SELINUS UNIVERSITY
OF SCIENCES AND LITERATURE

Design and Implementation of Big Data and Analytics to Enhance Dynamic Price Comparison Analytics of an e-Commerce Enterprise Company in Australia

By Bryan Dadiz

A DISSERTATION

Presented to the Department of
Information Technology
program at Selinus University

Faculty of Computer Science
in fulfillment of the requirements
for the degree of Doctor of Philosophy
in Information Technology

2024

Table of Contents

The List of Figures

The List of Tables

Abstract

The rapid proliferation of digital technologies has fueled the unprecedented growth of e-commerce, revolutionizing the way individuals and businesses engage in commercial transactions. This abstract explores the multifaceted landscape of e-commerce, its evolutionary trajectory, and its profound impacts on consumer behavior-commerce has transcended geographical boundaries and temporal limitations, creating a borderless marketplace accessible around the clock. Advances in technology, shifts in client tastes, and advances in supply chain management have all contributed to the expansion of e-commerce, from the early days of online retail to the contemporary era of social commerce and augmented reality purchases The alteration of customer behavior in reaction to e-commerce is the primary emphasis of this abstract. The convenience of online shopping, personalized recommendations, and one-click purchasing have changed the retail experience. Additionally, the growth of user-generated reviews and social media impact has shaped purchasing decisions, under scoring the crucial role of trust and authenticity in the digital marketplace.

The abstract also examines the challenges and opportunities presented by the e-commerce ecosystem. Privacy concerns, data security issues, and the digital divide have emerged as important considerations in an increasingly interconnected world. On the other side, e-commerce has made it possible for entrepreneurs and small enterprises to reach a worldwide audience, which has stimulated innovation and economic advancement. This abstract concludes by highlighting the manner in which

e-commerce has altered consumer discovery, evaluation, and purchase processes in addition to upending old retail structures. As e-commerce continues to evolve, understanding its impact on consumer behavior and the broader socioeconomic landscape will be pivotal for businesses, policymakers, and researchers alike.

Acknowledgement

I want to sincerely thank the people and groups listed below for their significant assistance and contributions in helping me finish this research paper/thesis.

My Supervisor/Advisor: [Name of your supervisor/advisor] – for their direction, mentorship, and unfailing assistance with our endeavor. They were really helpful in molding this work with their knowledge and perceptive criticism.

Family: I am incredibly appreciative of my family's support, tolerance, and unfailing faith in me. I was emotionally supported by them during the trying parts of this journey.

Friends: My friends' support, encouragement, and several conversations that helped me hone my ideas are greatly appreciated.

Colleagues: I would like to thank my colleagues for their collaboration, thought-provoking discussions, and constructive criticism.

Funding Agencies: I am grateful to [Name of Funding Agency/Grant] for their financial assistance in making this study feasible.

Institution: I appreciate [Name of Your Institution] because it offers the tools, libraries, and settings that make research possible.

Participants/Respondents: I acknowledge and thank the individuals who participated in surveys or interviews for this study. Your valuable insights were crucial to the success of this project. Technical Support: The assistance provided by [Name of

Technical Support or Service Providers] in various technical aspects of this work is deeply appreciated. Literature Sources: I want to acknowledge the authors and researchers whose work has been a cornerstone of my research. Their contributions have been instrumental in shaping my understanding.

Any other relevant acknowledgments: If there are specific individuals, organizations, or entities that have played a significant role in your project, be sure to include them here.

Dedication

This study or thesis is in honor of my [name of a loved one] who [describes them].

importance or influence on your life]. Their [characteristics, traits, or power] continue to Encourage me to pursue [your particular career or endeavor] and knowledge.

and [your specific field or endeavor].

I also want to dedicate my work to [name other notable individuals] as a means of thanking them for their encouragement and support over the years. Their belief in me and ability to keep me motivated enabled me to complete this [book,thesis,or research project].

To all those who have walked alongside me in this journey and to the countless individuals who believe in the power of [your field or endeavor], this work is a tribute to your enduring faith and commitment.

CHAPTER 1

Introduction

E-commerce websites are incredibly common and are expanding at a rate that has never been witnessed before. Based on research conducted in 2019 by Esther Shaulova, Lodovica Biagi, and Statista [1], the e-commerce sector produced $3.53 trillion in revenue. e-commerce report 2019, implying that the quantity of e-commerce websites across the globe is increasing. According to the aforementioned statistics, a sizable percentage of consumers prefer online shopping to traditional brick and mortar establishments. E-commerce website owners intentionally list prices that are higher than their actual costs in an attempt to profit from customer demand. With this approach, the same commodities are sold to various clients at varying prices. Recent years have seen an increase in pricing discrimination in e-commerce due to the ease with which individuals' browsing and purchase histories may be traced. There are concerns about how the practice may affect consumer welfare, market effectiveness, and competitiveness. As a result, customers who make purchases without knowing the upfront costs of such products waste their money. As a result, customers are tricked into paying more than necessary for items. Given the aforementioned information, the price comparison tool's importance is evident. However, a lot of study has been done, and browser add-ons are widely accessible and recommended in several languages in many countries, including the Philippines.

Having stated that, there are two well-known price comparison tools that have been released in the Philippines: iprice.ph and smartprix.com. In a manner similar to

these two websites, secprf.com serves the same function as a price comparison tool in our nation.

Only textual data searches, like those on our price comparison tool, are offered by this website; it does not, however, offer efficient results or statistics. This inspires me to create a pricing comparison tool for a particular Australian online retailer called a hobby shop that specializes in carrying a broad selection of brands for all different kinds of hobby fans. This business, which is a classic brick-and-mortar store that depends on foot traffic, has been in operation on Flinders Street for 45 years. The business was forced by the new realities of 2020 to significantly rely on its online platforms and look for ways to increase the sales of their eCommerce store.

My proposal intends to create and install an analytics system to improve dynamic pricing comparison for an Australian online retailer that sells hobbies. Speaking in comparison to J. In 2015, Nakash et al. compared the products offered on several websites using an inverted indexing method. [2] Amir Mohd, Siddiqi Muzammil Ahmad, Jawahire Nakash, and Shaikh Anas. BothTabrez Khan and Azam. Instantaneous product analysis is accomplished by data mining. The International Journal of Advanced Research in Computer Engineering issued an issue in March 2015. I've already included an SKU and a Regular Expression in my technique to solve the Fixing ofthe SKU's problem. A natural language processing algorithm will also be used to determine a product's price, maybe in relation to daily price changes.

Another piece that uses our methodology was released in 2019 by S. Mehak and others. In order to develop a tool for price comparison, Sohail Masood Bhatti, Sharaz Aslam, Rabia Zafar, Shakra Mehak, and others employed handmade scrap

data storage and Python-based site scraping tools. Using a filtering strategy and site scraping are two aspects of smart online shopping.

Mathema is the name of the Second International Conference on Computing in 2019. In 2018, K. Pradeep et al. used data mining to examine purchasing habits and produced a pattern analysis recommender system. [4].In 2016, R. Shah and colleagues created a website for price comparison using web crawling using the Django framework and MongoDB. For online scraping, they additionally utilized requests and the BeautifulSoup4 module. [5]. In 2015, C. Using the Beautiful Soup package, Zheng et al. created a mechanism for obtaining data from websites [6]. That same year, A. A technique for finding and automatically obtaining product price data from random e-commerce websites was released by Horch et al. [7]. 2014 saw Y. Using web mining techniques, Ming-Hsiung et al. developed a way to identify the best deal and let the user know so they can be more satisfied [8].

In this essay, I offer a solid pricing comparison tool for a hobbyist-focused Australian e-commerce company. Building spiders with the Selenium framework to browse webpages and collect or scrape information provided there [9]. The CSV database contains the data that was retrieved. The cosine similarity technique is used to vectorize the titles and determine the degree of similarity between the search text and the dataset text. [10] Using the Django Framework and Power BI, I will develop a prototype website that can be easily shared with the general public.[11]. We created the user interface to allow for easy interaction when searching for a query and to display results relevant to that query.

The organization of this work is as follows: in Part 2, we thoroughly explore the suggested system as well as our approach and algorithm. The experiment's results

and a performance analysis are reported in Section 3 of the essay. The system's shortcomings are discussed in Section 4 of the paper, along with our plan for further research. This aids companies in adjusting inventory, pricing, and marketing plans to match demand. Visual Search: AI-powered visual search enables users to look for products using photos, simplifying the purchasing process. This technology uses patterns, colors, and forms to locate items in the e-commerce catalog that are similar to each other. Recommendation Systems: AI-powered recommendation systems use consumer data analysis to recommend related or complimentary products, increasing the likelihood of cross-selling and up selling. Dynamic Pricing: Artificial intelligence (AI) helps organizations optimize revenue and preserve competitiveness by adjusting prices in reaction to market conditions, rival pricing, and customer behavior. Supply Chain and Inventory Management: Artificial intelligence enhances supply chain operations by optimizing distribution logistics, enhancing inventory management, and anticipating shifts in demand.

The rapid improvements in technology and the internet have significantly changed the landscape of commerce in the modern digital era. Electronic commerce, or e-commerce, has become a disruptive force that is reshaping how firms run and how customers conduct business. For both established businesses and new startups, this paradigm shift has created opportunities and difficulties that have never before existed. As a result, research on e-commerce has become more important and is currently piquing the interest of academics, business leaders, and policymakers. The traditional retail paradigm, which depends on actual stores and in-person interactions, has been upended by the rise of e-commerce. Digital platforms enable e-commerce, which makes it simpler to buy and sell goods and services online.

Consumer shopping patterns have changed as a result of this transformation, which has also forced businesses to reevaluate their strategy, supply chains, and methods for engaging with customers. E-commerce has democratized market access and encouraged innovation, enabling even the smallest companies to reach a global audience. From global marketplaces to tiny boutiques.

This work is structured as follows: in part 2, we provide a detailed description of the proposed system along with a thorough explanation of our methodology and algorithm. The results of the experiment are presented in the third section of the paperalong with a performance analysis. The paper's fourth section looks at the system's limitations and our research strategy.

1.2 Background of the Study

E-commerce websites are crucial to businesses because they give them a strong platform to connect with customers around the world, offer convenient shopping options, and adjust to shifting consumer trends. They also make it possible for businesses and consumers to make decisions that are more accessible, data-driven, and cost-effective.

It is crucial for businesses to maintain an advantage in pricing strategy in the fiercely competitive e-commerce market of today. E-commerce businesses benefit greatly from dynamic price comparison analytic, which entails real-time competitor pricing monitoring. The amalgamation of Big Data and Analytics presents a promising opportunity to revolutionize pricing tactics, bolster competitiveness, and propel business expansion.

Over the past ten years, e-commerce has grown at an unprecedented rate as more and more customers turn to online platforms for their shopping needs. Due to this growth, pricing is now a crucial part of the e-commerce ecosystem, influencing consumer decisions and ultimately determining a company's ability to survive. An e-commerce business can benefit from dynamic price comparison analytic in a number of ways.

Price optimization: Businesses may adapt their pricing strategies to competitors' price changes, shifts in the market, and variations in demand by using real-time data analysis. Customer retention: Effective dynamic price comparison analytic enables businesses to provide customers with attractive and consistent pricing, fostering trust and loyalty. Profit maximization: By optimizing pricing, businesses can enhance profitability while ensuring competitiveness. Market positioning: Data-driven insights can inform strategic decisions about market positioning and segmentation. Competitive advantage: Businesses can set competitive prices and get a larger market share and revenue by using fast and accurate pricing insights. Research Gap: More study is required on the careful design and implementation of a Big Data and analytic system in order to enhance this vital e-commerce component, even in spite of the growing popularity of dynamic pricing comparison analytics. The efficacy of current practices in a fast-paced, data-rich environment may be limited by their frequent lack of integration of advanced data processing and analytics techniques.

1.3 Objectives of the Study

General Objectives

To create and execute a full data-driven solution that includes web scraping, data warehousing, business intelligence, and machine learning capabilities for e-commerce analysis and decision-making. The objective is to offer insightful information on price comparisons, product pricing forecasts, and product retention, improving the e-commerce industry's overall effectiveness and competitiveness.

Specific Objectives

1. Design a web-scrape script to gather Data of e-commerce websites.
2. Design the model of the data-warehouse to house data from different sources.
3. Develop a Business Intelligence tool to provide insights on price comparisons of different e-commerce websites.
4. Implement Machine Learning to predict the possible price of the product.
5. Predict product retention based on stock prices.

1.4 Significance of the study

The development and implementation of a complete data-driven solution with web scraping, data warehousing, business intelligence, and machine learning capabilities for e-commerce analysis and decision-making. To increase the e-commerce sector's overall effectiveness and competitiveness, the goal is to provide analytical information on price comparisons, product pricing forecasts, and product retention.

1.5 Scope and Delimitation

Ecommerce Platform Development: The project entails the creation of a user-friendly and secure ecommerce platform that enables clients to explore products, add items to their shopping carts, continue to the checkout process, and process payments online.Product Catalog: The site will provide a large variety of products in several categories, each with a top-notch picture, a detailed explanation, and a price.

Customers can create an account, log in, and manage their profiles, which include their order history, wish lists, and saved payment methods. Shopping Cart and Checkout: The platform will feature a functional shopping cart that aggregates selected items and a secure checkout process that supports multiple payment options.Payment Gateway Integration: Integration with reputable payment gateways will enable secure online transactions, including credit or debit cards, digital wallets, and other relevant methods.Order Management: The system will allow users to track their orders, receive order status notifications, and contact customer support if needed. Delimitation:Physical Store Integration: The project will not include the integration of the e-commerce platform with physical store inventory management systems. Offline inventory and online inventory will be separate. Mobile App Development: While the platform will be accessible through mobile browsers, the project will not involve the development of dedicated mobile apps for iOS or Android devices. International Shipping: Initially, the project will focus on domestic shipping only. International shipping and associated customs regulations will not be within the project scope. Complex Customizations: The initial project phase will not include highly complex changes to support particular business processes or industry-specific requirements. Marketing and SEO: While basic search engine optimization (SEO)

principles will be implemented, the project will not involve extensive marketing strategies or campaigns. Third-Party Integrations: Integrations with third-party services beyond core functionalities (e.g., social media sharing, advanced analytics) will be limited during the initial implementation. Data Migration: Data migration from existing systems, if applicable, will not be part of the initial scope. The focus will be on setting up the new platform. Ongoing Maintenance and Support: While the project will ensure a stable launch, ongoing maintenance, updates, and support considerations will be addressed separately.

CHAPTER II

Theoretical Framework

Comparing the pricing of products from several vendors in an effort to find the best bargain is what the customer refers to as comparison shopping. Customers can find the best bargain by comparing product prices from multiple sellers using online price comparison tools. Users can filter and compare products based on price, features, customer ratings, and other factors on a comparison shopping website, also known as a pricing comparison website, pricing analysis tool, comparison shopping agent, shop bot, aggregator, or comparison shopping engine.

