5-2024

DATA-DRIVEN FORECASTING FOR EFFECTIVE DEMAND MANAGEMENT IN THE E-COMMERCE ECOSYSTEM

Chirag Pandey

DATA-DRIVEN FORECASTING FOR

EFFECTIVE DEMAND MANAGEMENT IN THE E-COMMERCE ECOSYSTEM

A Project

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Information Systems and Technology:

Business Intelligence

---

by

Chirag Pandey

May 2024

DATA-DRIVEN FORECASTING FOR

EFFECTIVE DEMAND MANAGEMENT IN THE E-COMMERCE ECOSYSTEM

A Project

Presented to the

Faculty of

California State University,

San Bernardino

by

Chirag Pandey

May 2024

Approved by:

Dr. Lewis Njualem, Committee Member, Chair

Dr. Conrad Shayo, Committee Member, Reader & Chair, Department of Information and Decision Sciences

ABSTRACT

This culmination project investigated and analyzed the impact of factors influencing sales forecasting models in the e-commerce ecosystem. The research questions are: Q1) To what extent are sales forecasting models in the e-commerce ecosystems influenced by customer demographic variables such as gender, age, and geographic locations? Q2) To what extent are sales forecasting models in the e-commerce ecosystems influenced by product specific factors? The datasets used were from Kaggle, and the Worldometer websites. The findings are: Q1) Individuals aged '55 or over' significantly impact total sales in both the USA and Brazil. Male consumers consistently accounted for a higher proportion of total sales compared to female consumers and were primary contributors across all age brackets in both countries. The concentration of top-performing states and customers is in regions of China and the USA. Q2) The top 10 product categories, including "Outwear and Coats", "Jeans", and "Sweaters", significantly contribute to revenue generation, accounting for 68.21% of total sales. The top 20 recognizable brands such as "Diesel", "Calvin Klein", and "The North Face" collectively contribute 22.04% of total sales. The XGBoost model outperformed the linear regression model with an R-square value of 0.98. The conclusions are: Q1) Customer demographic factors significantly impact e-commerce sales forecasting models. Analyzing sales across age, gender, and geography unveils crucial consumer behavior trends. This underscores the necessity of targeting specific demographic segments and tailoring marketing

strategies accordingly. Q2) Customer demographics significantly impact e-commerce sales forecasting. Analyzing sales across age, gender, and location reveals valuable consumer behavior patterns, necessitating targeted marketing and inventory strategies. Further studies include exploration of: (a) customer behavior dynamics in response to segmentation variables, and cultural influences on preferences. (b) Investigation of cyclical trends in sales forecasting models, and (c) utilizing advanced machine learning techniques, like deep learning algorithms to improve predictive accuracy.

ACKNOWLEDGEMENTS

It is with immense gratitude and sincerity that I extend my heartfelt appreciation to Dr. Lewis Njualem and Dr. Conrad Shayo for their invaluable support throughout the completion of this project. Their unwavering guidance, expertise, and profound knowledge in the field of supply chain management have been instrumental in shaping the successful completion of this project. I am deeply grateful for the countless hours they dedicated to providing guidance, reviewing drafts, and offering invaluable insights that significantly enhanced the project's outcomes. Their mentorship has not only contributed to the academic rigor of this work but has also left an indelible mark on my personal and professional growth.

DEDICATION

I dedicate this project to my beloved family and friends, whose constant support, understanding, and patience have been the cornerstone of my journey. Their boundless encouragement and belief in my abilities have served as a guiding light, inspiring me to navigate through obstacles with resilience and to savor every triumph with gratitude. In their steadfast presence, I have found solace, inspiration, and the courage to pursue my dreams relentlessly.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Brief Background

E-commerce refers to the electronic buying and selling of goods and services over the internet, revolutionizing global trade by offering consumers convenience and businesses cost-effective avenues for commerce (Jain et al., 2021). The major different types of e-commerce are Business-to-business (B2B), Business to-consumer (B2C), Business-to-government (B2G) and Consumer-to-consumer (C2C) (Gupta, 2014). Over time, e-commerce has evolved into a pivotal force within the economy (Taher, 2021). Its capacity to enable businesses to reach a broader spectrum of consumers, coupled with its convenience for customers to shop at their leisure, has played a crucial role in driving its expansion and importance (Taher, 2021). Even during periods of global pandemics and widespread lockdowns, the e-commerce ecosystem has experienced rapid growth. It has expanded exponentially, demonstrating remarkable evolution and adaptation (Singh, 2021).

In the rapidly evolving landscape of e-commerce, driven by advancements in technology and shifts in consumer behavior, the industry has experienced unprecedented growth (ChatGPT 3.5, September 2022). As the e-commerce ecosystem continues to evolve, the management of demand presents complex challenges (ChatGPT 3.5, September 2022). With the global retail sector valued at approximately 25 trillion dollars, ensuring the seamless availability of products in the right place and at the right time is paramount (Wellens et al., 2024). Demand management entails the orchestration of the supply chain to align customer demands with the capabilities of the supply chain (Croxton et al., 2002). Effectively managing the movement of products from suppliers to customers poses a substantial challenge in optimizing the supply chain network, underscoring the critical need for efficient demand management strategies to meet evolving consumer demands while maintaining operational efficiency and competitiveness (Wellens et al., 2024).

In today's dynamic business landscape, characterized by heightened consumer expectations and evolving market dynamics, the imperative for highly efficient delivery systems and robust supply chain management has become paramount, especially for large corporations (Aro-Gordon & Gupte, 2016). Navigating the challenges in e-commerce demand management requires addressing various complexities, including unpredictable consumer behavior, global market dynamics, omnichannel integration, supply chain disruptions, and rapid technological advancements (ChatGPT 3.5, September 2022). Other major challenges revealed in a study includes demand fluctuations, reverse logistics, stock outs, managing SKU’s, keeping count of the inventory, multi-channel shoppers, bottleneck and weak points, bullwhip effect and distressed stocks

(Patil & Divekar, 2014). These challenges underscore the importance of accurate forecasting, agile decision-making, and innovative approaches to optimize operations in the dynamic e-commerce landscape. Effectively managing demand requires a proactive approach that integrates advanced analytics, strategic planning, and responsive strategies to meet evolving customer needs and market demands (ChatGPT 3.5, September 2022).

Accurate predictions of demand serve as the cornerstone for various planning endeavors. Predictive analytics, encompassing demand and sales forecasting, represent pivotal subdomains impacting all enterprises and are instrumental in mitigating inefficiencies within the supply chain (Kumar et al., 2020). Sales forecasting holds significant importance within the business management process, as it plays a crucial role in resource allocation, marketing, and financial decision-making. Inaccurate sales forecasting (SF) results in either excess inventory or stock shortages, contributing to heightened inventory expenses and subsequently diminishing profits and return on investment (Gustriansyah et al., 2022). This encourages the utilization of cutting-edge technologies to enhance sales forecasting (Fisher & Raman, 2018). At the heart of this paradigm shift lies the transformative power of data-driven forecasting (ChatGPT 3.5, September 2022). The incorporation of cutting-edge machine learning algorithms, holds the potential to redefine the precision and efficacy of demand projections, paving the way for enhanced operational performance and strategic decision-making (ChatGPT 3.5, September 2022)

Study conducted by (Weller & Crone, 2012) examined the practices of 200 demand planning experts. Their research revealed that a significant majority of respondents, totaling 82.1%, relied on statistical forecasting methods such as exponential smoothing, moving average, and the naive method. In contrast, only a minority, comprising 13.5% of the participants, utilized advanced time series models like econometric models (6.9%), autoregressive integrated moving average (ARIMA; 3.5%), and neural networks (NNs; 1.5%). However, recent research and forecasting competitions strongly advocate for the use of machine learning (ML) techniques (Huber & Stuckenschmidt, 2020). Since 2015, machine learning has been the predominant approach in all significant large-scale retail forecasting competitions hosted on Kaggle (Bojer & Meldgaard, 2021). The integration of machine learning algorithms into demand forecasting models is steadily growing, highlighting their ability to adapt to dynamic market conditions and provide a more accurate and responsive approach to predicting consumer behavior (ChatGPT 3.5, September 2022). This project will explore the implications of integrating advanced data analytics in demand management within the e-commerce ecosystem.

Problem Statement

According to (Biswas et al., 2023), the development of a dependable and precise sales forecasting system necessitates the consideration of various variables encompassing gender, geographic location, occupation, and pricing discounts. Additionally, there are numerous satisfied buyers who may not provide reviews for the products. This project aims to investigate the influence of various factors on sales forecasting models within the e-commerce ecosystem, including customer demographic factors and product-specific factors. And to further explore the integration of advanced analytics and machine learning for more accurate forecasting, optimizing operations in the e-commerce ecosystem.

