AMAZON SALES TREND ANALYSIS

A PROJECT REPORT

Table of Contents

Table OF Figures

ABSTRACT

In the face of intensifying competition, sales management has emerged as a crucial aspect of commercial and business operations. The need for effective sales management is driven by the dual objectives of reducing costs and increasing profits through the optimization of distribution methods. This project delves into the importance of sales management as a central function in business enterprises, emphasizing its role in navigating the complexities of the modern marketplace.

To achieve these objectives, the study employs a range of analytical techniques, including descriptive analysis, trend analysis, linear regression modelling, and data visualization. These methods are applied to sales trend data to uncover patterns, forecast future sales, and identify key areas for improvement. By examining various strategies and approaches, the project highlights how improved sales management practices, supported by robust analytical techniques, can lead to more efficient distribution, ultimately contributing to enhanced profitability and sustained business growth. Through a comprehensive analysis of sales management techniques, this study seeks to provide insights into how businesses can better meet the demands of a competitive environment and achieve long-term success.

CHAPTER 1

INTRODUCTION

1.1 Introduction:

In the rapidly evolving global market, sales management has become a critical function for businesses to thrive. As competition intensifies across industries, companies must adopt innovative distribution methods to lower operational costs while simultaneously increasing profitability. Sales management plays a vital role in navigating these challenges by providing strategies to optimize revenue streams and improve market responsiveness. This project focuses on analyzing sales trends for Amazon, a global leader in e-commerce, with the objective of identifying key sales patterns and developing actionable insights to drive operational efficiency and profitability.

1.1.1 About Amazon

Amazon :"Work Hard. Have Fun. Make History."

A Global E-Commerce Powerhouse

Amazon, Inc., founded by Jeff Bezos in 1994, has evolved from a modest online bookstore into a major global force in e-commerce. Initially launched as a digital platform for books, Amazon's rapid expansion has turned it into a comprehensive global enterprise, offering a wide range of products and services, including cloud computing, artificial intelligence, digital streaming, and logistics.

Evolution and Expansion

Amazon's transformation from a startup in Bezos's garage to a global leader in online retail reflects its innovative strategy and strategic growth. The company's early success was driven by its focus on providing a broad selection of books at competitive prices through a user-friendly online experience. This commitment to customer satisfaction set the stage for Amazon's expansion into other categories such as electronics, clothing, and household goods. The early 2000s saw Amazon diversify rapidly, launching Amazon Web Services (AWS) in 2006, which revolutionized cloud computing and became a major revenue source. The acquisition of Whole Foods Market in 2017 marked Amazon's strategic entry into physical retail, merging online and offline shopping experiences.

Technological Innovations

Amazon's success is deeply tied to its relentless pursuit of technological advancement. The company has introduced significant innovations in e-commerce technology, such as one-click purchasing, personalized recommendations, and sophisticated search algorithms. Investments in artificial intelligence (AI) and machine learning have transformed online shopping by improving recommendation accuracy and optimizing inventory management. Amazon Echo and the Alexa voice assistant exemplify the company’s prowess in consumer electronics, setting new standards for smart home devices and integrating voice technology into everyday life.

Business Model and Strategies

Amazon's success is rooted in its unique business model, which features a customer-centric approach, a vast product range, and competitive pricing. The Prime membership program is a key example, offering benefits like free shipping, exclusive deals, and access to streaming services, which not only generates recurring revenue but also strengthens customer loyalty. Amazon’s marketplace platform allows third-party sellers to reach a global audience, enhancing the company’s product diversity and revenue while creating a vibrant ecosystem of continuous growth and innovation.

Customer-Centric Approach: Amazon’s focus is on delivering an outstanding customer experience through a user-friendly website, a wide array of products, competitive pricing, and quick shipping.

The Prime membership program highlights this commitment with benefits such as free shipping, exclusive deals, and streaming access.

Marketplace Platform: Amazon's marketplace enables third-party sellers to list products,broadening the available product range and creating a competitive environment. This benefits sellersby providing access to Amazon's extensive customer base and benefits Amazon through increased product diversity and seller fees.

Subscription Services: Beyond Prime, Amazon offers various subscription services like Amazon Music Unlimited and Amazon Fresh, contributing to recurring revenue and increased customer engagement.

Global Reach and Local Adaptation: Amazon's international presence is bolstered by localized operations that cater to specific markets. The company adapts its offerings and services to fit local preferences and regulations, demonstrating its ability to operate effectively across diverse global markets.

Amazon's impact extends beyond e-commerce, influencing many aspects of the global economy and society. Its logistical infrastructure, including an extensive network of fulfilment centres and delivery services, has set new standards for efficiency and speed, reshaping consumer expectations for delivery times and convenience. Additionally, Amazon's global reach affects international trade and retail practices, showcasing its strategic versatility. However, its dominance also raises concerns about market concentration, competitive practices, and the effects on small businesses.

Retail Industry: Amazon's influence in online retail has transformed consumer expectations and competition, forcing traditional retailers to adapt to the rise of e-commerce and blend online and offline retail experiences.

Technology Sector: Amazon's advancements in cloud computing, AI, and smart devices have positioned it as a key player in the tech sector, with AWS driving the growth of cloud services and influencing other technology companies.

Employment and Labor Practices: Amazon's large-scale operations have created many job opportunities but also led to debates over labour practices and working conditions. The company has faced scrutiny regarding warehouse conditions, wages, and labour rights.

Supply Chain and Logistics: Amazon’s innovations in logistics and supply chain management have set new benchmarks, with investments in its delivery network and partnerships improving order fulfilment efficiency.

Ethical and Social Considerations

As a major global corporation, Amazon faces scrutiny over its ethical and social responsibilities. Concerns include labour conditions, environmental sustainability, and data privacy. The company has undertaken efforts to address these issues, such as improving worker conditions and committing to reducing carbon emissions, reflecting its recognition of broader societal responsibilities.

Environmental Impact: Amazon’s operations have a substantial environmental footprint, including carbon emissions from transportation and logistics. The company has committed to achieving net-zero carbon by 2040 and investing in renewable energy projects.

Data Privacy and Security: The vast amount of data collected from customers raises concerns about privacy and security. While Amazon has implemented measures to protect user information, data breaches and privacy issues remain a concern.

Corporate Social Responsibility: Amazon has engaged in various initiatives to address social and community issues, including charitable contributions and local community support. However, these efforts are often evaluated in the context of the company's overall societal impact.

Future Prospects and Challenges

Looking ahead, Amazon’s trajectory will be influenced by technological advancements, changing consumer preferences, and regulatory developments. The company’s focus on innovation and diversification positions it for continued growth, though it must navigate challenges related to competition, market saturation, and regulatory pressures.

Innovation and Technology: Amazon's ability to maintain a competitive edge will depend on continued innovation and technological advancements. Progress in AI, robotics, and logistics is likely to drive future growth.

Regulatory Environment: The company's global operations face regulatory scrutiny and potential changes in laws related to antitrust, labour practices, and data privacy. Successfully navigating these challenges will be crucial for Amazon's ongoing success.

Market Competition: The e-commerce sector is increasingly competitive, with major players like Alibaba, Walmart, and Microsoft competing for market share. Amazon must adapt to evolving market conditions and consumer preferences to retain its leadership position.

Sustainability and Ethical Practices: Addressing environmental and social concerns will be essential for Amazon's long-term success and reputation. The company's commitment to sustainability and ethical practices will impact its relationships with customers, employees, and stakeholders.

1.1.2 About Unified Mentor

Unified Mentor's journey started with a steadfast ambition to transform the face of online Mentor and equip people to succeed in the fast-paced world of technology. Established in 2022, Unified Mentor is a shining example of how lofty goals and real accomplishments can be reconciled. Unified Mentor has assumed the role of a transformational catalyst by concentrating on web development, data science, and digital marketing and encouraging growth, innovation, and excellence.

At Unified Mentor, it is deeply believed that personal growth drives our collective success. They are always on the lookout for imaginative, talented individuals who love challenges and strive for excellence. If one is ambitious and passionate about your goals, they are the perfect fit for their vision of Leadership through Global Talent Development. Unified Mentor has a vibrant community which is a blend of skilled student mentors and experienced industry professionals. They work hand in hand to create top-notch online courses across a range of subjects. With their guidance, one receives an exceptional learning experience that's rooted in real-world insights.

Unified Mentor's mission is simple yet powerful: to enhance employability through personalized skill development programs

1.2 Identification of Client & Need:

The client, Amazon, is one of the most successful e-commerce platforms worldwide, serving millions of customers across different regions. As an organization operating in a highly competitive industry, Amazon's need for precise sales management is paramount. Effective management of sales trends is crucial to streamline operations, reduce distribution costs, and enhance customer satisfaction. The need for this analysis arises from Amazon's continuous pursuit of efficiency, where data-driven decisions are key to maintaining its market leadership.

1.3 Relevant Contemporary Issues:

Amazon faces several contemporary issues that make sales trend analysis essential. The increasing costs of logistics, global supply chain disruptions, and competition from other e-commerce platforms present significant challenges. Additionally, shifting consumer behaviour and the rise of digital transformation demand that Amazon optimize its sales strategies in real time. Understanding the month-wise, year-wise, and yearly month-wise trends will help Amazon enhance its distribution and sales approach, providing a competitive edge.

1.4 Problem Identification:

The problem revolves around Amazon's need to track sales trends over time to identify areas for cost reduction and profit maximization. By analysing historical sales data, the company can uncover seasonality patterns, region-specific trends, and fluctuations in consumer demand. These insights can guide Amazon in refining its sales and distribution models, ensuring resources are allocated efficiently.

1.5 Task Identification

1. **Data Extraction:** Extracting the relevant Amazon sales dataset.
2. **Data Transformation:** Cleaning and structuring the data to ensure accuracy and consistency.
3. **Data Analysis:** Analyzing sales trends across different time frames, including month-wise, year-wise, and yearly month-wise sales trends.