Instead of actually selling anything, the bulk of comparison shopping websites compile product lists from other businesses. To make money, they rely on affiliate marketing contracts. Between 1995 and 2000, when these businesses were just getting started, comparison shopping websites offered services for online sellers and items, including ratings and reviews in addition to price comparison. In 1998 and 1999, a number of companies developed systems that retrieved prices from retailer websites and saved them in a centralized database. Once a product was looked up, consumers could examine a list of retailers and their prices. Rather than receiving payment for each click on a price, advertisers were paid for each click (Kwarteng, Jibril, Botha, & Osakwe, 2020; Wikipedia, 2023).

In actuality, the Internet's quick development has had a big impact on the expansion of online trading. It is projected that 2.14 billion people will use digital devices globally in 2021, and e-commerce revenue will surpass $4.9 trillion in that same year (Pasquali, 2023). Because online platforms are easily accessible,

customers can obtain price information much more easily than they can in physical stores.

If customers have complete and perfect knowledge, the growth of online retailers should also result in optimum competition and price convergence. Users can filter and compare products based on features, cost, reviews, and other factors on comparison shopping websites. It is also known as an aggregator, shop bot, comparison-shopping engine, pricing analysis tool, comparison shopping agent, or pricing comparison website. Studies have shown that information asymmetries can lessen market competition (Stiglitz, 2000; Akerlof, 1970). Customers could have problems obtaining store and product information in the online market, which is not an exception.

Numerous studies that looked at how information asymmetry affects the online market have refuted the notion of friction less search(Hossain & Morgan, 2006; Thomas Blake, 2021). Although consumers may now more readily search for pricing online, Ellison and Ellison (2009) argue that this also gives firms the chance to utilize search-complicating tactics. Online shops, for instance, could advertise lower-priced goods on sites that compare prices in order to get people, but advertise higher-quality goods with higher margins on their own websites.

According to Hossain, Morgan, and Blake et al., online merchants frequently withhold shipping costs until the final checkout page, raising the cost of comparison for customers. Online purchasing presents more risks in terms of product quality and service compared to traditional brick-and-mortar retailers. Consumers must pay more to compare items and stores as more businesses enter the market (Yuxin Chen, 2010; Kutlu, 2015).

Online shops, for instance, could advertise lower-priced goods on sites that compare prices in order to get people, but advertise higher-quality goods with higher margins on their own websites. According to Hossain, Morgan, and Blake et al., online merchants frequently withhold shipping costs until the final checkout page, raising the cost of comparison for customers. Online purchasing presents more risks in terms of product quality and service compared to traditional brick-and-mortar retailers. Consumers must pay more to compare items and stores as more businesses enter the market (Yuxin Chen, 2010; Kutlu, 2015).

The current study takes into account the results of earlier research as it examines the strategic methods employed by online sellers and looks into the connection between market competitiveness and the number of businesses.

The study demonstrates that although price information can be accessed for considerably less money in online markets, the cost of learning about products and businesses has not decreased and, in certain situations, may even go up due to retailers providing false or inconsistent information about their stock. Contrary to expectations from standard economic theory, according to the study, costs do not always go down as more businesses carry a particular good. There is a counterbalanced effect on prices from increased competition and higher customer search costs resulting from the entry of more businesses into the market. Competition is only evident when a market transitions from a monopoly to a duopoly. As more businesses compete for customers and consumer behavior becomes less predictable, the total amount spent on shopping increases. As a result, businesses usually raise their prices rather than lower them. Additionally, the research shows that price dispersion increases with the number of businesses, which is consistent

with both the search theory's predictions and the actual findings reported in the literature.

The study's findings are in line with past research that hypothesized certain frictions existed in the internet market's information structure and that information asymmetry may increase search difficulty and decrease competition. Examining the connection between market competitiveness and the quantity of businesses in the online market, the study (A. Chandra and M. Tapatta, 2011; Ali Hortaçsu, 2003; Maarten C. W. Janssen, 2004) highlights the importance of accounting for the cost of obtaining comprehensive and accurate product and store information.

2.1 REVIEW OF RELATED LITERATURES AND STUDIES

Literature Review

The concept of perfect knowledge and how it impacts competition is heavily stressed in traditional models of market competition. However, because information is so important in determining the level of competition, these models fall short of explaining the observed price dispersion in real-world markets (Stiglitz J. E., 1979). When there is an information asymmetry, markets are known to behave less competitively (Scitovsky, 1950; Salop, 1976). In markets where customers must pay a premium to learn about products, the search theory challenges the notion that having more firms will foster market competition (Diamond, 1971; Stiglitz, 1987).

The internet market, on the other hand, offers what appears to be the ideal information environment, making it simple to find out about prices and products by performing a quick search. However, there is continuous debate in the business and economic literature over the existence of frictionless internet search. (Yuxin Chen,

2010; Ellison, 2009; Kutlu, 2015). The number of businesses in a market influences the true level of competition when information gathering is minimal or nonexistent. Realized pricing and price dispersion reflect the effects of search costs and competition (Pereira, 2005). Many studies have been conducted on this topic in an effort to compare prices between online and offline sources as well as provide theoretical and practical justifications for the empirical data on prices and price dispersion.

Real-world markets show price dispersion because of information asymmetry, in contrast to classic models of competition that assume perfect knowledge. The cost of receiving information defines the actual level of competition on the internet market, and the convenience of obtaining information affects the market's level of competition. The ongoing debate over friction less internet search demonstrates how important it is to comprehend the role that information plays in market competition. Anyone interested in understanding the dynamics of online markets should read up on the literature on price comparison between online and offline sources as well as the theoretical foundations of online pricing dispersion.

2.2 CONCEPTS OF THE STUDY

Online and offline prices contrasted

Online markets, which differ from physical markets in their information architecture, have been made possible by the advancement of Internet technology. This information structure influences consumer search behavior, which influences price and price dispersion. Data collection and comparison are regarded as more difficult for physical marketplaces than for online ones. Achieving the economic goal

of perfect competition is also predicted to be made possible by the Internet's lower entry and transaction costs (Joseph Alba et al., 1997) (Bakos, 1997).

In numerous empirical studies, reduced online costs in a variety of markets have been discovered. The first study of online prices was conducted by Brynjolfsson and Smith (Erik Brynjolfsson, 2000), who discovered significantly reduced costs for books and CDs there. In the insurance markets, Brown and Goolsbee backed up these findings. For the sale of cars, Pan et al. discovered lower prices, Orlov for the airline industry, and Morton et al. for desktop and laptop computers. The contrary, according to several investigations, is the case. When looking at a selection of books, CDs, and software, Bailey discovered greater costs online. In their investigation of books and software, Pan et al. also discovered comparable outcomes. Shankar and Ancarani assert that in Italy, the cost of books and CDs is higher online than it is in physical bookstores. Erevelles et al. discovered discrepancies in vitamin prices online. Still, several studies have failed to find a pricing difference between offline and online sources. For instance, Clay et al. found no discernible price difference between books purchased online and offline. Pan et al. discovered comparable prices for PDAs and electrical devices, both offline and online.

A direct comparison of online and offline costs yields contradictory empirical evidence, which is consistent with the significant degree of price variation that exists on the internet. Support for this is provided by the price dispersion online results. There is notable and continuous online price dispersion in the book, airline, and electronic device markets, as supported by empirical data from Eric Brynjolfsson (2000), Eric K. Clemons (2002), Orlov (2011), and Michael R. Baye (2004). Pathak found that there is price dispersion even with online comparisonshopping services

due to these businesses' shoddy selection practices and sluggish information updates.

The growth of online technology has made it easier for online markets to grow, but it is yet uncertain how these markets will affect prices and price dispersion. While some studies show lower prices online, others either show the opposite or show little to no difference in prices between online and offline. It appears that there might not be the best kind of competition in the online market, despite the fact that numerous marketplaces have shown a sizable amount of online price variation. Further study is necessary to completely comprehend the reasons behind and remedies for online price dispersion.

An explanation of empirical data on online prices and price distribution

The development of online markets, which are different from physical markets in their information structure and have an impact on consumer search behavior and pricing, has been made possible by the development of Internet technologies. Online information comparison is thought to be simpler, leading to a convergence toward perfect competition because entrance and transaction costs are lower. Cost savings should result in lower pricing and less price variation (Joseph Alba, 1997; Bakos, 1997). However, conflicting results have been found in several empirical investigations on online pricing(Bailey, 1998; Jeffrey R. Brown, 2002) (Erik Brynjolfsson Friction less Commerce? A Comparison of Internet and Conventional Retailers Erik Brynjolfsson and Michael D. Smith 2000), (Xing Pan, 2002), and (Joan Lindsey-Mullikin, 2006). Some have reported lower prices than those found in physical stores, while others have found higher prices.

Researchers have looked into why price dispersion continues despite the nearly free nature of internet price comparison to understand the causes behind these contradictory results. There have been two basic hypotheses put forth. First off, online merchants might offer distinct service traits that are not noted by academics (Joan Lindsey-Mullikin, 2006). Disparities in pricing between retailers could be explained by differences in service quality, such as shipping and return policies, delivery schedules, and customer support responses, since better service providers establish strong brands and charge more(Joel Waldfogel, 2006). Thus, unobservevariability in seller attributes, like service quality, reputation, and consumer awareness, could account for the continuous online price dispersion as well as the inconsistent results between online and offline prices (Sulin Ba, 2012; Pathak, 2012).

Another justification for the longevity of online pricing dispersion is the strategic behavior of sellers, which may have an effect on consumer search costs. Online businesses are able to change prices more swiftly and affordably than conventional brick-and-mortar stores (Jean Boivin, 2012; Alberto Cavallo, 2014; Pathak B., 2010). The cost of a price search is increased because consumers may find it costly to compare prices because online vendors often alter their offerings.
The quantity of product and service information that retailers provide may have an impact on the price that customers pay to locate the best product and retailer.

Customers may easily compare the service quality of various businesses online; however, Joan Lindsey-Mullikin (2006) points out that it is sometimes difficult to find information regarding product attributes. As a result, the two hypotheses put forth—unobserved variability in seller traits and sellers' strategic behavior—might aid in our understanding of why price dispersion persists in online markets. Despite

conflicting results regarding online prices, studies have repeatedly found significant online price dispersion in markets like books, airline tickets, and electronic products (Michael R. Baye, 2004; Erik Brynjolfsson, 2000; Eric. K. Clemons, 2002; Orlov, 2011), demonstrating that while price searching online may be more convenient, it is not always free.

Online price comparison is a difficult procedure that involves more than just comparing the prices of a single product from many dealers. Consumers must also consider a number of factors before making a purchase, such as the goods' quality, the service, the store's reputation, and any additional costs. Online retailers now have the chance to affect consumer information and the search process as a result of this multi-dimensional evaluation (Kirk L. Wakefield, 1993). According to Lynch and Ariely, the presence of pricing information on the first page affects consumer behavior. When price information is readily available, consumers frequently focus solely on price; however, if consumers must click through several links to view pricing information, other product features take on greater significance.

Despite the widespread belief that there are no costs associated with online price comparisons, consumers do incur costs when they click more than once (Yuriy Gorodnichenko, 2018; Honka, 2014; Babur De Los Santos, 2012). Retailers may be able to influence consumers' search behaviors and how they receive and process information by causing them to quit searching earlier than they had anticipated. For example, as demonstrated by Morgan and Hossain (2006) and Blake et al. (2021), Shipping costs are typically withheld by retailers until the very end of the purchasing process, making it challenging for customers to compare prices impartially.Dinersteiet

al. speculate that sellers' markup may be influenced by the way eBay's search algorithm is designed.

Therefore, it's critical to understand that internet pricing searches involve more than just price comparison. Retailers have the power to influence how consumers search and how they receive and process information. By concealing information, such as shipping costs or requiring extra clicks to obtain pricing information, retailers can impact consumer behavior and potentially increase their profits. Understanding these factors is crucial for both researchers and consumers to make informed decisions when conducting online price searches.

Factors Affecting Online Price: Lower Overhead Costs: Compared to traditional merchants, online ones frequently have fewer overhead expenses. They don't need to maintain physical storefronts, pay for utilities, or employ as many staff members, which can lead to cost savings that are sometimes reflected in lower online prices. Direct-to-Consumer (DTC) Model: Many internet retailers use a direct-to-consumer business strategy, which cuts out middlemen and enables them to sell goods for less than they would at a physical store.E-commerce platforms can instantly adjust prices in response to shifts in the market, competitors, and customer behavior because of dynamic pricing algorithms. This may lead to more frequent price adjustments and potential reductions. Geographical Expansion: Online sellers may be able to reduce prices because they can distribute their fixed costs over a larger customer base and supply in larger quantities. Because it is easier for customers to compare prices across multiple online retailers, businesses are encouraged to offer attractive prices in order to attract customers.

Empirical findings on online pricing and price dispersion provide insights into how prices vary across different online retailers for the same product or service.These results provide insight into consumer behavior, market dynamics, and pricing tactics used by internet businesses. Let's examine the justification for these empirical discoveries.

Price Dispersion Price dispersion is the phenomenon where multiple vendors or retailers in the online marketplace offer the same goods at various rates. This is evident in a variety of sectors and product categories. Empirical research has shown several factors that contribute to price dispersion. Competitive Landscape: A highly competitive market can lead to more price dispersion as retailers try to undercut each other to attract customers.

Search Costs: Consumers might not be aware of all available prices, and sellers with lower prices might attract less attention. This lack of perfect information allows for price differences to exist. Brand Perception: Established or premium brands may charge higher prices, leading to price dispersion as other sellers offer similar products at lower prices. Differentiation: Sellers might differentiate themselves based on factors like shipping speed, customer service, or return policies, which can justify price differences. Consumer Heterogeneity: Based on their tastes, needs, and readiness to pay, different consumers are prepared to pay varying rates for the same goods. Online price variation: Online price variation describes shifts in a product's price over time, often even in a matter of seconds. Empirical findings on online price variation highlight several contributing factors.Online merchants frequently employ dynamic pricing algorithms, which modify prices in response to real-time data such as demand, rival prices, and inventory levels.

Geographical Differences: Depending on the customer's location, online shops may change their prices to reflect local purchasing power or competitive environments. Seasonal and Event-Based Changes: Prices can change during peak shopping seasons, holidays, or special sales events to take advantage of increased demand.

Supply and Demand Fluctuations: Rapid changes in supply and demand, as well as changes in market conditions, can lead to price variations.A/B Testing: Retailers might experiment with different price points to gauge customer reactions and optimize revenue. Consumer Behavior: Empirical findings related to consumer behavior provide insights into how shoppers respond to online price dispersion and variation. Price Sensitivity: Consumers may choose the lowest available price, leading to competition among retailers to offer the best deal. Search Behavior: Consumers often engage in comparison shopping, visiting multiple websites before making a purchase decision, leading to increased price dispersion awareness. Trust and familiarity:consumers might prefer to buy from familiar or reputable online retailers, even if the prices are slightly higher, contributing to price dispersion. In conclusion, empirical findings on online price dispersion and variation highlight the complex interplay of factors that influence how prices are set and how consumers respond to these prices in the dynamic online marketplace.

