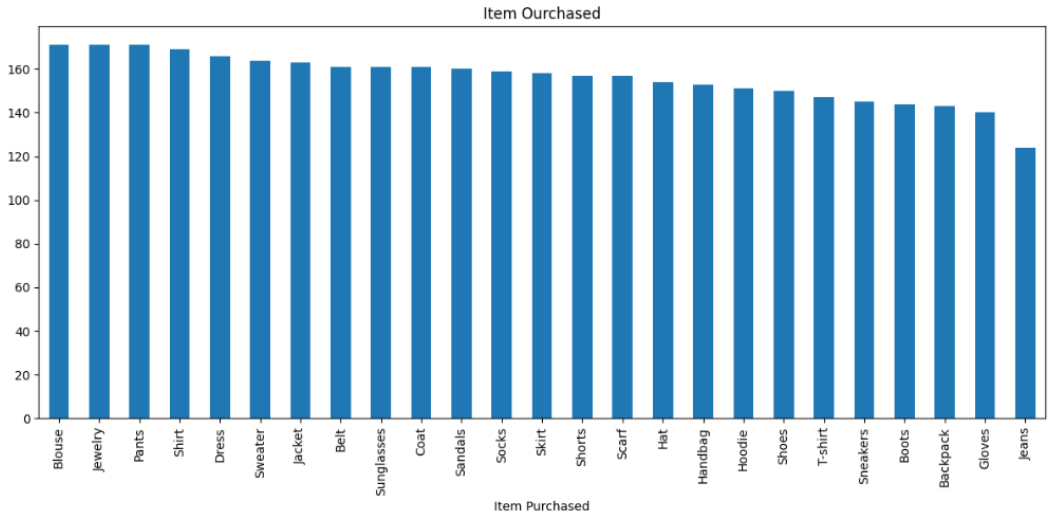


Fig.6 season vise customer

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()
```

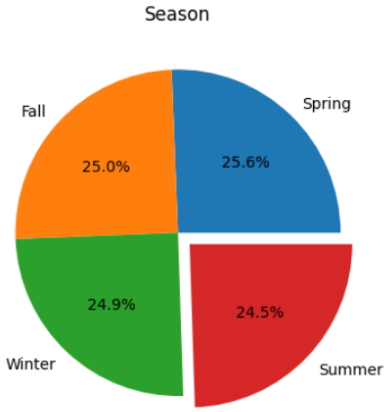


Fig.7 Frequency of Purchases

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()
```

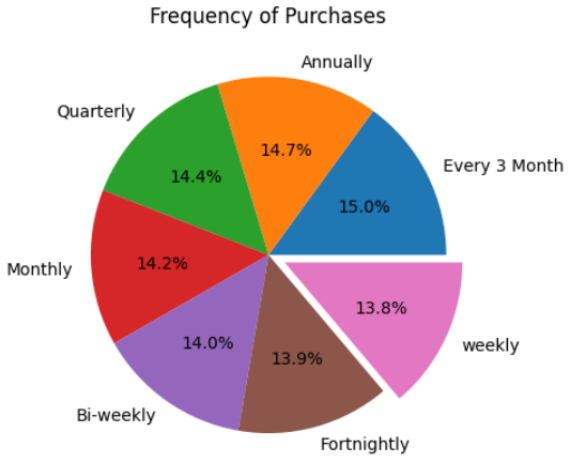


Fig.8 Payment Method

Payment Method

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

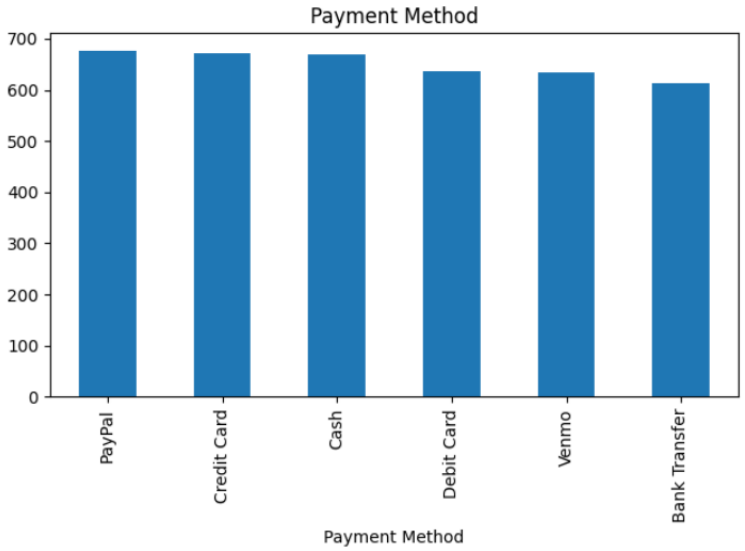


Fig.9 Information of customer dataset

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.