4. **Sales Forecasting:** Using statistical models to forecast future sales trends and highlight areas for optimization.
5. **Recommendations:** Developing recommendations to improve Amazon's distribution and sales strategies based on the analysis.

1.6 Timeline:

- **Week 1:** Data extraction and cleaning (ETL process).
- **Week 2-3:** Analysis of month-wise, year-wise, and yearly month-wise sales trends.
- **Week 4:** Forecasting and preparing final insights.
- **Week 5:** Presentation and report finalization.

1.7 Organization of the Report:

1. **Chapter 1: Introduction**
2. **Chapter 2: Literature Survey**
3. **Chapter 3: Methodology**
4. **Chapter 4: Analysis and interpretation**
5. **Chapter 5: Findings And Suggestions**
6. **Conclusions**
7. **References / Bibliography**

CHAPTER 2

LITERATURE SURVEY

The literature survey is a crucial part of any research project, offering a comprehensive understanding of the current state of knowledge related to the problem. It helps to identify gaps, trends, and the relevance of the study within the global research landscape. In the context of Amazon sales trend analysis, the literature survey will focus on how similar issues of sales management, competition, distribution optimization, and sales forecasting have been investigated globally, along with solutions proposed by different researchers. This chapter also highlights a bibliometric analysis of relevant research, a problem definition, and the project's goals and objectives.

2.1 Timeline of the Reported Problem as Investigated Globally

The global e-commerce industry has undergone significant transformations in recent years, driven by technological advancements and shifts in consumer behaviour. Sales trend analysis and forecasting have been key areas of research due to the growing need for businesses to respond quickly to market changes and consumer demands. Below is an outline of how this problem has been investigated over time:

1. **Early 2000s:**

   Initial research focused on basic statistical methods like linear regression for sales trend analysis. E-commerce platforms, including Amazon, began to explore data-driven approaches to optimize sales and distribution.

2. **2010–2015:**

   The rise of big data analytics and machine learning allowed researchers to improve sales forecasting accuracy. Models like ARIMA were widely adopted to capture seasonality and sales fluctuations.

3. **2016–2020:**

   Advanced models such as Seasonal ARIMA (SARIMA) and neural networks became popular for handling complex e-commerce sales trends, including cyclical and seasonal variations.

4. **2021–Present:**

   The focus has shifted to real-time sales data analysis, combining AI with cloud computing. Modern solutions involve real-time optimization of sales strategies and distribution methods to reduce costs and increase efficiency in the highly competitive e-commerce landscape.

2.2 Bibliometric Analysis

A bibliometric analysis involves the quantitative assessment of published research on a specific topic, revealing trends, key areas of focus, and influential studies in the field. For sales trend analysis, a bibliometric analysis reveals:

- **Keywords:** Commonly used terms in research include "sales forecasting," "time series analysis," "ARIMA," "machine learning," and "e-commerce trends."

- **Journals:** Key research is published in high-impact journals related to business analytics, supply chain management, and artificial intelligence.

- **Research Output:** Over the past decade, there has been a sharp increase in the number of research papers on sales forecasting and trend analysis, reflecting the importance of data-driven sales management in e-commerce.

- **Citation Trends:** Studies on machine learning and AI-driven sales forecasting models receive high citations, indicating their relevance in improving sales performance and profitability in globalmarkets.

2.3 Proposed Solutions by Different Researchers

1. **Time Series Forecasting:** Many researchers have proposed time series forecasting models, including ARIMA, SARIMA, and exponential smoothing, to analyze sales trends. These models help in identifying seasonality and cyclic patterns in sales data.

2. **Machine Learning Models:** With the advancement of machine learning, algorithms like Random Forest, Gradient Boosting, and Neural Networks have been introduced to improve the accuracy of sales trend predictions by learning from historical data**.**
3. **Optimization Models:** Some researchers have combined sales trend forecasting with optimization techniques to minimize costs and maximize profits. This involves refining supply chain strategies based on predicted sales data.
4. **Real-Time Data Analysis:** The use of real-time data analytics platforms has been explored by researchers to provide continuous monitoring of sales patterns, allowing for quicker decision-making and more dynamic sales management.

2.4 Review of Literature

1. **Brown & Martinez (2020)**

- Title: "Real-Time Sales Data Analysis for E-commerce Optimization"
- Proposed Solution: This paper proposes the implementation of real-time data analytics platforms to continuously monitor and analyze sales data. The solution aims to enable dynamic adjustments to sales strategies and distribution plans.

2. **Choi & Huang (2018)**

- Title: "Time Series Forecasting in Retail: An Overview"
- Proposed Solution: The authors recommend applying time series forecasting models such as ARIMA and its variants to retail sales data. They emphasize the importance of incorporating seasonality and trend components into the models.

3. **Green & Davis (2019)**

- Title: "Advanced Forecasting Techniques for Improving Supply Chain Management in Retail"
- Proposed Solution: This review suggests using advanced forecasting techniques, including SARIMA and hybrid models, to improve supply chain management. The approach focuses on optimizing inventory levels and reducing costs based on accurate sales predictions.

4. **Johnson & Wang (2020)**

- Title: "Data-Driven Approaches to Enhancing Sales Performance: A Review"
- Proposed Solution: The paper advocates for a data-driven approach to sales performance enhancement, recommending the use of comprehensive analytics frameworks that integrate various data sources for a holistic view of sales trends.

5. **Khan & Lee (2019)**

- Title: "Big Data Analytics in E-commerce: A Comprehensive Review"
- Proposed Solution: This review suggests leveraging big data analytics to enhance sales forecasting. The authors propose using large-scale datasets and advanced analytical tools to capture complex sales patterns and improve predictive accuracy.

6. **Kumar & Rao (2019)**

- Title: "Predictive Analytics in Retail: Enhancing Sales Forecasting"
- Proposed Solution: The authors propose the use of predictive analytics, combining traditional statistical methods with machine learning algorithms. They emphasize the importance of integrating customer behaviour data and market dynamics into forecasting models to improve accuracy and responsiveness.

7. **Li & Smith (2020)**

- Title: "The Role of External Factors in Sales Forecasting: A Multi-Model Approach"
- Proposed Solution: The paper recommends using a multi-model approach that integrates traditional forecasting models with external data sources, such as weather conditions, economic indicators, and social media trends. The authors argue that incorporating these external factors can significantly enhance the reliability of sales forecast.

8. **Ng & Cheung (2020)**

- Title: "Sales Forecasting and Planning: Methods and Challenges"
- Proposed Solution: The authors advocate for the use of advanced statistical models and machine learning techniques to improve forecasting accuracy. They recommend integrating historical sales data with external factors like market trends and economic indicators.

9. **Patel & Singh (2021)**

- Title: "Machine Learning Techniques for Sales Forecasting: A Review"
- Proposed Solution: The paper proposes the use of machine learning techniques, such as Random Forest and Gradient Boosting, to enhance sales forecasting. These methods are suggested to capture non-linear relationships and improve prediction accuracy.

10. **Roberts & Moore (2017)**

- Title: "Seasonal and Cyclical Pattern Recognition in Sales Data: Techniques and Applications"
- Proposed Solution: The authors suggest using statistical methods and machine learning algorithms to identify and model seasonal and cyclical patterns in sales data. They highlight the importance of these patterns in accurate trend forecasting.

11. **Thompson & Adams (2018)**

- Title: "Integrating Forecasting Models with Optimization Techniques for Sales Management"
- Proposed Solution: The authors propose integrating sales forecasting models with optimization techniques to improve sales management. They suggest using optimization algorithms to refine distribution strategies and resource allocation based on forecasted sales data.

12. **Wilson & White (2021)**

- Title: "The Impact of Machine Learning on Sales Forecasting in E-commerce"
- Proposed Solution: The authors recommend integrating machine learning models, such as neural networks, with traditional forecasting methods to enhance prediction accuracy and adapt to changing sales patterns in e-commerce.

13. Zhang & Chen (2021)

- Title: "Deep Learning Applications in Sales Forecasting: A Comprehensive Review"
- Proposed Solution: This review highlights the application of deep learning techniques, such as LSTM and RNN, for sales forecasting. The authors suggest that these models can effectively capture temporal dependencies and complex patterns in sales data, leading to more accurate predictions.

2.5 Summary Linking Literature Review with the Project

The literature review underscores the importance of effective data preparation, appropriate forecasting methods, and robust visualization techniques in sales analysis. For this project, working with a limited and unorganized dataset of 100 rows and 14 columns, these insights are particularly relevant. The review supports using linear regression for forecasting in small datasets, as it provides a straightforward approach to modelling trends despite data limitations. Data cleaning and transformation using Python's Pandas are essential to organize the dataset effectively, aligning with the literature's emphasis on data preparation. Visualization tools such as Seaborn, Matplotlib, and Power BI are critical for presenting the analysis results clearly and dynamically. By integrating these methods, the project leverages established best practices to overcome the dataset's constraints and deliver insightful sales trend analysis**.**

CHAPTER 3

METHODOLOGY

3.1 Need and Significance of the Study (GAP)

The analysis of the Amazon sales trend aims to fill gaps in understanding how different factors such as region, product type, and time periods impact sales and profitability. This study is significant as it provides insights into sales trends, helping to optimize distribution methods and improve sales management, ultimately leading to cost reduction and increased profits.

3.2 Objectives of the Study

To analyse monthly and yearly sales trends.

To identify top-selling products by region.

To evaluate profit by region.

To forecast sales trends using a linear regression model.

3.3 Hypothesis of the Study

There is a significant relationship between the region and the profitability of products

- H_o: Region and product profitability are independent (no association).
- H_1: Region and product profitability are dependent (there is an association).