While consumers' price sensitivity and search behavior also play a role in the observed pricing trends, retailers use a variety of techniques to draw customers and maximize income.When it comes to e-commerce and internet shopping, online price and price dispersion are significant economic phenomena that have been thoroughly explored. Empirical findings in this area can be explained by various factors and

theories. I'll summarize these empirical results and some of the main theories behind them here:

The disparity in prices among various platforms or vendors for the same or similar products is known as price dispersion. big Price Variations: Online markets frequently show a wide range in prices for the same or nearly the same goods. One empirical discovery on price dispersion in online markets is this one. Price Variations Over Time: Online markets are subject to frequent price variations. For instance, a product may be offered on one website at a price that is significantly higher or lower than on another. Numerous variables, such as shifts in supply chain dynamics, demand, and competition, can contribute to these oscillations.

Price disparity can be explained by market competitiveness, which is dependent on how competitive the internet market is. Higher competition tends to lead to narrower price dispersion as sellers compete to offer the lowest prices to attract customers. Seller Heterogeneity: Sellers on online platforms may have different cost structures, strategies, and pricing models. This heterogeneity can lead to price dispersion as each seller sets prices based on their own considerations. Search Costs. Buyers may not always find the lowest prices due to search costs, which can include the time and effort required to search multiple websites. Online Price Trends. Online prices are subject to various trends and patterns. For convenience, some customers might be willing to pay a little bit more. Empirical discoveries in this field include: Online shops usually utilize dynamic pricing algorithms to adjust rates on the fly based on customer profiles, competition prices, and demand. Pricing discrimination refers to the practice of charging different customers different amounts according to their location, browsing history, or other

personal information. Seasonal Variation: Prices for certain goods may vary depending on the season, with deals and discounts offered during holidays and other events.

Explanations for Online Price Trends Dynamic Pricing Algorithms: The use of algorithms and data analysis allows online retailers to respond quickly to changing market conditions, optimizing prices to maximize profits. Personalization: Online platforms can track user behavior and preferences to tailor prices to individual customers, potentially maximizing revenue while staying competitive. Supply and Demand: Seasonal price variations are often driven by changes in supply and demand. Retailers could decrease prices throughout the Christmas season or raise prices during periods of high demand. As a result of complicated interactions between a number of market dynamics, such as competition, seller heterogeneity, technology, and customer behavior, empirical data on online price and price dispersion are the outcome. Data analysis and taking into account the economic and technological forces that drive online markets are both necessary for comprehending these phenomena.

In order to analyze pricing patterns, trends, and variations of goods and services, empirical data on online prices and price distribution refers to real-world information gathered from online retail websites and platforms. Such information is frequently gathered by businesses and researchers to learn more about how prices are determined and distributed in online markets. Here is a description of a few typical features of empirical information on online prices and price distribution: Price Distribution: Data on price distribution reveals how a specific product or category's

prices are distributed among various price points. Histograms, box plots, and other graphical displays of this data are frequently used to present it.

If products exhibit significant price variations, tend to cluster around particular price points, or adhere to particular pricing strategies (such as premium pricing, economy pricing), the analysis of price distribution can show these trends. Price Fluctuations and Trends: Researchers gather data over time to spot trends in online prices. They can monitor how prices fluctuate in response to market forces, seasonality, and holidays. Businesses can use this information to inform strategic decisions such as when to run promotions or how to change prices to stay competitive. The degree to which prices for the same product vary among various online retailers is measured by price dispersion data. There may be more opportunities for consumers to find lower prices if there is greater price dispersion. This information can be used by retailers and customers to evaluate market competition and spot areas for cost-saving opportunities. Data on price elasticity examines how changes in price impact a product's demand. It gauges how responsive customer demand is to fluctuations in price. Consumers can use price elasticity data to assess the influence of price changes on their purchasing decisions, while retailers can use it to determine the best prices to maximize revenue. Retailers routinely compile data regarding the rates established by their competitors.

With this information, they can adjust their own pricing strategies to remain competitive in the market. You can also use competitor price information to identify pricing trends and comprehend market dynamics. Data from empirical studies can show how dynamic pricing strategies are used by online retailers. This includes modifying costs in response to variables like demand, client location, hour of the day,

or inventory levels. Businesses can use this information to evaluate the efficacy of dynamic pricing and adjust their strategies as necessary.

To spot patterns or anomalies, researchers and retailers may examine price data from various product categories or types. For instance, they might contrast the price distribution of clothing and electronics. Pricing models, promotional tactics, and product assortment decisions can all benefit from such analysis. Price information can help us understand how consumers react to promotions and price changes. Retailers, researchers, and policymakers can all benefit from empirical information on online prices and price distribution. Retailers can tailor their pricing and marketing tactics using this data in an effort to influence customer behavior. It can affect pricing strategies, increase market rivalry, and provide customers with greater choice and clarity. Several techniques are used to gather this data, including web scraping, API access and data mining methods. The terms and conditions of the websites from which the data is collected, as well as legal and ethical guidelines, must all be complied with when collecting data.

From a Retailers Point of View

Prior to the epidemic, having an internet presence was essential for businesses to promote their goods and connect with consumers. But as the pandemic spread across the globe, this requirement changed, spurring the creation of online communities and virtual linkages to support the real world. Being online is essential for creating a brand, increasing trust, and managing reputation now more than ever. [2]. The two most important online marketing techniques are local search optimization, also known as local SEO, and having a company website. By keeping

up a correct local business listing and adding it to important online directories and search engines like Google, a company may increase its online visibility. Known directories such as Yelp, Facebook, LinkedIn, Foursquare, MapQuest, Waze, and the Yellow Pages should also be considered in addition to Google. Unfortunately, most small business owners do not have access to the resources and technological know-how needed to establish an online presence.

To address these issues and attract more businesses to publish on the platform, a simple yet innovative web-based system has been proposed to make it easier and more economical for stores to advertise their product prices. It will eventually be regarded as one of the greatest tools for online marketing and competitiveness as a result.

With its simple interface and user-friendly design, the new system is made to make it simple and quick for businesses to publicize the prices of their items. Additionally, it is affordable, making it available to small business owners who may lack the funds to purchase expensive marketing tools. The system would compile and examine the information made public by firms and offer insightful information on market trends and rivals' pricing.

Businesses may enhance their internet presence and raise their chances of being found by potential clients by putting this strategy into place. In the end, this will aid in boosting their brand's legitimacy and visibility in the online world, resulting in better success and growth.

Cost Structure: Retailers must take into account their overall costs, which may include expenses for utilities, rent, employee salaries, and inventory. Between online and offline operations, these costs may vary. For instance, because they don't have

physical stores, online retailers might have lower overhead costs. Retailers must decide on their overarching pricing strategy. This tactic might entail maximizing profit margins, engaging in price competition, or charging more for goods or services of higher quality.

Pricing can vary both online and offline, depending on the strategy used. Retailers need to assess how competitive the offline and online markets are. This entails keeping an eye on the prices set by rivals, spotting market trends, and comprehending customer preferences. Pricing for Each Channel Depending on factors like demand, competition, and operational costs, retailers may make different price adjustments for online and offline channels. Online prices, for instance, may be more flexible and responsive to changing market conditions. Discounts and Promotions Retailers frequently use sales events, discounts, and promotions to draw customers. Between online and offline sales channels, these promotions may differ in type and frequency. Online retailers must take shipping and handling costs into account when determining prices. Depending on the shipping method, the destination, and the package size, these costs may change. Retailers should set up clear return procedures for both in-store and online purchases. These regulations may affect pricing choices as well as customer purchasing and return behavior. Retailers can use customer data and analytics to better understand both online and offline consumer behavior. This information can help with product assortment, marketing strategy optimization, and pricing decisions.

Retailers are required to abide by pertinent laws and regulations, which can differ between online and offline sales. This covers the collection of sales taxes, consumer protection laws, and data privacy laws. Customer Experience Retailers

must take into account the total customer experience across all channels, including online and offline ones. This includes things like the speed of fulfillment, the caliber of the customer service, and the ambiance of the store. Omnichannel Strategy: Some retailers may use an omnichannel strategy if they run both online and physical stores. To create a seamless shopping experience, this entails synchronizing prices, inventory, and promotions across all channels. Market Expansion: By entering new geographic regions or focusing on various customer segments, retailers may think about extending their market reach. This growth may have an impact on pricing policies for both online and offline sales.

About Web Scraping

(1) Definition

Although several related terms and concepts have been found to be ill-defined, the term web scrapingis used in this study according to Massimino (2016), Kielle et al. (2004), and Nakash et al. (2015). Automated access to web documents, along with the collection, organization, and storage of specific pre-defined data (such as price) from each is known as web scraping. Screen scraping and information scraping are other names for it. Web crawling, on the other hand, comprises obtaining content from the internet and indexing it through links. This demonstrates that only the URL is collected in terms of specific data. The information can be accessed in its entirety via the URL; however, it is normally not archived. By evaluating the information available and combining all of the results, search engines like Google scour the web for pertinent links. Crawlers and spiders are also used as measuring devices for price contrast. Because the authors are solely interested in

collecting information on food prices from publicly accessible websites for research and because Table I makes a conceptual distinction between scraping and crawling, this work exclusively focuses on online scraping. According to Hemenway and Calishain (2004), Shopbots are computer programs that scour websites in quest of the best deal by compiling pricing information from numerous providers.

There is no one technique that must be used; instead, there are numerous approaches that can be used to create web scrapers. While this paper does not aim to offer comprehensive guidance on building one, we will briefly describe the operation of a web scraper and its potential applications in producing food pricing data sets. It is frequently necessary to use a script to gain access to the websites where the data is kept, find the pertinent pre-defined portions, and then extract and store them systematically. It is essential to save the price, the name of the item, and a timestamp showing when the content was viewed in order to provide a consistent output across time. Extra data, like product size and customer rating, can also be preserved if it is available and necessary (Walden et al., 2017).

Table 1: Web crawling and web scraping have different purposes.

	Web Scraping	Web Crawling
Process	automatically locating documents on the internet and extracting information from them	Finding and obtaining links repeatedly by starting with a list of starting URLs
Target Information	Predefined information on particular websites	URLs that, depending on the search term, can lead

		to various types of information
Output	Structured data that was downloaded	indexed links kept in a database
Use	Data gathering, such as price series	indexed links kept in a database

A web scraper can be created in a variety of ways, and there is probably no one method that works for everyone (Papin et al., 2018). While creating detailed instructions on how to design a web scraper is not the aim of this work, we will briefly discuss the technical aspects of this tool and how food pricing data sets can be created using them.

To find the pertinent, previously identified bits, search the websites containing the data, download, save, and retrieve them as structured data sets from the websites, a script is usually needed. The purchase price, the product's brand, and the date the data was accessed must all be recorded in order to ensure the same result over time. Upon request and availability, further information about customer rating, labels, category, country of origin, package size, etc., can be given. Making note of the URL and the distinct product ID can help you find any anomalies. (Papin et al., 2018).

To retrieve the pre-defined information, the script essentially navigates the websites as a genuine online user would. This download should occur with a slight delay between requests in order to maintain a moderate level of website traffic

(Hemenway & Calishain, 2004). A graphic breakdown of the many phases involved in creating a web scraper is provided in Figure 1.

(2) Technical Procedure

The majority of popular programming languages can technically be used to code the script (Massimo, 2016; Hemenway and Calishain, 2004 For this reason, there are also pre-programmed libraries in many different languages, including Python or R. It is always required to adjust the script to the desired website, taking into account the configuration, the page layout, any necessary authentication, etc., despite the fact that such code parts are obviously valuable as building blocks. Simple HTTP requests cannot be used due to the prevalence of reactive elements on webpages. The lack of an API prevents consumers from quickly accessing the food pricing of large online food retailers, who frequently offer online catalogues (Papin et al., 2018). In order to navigate the webpage, the script must "click" using a "real" web browser.

Allow enough time for testing and troubleshooting as well. A scheduler can initiate the download at predetermined intervals to automate the entire script execution process after it has been written and successfully verified. But if a website modifies even a little bit in terms of design, selection of products, or other elements, the scraper might stop working. If a script is not completed or if the download size is surprisingly small, an alarm could be added to send an email, for example (Papin et al., 2018). Any format containing specified elements (name, ID, price, timestamp, etc.) can be used to save the gathered data. such as text or CSV files. Every new observation made over time merely adds new information to this file, producing a

single, reliable data set. Separate files may be created if multiple websites are scraped (for example, multiple stores or regions), but formatting should be consistent to make analysis easier (Papin et al., 2018).

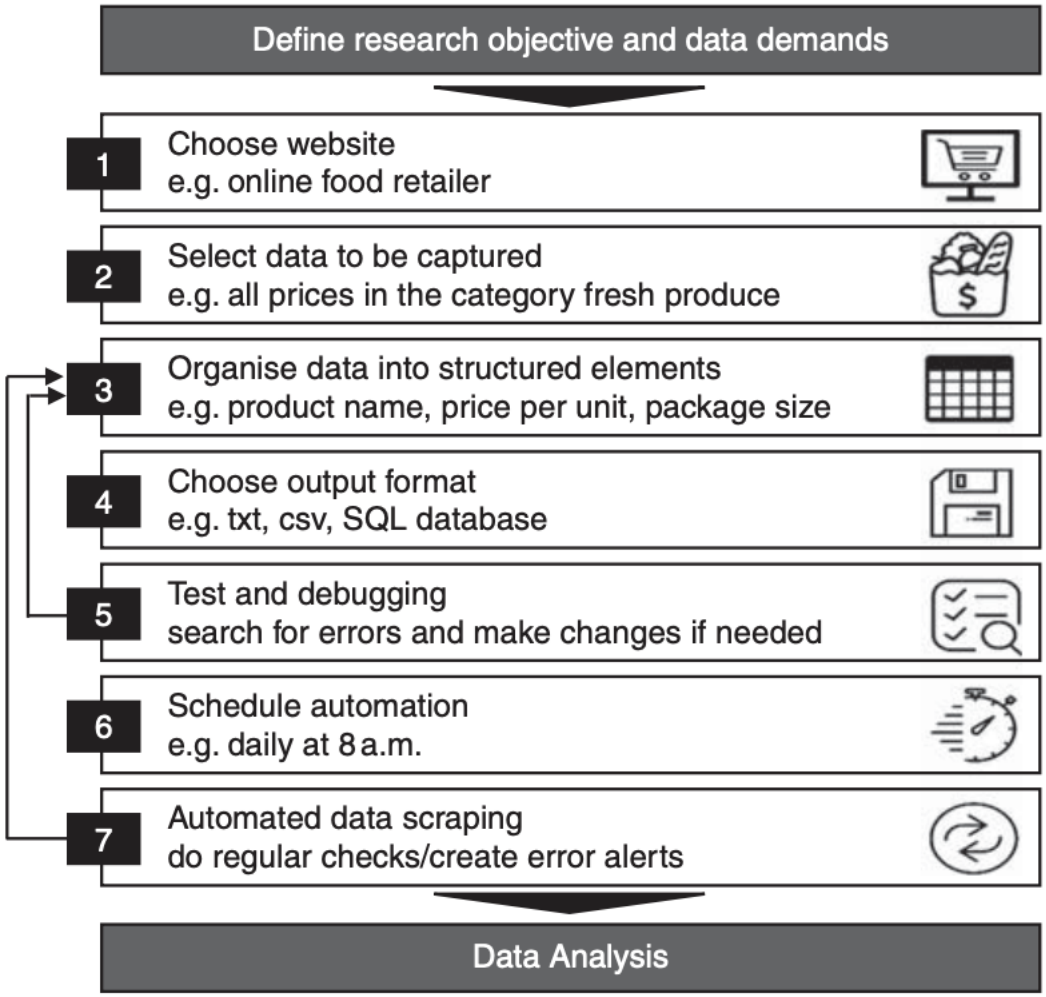

***Figure 1**. Diagram of the web scraping process*

In many cases, daily web scraping is required for various applications. However, it may not be desirable to have this task occupying a personal computer.An alternative is to use specialized single-board or private server computers like the Raspberry Pi. (Papin et al., 2018). This would enable the scraping process to proceed without being impeded by regular computer use.