Research Questions

1) To what extent are sales forecasting models in the e-commerce ecosystems influenced by customer demographic variables such as gender, age, and geographic locations? (Biswas et al., 2023)

2) To what extent are sales forecasting models in the e-commerce ecosystems influenced by product specific factors such as category and brand? (Biswas et al., 2023)

Objective

The overarching objective is to investigate innovative data-driven approaches, including advanced analytics and machine learning techniques, for demand forecasting within the e-commerce ecosystem.

Organization of this Project

This culminating project consists of five chapters structured as follows: Chapter 1 introduces the project, including the problem statement, research questions, and objectives. Chapter 2 is dedicated to conducting a literature review, scrutinizing existing research relevant to the project's subject matter. This section aims to furnish a comprehensive understanding of the current knowledge of where the research questions came from. Chapter 3 will cover the data collection and research methods used to answer the questions. Chapter 4 will cover data cleaning, analysis and present the research findings, incorporating visual aids such as tables and figures to facilitate understanding. Lastly, Chapter 5 will discuss the research findings, provide conclusions, and propose avenues for further exploration.

CHAPTER TWO

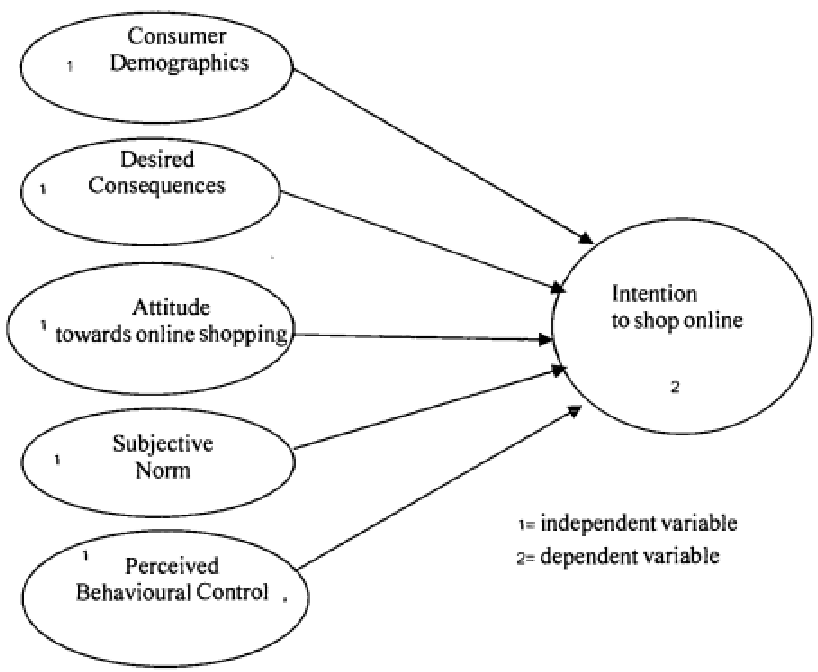
LITERATURE REVIEW

Chapter two provides a summary of where the three research questions came from based on the suggested areas for further studies.

Question 1: To what extent are sales forecasting models in the e-commerce ecosystems influenced by customer demographic variables such as gender, age and geographic locations? (Biswas et al., 2023)

Sales forecasting remains a prominent subject in current literature, yet recent studies have not identified a definitive predictive model with a guaranteed success rate (Biswas et al., 2023). Major techniques and models include time series analysis methods like ARIMA and exponential smoothing, machine learning algorithms such as linear regression, decision trees, and neural networks, market basket analysis techniques like association rule mining, and customer segmentation models such as RFM analysis and cluster analysis (ChatGPT 3.5, September 2022). Many factors impact sales, and the precise way they influence sales outcomes cannot be anticipated through a singular model (Liu et al., 2013). Despite the complexity of sales prediction, customer demographics, spanning variables like age, gender, income, household size, education, occupation, ethnicity, and geographic location, emerge as pivotal factors influencing online purchase decisions (Sari et al., 2016). For example, what are the characteristics of online shoppers? For instance, the characteristics, gender preferences, age distribution, and geographic locations of online shoppers contribute significantly to sales trends (ChatGPT 3.5, September 2022).

Figure 1: The Conceptual Model: Factors Affecting Online Shopping Behaviour (Thananuraksakul, 2007).

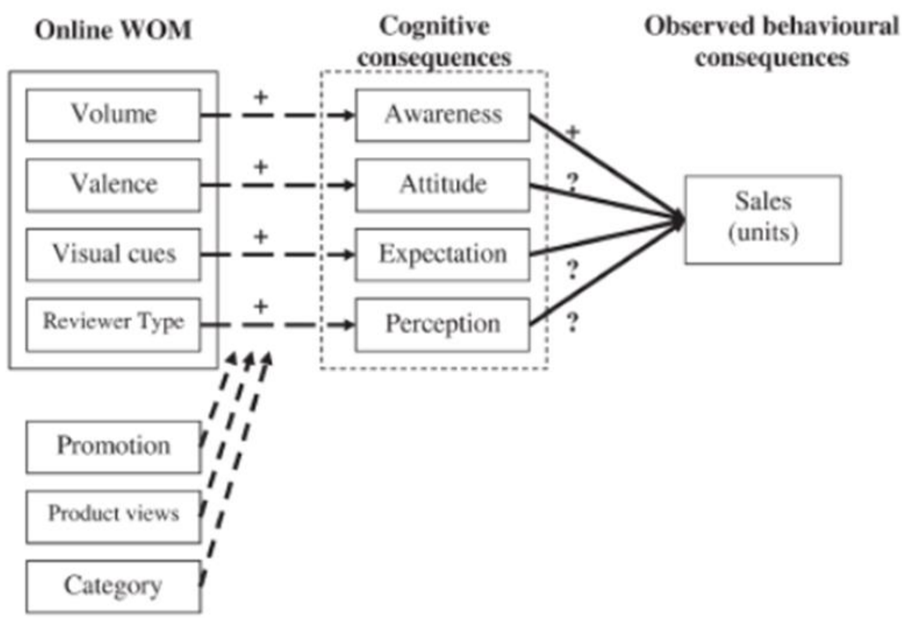


Drawing from the Theory of Planned Behavior, the conceptual model in Figure 1 integrates two additional constructs. First, consumer demographics encompassing gender, income, education, age, occupation, and online buying experience. Second, desired consequences involving convenience, timesaving, trust, and pricing (Thananuraksakul, 2007). For instance, research suggests that men exhibit a higher propensity for online purchasing and are more inclined to make repeat orders online (Buckinx & Poel, 2003). Conversely, female consumers show a greater tendency to shop online for automobiles, often seeking to avoid the pressure exerted by sales representatives at traditional car dealerships (Internet Business News,2006). Additionally, women are noted to engage in more online apparel purchases compared to men (Goldsmith & Flynn, 2004). According to Pew Research, online shopping preference doesn't follow a linear trend with age. Instead, interest is lowest among the youngest and oldest age groups, with 38% of online teens shopping compared to 56% of users aged 64–72 and 47% of those aged 73 and older 32 (Wan et al., 2012). However, interest is highest among the medium age range, with 80% for ages 33–44 and 71% for ages 18–32 (Wan et al., 2012). In geographic segmentation, customers are targeted based on their location, ensuring that product popularity is equally localized. This approach involves targeting specific customers in a particular area with items that are popular and in demand locally (Jha, n.d.). Therefore, demographic customer segmentation is crucial for sales forecasting as it enables businesses to make more precise predictions about future sales trends(The Role of Customer Segmentation in Sales Forecasting, n.d.). By analyzing the behaviors and purchasing patterns of individual customer segments, businesses can determine which segments are most inclined to make purchases in the future and can adjust their sales forecasts accordingly (The Role of Customer Segmentation in Sales Forecasting, n.d.). One such method, known as the Demographic Recommendation Technique, suggests items based on the demographic similarities among users. This strategy operates on the premise that individuals with similar demographic attributes will rate items similarly (Usmani et al., 2017).

Question 2: To what extent are sales forecasting models in the e-commerce ecosystems influenced by product specific factors such as category and brand? (Biswas et al., 2023)

Predicting sales accurately and promptly for each product is pivotal for enterprises. Inadequate forecasting can disrupt the entire supply chain, resulting in significant damage to the enterprise's operations (Biswas et al., 2023). The sales performance of a product is influenced by many factors, including user-generated ratings and reviews, as well as intrinsic product attributes such as brand and category (ChatGPT 3.5, September 2022). Extensive research has examined the influence of online word of mouth (WOM) in previous studies. WOM allows consumers to share opinions and experiences about products with individuals beyond their personal networks (Davis & Khazanchi, 2008). WOM contributes to heightened awareness and can shape consumers' perceptions, leading to either positive or negative attitudes toward a product, consequently impacting sales outcomes (Suárez Álvarez et al., 2007).