3.4 Scope of the Study

This study conducts an in-depth analysis of Amazon's sales data, focusing on key trends across different time frames and regions. Specifically, it examines sales patterns on a monthly and yearly basis and compares them across various regions. Additionally, the study identifies the top-selling products and evaluates profitability in different regions. The analysis is confined to a dataset

consisting of 100 rows and 14 columns of Amazon sales information, utilizing secondary data to derive insights.

3.5 Research Design

The research employs a comprehensive descriptive and analytical methodology, focusing on quantitative data analysis to thoroughly examine Amazon's sales trends. The study's design integrates data-driven insights through the use of advanced tools like Python and Power BI. Python is leveraged for its robust data manipulation capabilities, utilizing libraries such as Pandas for data structuring and cleaning, Matplotlib and Pyplot for detailed visualizations, Seaborn for enhancing the aesthetic and clarity of graphs, and Sklearn for applying machine learning techniques like linear regression. These tools facilitate an in-depth exploration of the sales data, allowing for precise identification of trends, patterns, and anomalies. Power BI complements this by providing dynamic and interactive visualizations, making it easier to communicate findings effectively. The combined approach ensures that the research is not only rigorous in its analytical process but also practical in its application**,** offering valuable insights into Amazon's sales performance across different regions and timeframes.

3.6 Sampling Unit, Size, and Techniques

Sampling Unit: Amazon sales data

Sample Size: 100 rows

Sampling Technique: Secondary data analysis from an existing dataset.

3.7 Data Collection Method

This project utilized secondary data provided by Unified Mentor Pvt. Ltd., specifically focusing on a dataset related to Amazon sales trends.

3.8 Tools and Techniques Use

- **Tools:** Python (data analysis, modeling), Power BI (visualization).
- **Techniques:** Descriptive analysis, trend analysis, linear regression modeling, data visualization.
- **Python Libraries:**
    - NumPy (import numpy as np)
    - Pandas (import pandas as pd)
    - Matplotlib (import matplotlib.pyplot as plt)
    - Seaborn (import seaborn as sns)
    - Scikit-learn (from sklearn.linear_model import LinearRegression)
    - Warnings (import warnings, warnings.filterwarnings('ignore'))

CHAPTER 4

DATA ANALYSIS AND INTERPRETATION

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise.

4.1.Understanding The Data

Importing the libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import warnings
import seaborn as sns
warnings.filterwarnings('ignore')
```

Reading the file

```
data = pd.read_csv(r'D:\DWNLD\Amazon Sales data.csv ')
# Retrieve the first 5 rows
data.head()
```

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	669165933	6/27/2010	9925	255.28	159.42	2533654.00	1582243.50	951410.50
1	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	963881480	9/15/2012	2804	205.70	117.11	576782.80	328376.44	248406.36
2	Europe	Russia	Office Supplies	Offline	L	05-02-2014	341417157	05-08-2014	1779	651.21	524.96	1158502.59	933903.84	224598.75
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	6/20/2014	514321792	07-05-2014	8102	9.33	6.92	75591.66	56065.84	19525.82
4	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	02-01-2013	115456712	02-06-2013	5062	651.21	524.96	3296425.02	2657347.52	639077.50

4.2 Preprocessing

```
# Checking the shape of data
data.shape
```

(100, 14)

```
# Checking the information about data
data.info()
```

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 100 entries, 0 to 99 Data columns

(total 14 columns):

#	Column	Non-Null Count	Dtype
0	Region	100 non-null	object
1	Country	100 non-null	object
2	Item Type	100 non-null	object
3	Sales Channel	100 non-null	object
4	Order Priority	100 non-null	object
5	Order Date	100 non-null	object
6	Order ID	100 non-null	int64
7	Ship Date	100 non-null	object
8	Units Sold	100 non-null	int64
9	Unit Price	100 non-null	float64
10	Unit Cost	100 non-null	float64
11	Total Revenue	100 non-null	float64
12	Total Cost	100 non-null	float64
13	Total Profit	100 non-null	float64

dtypes: float64(5), int64(2), object(7)

Converting the order date into unified format

```
##### Convertion of date into actual format
# Define the formats to attempt
formats = ['%m/%d/%Y', '%m-%d-%Y']

# Initialize an empty list to store converted dates
converted_orderdates = []

# Iterate through each date string
for orderdate_str in data['Order Date']:
    converted_orderdate = None
    for fmt in formats:
        try:
            converted_orderdate = pd.to_datetime(orderdate_str, format=fmt)
            break  # If successful, break out of the loop
        except ValueError:
            pass  # If ValueError, continue to next format
    if converted_orderdate is not None:
        converted_orderdates.append(converted_orderdate)
    else :
        converted_orderdates.append(pd.NaT)

# Assign the converted dates back to the DataFrame
data['Order Date'] =converted_orderdates
data.head()
```

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	2010-05-28	669165933	6/27/2010	9925	255.28	159.42	2533654.00	1582243.50	951410.50
1	Central America and the Caribbean	Grenada	Cereal	Online	C	2012-08-22	963881480	9/15/2012	2804	205.70	117.11	576782.80	328376.44	248406.36
2	Europe	Russia	Office Supplies	Offline	L	2014-05-02	341417157	05-08-2014	1779	651.21	524.96	1158502.59	933903.84	224598.75
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	2014-06-20	514321792	07-05-2014	8102	9.33	6.92	75591.66	56065.84	19525.82
4	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	2013-02-01	115456712	02-06-2013	5062	651.21	524.96	3296425.02	2657347.52	639077.50

Converting the ship date into unified format

```
# Ship Date
# Define the formats to attempt
formats = ['%m/%d/%Y', '%m-%d-%Y']

# Initialize an empty list to store converted dates
converted_shipdates = []

# Iterate through each date string
for shipdate_str in data['Ship Date']:
    converted_shipdate = None
    for fmt in formats:
        try:
            converted_shipdate = pd.to_datetime(shipdate_str, format=fmt)
            break  # If successful, break out of the loop
        except ValueError:
            pass  # If ValueError, continue to next format
    if converted_shipdate is not None:
        converted_shipdates.append(converted_shipdate)
    else :
        converted_shipdates.append(pd.NaT)

# Assign the converted dates back to the DataFrame
data['Ship Date'] =converted_shipdates
data.head()
```

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	2010-05-28	669165933	2010-06-27	9925	255.28	159.42	2533654.00	1582243.50	951410.50
1	Central America and the Caribbean	Grenada	Cereal	Online	C	2012-08-22	963881480	2012-09-15	2804	205.70	117.11	576782.80	328376.44	248406.36
2	Europe	Russia	Office Supplies	Offline	L	2014-05-02	341417157	2014-05-08	1779	651.21	524.96	1158502.59	933903.84	224598.75
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	2014-06-20	514321792	2014-07-05	8102	9.33	6.92	75591.66	56065.84	19525.82
4	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	2013-02-01	115456712	2013-02-06	5062	651.21	524.96	3296425.02	2657347.52	639077.50

Checking the uniqueness

```
# For a single column
unique_count = data['Order Date'].nunique()
print(f"Number of unique values: {unique_count}")
unique_count = data['Ship Date'].nunique()
print(f"Number of unique values: {unique_count}")

# For all columns
unique_counts = data.nunique()
print(unique_counts)
```

```
Number of unique values: 100
Number of unique values: 99
Region              7
Country            76
Item Type          12
Sales Channel       2
Order Priority      4
Order Date        100
Order ID          100
Ship Date          99
Units Sold         99
Unit Price         12
Unit Cost          12
Total Revenue     100
Total Cost        100
Total Profit      100
dtype: int64
```

Checking the missing value

```
# Checkimg the missing values
missing_values = data.isnull().sum()

print("Missing Values of all columns")
print(missing_values)
```

Country	0
Item Type	0
Sales Channel	0
Order Priority	0
Order Date	0
Order ID	0
Ship Date	0
Units Sold	0
Unit Price	0
Unit Cost	0
Total Revenue	0
Total Cost	0
Total Profit dtype: int64	0

4.3. Extracting 'Data

Year' and 'Month' from 'Order Date'

```
# Extract the year from the 'Order Date' column and create a new 'Year' column
data['Year'] = data['Order Date'].dt.year

# Extract the month from the 'Order Date' column and create a new 'Month' column
data['Month'] = data['Order Date'].dt.month

# Display the first few rows of the dataframe to verify the new columns
data.head()
```

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit	Year	Month
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	2010-05-28	669165933	2010-06-27	9925	255.28	159.42	2533654.00	1582243.50	951410.50	2010	5
1	Central America and the Caribbean	Grenada	Cereal	Online	C	2012-08-22	963881480	2012-09-15	2804	205.70	117.11	576782.80	328376.44	248406.36	2012	8
2	Europe	Russia	Office Supplies	Offline	L	2014-05-02	341417157	2014-05-08	1779	651.21	524.96	1158502.59	933903.84	224598.75	2014	5
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	2014-06-20	514321792	2014-07-05	8102	9.33	6.92	75591.66	56065.84	19525.82	2014	6
4	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	2013-02-01	115456712	2013-02-06	5062	651.21	524.96	3296425.02	2657347.52	639077.50	2013	2

Converting the date into actual format

```
# convertion of date time into actual fromat
data['Order Date'] = pd.to_datetime(data['Order Date'])
data['Ship Date'] = pd.to_datetime(data['Ship Date'])
```

Convert numerical columns to appropriate data types :

```
# Convertion of numerical column into list:
numerical_columns = ['Units Sold', 'Unit Price', 'Unit Cost', 'Total Revenue', 'Total Cost', 'Total Profit']
data[numerical_columns] = data[numerical_columns].apply(pd.to_numeric)
```

Extract and Transform Specific Information