Commercial online scraping service providers can be used by people who lack the technical resources or coding expertise necessary to create their own web scraping solutions. It is important to keep in mind that these commercial choices might not provide the amount of opensource documentation and transparency necessary for scientific research and peer-reviewed publications (Papin et al., 2018).

In certain circumstances, using opensource libraries that may be customized to match particular requirements might be more appropriate.By utilizing open source libraries, researchers can access the underlying code and make any necessary modifications to it to meet their unique needs. In contrast, the underlying code used by commercial web scraping services is frequently proprietary and not publicly available. Additionally, making use of open source libraries can help to guarantee that the research is transparent and reproducible, both of which are crucial for scientific research and peer-reviewed publications (Papin et al., 2018).

In conclusion, although personal computers can be used for web scraping, there are also alternativessuch as dedicated single-board PCs or private servers. Commercial online scraping service providers can be used by people who lack the technical resources or coding expertise to create their own web scraping solutions. It is crucial to remember that these commercial choices might not provide the open-source documentation and transparency needed for scholarly research and peer-reviewed publications. Because opensource libraries allow for openness, repeatability, and the freedom to adapt the code to suit particular needs, using them is frequently the preferred choice.

(3) Advantages

Boulanger et al. (2020) claim that web scraping is a helpful method for learning about food prices because it solves a lot of issues with traditional data sources, like government statistics and scanner data.

Low Cost

The cheap cost of data collection is one of web scrapings key benefits. It can be costly to obtain regularly occurring dis-aggregated data, especially for retail purposes. Although writing and testing the code, as well as the cost of electricity and an internet connection, are additional costs, web scraping with open source software is essentially free. A negligible marginal cost per observation counteracts the high scalability across nations and products (Cavallo, 2018).If funding is available, interacting with for-profit companies that offer data as a service would be an additional option (Massimino, 2016).

Frequent Real-Time Sampling

Another advantage of web scraping is the regular, in-the-moment sampling of data it produces. Depending on the study problem, the user can select the data collection frequency after the script has been created, which can vary from hourly to monthly. Daily data should be sufficient for the majority of food price research applications because they allow for a more thorough analysis of price dynamics and the application of additional statistical techniques. (Edelman, 2012). Real-time data collection not only makes information instantly available but also makes it possible to analyze recent occurrences or changes in policy, which is useful for forecasting.

Product Range Details

Not to mention, online scraping makes it possible to collect thorough product data that is normally missing from official statistics.Official consumer price statistics typically include prices for products that fall into specific categories or commodities; however, they sometimes leave out other details like package size, brand, and variance in quality. On the other hand, prices for particular products as well as all available product attributes can be obtained through web scraping. Some study areas might find this information useful. If not, the researchers themselves can use the techniques they deem appropriate to aggregate the data to the necessary level.

Because it has the ability to overcome some of the drawbacks of conventional data sources, such as cost, frequency of data collection, and degree of product detail, web scraping is a valuable technique for obtaining information on food costs. Web scraping has some costs, such as the time spent writing and testing the code, electricity, and online access, but because of its scalability and adaptability, it is a popular choice among academics. Researchers that demand open source documentation and transparency for their scientific research and peer-reviewed publications find web scraping particularly appealing.

Table 2: Alternative Data Comparison - Notes: Source: Own representation based on publicly available price indices from Rigobon and Cavallo (2016, p. 156) that generally takes into account download time and any errors.

	Scraped Data	Scanner Data	National Statistics Data
Cost per observation	Low	High	Free[b]
Data frequency	Daily	Weekly	Monthly
Real-time data	Yes[c]	No	No
Full product range	Yes[c]	No	No
Product details	Yes	Yes	Limited
International comparability	Yes	Limited	Limited
Transaction data	No	Yes	Yes

Sort type

The term "retail level" is frequently used to describe the type of store from which secondary data is acquired, with infrequent exceptions made for supermarkets, discounters, and small single retailers. Instead of using a masked aggregate, web scraping enables researchers to selectively collect pricing from specific stores, merchants, wholesalers, or online delivery providers. In order to collect data from different areas and countries, self-identification may be necessary. This can be accomplished by doing things like selecting a country or supplying a postal code that

can be put into the web scraper code. a comparison of the methods used by several online pricing indexes to collect prices.

Transparency and Customizing

Government statistics and scanner data are generally reliable and well-organized, but they don't always include all the details that researchers require for their work. Comparing data across borders may be challenging because these data may only be partially aggregated, may not come from the same source, or may not be available in the same format for all relevant geographic areas.

Web scraping, on the other hand, enables the creation of specialized datasets that are tailored to specific needs and transparent regarding the data's collection process, with no omitted variables or black boxes. The scientific community can benefit from this transparency and repeatability if open-source software is used and shared. By allowing researchers to collect their own data as part of their work rather than relying solely on secondary data, empirical research can be conducted with more efficacy and precision (Cavallo & Rigobon, 2016). web scraping for conducting commercial research.

(4) Limitations

Web scraping has a lot of benefits, but there are some drawbacks as well. The absence of historical data accessible through this method is a key drawback. Since online scraping concentrates on gathering real-time data, it is required to start collecting data on the first day of the pertinent time period in order to produce a lengthy time series. Ad hoc analyses, however, could not always make this possible.

This disadvantage highlights the significance of thorough planning and preparation for web scraping data acquisition.

Too Big Data

Given the low marginal cost and effort needed to scrape data, it might be alluring to gather much more information than is required to fully address a given research question. Instead of conceptually motivated research, this may lead to purely exploratory data mining (Massimino, 2016). Large data sets are now easier to handle thanks to developments in processing and storage, but evaluating the data is still difficult, especially for professionals and researchers who are not familiar with big data. Therefore, it is imperative to carefully consider what is sufficient for the planned research when thinking about web scraping, including time duration, product scope, frequency, and information level. For instance, when analyzing the general trend of food prices in a nation or region, the research process significantly depends on the selection of representative product categories and retailers.

No Instruction Data

Web scraping typically only gathers pricing data from websites; as this data is confidential, it does not disclose the frequency of clicks or purchases. The absence of transaction data poses a significant problem since it is probable that high prices willdeter customers from purchasing products, rendering them meaningless to them(Goolsbee and Chevalier, 2003). Online markets, like any other market, are likely to have a large selection of infrequently purchased items in addition to hot sellers. Users also tend to buy fewer products or brands with fewer followers less

frequently. Transaction data is necessary to evaluate this connection, which is sometimes referred to as double danger. (Ehrenberg et al., 1990). Unweighted prices, however, were identical to price quotes weighted by clicks, according to research by Gorodnichenko et al. (2018). If accessible on a particular website, sorting by "bestsellers" or "most popular" categories can give some indication of frequently bought products.

Online Availability

Online pricing data may not be standardized or easily accessible for the majority of transactions, even though there are more prices available online. This is especially true in developing countries where there are many unregulated markets. Official statistics may not be as trustworthy as the scant amount of online price data available for developing nations (Cavallo, 2013). Globally, online pricing is becoming more widely available. The Asia-Pacific region is leading the online grocery shopping market, and Chinese consumers are embracing this trend rapidly (Somogyi and Wang, 2018; Nielsen, 2017). With rising rates of smartphone and internet penetration, particularly in cities, developing markets are catching up globally. (Nielsen, 2017).

Legal and Ethical Limitations

Legal and moral restrictions. Web scraping is neither inherently illegal nor legal because it is a technology. Instead, for each particular application, one must carefully evaluate the legal circumstances. Naturally, only content that is open to the public should be viewed, and copyright laws must be observed. You must abide by

the relevant Terms of Service or Terms of Use in order to download and use third-party data. Kienle et al. (2004) advise that someone read the Robots Exclusion Protocol and the Terms of Use. (robots.txt file) to ascertain whether or not data scraping from a website is allowed. The robots.txt file, which may be read via the URL, includes an access policy that is consistently and clearly stated. http://[www.domain.com]/robots.txt (Calishain and Hemenway, 2004).Users may automatically accept the Terms of Use even though they are still listed somewhere on the website. (referred to as a "browse-wrap agreement"). Depending on the situation, courts have examined whether they are enforceable and legally binding contracts, with various outcomes (Toto and Buffington, 2016).

Although web scraping has grown in popularity, there are still legal considerations to be made, and this area of law is constantly changing. The legal assessment frequently hinges on the justifications for online scraping and the intended use of the collected data (Hirschey, 2014; Zhu and Madnick, 2010). As a result, it may be wise to get legal counsel before starting a research project to steer clear of any potential conflicts with the law.

Despite the technical advantages of online scraping, ethical issues must be discussed because they cannot be dismissed. Numerous sizable websites need registration and come with scraper blocking software, Papin et al. (2018) give a method for web scraping that makes use of the TOR network and a headless browser. (Dingledine et al., 2004) to remain anonymous and circumvent website security measures. It does, however, bring up ethical questions about whether or not these techniques belong in publicly funded scientific studies. Asking the website

owner directly for their permission would be the best course of action. Thus, when doing research on internet scraping, it is imperative to consider ethical issues.

As it continues to grow, the Internet is quickly emerging as one of the primary global sources of open data. Internet users publish and manage several petabytes of data every day using a variety of file formats, including CSV (Comma Separated Values), PDF (Portable Document Format), XML (Extensible Markup Language), and HTML (HyperText Markup Language), documents [1] due to the rapid advancement of technology. The concept of managing large volumes of generated data is known as "big data," and it is defined by volume, variety, and velocity, or the "3Vs." [2] This massive amount of data presents a significant challenge to businesses and academics interested in data mining and data analysis issues because of its volume, variety, and velocity. A Structure for Web Scraping to Enable Descriptive Evaluation of Meteorological Big Data to Assist in Making Decisions [1] [3]. Obtaining, preserving, and evaluating vast volumes of sparse, unstructured, and heterogeneous data in a timely way for use in prediction and decision support presents a number of difficulties for researchers. Here, "big data" refers to an abstract idea [4] that emerged to characterize the vast amount of created data that is often inconsistent and disorganized in the world. Globally, researchers and decision-makers are using big data more and more frequently. [2]. In recent years, big data has become increasingly prevalent in business and public policy decisions. [5]. Furthermore, big data statistics are used by many organizations in a variety of industries, including transportation, to inform their decisions. [6], education [8], health [7], agriculture [9], etc.

In order to benefit from big data, Morocco has recently started a number of projects. Despite the efforts, big data in Morocco is still in its infancy. [10][11]. Because of this, the phrase "big data" is now frequently used to describe a range of big data tools and techniques that make it easier to gather data and transform it into structured data. A big data strategy with a lot of potential is web scraping, which is one of the primary technologies used to access the data. Web scraping is one modern data access method that is becoming more and more popular [12] [13]. It focuses on transforming unstructured data from the internet into structured data that can be saved and utilized in a central spreadsheet or local database for analysis. [14]. One big data technique with a lot of potential for automatically extracting data from online sources is web scraping. Large datasets may be produced through big data scraping, but there are also great opportunities for data analysis and visualization that can help with forecasting and decision-making. In this context, a number of research projects have been realized that use web scraping for decision support.

For instance, we highlight how simple it is to make decisions regarding urban design. [15], the assistance in making decisions about tourism [16], the assistance in adapting to climate change [17], etc. For this reason, web scraping is one of the main methods for obtaining unstructured data from the Internet. [18] This paper is especially interested in web scraping as it relates to big data. The aim is to make it feasible to gather and organize large amounts of meteorological data from internet sources, present it in an easier-to-read format for analysis, and use it for forecasting and decision-making. This weather data can help with both large-scale emergency responses (medium and long-term) and basic daily decisions about various needs

(short-term). For instance, this data is frequently utilized as a decision-support tool in the near future. Those with an interest in weather-dependent socioeconomic pursuits and farmers who meticulously schedule their daily tasks in accordance with weather forecasts could be among the beneficiaries. (for example, monitoring precipitation before deciding to water), those who plan their weekend activities based on weather forecasts, or those who want advice on what to wear (sunshine, temperature, or precipitation), and so forth. These dynamic and statistical data can be used to determine the level of vigilance needed for medium- and long-term planning, as well as to characterize the risks associated with hazardous events.

The study's recommendation was to develop a web scraping framework in order to facilitate the descriptive analysis of large-scale meteorological data sets. The project intends to provide decision support for weather forecasts and various prediction capabilities in addition to providing meteorological data analysis (of the weather-related variables mentioned above, like humidity, temperature, and precipitation). Restoring data that has been made available as statistical models in a dashboard is made feasible by the suggested framework. Data extraction, data archiving in a data warehouse, and data presentation and description are the procedures involved in web scraping.

The feasibility of the proposed framework was assessed by extracting meteorological data from the internet using a web scraping system and presenting it in an easily readable format for decision support. The essay is organized as follows for the remaining portion: the second section discusses pertinent literature and contains a few recent studies on big data concepts, web scraping tools, and related topics.

The third section examines the web-scarping framework's planning and modeling procedures, which are utilized in the descriptive analysis of large-scale meteorological data. The usage of the web scraping framework in a meteorological context is also covered in this section.

The results are presented and discussed inthe following two sections. The final section concludes the research and lists upcoming projects.