Figure 2: The Impact of Online WOM Measures on Potential Sales: Initial Conceptual Model (Davis & Khazanchi, 2008).



According to the Social Media Report in 2012, a survey revealed that 70% of customers relied on User Generated Content (UGC) when making purchasing decisions, highlighting its importance in shaping consumer behavior. (Chong & Zhou, 2014). However, due to the lack of reviews or ratings from many satisfied buyers, we will focus our examination on the impact of product-related factors such as category and brand (ChatGPT 3.5, September 2022). Additionally, the abundance of user-generated reviews on the internet poses a challenge for both consumers and merchants, as sifting through thousands or millions of reviews to extract meaningful insights, such as public sentiment towards products, can be arduous and time-consuming (Imtiaz & Ben Islam, 2020).

The rapid growth of e-commerce is driven in part by the extensive adoption of product recommender systems, which assist consumers in making purchase decisions. A notable example is Amazon's co-purchase recommender system, which suggests additional products under the title "customers who bought this item also bought" (Lin & Wang, 2018). One technique employed in product recommendation is content-based filtering, which focuses on analyzing product attributes to generate predictions (Isinkaye et al., 2015).

Figure 3: Amazon Product Recommendation Network (Lin & Wang, 2018).

This method emphasizes the characteristics of products rather than user ratings and can recommend new items even without user ratings (Isinkaye et al., 2015). Understanding the influence of product attributes on customer satisfaction is a key priority for companies seeking to develop more effective product strategies. It is crucial to identify which attributes of a product have the greatest impact on customer satisfaction (Imtiaz & Ben Islam, 2020). In summary, Chapter Two elucidates the nuanced interplay between customer and product variables and sales forecasting models in the e-commerce landscape. The subsequent chapter will delve into the methodological approaches employed to investigate these intricate relationships.

CHAPTER THREE

RESEARCH METHODS

This chapter describes the procedures and methods used for data collection and analysis for this culminating experience project. Through systematic data collection and analysis, we delve into the intricate dynamics of sales forecasting models in e-commerce. Leveraging tools like Power BI and R, we aim to uncover the influence of demographic variables and product-specific factors on sales predictions, contributing to a deeper understanding of online retail dynamics.

Dataset

Project utilizes the Looker Ecommerce Dataset sourced from Kaggle.com. TheLook, an Ecommerce clothing platform, serves as the basis for this dataset, curated by the Looker team. It encompasses a comprehensive range of information, including customer details, product listings, order records, logistics data, web events, and digital marketing campaigns. The dataset contains 7 CSV (Comma-Separated Values) files, comprising orders, products, users, order items, inventory items, events, and distribution centers. These datasets collectively encompass information pertaining to 125,226 unique orders placed by users across 14 different countries between the years 2019 to 2024. To address the research questions, the data offers pertinent variables encompassing customer demographic information such as gender, age, and geographic locations, alongside product-specific attributes including category and brand. Another table is created using data from the Worldometer website, which provides GDP per capita information for the 14 countries in TheLook dataset. It will help in performing comparative analysis between High GDP per capita and low GDP per capita countries in research question 1. The information about both datasets is shown in Figure 4 and Table1.

Figure 4: TheLook Dataset

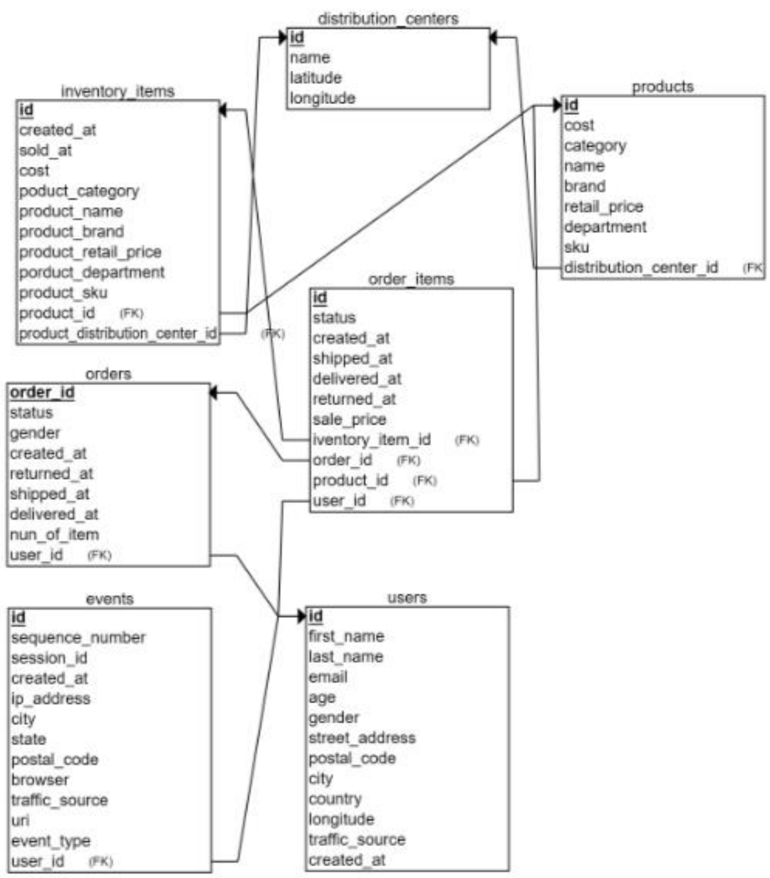


Table 1: Country Wise GDP per Capita (2022) Data

Country	GDP per capita(2022)	Rank
United States	$76,399	8
Austria	$67,936	14
Belgium	$65,027	15
Germany	$63,150	17
Australia	$62,625	18
France	$55,493	25
United Kingdom	$54,603	27
South Korea	$50,070	30
Spain	$45,825	37
Japan	$45,573	38
Poland	$43,269	39
China	$21,476	72
Colombia	$20,287	75
Brasil	$17,822	83

Source: [Worldometer]

Tools and Technologies

In this project, Power BI and R were utilized for data analysis. Power BI is a robust business analytics tool designed for efficient data visualization and insights generation. Its user-friendly interface facilitates easy exploration of data and creation of interactive dashboards, making it suitable for analyzing large datasets across various industries. Additionally, Power BI offers features such as drag-and-drop functionality and comprehensive data modeling capabilities, enhancing its usability for diverse analytical tasks. On the other hand, R is a programming language and software environment known for its strong statistical computing and graphics capabilities. With a wide range of statistical techniques and visualization tools, R proves to be a versatile option for data analysis in various fields. Its extensive library of packages enables users to access advanced statistical modeling, machine learning algorithms, and customizable graphics, allowing for effective handling of complex analytical tasks.

Methodology

Question 1: To what extent are sales forecasting models in the e-commerce ecosystems influenced by customer demographic variables such as gender, age and geographic locations?

Previous research has primarily focused on leveraging sentiment analysis to construct sales forecasting models in e-commerce ecosystems (Biswas et al., 2023). However, a notable gap exists in addressing the behavior of highly satisfied buyers who abstain from providing reviews. To address this gap, this project embarks on a comprehensive analysis of sales patterns, delving into the influence of customer demographic variables such as age, gender, and geographical location. The methodology involves employing descriptive analysis with visuals such as Tables, Matrix, Bar charts and Pie charts within Power BI to gain insights at various levels of granularity. Descriptive analysis defines the world or a situation by answering questions like who, what, where, when, and how much. It's crucial whether you're spotting trends in groups, inventing new ways to measure important things, or detailing samples for studies trying to find out causes (Loeb et al., 2017). By analyzing metrics such as total sales, order count, number of customers, and item count across different age groups and genders, the aim is to compare countries with varying economic statuses, specifically the USA (ranked 8th in GDP per capita) and Brazil (ranked 25th). Furthermore, this study investigates the sales contribution of the top 10 states, examining the demographic characteristics of the highest-spending customers and elucidating how sales are distributed across different age groups for each gender.

Question 2: To what extent are sales forecasting models in the e-commerce ecosystems influenced by product specific factors such as category and brand?

The previous study utilized data from 120 users, focusing specifically on the e-commerce product category of mobile phones. They developed a sales prediction system using an artificial neural network (ANN) computational model, making use of the ANN toolbox within SPSS (Biswas et al., 2023). Building on it, the project involves examining 26 product categories and 2,480 brands within a large customer base of 27,703 individuals. The approach involved employing descriptive analysis within Power BI to delve into total sales and profit percentage across various product categories and brands. By identifying the top 10 product categories and brands by total sales and assessing the percentage of total sales from these categories and brands, we aimed to gain insights into the distribution of sales within our dataset.