```
# Calculate shipping duration in days
data['Shipping Duration '] = (data['Ship Date'] - data['Order Date']).dt.days

data.head()
```

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit	Year	Month	Shipping Duration
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	2010-05-28	669165933	2010-06-27	9925	255.28	159.42	2533654.00	1582243.50	951410.50	2010	5	30
1	Central America and the Caribbean	Grenada	Cereal	Online	C	2012-08-22	963881480	2012-09-15	2804	205.70	117.11	576782.80	328376.44	248406.36	2012	8	24
2	Europe	Russia	Office Supplies	Offline	L	2014-05-02	341417157	2014-05-08	1779	651.21	524.96	1158502.59	933903.84	224598.75	2014	5	6
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	2014-06-20	514321792	2014-07-05	8102	9.33	6.92	75591.66	56065.84	19525.82	2014	6	15
4	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	2013-02-01	115456712	2013-02-06	5062	651.21	524.96	3296425.02	2657347.52	639077.50	2013	2	5

4.4 Analysis &
4.5 Visualizations

4.4.1 Calculating total sales by region

```
# Finding total sales by groupby function
total_sales_by_region = data.groupby('Region')['Total Revenue'].sum().reset_index()
print("Total Sales by Region:")
print(total_sales_by_region)
```

```
Total Sales by Region:
                            Region  Total Revenue
0                             Asia    21347091.02
1             Australia and Oceania    14094265.13
2  Central America and the Caribbean     9170385.49
3                           Europe    33368932.11
4      Middle East and North Africa    14052706.58
5                     North America     5643356.55
6                Sub-Saharan Africa    39672031.43
```

4.5.1 Visualizing the sales by region

```
# Create the bar plot
plt.figure(figsize=(10, 6))

# Plotting the bar chart
plt.bar(total_sales_by_region['Region'], total_sales_by_region['Total Revenue'], color='#4C72B0', edgecolor='none')

# Adding title and labels
plt.title('Total Sales by Region', fontsize=12, fontweight='bold', pad=10)
plt.xlabel('Region', fontsize=12)
plt.ylabel('Total Revenue ($)', fontsize=12)

# Adding grid for better readability
plt.grid(axis='y', linestyle='--', alpha=0.6)

# Customize the x-ticks for better readability
plt.xticks(rotation=45, ha='right', fontsize=10)
plt.yticks(fontsize=10)

# Improve layout
plt.tight_layout()

# Show the plot
plt.show()
```

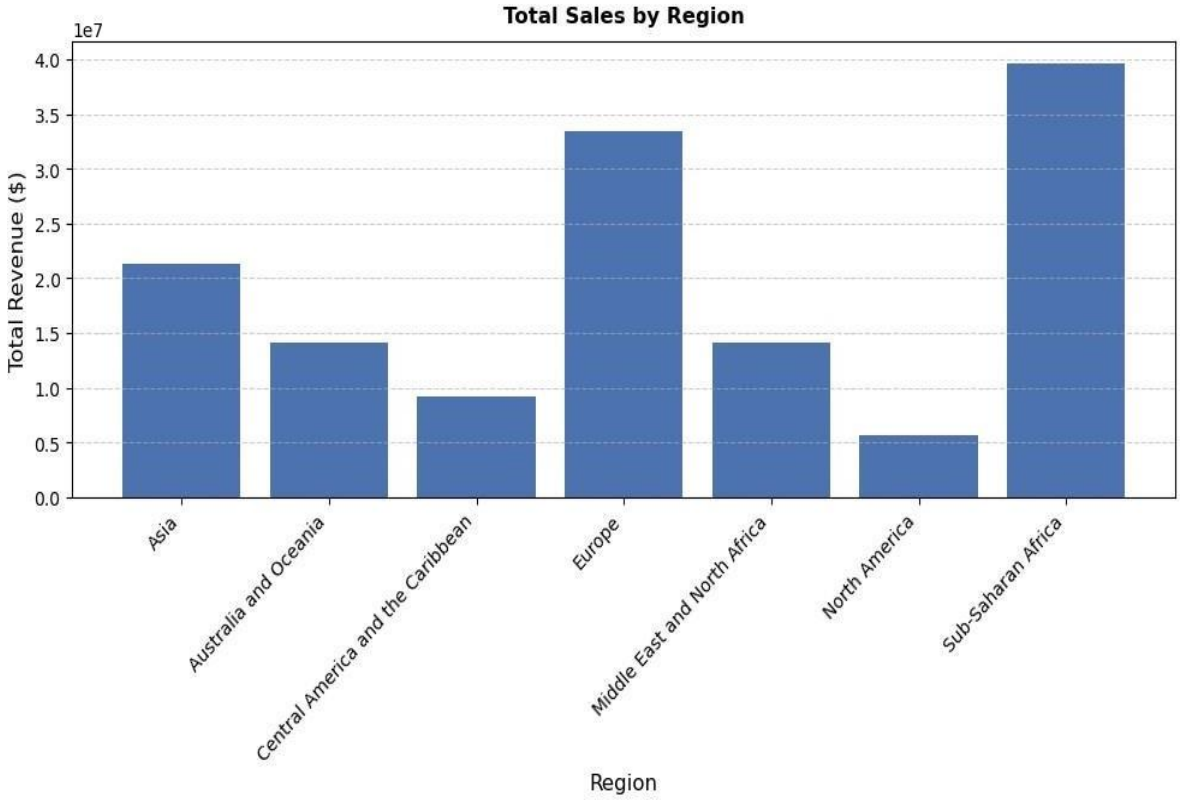

Fig 4.5 1 Total Sales by Region

Inference

1. Europe and Sub-Saharan Africa lead in total revenue, indicating strong market performance and potential for further investment in these regions.
2. Asia shows significant revenue, reflecting a substantial customer base and potential for continued growth.
3. North America has the lowest revenue, suggesting a need for strategic review to address potential market challenges or saturation.

4.4.2 Calculate total profit by region

```
# calculating total profit by group by function:
total_profit_by_region = data.groupby('Region')['Total Profit'].sum().reset_index()
print("Total Profit by Region:")
print(total_profit_by_region)
```

Total Profit by Region:

	Region	Total Profit
0	Asia	6113845.87
1	Australia and Oceania	4722160.03
2	Central America and the Caribbean	2846907.85
3	Europe	11082938.63
4	Middle East and North Africa	5761191.86
5	North America	1457942.76
6	Sub-Saharan Africa	12183211.40

4.5.2 Visualizing Calculate total profit by region

```
plt.figure(figsize=(10, 6))

# Apply the colormap
colors = plt.get_cmap('viridis')(np.linspace(0, 1, len(total_profit_by_region)))

# Create the bar plot with colormap
plt.bar(total_profit_by_region['Region'], total_profit_by_region['Total Profit'], color=colors,)

# Adding title and labels
plt.title('Total Profit by Region', fontsize=12, pad=10)
plt.xlabel('Region', fontsize=12)
plt.ylabel('Total Profit ($)', fontsize=12)

# Adding grid for better readability
plt.grid(axis='y', linestyle='--', alpha=0.6)

# Customize the x-ticks
plt.xticks(rotation=45, ha='right', fontsize=10)
plt.yticks(fontsize=10)

# Improve layout
plt.tight_layout()

plt.show()
```

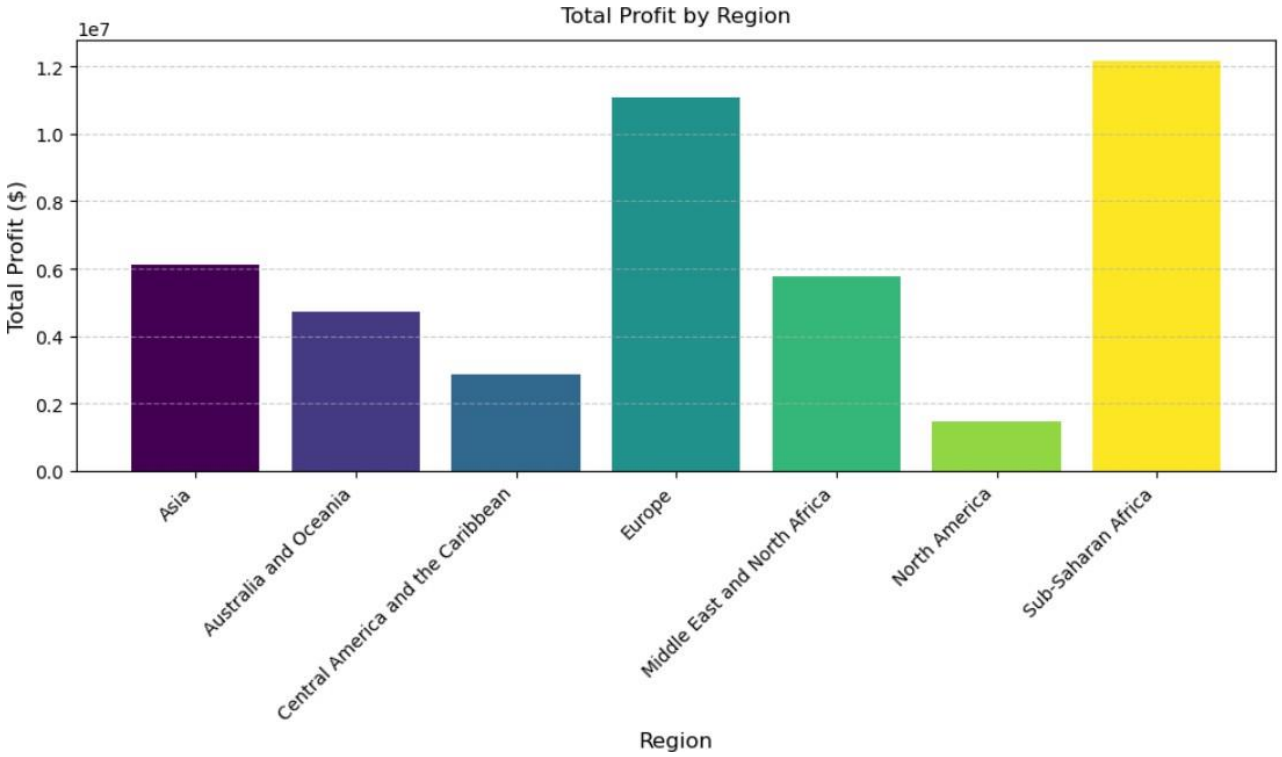

Fig 4.5 2 Total Profit by Region

Inference

1. Europe and Sub-Saharan Africa have the highest total profits, highlighting their strong profitability and effective cost management.
2. Asia and Middle East and North Africa show substantial profits, indicating healthy market conditions and operational efficiency.
3. North America generates the lowest profit, suggesting potential issues with cost control or competitive pressures that may need addressing.

4.4.3 Calculating top-selling products by total units sold