Web Scraping Techniques

Occasionally, the most efficient and practical webscraping methods are copy and paste and manual inspection. There are two easy and effective ways to extract data from web pages: regular expressions and text grabs. However, this method is painful and time-consuming. This technique depends on the ability of programming languages or UNIX commands to match regular expressions. ("Web scraping," 2015b). C. Programming with the Hypertext Transfer Protocol (HTTP): Information can be taken from both static and dynamic web pages using this technique. Posting HTTP requests and retrieving data from a remote web server are both possible with socket programming. (Source: "Web scraping," 2015b). ("Web scraping," 2015b). C. Programming with the Hypertext Transfer Protocol (HTTP): It is possible to extract data from both static and dynamic web pages using this method. Sending and receiving data from a remote web server is possible with socket programming. Client-side scripts produce dynamic content that programs can access by utilizing feature-rich web browsers such as Mozilla Browser Control or Internet Explorer. When web pages are parsed (a process known as "web scraping," 2015b), these browser controls create a DOM tree that programs can use to retrieve portions of the web

pages. H. Applications for Web Scraping: A multitude of software tools can be used to create creative solutions for web scraping. Apart from functions to record web pages, database interfaces to store the scraped data in local databases, and some scripting functions for content extraction and transformation, this software may try to automatically identify a page's data structure. ("Web scraping," 2015b). G. Systems of vertical aggregation: A number of companies have developed harvesting systems, especially for the vertical market.

These platforms generate and oversee an enormous quantity of "bots" for various verticals without requiring human intervention or carrying out any task particularly associated with a target site. Following the initial phase of knowledge base creation for the entire vertical, the platform automatically generates the bots. How quickly the platform can expand to hundreds or thousands of sites, how well the data is retrieved, and how many fields it can retrieve are what define its resilience. This scalability implies that the main target of "Web scraping" (2015b, G). is the long tail of websites from which content is difficult or time-consuming for traditional aggregators to pull. To locate particular data snippets on the pages being scraped, one can utilize annotations, semantic markups, and metadata. With the annotations that it incorporates into the pages, this method can be considered a particular kind of DOM parsing, similar to what Microformat does. In an alternative scenario, web scrapers can retrieve data schema and instructions from the semantic layer containing the annotations before scraping the pages because it is managed and stored differently from the web pages. ("Web scraping," 2015b). Web page analyzers that use computer vision see websites as humans would in order to recognize and

extract information from them ("Web scraping," 2015b). These projects make use of machine learning and computer vision.

A set of tools called web scraping software is used to collect large amounts of data from websites, including job boards, real estate websites, and directory websites, in place of the laborious copy-and-paste process. Should you wish to scrape data on real estate properties in the United Kingdom, you would need to hire a few men to visit each property page and copy and paste information from websites into an Excel spreadsheet. It may take several days or even months to prepare your property data for this kind of use. Web scraping visits each page, extracts data from it, and parses the HTML code in order to programmatically automate manual tasks.

With one of the numerous web scraping tools currently on the market, you can obtain data from any website you desire. The list of scraping tools is below. The features, support, and upgrade cycle that web scraping software offers determine its price. To ensure that the scraping features you require are included, download the demo version. ("List of Web Harvester, Data Scraper, Web Scraping Software and Tools," n.d.). Visual Web Ripper, developed by Sequentum Group in 2006, is one of the most advanced web scraping tools on the market today. Because of these features, you can scrape data from a range of websites, including e-commerce ones. (such as Magento, Amazon, and eBay sites), business directories, straightforward webpages, discussion boards, and classified ads. is a helpful resource for those who are unfamiliar with web scraping. There are three versions of Screen Scraper, a web scraping program: Enterprise, Professional, and Basic. (see "List of Web Harvester, Data Scraper, Web Scraping Software and Tools," n.d.). The Web Content Extractor (WCE) from Newprosoft is an easy-to-use, uncomplicated program. It has a great

wizard that assists users in configuring scrapers. For importing data into Excel, text, HTML, SQL script files, MySQL script files, Microsoft Access databases, XML files, HTTP ODBC data sources, and submitting forms, Web Content Extractor is a great tool. To extract data from websites, a few clicks are all it takes. ("Software for Web Scraping," n.d.) ("List of Web Harvester, Data Scraper, and Web Scraping Software and Tools," n.d.). C. Mozanda: A useful tool for information mining on the internet is a web scraper. It can retrieve information from webpages and PDFs. Even non-technical users can perform basic scraping with its straightforward Point and Selection interface. The primary distinction between Mozanda and other scrapers is that your agent, or scraping project, is operated in their cloud environment. "List of Web Harvester, Data Scraper, and Web Scraping Software and Tools" at the time. Robotic process automation software such as UiPath can visit a website automatically, collect data from multiple webpages for the purpose of integrating data into another online service or application, filter the data, and then transform it into the format that the user wants. Because UiPath emulates a real browser used by a real user, it can extract data that is invisible to most automation systems. (Savinkin, n.d.). With no need for programming experience, its drag-and-drop graphical designer lets you create complex web agents, and your inner.NET hacker gives you total control over the data in the "List of Web Harvester, Data Scraper, Web Scraping Software and Tools," even without a date. E. Out Wit Hub: A Tool for Everyone: OutWit Hub is a robust Firefox plugin. The contents of an extracted Web page are presented in an understandable manner without requiring a great deal of technological expertise or sophisticated programming knowledge. Users can easily extract text from a group of pages, including links, images, email addresses, data tables, and other content,

without ever having to view the source code. Documents and images are kept in SQL, HTML, CSV, or Excel databases; once extracted, the data can be saved directly to your hard drive. The OutWit Hub ("Software for Web Scraping," n.d.) is a useful tool for people who don't know much about web scraping. Three versions of the web scraping software Screen Scraper are available: Enterprise, Professional, and Basic. You can download and use the free basic version of "List of Web Harvester, Data Scraper, and Web Scraping Software and Tools," n.d.; it only includes the most basic scraping features. Before becoming proficient, a novice user must practice a lot with different versions. Screen Scraper's integration with other platforms is essential, and its compatibility with Java enables the creation of intricate scripts for intricate programs (Savinkin, n.d.). H. WebHarvy's point-and-click scraping interface is visually appealing. The shortest learning curve and data extraction time are achieved with WebHarvy, the best tool for extracting text, URLs, and images from websites. In line with SysNucleus (n.d.), "extracted data can be stored into popular formats (CSV, tab-separated values) for database entry." (TSV), XML, and SQL). It can extract properly formatted data from HTML as well as tabular data, which is its primary use. Ajax-based data scraping and deep crawling cannot extract data from the web. ("List of Web Harvester, Data Scraper, Web Scraping Software and Tools," n.d.). Easy Web Extract was developed in 2009 by Web2Mine with the intention of providing quick and simple data extraction. With the help of this scrape tool, which was developed with.NET technology, you can transform data using the C#, VB, and JS built-in scripts. When starting a scrape project, one disadvantage is that loading the URL can occasionally take a very long time (Savinkin, n.d.). 1. WebSundew. With its simple point-and-click interface, Web Sundew is an intuitive web scraping tool that

lets you select which fields on websites you want to extract. Rapid data ripping and high productivity are the design features of this screen scraper. The Enterprise edition allows for remote scraping execution and File Transfer Protocol (FTP) (Savinkin, n.d.) publication of the extracted data. File and image extraction are both supported. It performs multilevel web extraction using a technique called deep crawling ("List of Web Harvester, Data Scraper, Web Scraping Software and Tools," n.d.). Automation Anywhere has been developing Web Data Extractor, a web scraping tool, in the United States since 2003. Its primary focus areas include email, phone, fax number scraping, body text, link extraction, and meta tag extraction. Rule-based online scraping is impacted. At that time, there was a "List of Web Harvester, Data Scraper, and Web Scraping Software and Tools." C. One of the greatest web scraping tools out there is Helium, which has all the features required to get data from any website.

The point-and-click user interface can be used to define the scraping fields. According to "List of Web Harvester, Data Scraper, and Web Scraping Software and Tools," n.d., it offers proxy support, Ajax-based scraping, and CAPTCHA-based scraping. A. An open-source web scraper is called WebExtractor 360. Data scraping from websites is done using regular expressions. Regular expressions are the foundation of this scraping tool, so familiarity with them is required. Source code is available for free download from this scrape. FMiner is a top-notch Python-based visual web scraping utility. In a neat diagrammatic style, it illustrates the movements and processes involved in scraping. It also makes it possible for special Python code to run (List of Web Harvester, Data Scraper,Web Scraping Software and Tools," n.d.). N. Scrapy is a free, open-source platform for gathering information from

websites. The Python-based Scrapy is compatible with Windows, Mac, and Linux. ImportIO is a free online web scraping tool that allows you to gather different kinds of data and group them into data sets. February 2012 saw its founding. The scraper does not need to be run on your computer because ImportIO is a cloud-based platform and all of your data is kept on cloud servers. Regardless of your level of technical expertise, you can use import io. ("Software for Web Scraping," n.d.). A. Web scrapers present users with two excellent options.

The Google Data Extraction Service and the free Google Chrome Extension are these. A sitemap that describes how to navigate and retrieve data from a website can be created by users with the Google Chrome Extension. The web scraper will be able to navigate the website and extract all data in an appropriate manner with the aid of these sitemaps. After scraping, the data can be exported as a CSV file. Depending on your requirements, Enterprise Data Extraction Service provides excellent results. This feature lets you scrape a lot of data, schedule multiple scrapes to run at specific times, and perform multiple simultaneous scrapes.

CHAPTER III

RESEARCH METHODOLOGY

An overview of the project's development technique is provided in this document. It includes details on the procedures followed and the project's research paradigm and design. It comprises the research process and system structure used to describe the phases of development that have been applied accordingly.

Research Method

The developers have used various methods to produce detailed and essential information on the significant requirements that are used in performing the proper process and systematic approach throughout the development life cycle.

Descriptive Method

The descriptive technique was utilized by the creators to gather in-depth and extensive information about the study's components, including information about how they are used, what they do, what they stand for, and how society is affected by them. Additionally, it is utilized to review research and applications that are linked to the project that has been created. The fundamental objective of this approach is to fully understand the project. This information has been gathered to create a foundation and framework that will aid the project's developers as they work to complete the project.

The internet was a resource that the project's developers used to obtain information. The data searched and collected serves as the fundamental knowledge

for building the project. In general, it delivers an overall view of how the project is developed and its coverage, allowing the client to understand and identify the core background of the project.

Developmental Method

The development methodology has been adopted by the developers to track the project's development step-by-step. It controls the development flow and integrates the implementation processes necessary for the stable delivery of each phase. The benefit of the endeavor

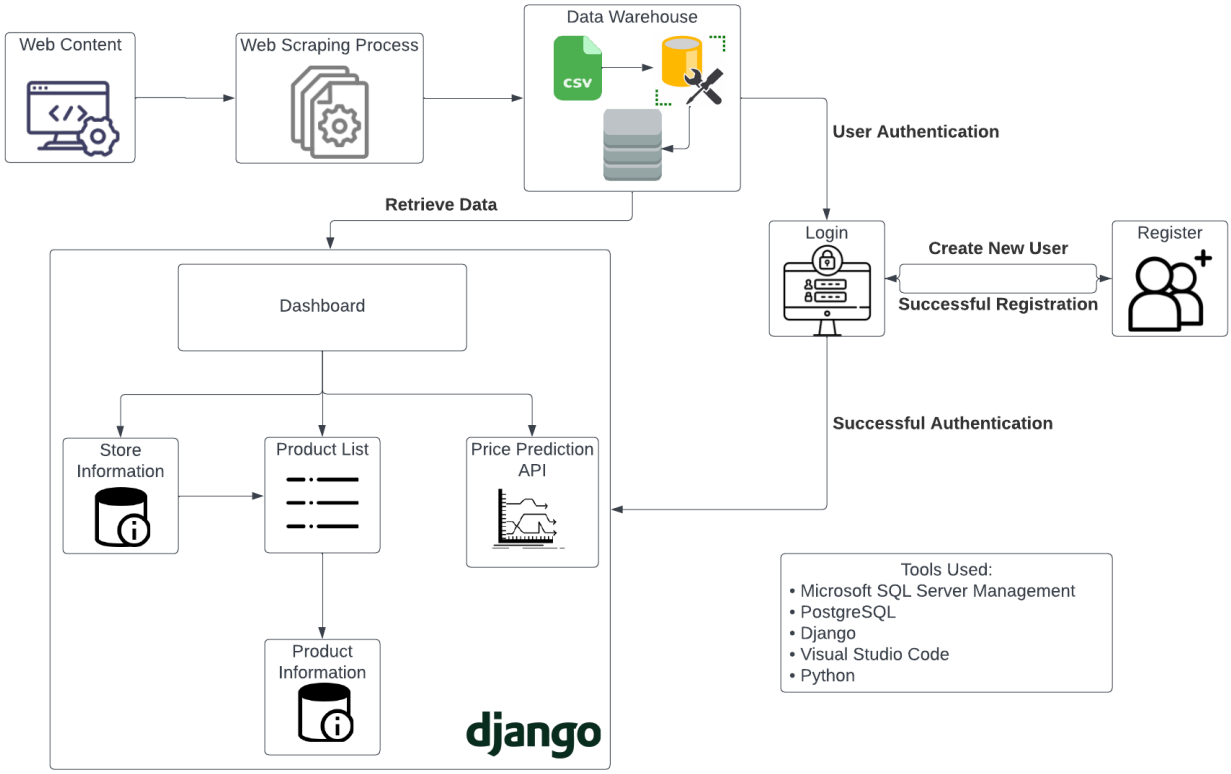

Figure 2: System Architecture and Design

The developers' recommended system architecture is displayed in Figure 2. The project consists of a dashboard of the database built by our data analysts with the "Price Prediction" feature added to forecast product prices based on collected data.

To build the application, the developers used Python as the programming language, Django as the environment, and both SQL Server Management Studio 19 and PostgreSQL as the SQL server. Through the use of a browser, the client can log in or register to access the dashboard to view information about the database, the list of both products and stores, and the price prediction API. The project is locally hosted for now but will be online in the future.

The price of a product entered by the client is predicted using machine learning and a linear regression model, both of which are used and trained by the Price Prediction API.

Research Design

To eliminate discrepancies throughout the development phase, The developers used the agile approach, a software development life cycle (SDLC) that requires them to follow protocols in a linear fashion. Continuous improvement, flexibility, and teamwork are highly valued aspects of the Agile Approach. As such, our project was successful in utilizing the agile methodology.

Figure 3: Agile Approach - System Development

Image Source: https://www.researchgate.net/figure/Agile-Methodology-in-System-Development-source-Okeke2021-retrieved-from_fig1_354310848

Figure 3 shows the chart for a software development life cycle called the Agile Approach for its efficient approach to handling the development process on a consistent and steady basis to avoid inconsistencies and errors.