Moving forward, a deeper analysis was conducted by employing a multiple linear regression model to forecast sales using R programming. In statistics, linear regression is a method used to model the connection between a single dependent variable y and one or several explanatory variables represented as X. When there's only one explanatory variable, it's termed as simple linear regression, while having more than one explanatory variable is termed as multiple linear regression (Su et al., 2012). This necessitated preprocessing steps such as converting categorical variables into factors and removing outliers using the interquartile range (IQR) technique. Subsequently, the data was divided into a 70:30 split between training and test sets. The independent variables considered included number of items, age, gender, state, country, product category, and product brand, with sales serving as the target variable. To evaluate the performance of the regression model, an XGBoost model was employed, and their performance was compared using R-squared values. XGBoost, short for Extreme Gradient Boosting, is a powerful and scalable machine learning technique based on tree boosting, a highly effective method in the field. Developed to address various machine learning challenges, XGBoost offers an end-to-end solution that has gained widespread adoption among data scientists (Chen & Guestrin, 2016). Additionally, the built-in functionality of XGBoost was utilized to determine feature importance, providing further insights into the factors driving sales forecasts within the e-commerce ecosystem. Through this comprehensive approach, the aim was to shed light on the extent to which sales forecasting models in e-commerce are influenced by product-specific factors, ultimately contributing to a deeper understanding of the dynamics at play in this domain. Transitioning from chapter three, where we meticulously outlined our research methods, chapter four delves into the culmination of our efforts as we analyze the data and present our findings.

CHAPTER FOUR

ANALYSIS AND FINDINGS

Chapter four, analysis, and findings, represents a pivotal stage in our project, where we meticulously analyze the data collected and present our findings. Through detailed examination and interpretation, we aim to provide insights that address the research questions comprehensively.

Data Cleaning

In this stage, all the CSV data files were imported into Power BI, and the model view was utilized to build the data model. Leveraging the Power BI query editor, meticulous renaming of columns and correction of data types was carried out to ensure consistency and accuracy. To streamline the dataset, unnecessary columns were removed while retaining those providing relevant information about products, items, orders, and users. Subsequently, all the tables were merged using merge queries to consolidate the data effectively. By utilizing the 'order status' column, the dataset was filtered to include only completed transactions, eliminating any incomplete orders. Upon inspection, no null values were found in the dataset. However, incorrect country names such as 'España' and 'Deutschland' were encountered, promptly replaced with their accurate counterparts: 'Spain' and 'Germany,' respectively. DAX (Data Expression Language) was employed for feature engineering, resulting in the creation of new columns including age groups, product profit, and profit percentage, thereby enhancing the analytical capabilities of the dataset. Following the execution of these comprehensive data cleaning procedures, the final dataset comprised 45,609 rows and 38 columns, primed for further analysis. Figure 5 displays the final overview of data after the implementation of cleaning steps.

Figure 5: Overview of Cleaned Data within the Interactive Power BI Report

Total Sales by Region

Country	GDP per capita(2022)	Rank
United States	$76399	8
Austria	$67936	14
Belgium	$65027	15
Germany	$63150	17
Australia	$62625	18
France	$55493	25
United Kingdom	$54603	27
South Korea	$50070	30
Spain	$45825	37
Japan	$45573	38
Poland	$43269	39
China	$21476	72
Colombia	$20287	75
Brasil	$17822	83

Country	Order Count	Item Count	Customer Count	Total Sales
China	10713	15653	9477	942,464.34
United States	7102	10276	6257	605,457.27
Brasil	4612	6719	4074	395,939.35
South Korea	1658	2348	1457	137,813.06
United Kingdom	1410	2100	1261	132,013.89
France	1460	2118	1279	124,353.80
Germany	1320	1975	1166	115,933.49
Spain	1219	1774	1083	108,090.18
Japan	763	1060	676	59,151.83
Australia	663	982	586	56,604.80
Belgium	349	487	311	31,868.28
Poland	78	109	69	5,686.49
Colombia	6	7	6	396.97
Austria	1	1	1	9.99

Total Countries	Total Orders
14	31.35K
Total States	**Total Items Ordered**
221	45.61K
Total Customers	**Total Sales**
27.70K	2.72M

Analysis and Findings

Question 1: To what extent are sales forecasting models in the e-commerce ecosystems influenced by customer demographic variables such as gender, age and geographic locations?

Figure 6 shows the examination of sales distribution across age and top-selling product categories. It unveils insightful patterns in consumer behavior within the USA and Brazil. Notably, individuals aged '55 or over' emerge as pivotal contributors, constituting 27.6% and 26.7% of total sales in the USA and Brazil, respectively. Furthermore, the dominance of 'Outwear and Coats' and 'Jeans' in sales contribution across all age groups underscores their universal appeal in both countries.

Figure 6: Cross-Country Age and Product Sales Analysis

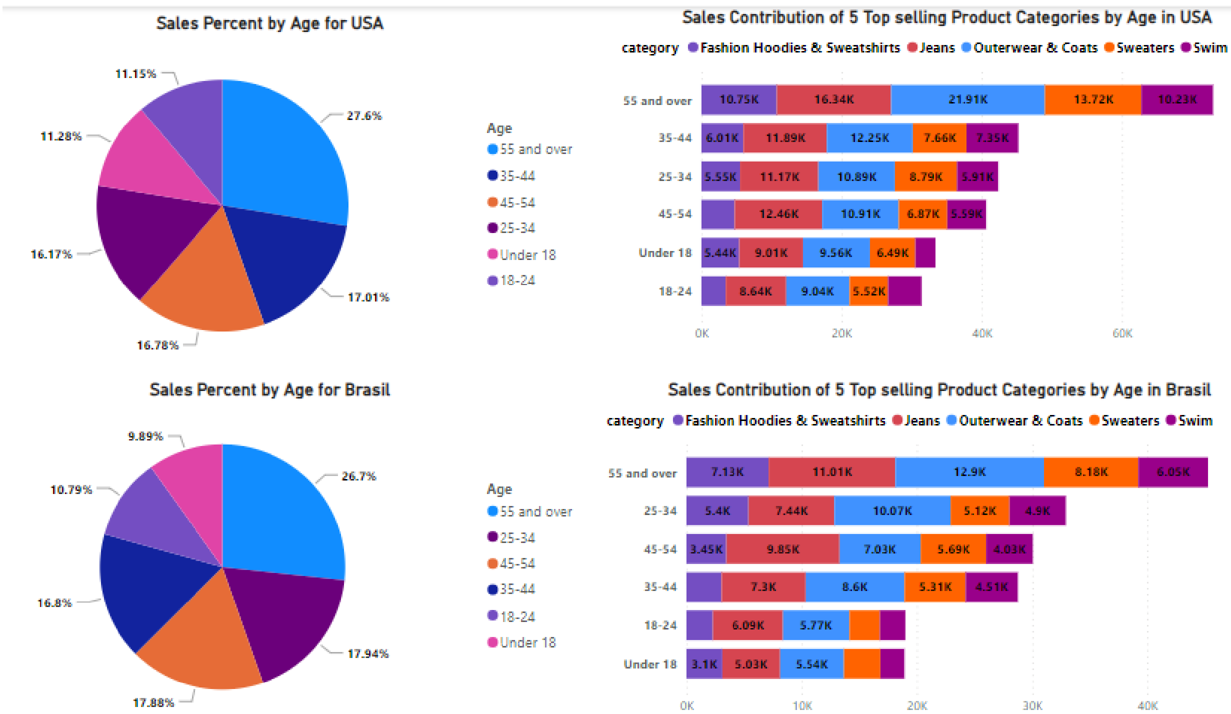
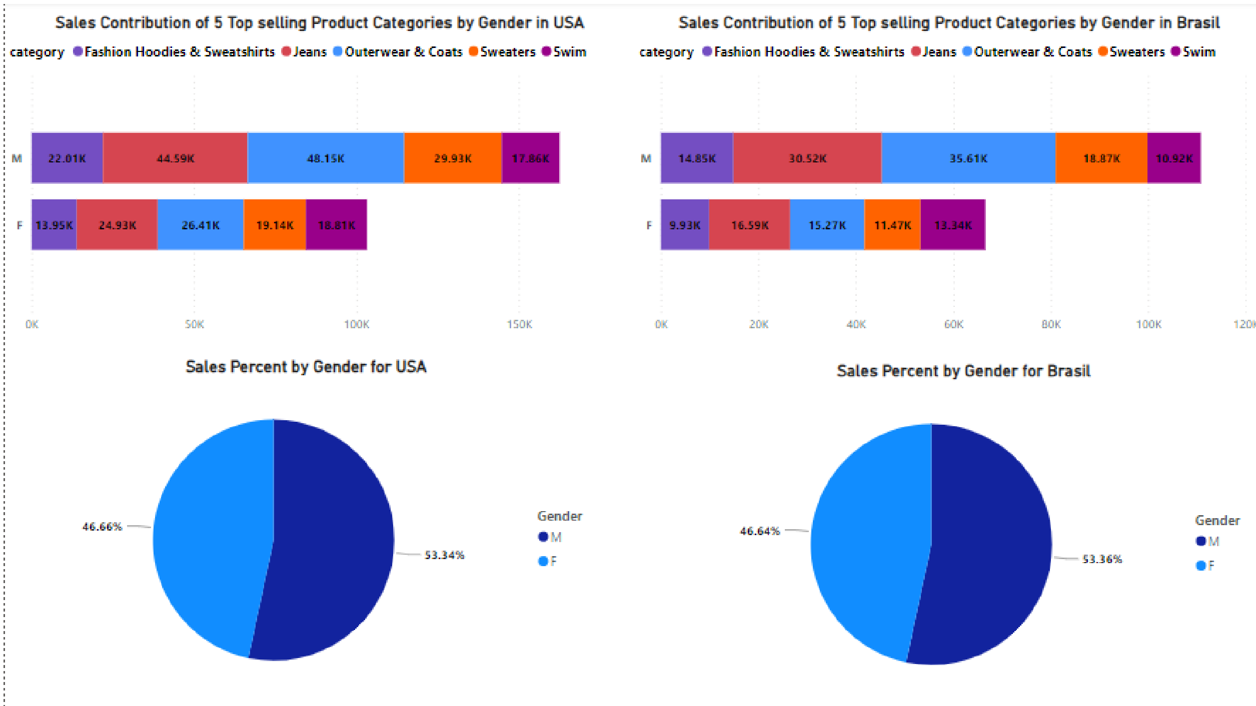