```
# calculating Top Selling ptoduct usinf groupby,sum and sort function:

top_selling_products = data.groupby('Item Type')['Units Sold'].sum().reset_index().sort_values(by='Units Sold', ascending=False)
print("Top-Selling Products:")
print(top_selling_products)
```

Top-Selling Products:

	Item Type	Units Sold
4	Cosmetics	83718
3	Clothes	71260
1	Beverages	56708
5	Fruits	49998
9	Personal Care	48708
8	Office Supplies	46967
6	Household	44727
0	Baby Food	40545
2	Cereal	25877
11	Vegetables	20051
10	Snacks	13637
7	Meat	10675

4.5.3 Visualizing the Top-Selling Product

```
# Retrieving the top selling product
top_selling_products = top_selling_products.sort_values(by='Units Sold', ascending=True)
plt.figure(figsize=(10, 8))

colors = sns.color_palette("mako", len(top_selling_products))
colors = colors[::-1]  # Reverse the color palette

# Create horizontal bar plot
bars = plt.barh(top_selling_products['Item Type'], top_selling_products['Units Sold'], color=colors, edgecolor='black')

# Adding title and labels
plt.title('Top-Selling Products by Total Units Sold', fontsize=12, pad=10)
plt.xlabel('Total Units Sold', fontsize=12)
plt.ylabel('Item Type', fontsize=12)

# Adding grid for better readability and Customize the x-ticks and y-ticks
plt.grid(axis='x', linestyle='--', alpha=0.6)
plt.xticks(fontsize=9)
plt.yticks(fontsize=9)

# Annotate bars with the value
for index, value in enumerate(top_selling_products['Units Sold']):
    plt.text(value,index, f'{value:,.0f}',ha="left", fontsize=8)
plt.tight_layout()
plt.show()
```

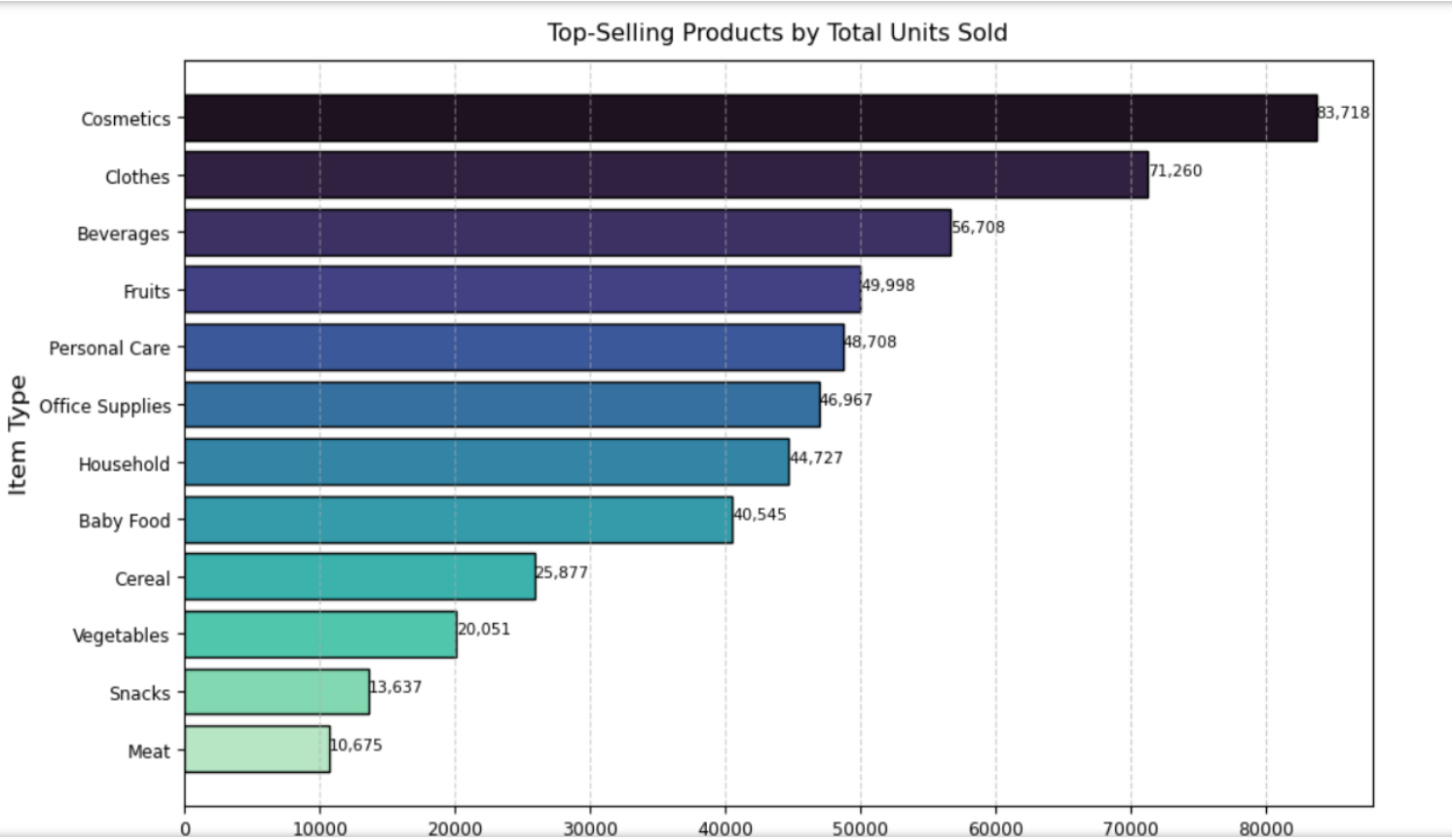

Fig 4.5 3 Top Selling Product

Inference

1. Cosmetics and Clothes lead in sales, indicating high consumer demand and market popularity in these categories.
2. Beverages and Fruits show significant sales, reflecting steady consumption and market presence.
3. Meat and Snacks are the least sold, suggesting lower consumer interest or higher competition in these categories.

4.4.4 computing monthly sales

```
# Computing month-wise-sales
month_wise_sales = data.groupby('Month')['Total Revenue'].sum().reset_index()
print("Month-wise Sales Trend:")
print(month_wise_sales)
```

```
Month-wise Sales Trend:
    Month  Total Revenue
0       1    10482467.12
1       2    24740517.77
2       3     2274823.87
3       4    16187186.33
4       5    13215739.99
5       6     5230325.77
6       7    15669518.50
7       8     1128164.91
8       9     5314762.56
9      10    15287576.61
10     11    20568222.76
11     12     7249462.12
```

4.5.4 Visualizing the Month-wise Sales Trend

```
plt.figure(figsize=(10, 6))

# Use a single color from the palette or define your color directly
color = sns.color_palette("mako", len(month_wise_sales))[1]

# Create the line plot
plt.plot(month_wise_sales['Month'], month_wise_sales['Total Revenue'],
 ↪marker='o', color=color)

# Adding title and labels
plt.title('Month-wise Sales Trend', fontsize=12,
plt.xlabel('Month', fontsize=12)
plt.ylabel('Total Revenue', fontsize=12)

# Adding grid for better readability
plt.grid(True)

# Customize the x-ticks (assuming your months are in a 1-12 range)
plt.xticks(range(1, 13), fontsize=8)

plt.tight_layout()
plt.show()
```

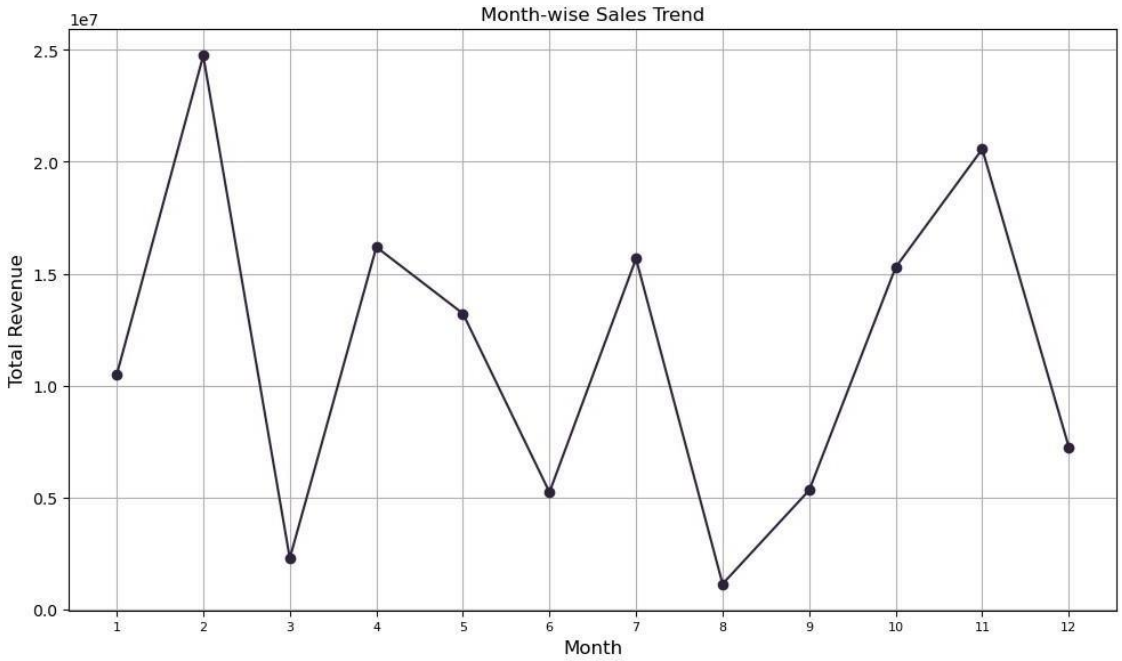

Fig 4.5 4 Month-Wise Sales

Inference

1. February and March experience the highest sales revenue, indicating strong performance during these months, possibly due to seasonal factors or promotions.
2. August shows the lowest revenue, which might reflect a seasonal drop or reduced consumer spending during the summer.
3. November has a significant increase in sales, potentially linked to end-of-year promotions or holiday shopping.

4.4.5 Computing year wise sales