Requirements

The gathering and defining of the project requirements is the first phase in the Agile approach. It entails comprehending the needs and preferences of the client and recording them in the product backlog, which is a prioritized list. The intended functionality is described from the perspective of the end user in user stories, which are widely used to document requirements.

In gathering data, ideas, and concepts, the data analysts were assigned shops they had to scrape data from and import them into the database. While the developer of the web application used the internet to gather different designs and APIs to implement in the project, The web app developer searches through different sources to provide an idea or topic for the dashboard to be completed.

Design

The design phase follows the establishment of the requirements. Design is a continuous, iterative process under the Agile framework. Agile promotes iterative design, where the design changes as the development goes along rather than producing a finished product up front. This enables quick input and adjustment to shifting demands.

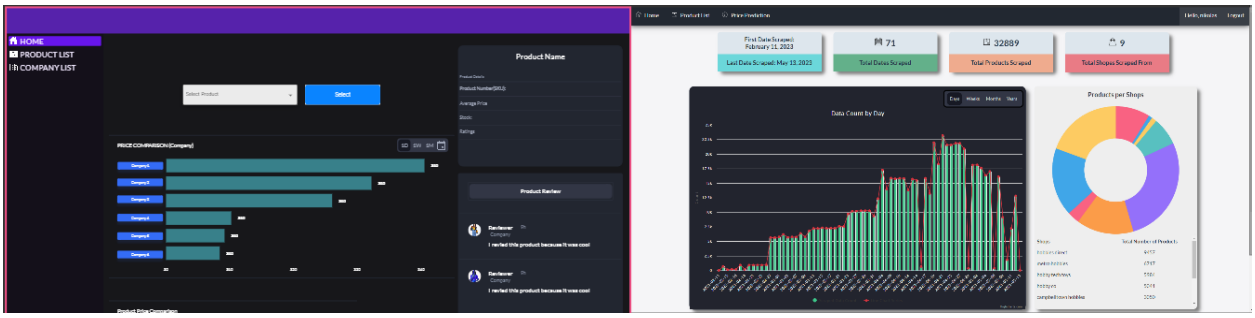

Figure 4: Sample Project Design from its First Design (Left) to its Final Design (Right)

For their front end and back end, the developers chose Python and the Django Web Framework.Microsoft SQL Server Management Studio was used for the database. Because the Server Management Studio required an internet connection in

order to retrieve the data, the Django Framework logs data more slowly as a result. For the local database, PostgreSQL was employed.

Development

The actual coding and implementation of the software are done during the development phase. Agile encourages a cooperative and cross-functional approach, where the development team collaborates to produce small, functional increments of the product known as "sprints" or "iterations." Each sprint typically lasts for a fixed duration, such as two weeks, and the development team concentrates on producing a specific set of user stories within that time frame.

The developers started creating the Web Application. Weekly meetings meant weekly checking. As the project progresses. The project changes direction. When a feature is approved, it is implemented in the production environment.

Testing

Testing is integrated throughout the entire project life cycle in agile development. Agile teams ensure the quality of the product by utilizing a range of testing methodologies, such as acceptance, integration, and unit testing. At every step of the testing process, new functionality that requires testing is added. New functionality that needs to be tested is introduced at each stage of the testing process. Early detection and resolution of any flaws or problems will lower the likelihood of a defective product being delivered. It is presented to our project lead to sample and know whether or not it has fulfilled its purpose and met expectations.

Deployment

The software is ready for deployment once a sprint's development and testing are finished. With Agile, deployment is a continuous process that may result in a releasable product increment at the end of each iteration. This allows for continuous integration and deployment, enabling stakeholders to provide feedback and start using the product as soon as possible.

Review

After the deployment of each increment, a review or retrospective takes place. The team reflects on the completed work, gathers feedback from clients, and assesses the effectiveness of the Agile processes and practices. This feedback is used to make improvements, adjust priorities, and plan for the next iteration.

The review phase ensures continuous learning and adaptation, which is a core principle of Agile methodology.Along with the review of the system, our lead also reviewed the data that we scraped from different websites, like the change in prices and how many toy models are sold on different websites.

Figure 5: Example of a Store Website and the Manufactures being sold displayed in a bar chart

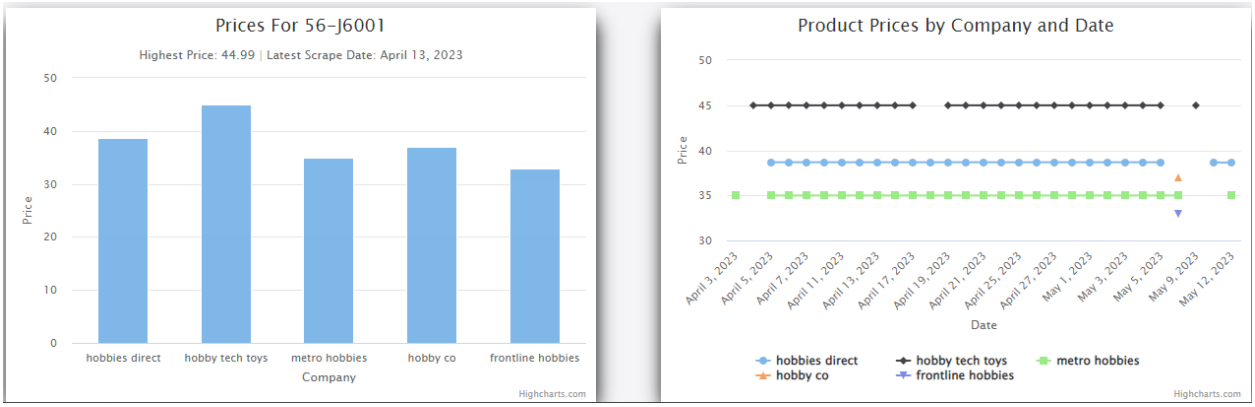

Figure 6: Example of a Toy Model having different prices per store website displayed in both bar chart and Multiple lines chart

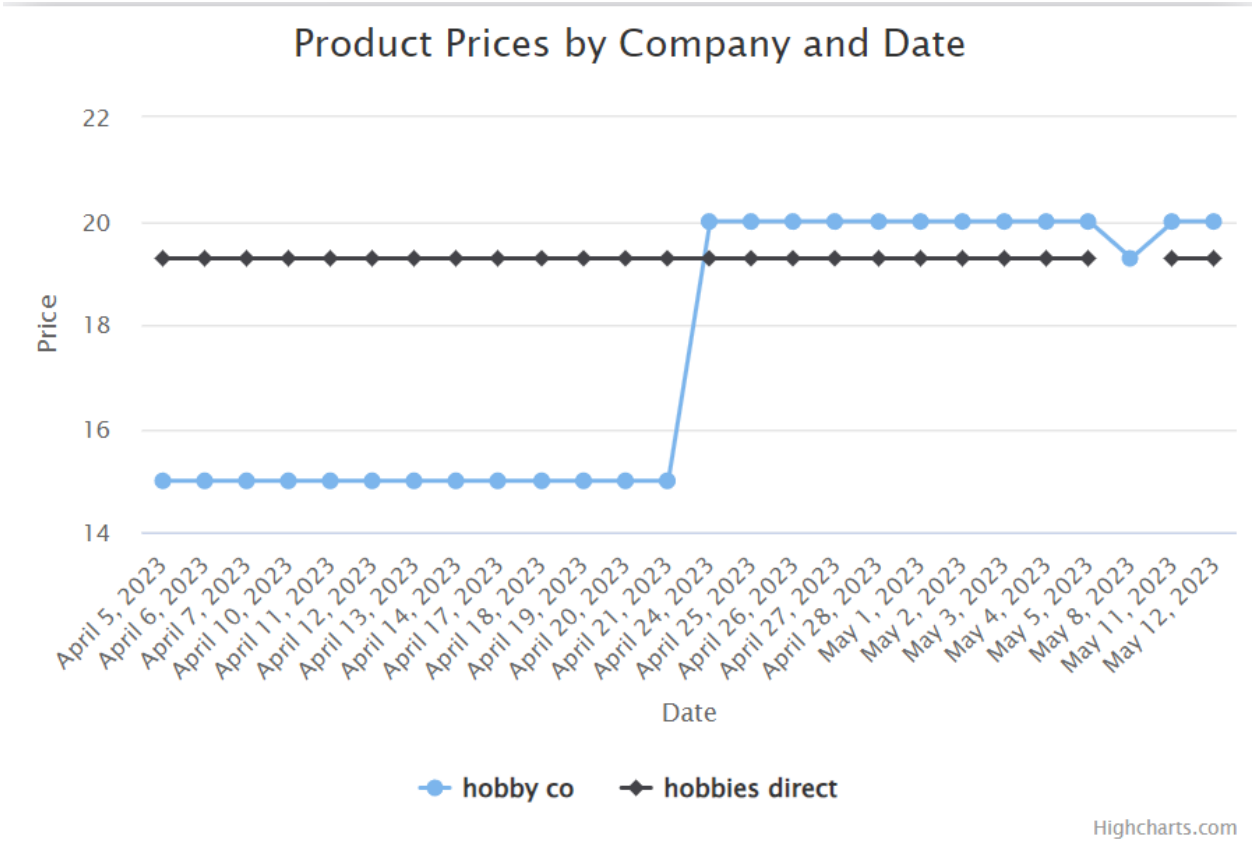

Figure 7: Example of a Toy Model having inflation in prices

To evaluate the direction throughout the development life cycle, each of these phases is repeated. A cross-functional team does all of the following tasks during each iteration: requirements analysis, design, coding and implementation, unit

testing, and review. It's used to build a proper product that would optimize its value throughout the development process and improve its overall specification

Tools, Methods & Algorithms

The developers now explain the acquired and used tools, methods, and algorithms in the development of the project.

Figure 8: Anaconda

In Figure 8, the developers used Anaconda to create the project's environment. Anaconda is a well-known Python distribution for data science and machine learning. With its Conda package manager, managing packages is easier. It includes popular libraries like NumPy, Pandas, and Scikit-Learn, along with tools like Jupyter Notebook and Spyder IDE. Anaconda offers a comprehensive and user-friendly environment for developing and executing Python code, making it a valuable resource for the project.

Figure 9: PostgreSQL and Microsoft SQL Server Management

In Figure 9, Microsoft SQL Server Management Studio (SSMS) is used to manage the database, as well as PostgreSQL. Our data analysts save the scraped data they have collected in SSMS. The information from SSMS is then exported and imported for local accessibility into PostgreSQL. Powerful database management systems are present in both tools.

PostgreSQL, a robust relational database management system (RDBMS) with a large feature set, is available for free. It has an active contributor community, advanced features, and extensibility.

Microsoft SQL Server Management Studio (SSMS), a graphical user interface (GUI) application, is used to manage Microsoft SQL Server databases. It connects effectively with other Microsoft tools and services and provides management and administration tools, query editing, and visual database design.

Figure 10: Visual Studio Code (User)

Visual Studio Code is a handy and portable source code editor (VS Code). For the purpose of creating, editing, and debugging programs in various programming languages, VS Code offers a stable and versatile environment. Thanks to its straightforward user interface, extensive collection of plugins, and built-in support for Git version control, it was the perfect code editor to use. Because it contains features like syntax highlighting, code snippets, intelligent code completion, debugging tools, and an integrated terminal, it is a favorite choice for engineers like us.

Frameworks Client/ Server Computing

Developing dependable and scalable web applications is made easier with the help of the high-level Python web framework Django. The Model-View-Controller (MVC) architectural pattern is utilized, which comprises numerous built-in features and tools to accelerate development.

With Django, developers can focus on writing their application's logic while leveraging the framework's features for handling common web development tasks.

These include URL routing, HTML template rendering, database ORM (Object-Relational Mapping) for interacting with databases, form handling, authentication, and session management.

Django additionally has a management interface known as Django Admin, which enables programmers to easily create an admin dashboard for handling data models and carrying out CRUD (Create, Read, Update, Delete) activities.

Scalability, maintainability, and compliance with best practices are hallmarks of the Django framework. Developers typically use it to build a variety of internet applications, from simple personal projects to complex enterprise systems.

Linear Regression

The equation of a linear regression line is Y = a + bX, where X is the explanatory variable and Y is the dependent variable. When x = 0, the value of y is called the intercept, and the slope of the line is called b.

Feature Selection

Consider employing more sophisticated feature selection methods. As an alternative to the bag-of-words method, consider using word embeddings or the TF-IDF (Term Frequency-Inverse Document Frequency) technique. These methods can extract more significant and pertinent information from the product names, which could improve prediction accuracy and speed up computation.

The difference or uncertainty between the projected values and the actual values is referred to as the margin of error. It displays the variability or dispersion of the projected values relative to the actual values.

The margin of error measures how close or far the anticipated price is from the actual or original price when using linear regression to forecast a product's pricing. A prediction with a smaller margin of error is more likely to be accurate than one with a larger margin of error.

The margin of error can be calculated using the absolute difference between each item's original price and its anticipated price. By adding up all of these differences and dividing by the total number of items, one can determine the average margin of error.

It is imperative to acknowledge that the margin of error in linear regression is contingent upon several factors, such as the selection of the appropriate regression model, the existence of outliers or noteworthy data points, the fundamental assumptions of the linear regression, the caliber and representativeness of the dataset, and additional factors.

A larger margin of error can be an indication of potential issues in the predictive model or the data used for training. To increase the precision of forecasts and decrease the margin of error, it may be required to investigate alternative models, take into account additional variables, or enhance the dataset in such circumstances.

When feature selection is applied in linear regression, the margin of error is still calculated in the same way as mentioned earlier. The margin of error represents the discrepancy between the predicted prices obtained from the selected features and the actual or original prices.

Price Prediction Test Results			
FixedSKU	**Original Price**	**Suggested/Predicted Price**	**Margin of Error**
C8247	• 23.1	22.72	0.38
T58612	• 296.9 • 329.99 • 279 • 289.99	275.06	12.69
56-J6001	• 34.99 • 44.99 • 38.67 • 38.65 • 32.99	35.93	0.86
C8440	• 47.48	-75.03	123.80
T47304	• 499.99 • 487.22 • 479 • 574.99	422.62	36.10
HOR-R8228	• 109.99 • 124.99	69.49	22.37

P9302W	• 113.33	106.70	23.13
C8333	• 47.48 • 29	36.84	13.42
T58685	• 299.99 • 309.02 • 289	240.34	94.40
56-J6000	• 33.99 • 37.99 • 38.67 • 36.99	14.39	33.05

Table 3: Price Prediction Test Results

To get the average margin error, we add up all the margin errors and divide by the quantity of items (10 in this case).:

Total Margin Error = 0.38 + 12.69 + 0.86 + 123.80 + 36.10 + 22.37 + 23.13 + 13.42 + 94.40 + 33.05 = 359.20

Average Margin Error = 359.20 / 10 = 35.92

This means that, on average, the predicted price differs from the original price by approximately $35.92.