Figure 7: Cross-Country Gender and Product Sales Analysis



Transitioning to Figure 7, sales distribution by gender and top-selling product categories is explored, revealing noteworthy trends in both the USA and Brazil. Male consumers account for 53.34% of total sales in the USA and 53.36% in Brazil, whereas female consumers contribute 46.66% and 46.64% in the respective countries. Across genders, 'Outwear and Coats' along with 'Jeans' emerge as the predominant product categories of choice, mirroring trends observed across age groups. Notably, 'Swim' ranks as the least contributing product category among men in both countries, while 'Fashion Hoodies & Sweatshirts' hold a similar position among women.

Figure 8: Consumer Demographics: Age and Gender Sales Insights Across Countries

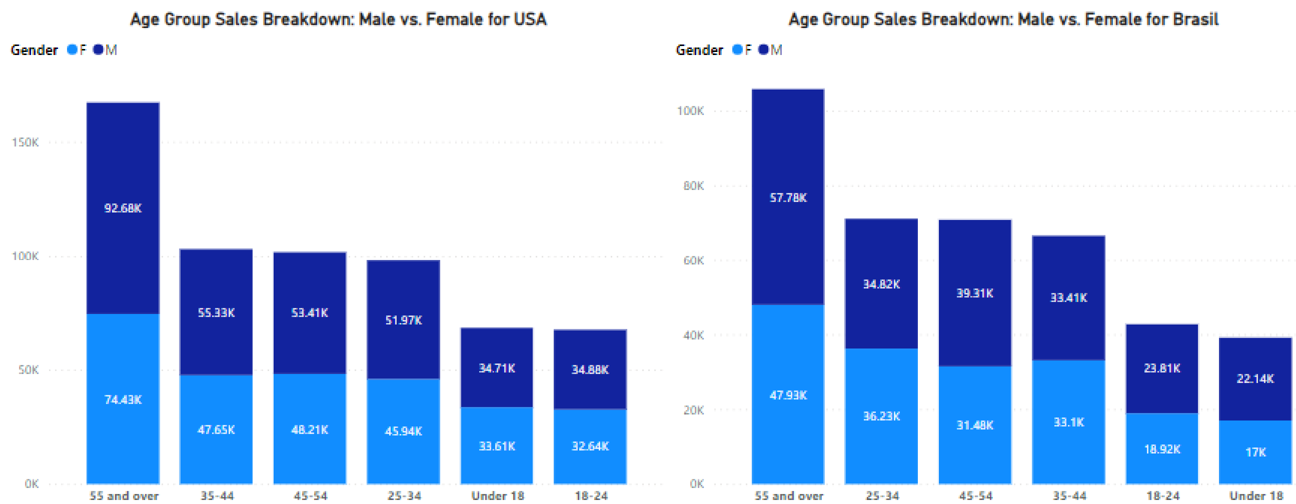


Figure 8 further dissects sales by age group and gender, highlighting male consumers as primary contributors across all age brackets in both countries. In the USA, the age groups '55 and over' and '35-44' demonstrate the highest sales contributions, while the '18-24' age group exhibits the lowest contribution. Conversely, in Brazil, the '55 and over' and '25-34' age groups top the sales charts, with the 'under 18' category showing the least contribution.

Figure 9: Top 10 Sales Performances: States and Customers Analysis

Top 10 States by Total Sales

Country	State	Total Sales	Total Orders	Total Customers
China	Guangdong	145,365.17	1610	1429
United Kingdom	England	116,474.36	1254	1122
United States	California	101,385.51	1143	1015
China	Shanghai	69,586.80	860	746
China	Zhejiang	64,642.34	707	609
China	Beijing	61,685.01	681	599
United States	Texas	60,139.27	725	637
China	Hebei	59,384.29	657	589
Brasil	São Paulo	58,317.55	699	606
China	Jiangsu	49,708.39	570	509

Percentage of Total Sales from Top 10 States

28.97%

Top 10 Customers by Total Sales

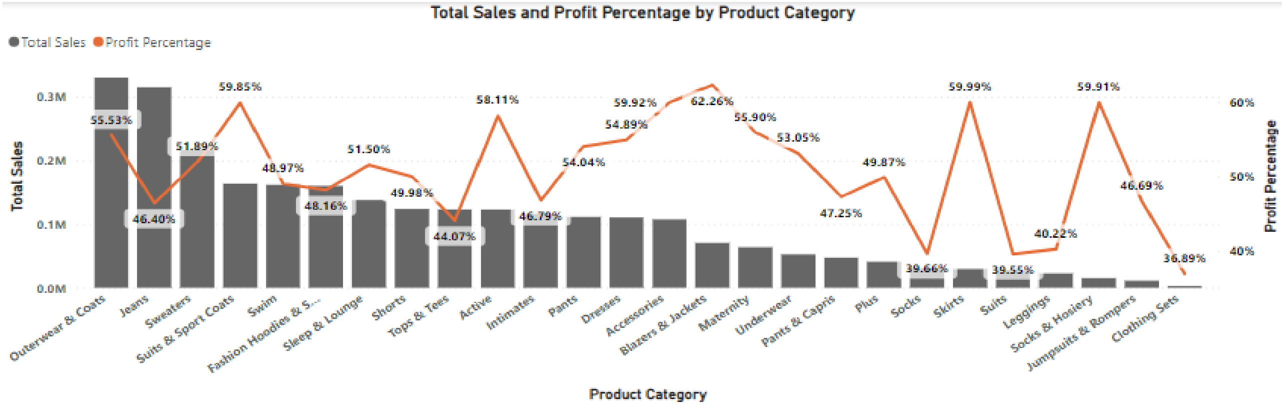
Name	Country	State	City	Gender	Age	Total Sales	Total Orders
Christian Jones	China	Guangdong	Chengdu	M	25-34	1,312.57	1
Derrick Brown	China	Guangdong	Nanjing	M	45-54	1,283.49	1
Tanner Becker	Spain	Canarias	Santa Cruz de Tenerife	M	55 and over	1,275.33	1
James Brown	United States	New York	Webster	M	55 and over	1,273.25	1
Michael Kim	China	Shanxi	Jingdezhen	M	35-44	1,264.99	2
Timothy Smith	United States	Mississippi	Madison	M	Under 18	1,179.32	2
Juan Miller	Japan	Chiba	Chiba	M	25-34	1,176.99	2
Ryan Johnson	China	Jiangxi	Chengdu	M	55 and over	1,174.49	2
Michael Lopez	Brasil	Distrito Federal	Brasília	M	25-34	1,160.99	1
Belinda Combs	China	Zhejiang	Beijing	F	35-44	1,141.50	1

Figure 9 presents insights on the top-performing states and customers based on total sales. Notably, 8 of the top states hail from China and the USA, with a sole representative from Brazil. Leading the pack is Guangdong, China, boasting the highest sales figure of $145,365.17. The combined sales from the top 10 states account for 28.97% of the total sales. Shifting focus to the top 10 customers, a similar trend emerges, with 7 of them originating from China and the USA, and 9 out of 10 being male. Remarkably, 6 of these top customers fall within the age groups of '55 and over' and '25-34’.