```
# Computing the Year-wise sales
year_wise_sales = data.groupby('Year')['Total Revenue'].sum().reset_index()
print("Year-wise Sales Trend:")
print(year_wise_sales)
```

Year-wise Sales Trend:

	Year	Total Revenue
0	2010	19186024.92
1	2011	11129166.07
2	2012	31898644.52
3	2013	20330448.66
4	2014	16630214.43
5	2015	12427982.86
6	2016	12372867.22
7	2017	13373419.63

4.5.5 Visualizing the year-wise Sales Trend

```
plt.figure(figsize=(10, 6))

# Plot the 'Year' on the x-axis and 'Total Revenue' on the y-axis
plt.plot(year_wise_sales['Year'], year_wise_sales['Total Revenue'], marker='o')

# Add a title to the plot for clarity
plt.title('Year-wise Sales Trend')

# labeling X & Y axis
plt.xlabel('Year')
plt.ylabel('Total Revenue')

# Add a grid to the plot for better readability of the data points
plt.grid(True)

# Set custom x-tick labels using the 'Year' values, ensuring they match the data
plt.xticks(year_wise_sales['Year'])

# Display the plot
plt.show()
```

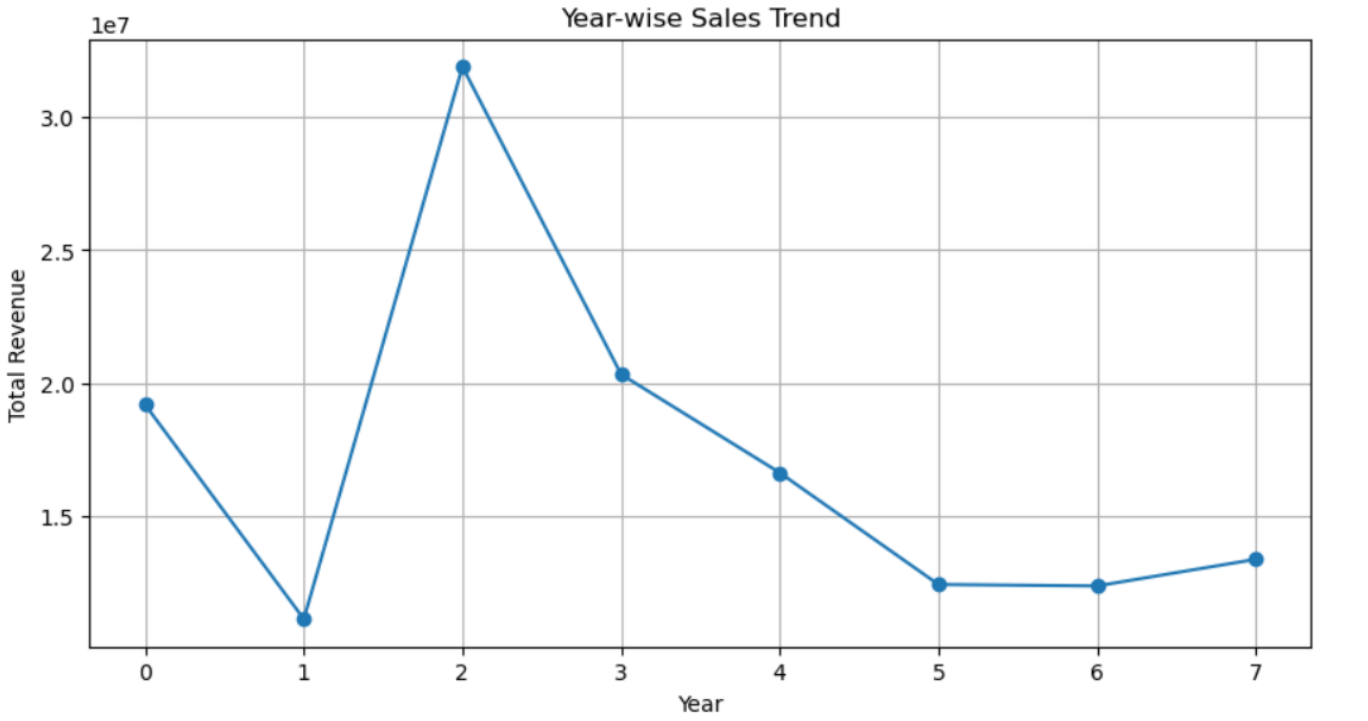

Fig 4.5 5 Year Wise Sales

Inference

1. 2012 shows the highest total revenue, indicating a peak year with strong sales performance.
2. 2011 and 2016 experienced lower sales compared to surrounding years, which might suggest potential market challenges or reduced demand during these periods.
3. 2017 reflects a moderate increase in revenue compared to previous years, suggesting a recovery or improvement in sales trends.

4.4.6 Computing Yearly Month-wise Sales Trend

```
# Computing the Yearly Month-wise Sales Trend
yearly_month_wise_sales = data.groupby(['Year', 'Month'])['Total Revenue'].sum().reset_index()
print("Yearly Month-wise Sales Trend:")
print(yearly_month_wise_sales)
```

Yearly Month-wise Sales Trend:

	Year	Month	Total Revenue
0	2010	2	3410661.12
1	2010	5	2587973.26
2	2010	6	1082418.40
3	2010	10	6064933.75
4	2010	11	3458252.00
5	2010	12	2581786.39
6	2011	1	1042225.35
7	2011	2	387002.20
8	2011	4	2798046.49
9	2011	5	272410.45
10	2011	6	19103.44
12	2011	9	574951.92
13	2011	11	5938385.58
14	2012	1	1012884.00
15	2012	2	6707849.42
16	2012	3	994765.42
17	2012	4	4556012.38
18	2012	5	3782781.82
19	2012	6	2132075.27
20	2012	7	4445093.92
21	2012	8	576782.80
22	2012	9	4648152.72
23	2012	10	3042246.77
24	2013	2	3296425.02
25	2013	3	835759.10
26	2013	4	3262562.10
27	2013	6	1352867.40
28	2013	7	8545511.20
29	2013	8	89623.98
30	2013	9	71253.21
31	2013	10	2702770.40
32	2013	12	173676.25
33	2014	2	1819660.25
34	2014	4	4510578.10
35	2014	5	3060338.59

36	2014	6	75591.66
37	2014	7	688641.85
38	2014	8	455479.04
39	2014	9	20404.71
40	2014	10	1352370.65
41	2014	11	4647149.58
42	2015	1	5513227.50
43	2015	2	2003911.12
44	2015	4	1059987.26
45	2015	7	1292409.45
46	2015	8	6279.09
47	2015	10	1904138.04
48	2015	11	648030.40
49	2016	3	197883.40
50	2016	5	414371.10
51	2016	6	568269.60
52	2016	7	600821.44
53	2016	10	221117.00
54	2016	11	5876405.20
55	2016	12	4493999.48
56	2017	1	2914130.27
57	2017	2	7115008.64
58	2017	3	246415.95
59	2017	5	3097864.77

4.5.6 Visualizing Yearly Month-wise Sales

```
plt.figure(figsize=(12, 8))

# Iterate through each unique year in the 'Year' column
for year in yearly_month_wise_sales['Year'].unique():
    subset = yearly_month_wise_sales[yearly_month_wise_sales[Year'] == year]
    plt.plot(subset['Month'], subset['Total Revenue'], marker='o',
 ↪label=str(year))