Overall, Linear Regression enables the system to make price predictions by learning the relationships between product names and prices from the training data.

Web Scraping

Using Python, we conducted web scraping on various websites such as Campbelltown Hobbies, Crazy Hobbies, Frontline Hobbies, Hobby Co., Hobbies Direct, Hobbytech Toys, Hobby One, and Metro Hobbies. These websites are dedicated to selling toys, including toy replicas.

During the web scraping process, we collected data on manufacturers specializing in toy replicas. The manufacturers we scraped include Traxxas, Tamiya, Woodland, Airfix, Italeri, Revell, Hornby, DCC Concepts, Metcalfe, KATO, Scalextric, and Bachmann.

We transformed all of the raw data into a CSV file format in order to manage the scraped data. Using SQL Server Management Studio 19, we then imported this data into a SQL server. We exported the data from the SQL server once it had been compiled and then imported it into the Postgresql local server as a CSV file.

In conclusion, we used Python to web scrape toy-selling websites, concentrating on companies that produce toy copies. The gathered information was converted into a CSV file format, put into a SQL server, and then exported to a nearby Postgresql server. The web application that was being developed then retrieves the data and displays it.

Software and Hardware Requirements

The following is what the system requires to run

Software Requirements	
Windows 10 or higher	Operating System
Anaconda Navigator	Python Distributor
Python	Programing Language version 3.11
Code Editor	Any Code editor but highly suggest Visual Studio Code
SQL Server Management Studio 19	For the Online Database
PostgreSQL(with PgAdmin)	For the Local Database

Table 4: Software Requirements

Note that you need the login credentials to access the database using SQL Server Management Studio and export the database from there to PostgreSQL using PgAdmin (you need to create your database inside the PgAdmin and import the exported database there). You also need to sync the Web Application to the PostgreSQL database by running **“python manage.py makemigrations”** to sync the web app and **“python manage.py migrate”** to migrate the model to PostgreSQL. After migrating, you can import the database.

Python Dependencies	
Dependencies	Versions
Django	4.2.1

Djangorestframework	3.14.0
django-cors-headers	4.0.0
scikit_learn	1.2.2
Numpy	1.24.3
Pip	23.0.1

Table 5: Python Dependencies

In case of hardware requirements, any Computer or Laptop that could run the mentioned software requirements above can be used to run the system.

CHAPTER IV

PRESENTATION, ANALYSIS, AND INTERPRETATION OF DATA

The software evaluation, results analysis, and corresponding interpretation are all presented in this chapter.

Presentation of Objective No. 1:

To show the Backend of the Project.

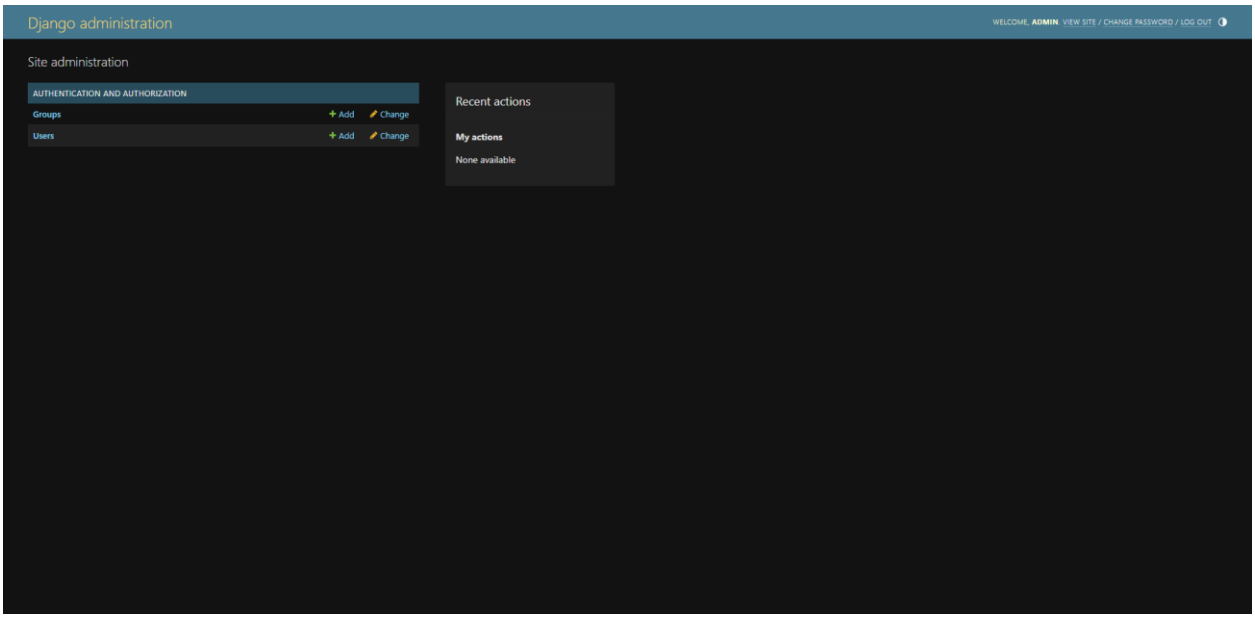

Figure 11: Django Admin Side

Figure 11shows the admin sites of the Django Web Framework. Here, developers and users can easily view users that were registered. Django provides an administrative interface called Django. For managing data models and carrying out CRUD (create, read, update, delete) activities, developers can rapidly create an admindashboard using the admin command.

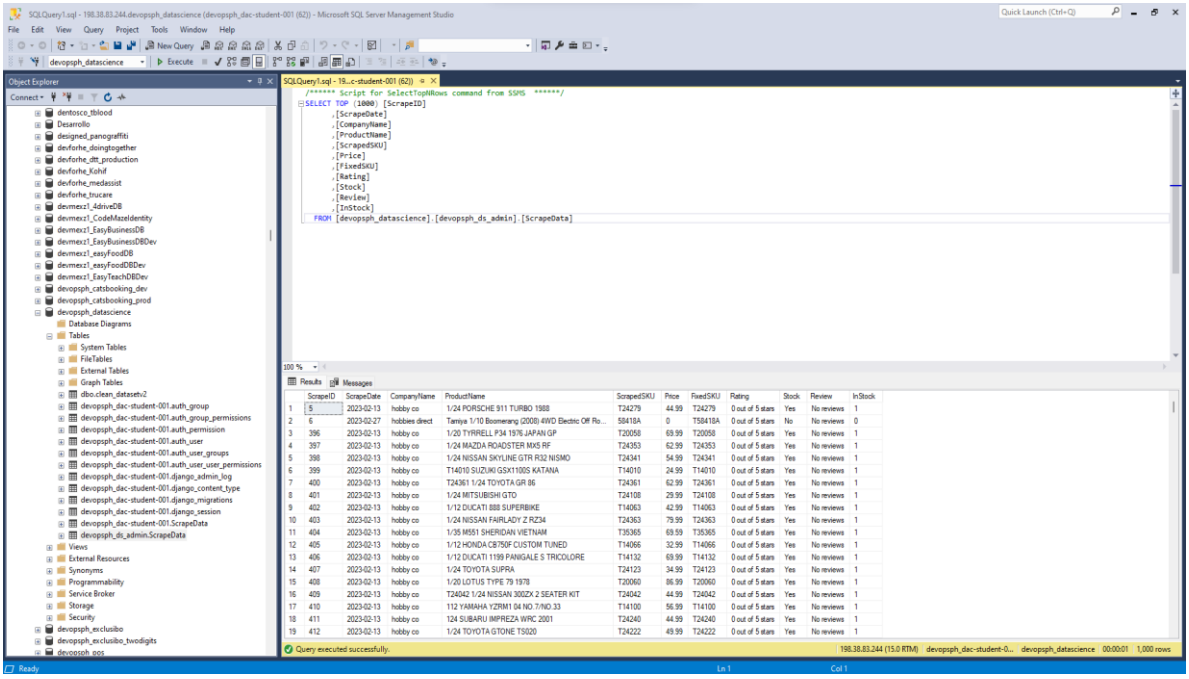

Figure 12: SSMS Interface

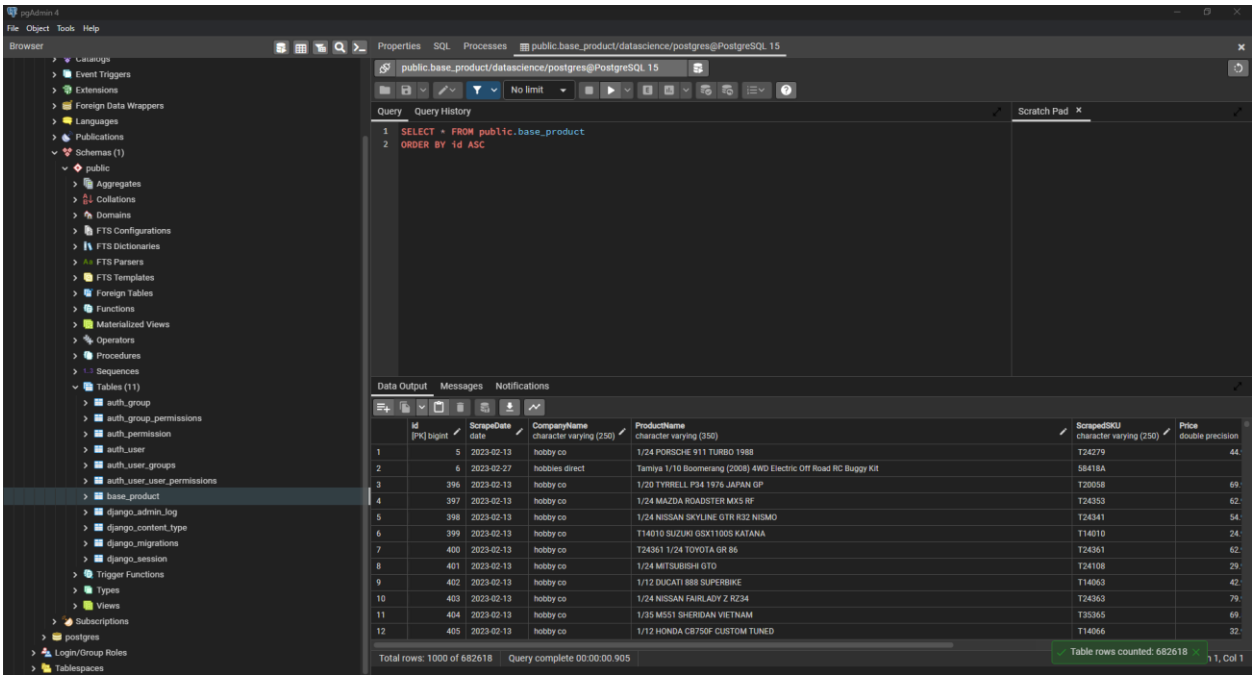

Figure 13: PostgreSQL(pgAdmin4) Interface

Both Figure 12 and Figure 13 show the User interface of the database. Here we can view the whole database and all the data in it.

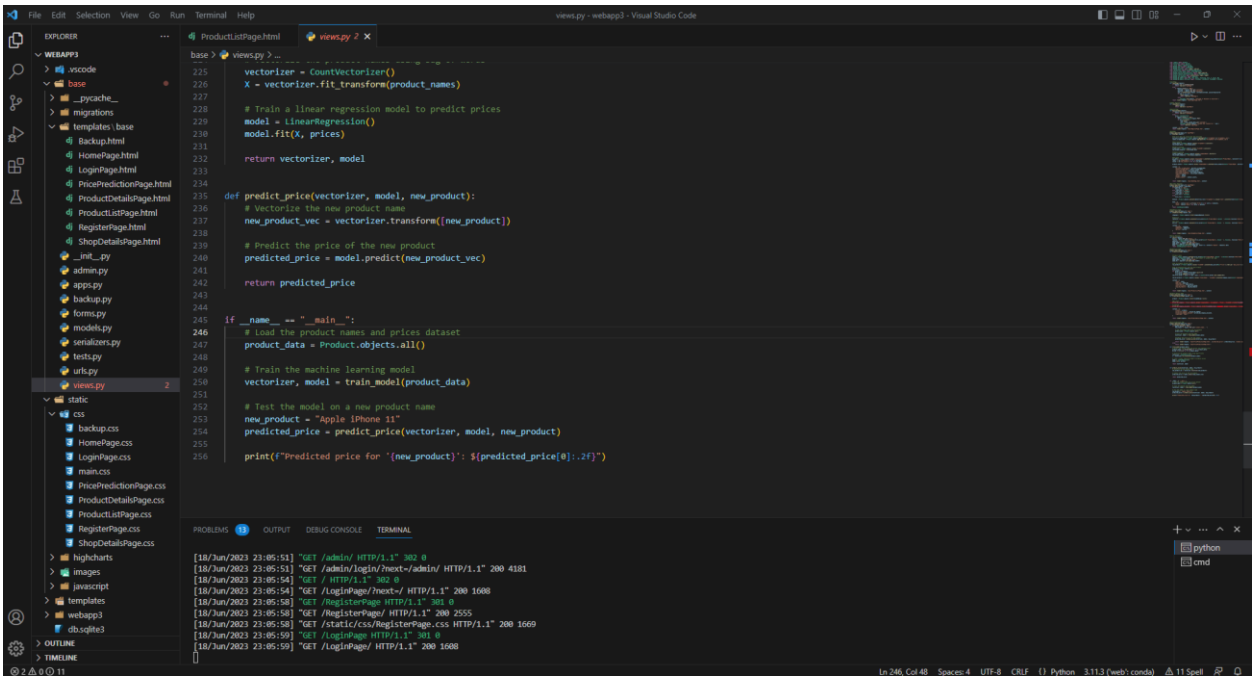

Figure 14: Visual Studio Code Application

Figure 14 shows the VS Code editor, where the developers utilize its features like syntax highlighting, code snippets, intelligent code completion, debugging tools, and an integrated terminal. The developers used this code editor to create the whole web application.