Question 2: To what extent are sales forecasting models in the e-commerce ecosystems influenced by product specific factors such as category and brand?

Continuing the exploration into the influences on sales forecasting models in e-commerce ecosystems, attention is now directed towards product-specific factors, particularly category and brand. Figure 10 offers valuable insights into the contribution of product category attributes to sales. Analyzing total sales and profit percentages across all product categories, it is observed that the top 10 product categories by total sales, out of 26 in total, play a significant role in revenue generation. These categories include "Outwear and Coats", "Jeans", "Sweaters", "Suits and Sports Coats", "Swim", "Fashion Hoodies and Sweatshirts", "Sleep and Lounge", "Shorts", "Top and Tees", and "Active", collectively contributing to 68.21% of total sales. Additionally, when considering profit percentage, "Blazers and Jackets" emerges as the most profitable category with a profit percentage of 62.26%, followed closely by "Skirts" at 59.99%.

Figure 10: Product Category Analysis: Total Sales and Profit Percentage



Total Categories

26

Percentage of Total Sales from Top 10 Product Categories

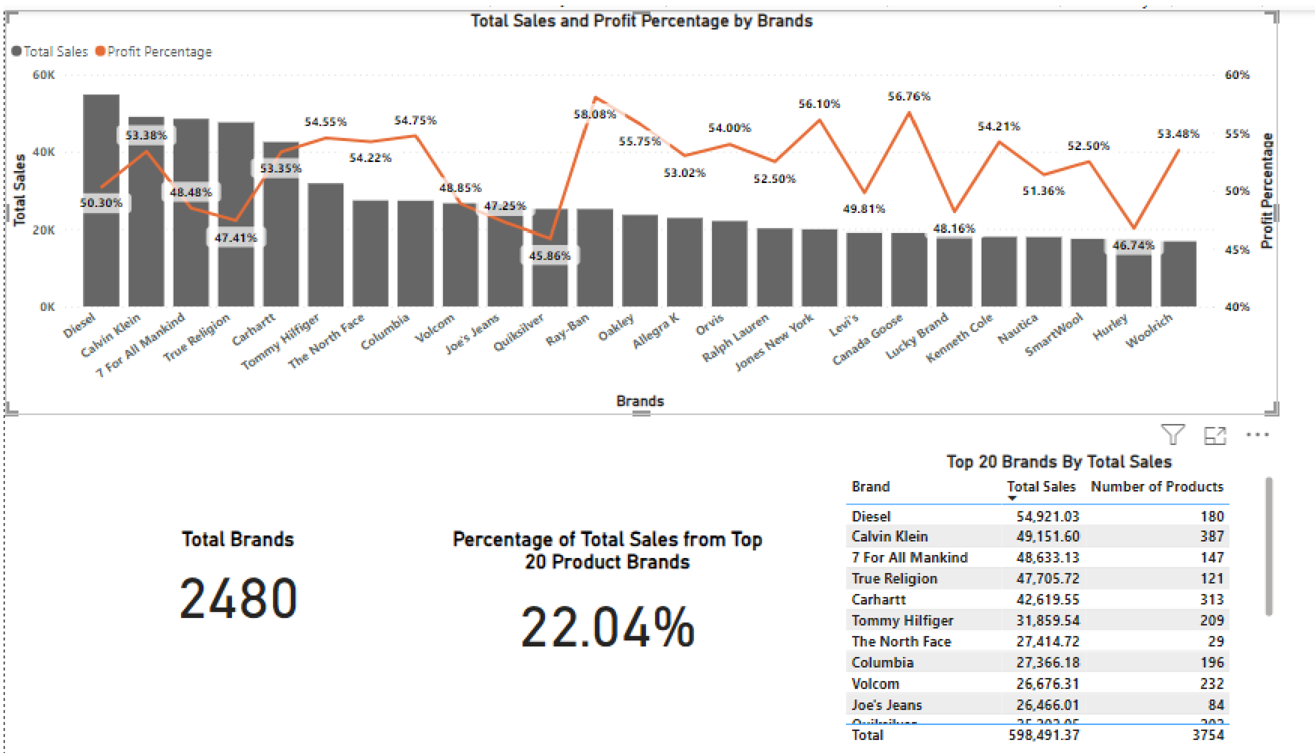
68.21%

Top 10 Product Categories By Total Sales

Category	Total Sales	Number of Products
Outerwear & Coats	329,564.11	1163
Jeans	314,433.31	1561
Sweaters	216,189.71	1387
Suits & Sport Coats	163,445.08	594
Swim	161,576.15	1408
Fashion Hoodies & Sweatshirts	159,889.72	1482
Sleep & Lounge	137,708.17	1419
Shorts	124,042.40	1395
Tops & Tees	122,837.34	1479
Active	122,688.52	1129
Total	1,852,374.51	13017

Transitioning to Figure 11, the influence of product brands on sales is examined. By analyzing total sales and profit percentages across all product brands, significant trends are uncovered among the top 20 product brands out of a total of 2480 brands. Recognizable names such as "Diesel", "Calvin Klein", "Tommy Hilfiger", and "The North Face" emerge as pivotal players in the market. "Diesel" leads the chart with total sales of $54,921, followed closely by "Calvin Klein". Collectively, these top 20 product brands contribute 22.04% of total sales. Furthermore, examining profit percentages within these top brands reveals intriguing insights. "Ray-Ban" stands out with the highest profit margin of 58.08%, followed by "Canada Goose" at 56.76%.

Figure 11: Product Brand Analysis: Total Sales and Profit Percentage



Brand	Total Sales	Number of Products
Diesel	54,921.03	180
Calvin Klein	49,151.60	387
7 For All Mankind	48,633.13	147
True Religion	47,705.72	121
Carhartt	42,619.55	313
Tommy Hilfiger	31,859.54	209
The North Face	27,414.72	29
Columbia	27,366.18	196
Volcom	26,676.31	232
Joe's Jeans	26,466.01	84
Total	598,491.37	3754

Transitioning to the implementation of machine learning techniques for sales forecasting using R, preprocessing steps and data cleaning were undertaken, as detailed in the methodologies section. Subsequently, two models were trained: a linear regression model and an XGBoost model. Figure 12 illustrates linear regression R-squared values of 0.9731 and 0.9737 on the training and testing datasets, respectively.

Figure 12: Linear Regression Model R Square Values on Train and Test Datasets

```
users.stateSouth Carolina                                       0.280162
users.stateSouth Dakota                                         0.366769
users.stateTasmania                                             0.074808 .
users.stateTennessee                                            0.202625
 [ reached getOption("max.print") -- omitted 2301 rows ]
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.84 on 26765 degrees of freedom
Multiple R-squared:  0.9754,    Adjusted R-squared:  0.9731
F-statistic: 427.2 on 2487 and 26765 DF,  p-value: < 2.2e-16
```

```
> error=test$product.sale_price - Predictions
> errorSq=error^2
> SSE=sum(errorSq)
> mue=mean(train$product.sale_price)
> error2=test$product.sale_price - mue
> error2Sq=error2^2
> SST=sum(error2Sq)
> R2=1-SSE/SST
> R2
[1] 0.9737699
```

Figure 13 showcases the variable importance within the XGBoost model, with product brand and category emerging as top influencers. Notably, the XGBoost model outperformed the linear regression model, achieving R-squared values of 0.9923 on the training dataset and 0.9891 on the testing dataset, as depicted in Figure 14.

Figure 13: Feature Importance using XGBoost Model

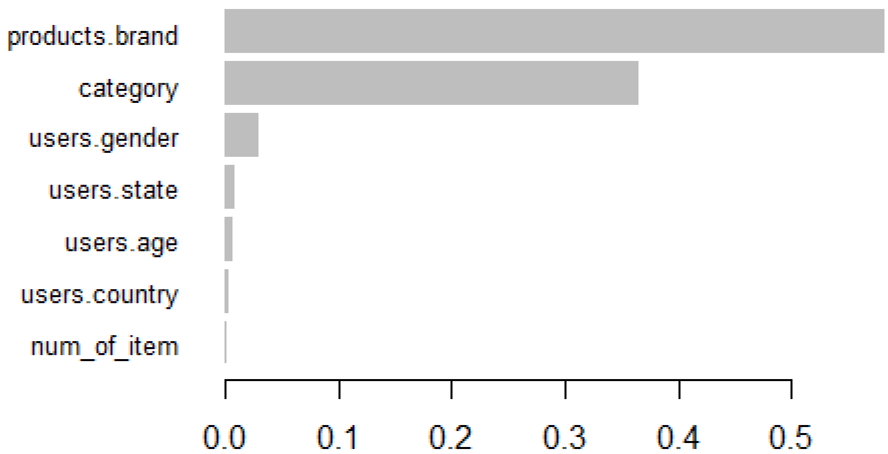


Figure 14: XGBoost Model: R Square Values on Train and Test Datasets

```
> print(paste("Training R-squared:", train_r_squared))
[1] "Training R-squared: 0.992349128608998"
> print(paste("Testing R-squared:", test_r_squared))
[1] "Testing R-squared: 0.989166507523019"
```

Chapter Four provided insights into sales distribution across demographics and product factors illuminate e-commerce dynamics. Transitioning to Chapter Five, we discuss conclusions and propose future study directions based on our findings.