# Set the title
plt.title('Yearly Month-wise Sales Trend')

# Label the x-axis and y-axis
plt.xlabel('Month')
plt.ylabel('Total Revenue')

# Add a legend
plt.legend(title='Year')

# Enable the grid on the plot for better readability of the data points
plt.grid(True)
plt.xticks(range(1, 13))

# Display the plot with all the added features and settings
plt.show()
```

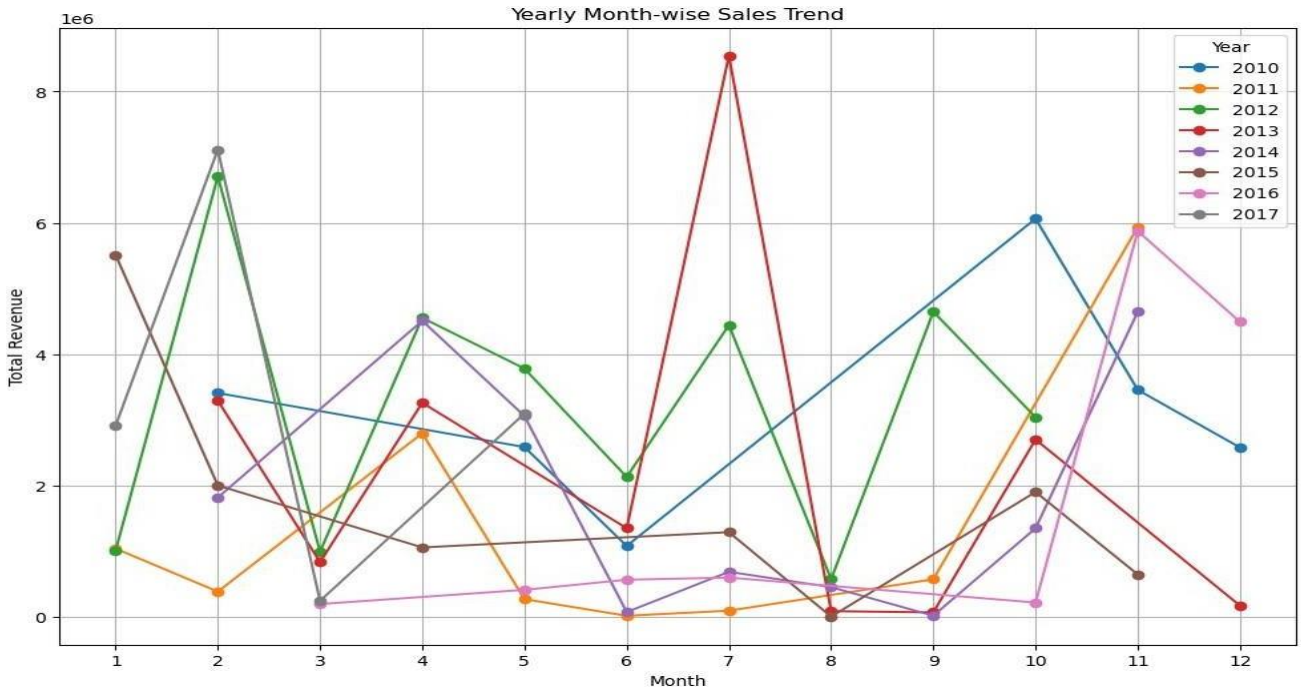

Fig 4.5 6 Yearly Month_Wise Sales

Inference

1. Sales Peaks in Certain Months:

   The highest sales revenue occurs in October 2010 and July 2013, indicating seasonal trends or promotions that drive higher sales in these months.

2. Revenue Fluctuations Over Years:

   Sales revenue shows variability from year to year, with notable increases in 2012 and 2013, but a decline in subsequent years, suggesting potential shifts in market conditions or business strategies.

3. Seasonal Trends:

   Certain months consistently exhibit higher revenue, such as February and December, reflecting possible seasonal purchasing patterns or holiday-related sales spikes.

Hypothesis

```
# Prepare data for ANOVA
regions = df['Region'].unique()
profit_by_region = [df[df['Region'] == region]['Total Profit'].dropna() for region in regions]

# Conduct ANOVA test
f_value, p_value = stats.f_oneway(*profit_by_region)

# Print results
print(f"F-value: {f_value}")
print(f"P-value: {p_value}")
```

```
F-value: 0.6876610763492621
P-value: 0.6600411699768258
```

Inference for Hypothesis H1:

Hypothesis H1: There is a significant relationship between the region and the profitability of products.

Result:

F-value: 0.6877

P-value: 0.6600

Since the p-value is greater than 0.05: We do not reject the null hypothesis and conclude that there is no significant relationship between the region and the profitability of products. This suggests that profitability does not vary significantly across different regions.

4.6 Modelling

```
year_wise_sales['Year'] = np.arange(len(year_wise_sales))

# Define the feature and target
X = year_wise_sales[['Year']].values
y = year_wise_sales['Total Revenue'].values

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create and fit the model
model = LinearRegression()
model.fit(X_train, y_train)

# Predict on the testing set
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R^2 Score: {r2}")

# Forecast for future years
future_years = np.arange(len(year_wise_sales), len(year_wise_sales) + 5).reshape(-1, 1)
forecast = model.predict(future_years)

# Round forecast values to the nearest integer
forecast = np.round(forecast).astype(int)

# Prepare forecast DataFrame
forecast_df = pd.DataFrame({
    'Year': np.arange(year_wise_sales['Year'].max() + 1, year_wise_sales['Year'].max() + 6),
    'Forecast': forecast
})

# Print the forecast DataFrame
print("Linear Regression Forecast:")
print(forecast_df)
```

```
Mean Squared Error: 86781818115989.1
R^2 Score: -204.7751597462832
Linear Regression Forecast:
   Year  Forecast
0     8  11415480
1     9   9673221
2    10   7930962
3    11   6188703
4    12   4446444
```

4.6.1 Model Visualization

```
plt.figure(figsize=(12, 6))

# Plot the historical data
plt.scatter(year_wise_sales['Year'], y, color='blue', label='Actual Data')

# Plot the regression line
plt.plot(X, model.predict(X), color='red', label='Regression Line')

# Plot the testing data
plt.scatter(X_test, y_test, color='green', marker='x', label='Test Data')

# Plot the forecast
plt.plot(forecast_df['Year'],
         forecast_df['Forecast'], color='orange', linestyle='--', marker='o', label='Forecast')

plt.xlabel('Year')
plt.ylabel('Total Revenue')
plt.title('Sales Forecasting with Linear Regression')
plt.legend()
plt.grid(True)
plt.show()
```

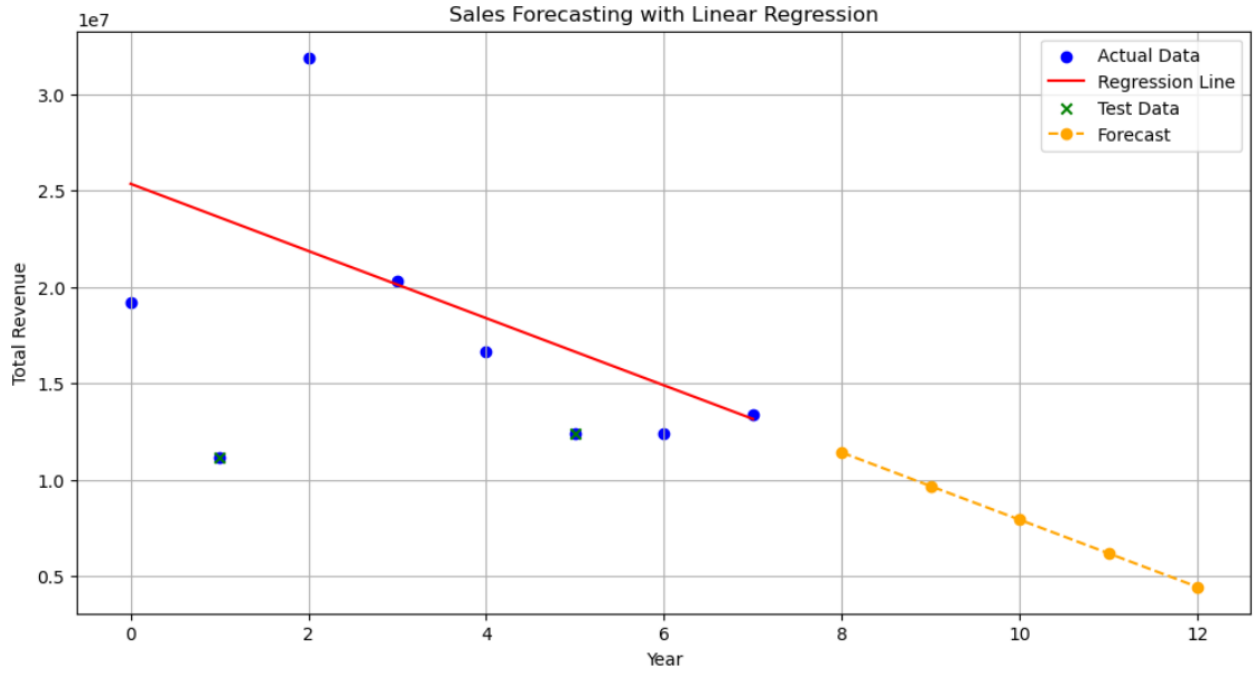

Fig 4.6 1 Model Visualization

Inference

1. Model Performance: The high Mean Squared Error (MSE) and negative R^2 score indicate that the linear regression model is not performing well. This suggests that the model may not be capturing the underlying trends in the data effectively.
2. Forecast Values: The forecasted values for the future years show a decreasing trend, which may reflect the model's inability to accurately capture the dynamics of the revenue changes over time.
3. Potential Issues: The negative R^2 score implies that the model's predictions are worse than a simple mean-based model, suggesting that the linear regression might not be the best fit for this data.
4. Next Steps: Consider exploring alternative forecasting methods, such as polynomial regression, time series analysis, or more advanced techniques, to improve accuracy and better capture the revenue trends

4.7 Dynamic Visualisation using Power Bi

4.7.1 Amazon Sales dashboard

Fig 4.7 1 Amazon Sales Dashboard

4.7.2 Dynamic Visualization based on yearly wise (2010)

Fig 4.7 2 Dynamic Visualization on Year

4.7.3 Dynamic Visualization s based on order priority (High)

Fig 4.7 3 Dynamic visualization on Order Priority

4.7.4 Dynamic Visualization based on item type (Cosmetics)

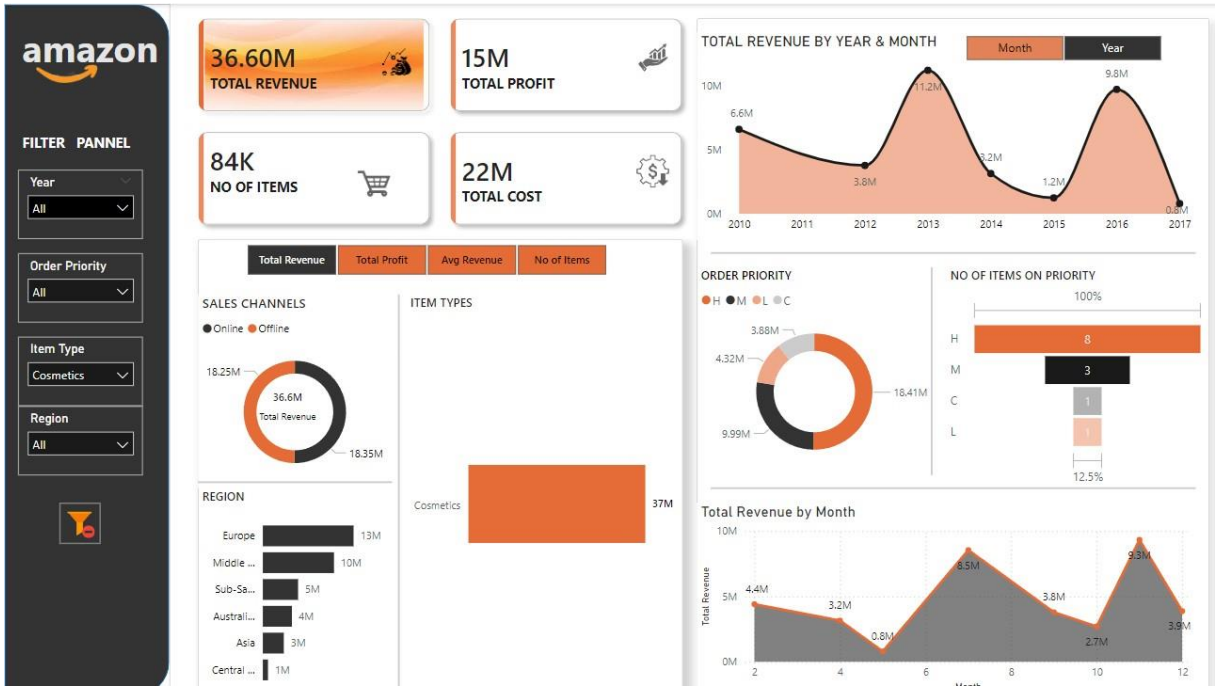

Fig 4.7 4 Dynamic Visualization on Item Type

4.7.5 Dynamic Visualization based on Region(Asia)

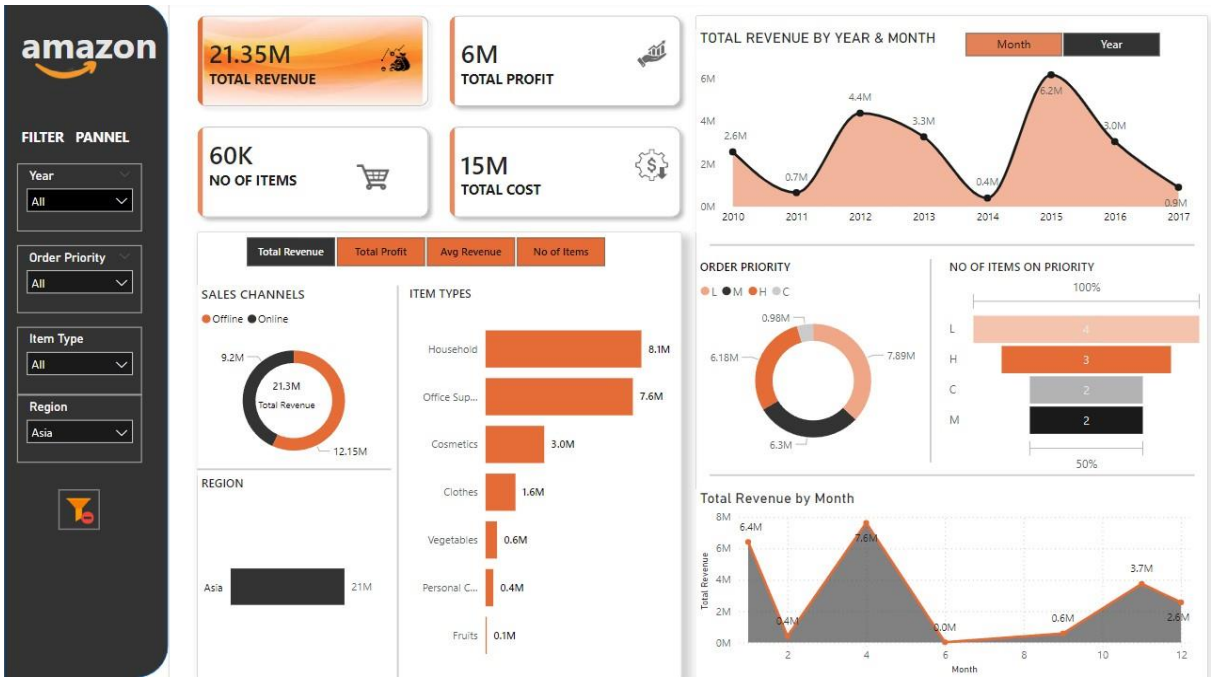

Fig 4.7 5 Dynamic Visualization on Region

CHAPTER 5

FINDINGS AND SUGGESTIONS OF THE STUDY

5.1 Findings

1. **Regional Sales Performance:**

   - **Highest Revenue Regions:** Europe, Asia, Australia, and Oceania appear to have the highest total revenue, indicating strong sales performance in these regions.

   - **Emerging Markets:** Sub-Saharan Africa shows significant sales, with a variety of item types sold, suggesting a growing market

2. **Item Type Trends:**

   - **Top Items by Revenue:** Items like Baby Food, Office Supplies, and Cosmetics have high total revenue, with Baby Food and Office Supplies being consistently popular across various regions.
   - **High Volume Sales:** Fruits and Vegetables, despite having lower unit prices, have high units sold, indicating high volume but lower profit margins

3. **Sales Channel Effectiveness:**

   - **Offline vs. Online Sales:** Online sales channels generate higher total revenue compared to offline channels. This suggests a growing trend towards e-commerce and its potential for higher profitability.

4. **Order Priority Impact:**
   - **High Priority Orders:** Orders with high priority (‘H’) generally show higher total revenue and profit, indicating that urgent orders are more valuable and potentially more profitable.

5. **Seasonal Trends:**

   - **Month-wise Trends:** Certain months, such as November and February, show higher sales volumes and revenues. For instance, high sales in December could be attributed to holiday season shopping.

6. **Shipping Duration:**

- **Impact on Sales:** Shipping durations vary significantly, from 1 day to over 45 days. Shorter shipping durations are generally associated with higher total revenue and profit, emphasizing the importance of efficient logistics.

7. **Country-Specific Insights:**

- **High Revenue Countries:** Countries like Australia, the UK, and the US (not listed but implied from regional data) are major contributors to the revenue, while smaller countries in Sub-Saharan Africa contribute less but show potential for growth.

8. **Yearly Insight:**

- **Total Revenue Trends:** Total revenue peaked in 2012 at $31.9 million, then declined significantly, with a slight recovery by 2017. This suggests 2012 was a high-performing year, followed by challenges or market shifts impacting revenue.

9. **Forecast Analysis:**

- Linear regression forecasts a steady decline in revenue over the next few years, highlighting a potential issue with future sales performance and necessitating a review of current strategies.

5.2 Suggestions

1. **Data Expansion:**

a. **Collect More Data:** To enhance modelling accuracy, gather additional data such as customer demographics, product reviews, and competitive analysis. This can provide deeper insights into consumer behaviour and market trends.

b. **Include External Factors**: Integrate data on external factors such as economic indicators, market trends, and industry reports to enrich the analysis and improve forecasting models.

2. **Enhance Forecasting Models**

a. **Incorporate Advanced Techniques:** Utilize advanced forecasting methods such as ARIMA or Prophet to improve the accuracy of sales predictions and account for seasonality and trends more effectively.

3. **Optimize Inventory Management**

a. **Adjust Inventory Levels:** Based on the sales trends, adjust inventory levels to ensure optimal stock availability, particularly for high-demand periods and regions.

4. **Improve Distribution and Logistics**

a. **Reduce Shipping Duration:** Implement strategies to shorten shipping times, as prolonged durations negatively impact sales. Consider optimizing logistics and exploring partnerships with faster delivery services.

5. **Focus on High-Performing Regions**

a. **Targeted Marketing**: Increase marketing efforts and investment in high-performing regions such as North America and Asia to capitalize on existing demand and further boost sales.

b. **Regional Strategies**: Develop tailored strategies for regions with lower performance to address specific challenges and enhance market presence.

6. **Leverage E-Commerce Growth**

a. **Expand Online Presence**: Invest in enhancing the online sales platform and marketing strategies to leverage the growing trend of online shopping and maximize sales opportunities.

b. **Omnichannel Strategy:** Integrate online and offline sales strategies to create a seamless customer experience and capture a broader market segment.

7. **Profit Margin Optimization**

a. **Review Pricing Strategies:** Analyse profit margins across different product categories and adjust pricing strategies accordingly to improve profitability.

b. **Cost Management:** Evaluate and optimize cost structures to enhance profit margins, especially for categories with lower profitability.

8. **Customer Segmentation and Personalization**

a. **Segment Customers**: Perform customer segmentation to better understand purchasing behaviour and tailor marketing and sales strategies to different customer groups.

b. **Personalized Offers:** Implement personalized promotions and offers based on customer preferences and buying history to increase engagement and sales.

These findings and suggestions aim to provide a comprehensive view of the sales performance and offer actionable strategies to improve sales, profitability, and operational efficiency.

6. CONCLUSION

This study on Amazon sales trend analysis has provided crucial insights into the dynamics of sales and profitability across different regions, product categories, and time periods. By identifying monthly and yearly sales trends, top-selling products by region, and profitability differences, the analysis has highlighted the importance of tailored marketing and distribution strategies. The effective use of a linear regression model for sales forecasting further emphasizes the potential for data-driven decision-making to optimize inventory management, reduce costs, and boost profits. Despite the study's limitations due to the dataset's size, the findings offer a strong foundation for enhancing Amazon's sales management strategies, with futureresearch opportunities to deepen the analysis by incorporating larger datasets and exploring additional variables. Overall, the study underscores the value of leveraging advanced analytics to maintain a competitive edge in the fast-paced e-commerce landscape.