CHAPTER FIVE

DISCUSSION, CONCLUSION, AND

AREAS FOR FURTHER STUDY

Chapter five marks the culmination of our project, wherein we delve into discussion, draw conclusions, and identify potential avenues for future investigation pertaining to both research questions.

Question 1: To what extent are sales forecasting models in the e-commerce ecosystems influenced by customer demographic variables such as gender, age, and geographic locations?

The analysis of sales distribution across different demographic variables revealed several insightful patterns that underscore the significance of demographic factors in shaping sales forecasts. Firstly, the examination of sales distribution across age groups highlighted distinct consumer behaviors within the USA and Brazil. Notably, individuals aged '55 or over' emerged as significant contributors to total sales in both countries, indicating the importance of targeting this demographic segment in sales forecasting models. Conversely, the younger age groups, particularly 'under 18' and '18-24', exhibited lower contributions to total sales, suggesting potential variations in consumer preferences and purchasing power across different age brackets. Secondly, the analysis of sales distribution by gender revealed notable trends in consumer behavior across genders. Male consumers consistently accounted for a higher proportion of total sales compared to female consumers in both the USA and Brazil. Moreover, the preference for certain product categories, such as 'Outwear and Coats' and 'Jeans', remained consistent across genders, indicating the universal appeal of these products irrespective of gender. Furthermore, the intersection of age groups and gender in sales distribution highlighted male consumers as primary contributors across all age brackets in both countries. This finding suggests the importance of considering the interaction between demographic variables, such as age and gender, in refining sales forecasting models to better capture consumer preferences and behaviors. Lastly, the analysis of top-performing states and customers based on total sales provided valuable insights into geographic variations in consumer behavior. The concentration of top-performing states and customers in regions like China and the USA underscores the influence of geographic locations on sales forecasts, emphasizing the need for localized approaches in demand management strategies. Overall, the findings from the data analysis demonstrate the significant impact of customer demographic variables, including age, gender, and geographic locations, on sales forecasting in the e-commerce ecosystem. By incorporating these insights into demand management strategies, businesses can enhance the accuracy and effectiveness of their sales forecasts, thereby optimizing resource allocation and improving customer satisfaction.

Conclusion

In conclusion, it is evident that customer demographic variables play a crucial role in influencing sales forecasting models in e-commerce ecosystems. The examination of sales distribution across age groups, genders, and geographic locations revealed insightful patterns in consumer behavior, highlighting the importance of targeting specific demographic segments in sales forecasting. Moreover, the findings emphasize the significance of tailoring marketing strategies and inventory management to different demographic segments, underscoring the need for personalized approaches in demand forecasting and resource allocation.

Areas for Further Study

Future research should explore the dynamics of customer behavior in response to segmentation variables, such as purchasing frequency, basket size, and customer lifetime value, discounts, and promotions on sales predictions. Additionally, another area is the impact of cultural and social factors on consumer preferences and purchasing decisions across different demographic segments.

Question 2: To what extent are sales forecasting models in the e-commerce ecosystems influenced by product specific factors such as category and brand?

The analysis of product-specific factors, including category and brand, in sales forecasting models within e-commerce ecosystems yields valuable insights crucial for businesses aiming to optimize their operations and maximize revenue. Starting with product categories, the examination reveals the significant contribution of top-selling categories such as "Outwear and Coats" and "Jeans" to total sales. Understanding the demand dynamics within these categories informs inventory management decisions, ensuring that businesses maintain sufficient stock levels to meet consumer demand and capitalize on revenue opportunities. Furthermore, the identification of highly profitable categories like "Blazers and Jackets" emphasizes the importance of considering profitability alongside sales volume when forecasting sales. Similarly, the analysis of top product brands underscores the influence of branding on consumer purchasing decisions. Recognizable brands such as "Diesel" and "Calvin Klein" emerge as key drivers of sales, indicating the importance of brand recognition and reputation in shaping consumer preferences. These insights highlight the necessity for businesses to cultivate strong brand identities and reputations to compete effectively in the market. By leveraging brand equity, businesses can enhance consumer trust and loyalty, thereby driving sales and maintaining a competitive edge. Transitioning to the implementation of machine learning techniques for sales forecasting, the significance of product categories and brands as independent variables in predicting sales outcomes is evident in Figure 13. The high R-squared values achieved by the trained models underscore the effectiveness of incorporating product-specific factors into forecasting methodologies. By integrating these insights into forecasting models, businesses can improve the accuracy and reliability of their sales forecasts, enabling better resource allocation, inventory management, and overall business performance. The findings highlight the intricate relationship between product-specific factors, sales forecasting models, and business operations within e-commerce ecosystems. By leveraging insights into product categories and brands, businesses can make informed decisions to optimize inventory management, capitalize on revenue opportunities, and cultivate strong brand identities, ultimately driving competitiveness and success in the market.

Conclusion

In conclusion, the analysis reveals the significant impact of product-specific factors such as category and brand on sales forecasting models in e-commerce ecosystems. It underscores the influence of consumer preferences and branding on sales outcomes, informing inventory management decisions and pricing strategies. Additionally, the implementation of machine learning techniques enhances the accuracy and effectiveness of sales forecasting, enabling businesses to make more informed decisions and optimize resource allocation.

Areas for Further Study

Future research should focus on the impact of temporal factors like cyclical and seasonal trends on sales forecasting models in e-commerce ecosystems. Additionally, exploring advanced machine learning techniques, including deep learning algorithms, could significantly enhance predictive accuracy and effectiveness.

APPENDIX

CODES

```r
## Importing libraries:

library(plotly)
library(dplyr)
library(tidyverse)
library(olsrr)
library(caTools)
library(xgboost)
library(data.table)
library(ggplot2)
library(caret)

## Loading data:
Looker_Ecommerce=read.csv("Data.csv")

nrow(Looker_Ecommerce)
ncol(Looker_Ecommerce)

summary(Looker_Ecommerce)

colnames(Looker_Ecommerce)

str(Looker_Ecommerce)
```

```r
## Variable selection:
Looker_Ecommerce = subset(Looker_Ecommerce,select = -c(1,2,3,4,5,6,7,8,9,11,12,16,17,19,20,21
colnames(Looker_Ecommerce)

## Checking NA values:
colSums(is.na(Looker_Ecommerce))

## Outlier removal:
boxplot(Looker_Ecommerce$product.sale_price)
range(Looker_Ecommerce$product.sale_price)
Q1 <- quantile(Looker_Ecommerce$product.sale_price, .25)
Q3 <- quantile(Looker_Ecommerce$product.sale_price, .75)
IQR <- IQR(Looker_Ecommerce$product.sale_price)

Looker_Ecommerce <- subset(Looker_Ecommerce, Looker_Ecommerce$product.sale_price>
                             (Q1 - 1.5*IQR) & Looker_Ecommerce$product.sale_price<
                             (Q3 + 1.5*IQR))
range(Looker_Ecommerce$product.sale_price)
str(Looker_Ecommerce)


## Converting categorical variables in Factor variable:
Looker_Ecommerce$users.gender = as.factor(Looker_Ecommerce$users.gender)
Looker_Ecommerce$users.state = as.factor(Looker_Ecommerce$users.state)
Looker_Ecommerce$users.country = as.factor(Looker_Ecommerce$users.country)
Looker_Ecommerce$category = as.factor(Looker_Ecommerce$category)
Looker_Ecommerce$products.brand = as.factor(Looker_Ecommerce$products.brand)
```

```r
## Splitting Data into Train and Test:
set.seed(1)
sample <- sample.split(Looker_Ecommerce$product.sale_price, SplitRatio = 0.7)
train  <- subset(Looker_Ecommerce, sample == TRUE)
test   <- subset(Looker_Ecommerce, sample == FALSE)
dim(train)
dim(test)


## Linear regression model:
model1 <- lm(product.sale_price~., data = train)

## Calculate R-squared on training data:
summary(model1)

test$products.brand[which(!(test$products.brand %in% unique(train$products.brand)))] <- NA
test = na.omit(test)

## Calculate R-squared on testing data:
Predictions = predict(model1,newdata = test)
error=test$product.sale_price - Predictions
errorSq=error^2
SSE=sum(errorSq)
mue=mean(train$product.sale_price)
error2=test$product.sale_price - mue
error2Sq=error2^2
SST=sum(error2Sq)
R2=1-SSE/SST
R2
```

```r
## XGBoost model:
train_x = data.matrix(train[, -7])
train_y = train[,7]

test_x = data.matrix(test[, -7])
test_y = test[, 7]

xgb_train = xgb.DMatrix(data = train_x, label = train_y)
xgb_test = xgb.DMatrix(data = test_x, label = test_y)

model4 = xgb.train(data = xgb_train, max.depth = 3, nrounds = 700, verbose = 0)

## Make predictions on the training data
train_pred <- predict(model4, xgb_train)

## Calculate R-squared on training data
train_r_squared <- caret::R2(pred = train_pred, obs = train_y)
print(paste("Training R-squared:", train_r_squared))

## Make predictions on the testing data
test_pred <- predict(model4, xgb_test)

## Calculate R-squared on testing data
test_r_squared <- caret::R2(pred = test_pred, obs = test_y)
print(paste("Testing R-squared:", test_r_squared))

importance_matrix = xgb.importance(colnames(xgb_train), model = model4)
xgb.plot.importance(importance_matrix[1:7,])

importance_matrix
```

REFERENCES

Aro-Gordon, S., & Gupte, J. (2016). Review of modern inventory management techniques. The Global Journal of Business and Management, 1, 1–22.

Biswas, B., Sanyal, M. K., & Mukherjee, T. (2023). AI-Based Sales Forecasting Model for Digital Marketing. International Journal of E-Business Research (IJEBR), 19(1), 1–14. https://doi.org/10.4018/IJEBR.317888

Bojer, C. S., & Meldgaard, J. P. (2021). Kaggle forecasting competitions: An overlooked learning opportunity. International Journal of Forecasting, 37(2), 587–603. https://doi.org/10.1016/j.ijforecast.2020.07.007

Buckinx, W. R., & Poel, D. V. D. (2003). Predicting Online Purchasing Behavior. Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium, Article 03/195. https://ideas.repec.org//p/rug/rugwps/03-195.html

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Chong, A. Y.-L., & Zhou, L. (2014a). Demand chain management: Relationships between external antecedents, web-based integration and service innovation performance. International Journal of Production Economics, 154, 48–58. https://doi.org/10.1016/j.ijpe.2014.04.005

Chong, A. Y.-L., & Zhou, L. (2014b). Demand chain management: Relationships between external antecedents, web-based integration and service innovation performance. International Journal of Production Economics, 154, 48–58.

Croxton, K. L., Lambert, D. M., García-Dastugue, S. J., & Rogers, D. S. (2002). The Demand Management Process. The International Journal of Logistics Management, 13(2), 51–66. https://doi.org/10.1108/09574090210806423

Davis, A., & Khazanchi, D. (2008). An Empirical Study of Online Word of Mouth as a Predictor for Multi-product Category e-Commerce Sales. Electronic Markets, 18(2), 130–141. https://doi.org/10.1080/10196780802044776

Fisher, M., & Raman, A. (2018). Using Data and Big Data in Retailing. Production and Operations Management, 27(9), 1665–1669. https://doi.org/10.1111/poms.12846

Goldsmith, R. E., & Flynn, L. R. (2004). Psychological and behavioral drivers of online clothing purchase. Journal of Fashion Marketing and Management: An International Journal, 8(1), 84–95. https://doi.org/10.1108/13612020410518718

Gupta, A. (2014). E-COMMERCE: ROLE OF E-COMMERCE IN TODAY’S BUSINESS. https://www.semanticscholar.org/paper/E-COMMERCE-%3A-ROLE-OF-E-COMMERCE-IN-TODAY%27S-BUSINESS-Gupta/8516fda8072235363b2399dd98044e535169c220

Gustriansyah, R., Ermatita, E., & Rini, D. P. (2022). An approach for sales forecasting. Expert Systems with Applications, 207, 118043. https://doi.org/10.1016/j.eswa.2022.118043

Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. International Journal of Forecasting, 36(4), 1420–1438. https://doi.org/10.1016/j.ijforecast.2020.02.005

Imtiaz, Md. N., & Ben Islam, Md. K. (2020). Identifying Significance of Product Features on Customer Satisfaction Recognizing Public Sentiment Polarity: Analysis of Smart Phone Industry Using Machine-Learning Approaches. Applied Artificial Intelligence, 34(11), 832–848. https://doi.org/10.1080/08839514.2020.1787676

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal, 16(3), 261–273.

Jain, V., Malviya, B., & Arya, S. (2021). An Overview of Electronic Commerce (e-Commerce). Journal of Contemporary Issues in Business and Government, 27, 665–670. https://doi.org/10.47750/cibg.2021.27.03.090

Jha, L. (n.d.). Performance Analysis and Optimization of Customer E-Commerce using Data Segmentation Algorithm. Retrieved February 26, 2024, from https://scholar.archive.org/work/xr73r2tuevchzjlifya7zyjwki/access/wayback/https://www.ijiccn.com/images/files/vol2-issue1/Performance-Analysis-and-Optimization-of-Customer-E-Commerce-using-Data-Segmentation-Algorithm.pdf

Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. Industrial Marketing Management, 90, 493–507. https://doi.org/10.1016/j.indmarman.2019.05.003

Lin, Z., & Wang, Q. (2018). E-commerce product networks, word-of-mouth convergence, and product sales. Journal of the Association for Information Systems, 19(1), 2.

Liu, N., Ren, S., Choi, T.-M., Hui, C.-L., & Ng, S.-F. (2013). Sales Forecasting for Fashion Retailing Service Industry: A Review. Mathematical Problems in Engineering, 2013, e738675. https://doi.org/10.1155/2013/738675

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). Descriptive Analysis in Education: A Guide for Researchers. NCEE 2017-4023. In National Center for Education Evaluation and Regional Assistance. National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED573325

Patil, H., & Divekar, Brig. R. (2014). Inventory Management Challenges for B2C E-commerce Retailers. Procedia Economics and Finance, 11, 561–571. https://doi.org/10.1016/S2212-5671(14)00221-4

Sari, J. N., Nugroho, L. E., Ferdiana, R., & Santosa, P. I. (2016). Review on Customer Segmentation Technique on Ecommerce. Advanced Science Letters, 22(10), 3018–3022. https://doi.org/10.1166/asl.2016.7985

Singh, R. (2021). A Study of Artificial Intelligence and E-Commerce Ecosystem – A Customer's Perspective. International Journal of Research in Engineering, Science and Management, 4(2), Article 2.

Su, X., Yan, X., & Tsai, C. (2012). Linear regression. WIREs Computational Statistics, 4(3), 275–294. https://doi.org/10.1002/wics.1198

Suárez Álvarez, L., Díaz Martín, A. M., & Casielles, R. V. (2007). Relationship Marketing and Information and Communication Technologies: Analysis of Retail Travel Agencies. Journal of Travel Research, 45(4), 453–463. https://doi.org/10.1177/0047287507299593

Taher, G. (2021). E-Commerce: Advantages and Limitations. International Journal of Academic Research in Accounting, Finance and Management Sciences, 11(1), Pages 153-165. https://doi.org/10.6007/IJARAFMS/v11-i1/8987

Thananuraksakul, S. (2007). Factors affecting online shopping behaviour: A study of Thai consumers. University of South Australia. https://www.academia.edu/download/55652213/reading.pdf

The Role of Customer Segmentation in Sales Forecasting. (n.d.). Retrieved February 26, 2024, from https://abmatic.ai/blog/role-of-customer-segmentation-in-sales-forecasting

Usmani, Z. A., Manchekar, S., Malim, T., & Mir, A. (2017). A predictive approach for improving the sales of products in e-commerce. 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 188–192. https://ieeexplore.ieee.org/abstract/document/7972409/?casa_token=JrvKpaiSNA8AAAAA:YAvRTzPHTe7Ak9Rq_dHX5kiIoCnaJ0ZjMx93xKEjgBTE9H83bh4vf3VPSA9jNkYOQBkzYlYwVlw

Wan, Y., Nakayama, M., & Sutcliffe, N. (2012). The impact of age and shopping experiences on the classification of search, experience, and credence goods in online shopping. Information Systems and E-Business Management, 10(1), 135–148. https://doi.org/10.1007/s10257-010-0156-y

Wellens, A. P., Boute, R. N., & Udenio, M. (2024). Simplifying tree-based methods for retail sales forecasting with explanatory variables. European Journal of Operational Research, 314(2), 523–539. https://doi.org/10.1016/j.ejor.2023.10.039

Weller, M., & Crone, S. (2012, November 1). Supply Chain Forecasting:Best Practices & Benchmarking Study. https://www.semanticscholar.org/paper/Supply-Chain-Forecasting%3ABest-Practices-%26-Study-Weller-Crone/9d0ee64331d241dc6821c9865f24af77d6da647b