Presentation of Objective No. 2:

To show the Authentication of the Web Application

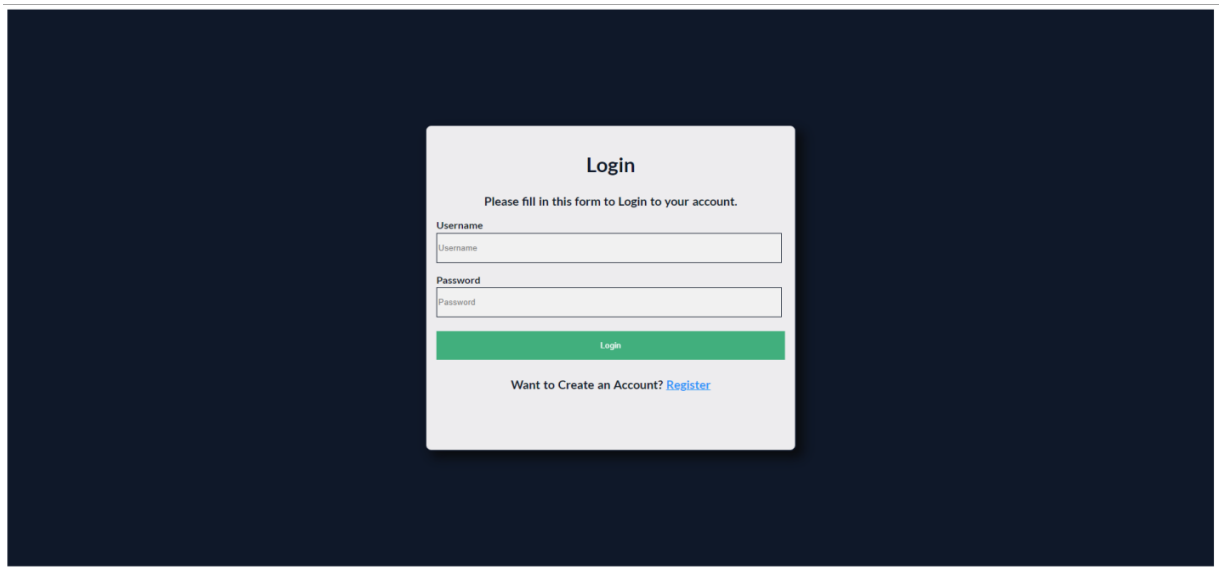

Figure 15: Login Page

Figure 15 depicts the login page, which offers users a safe and controlled entry point through which they can identify themselves using their login credentials and receive access to resources or restricted portions of a website or application.

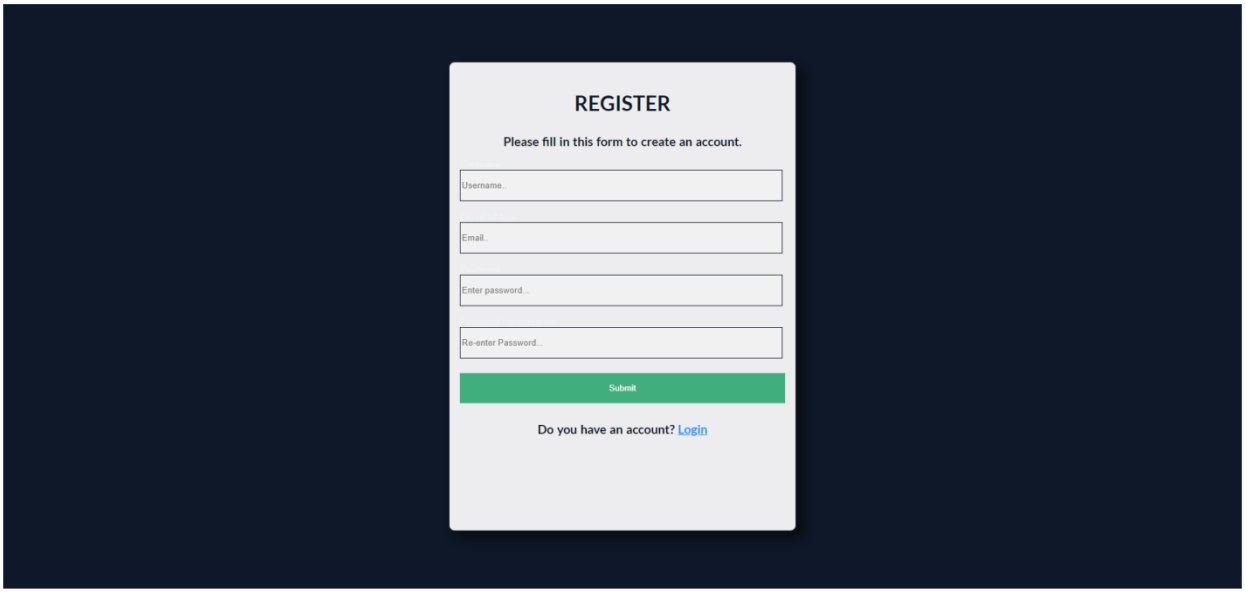

Figure 16: Registration Page

The registration page in Figure 16 demonstrates how new users can create accounts by supplying the required data, authenticating the input, storing account information securely, and laying the groundwork for user authentication. It is essential for establishing users' presence within a system and for onboarding them into a web application.

Presentation of Objective No. 3:

Shows the Information of the Database in the Web Application

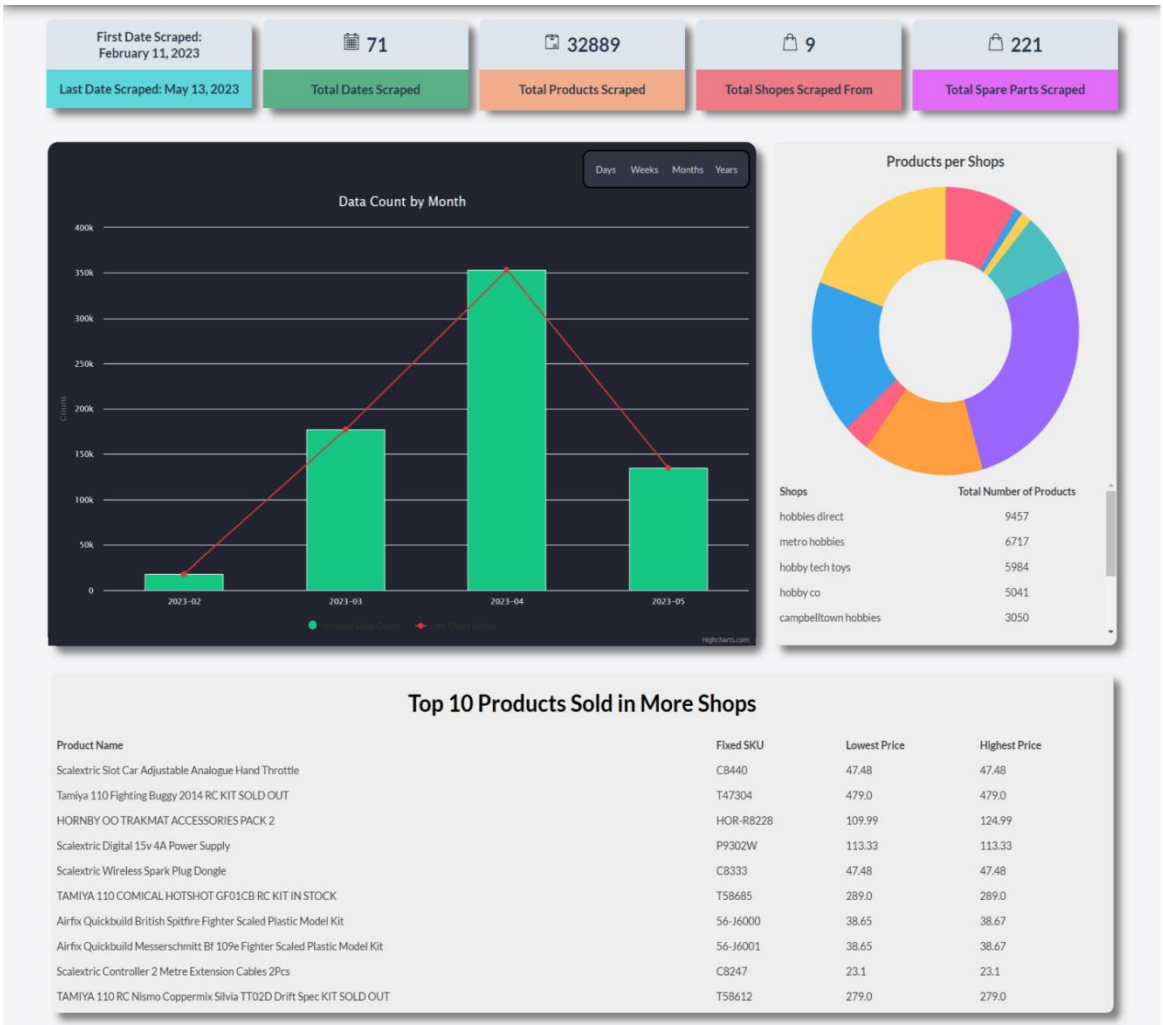

Figure 17: Home Page

In Figure 17, The Home page presents various statistics and visualizations related to the scraped data, providing insights into the data's timeframe distribution and the number of products per shop.

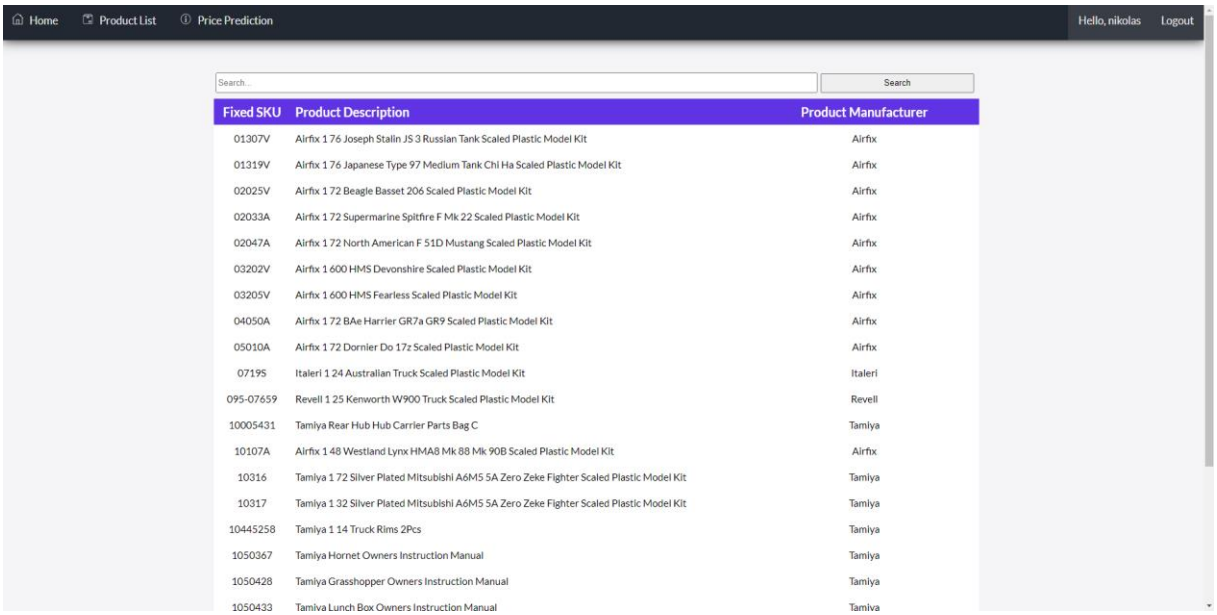

Figure 18: Product List Page

Figure 18 shows a product list page with a search bar, a table displaying the products, and pagination functionality. Users can search for specific products, view their details, and navigate through the list using pagination links.

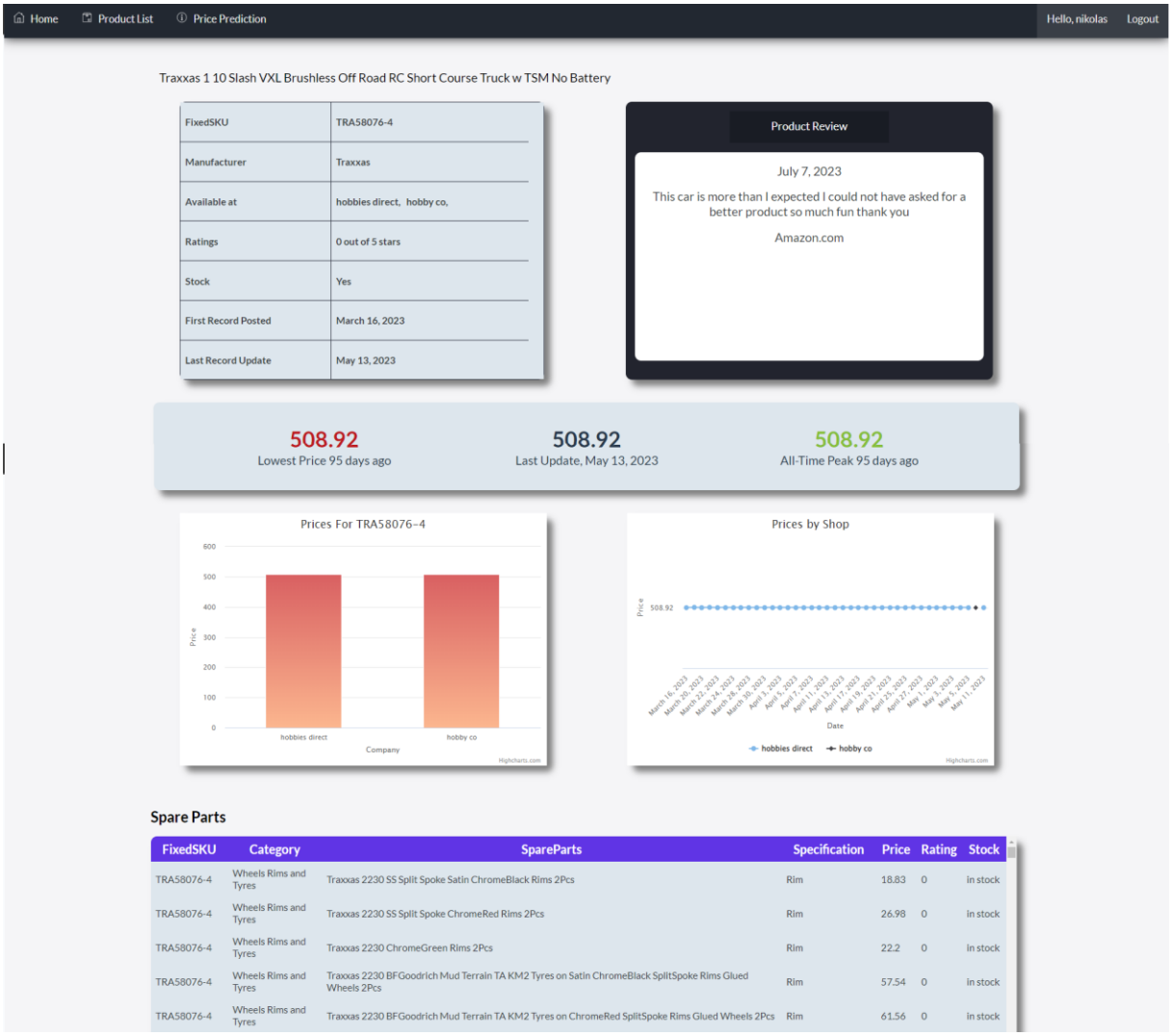

Figure 19: Product Details Page with SpareParts

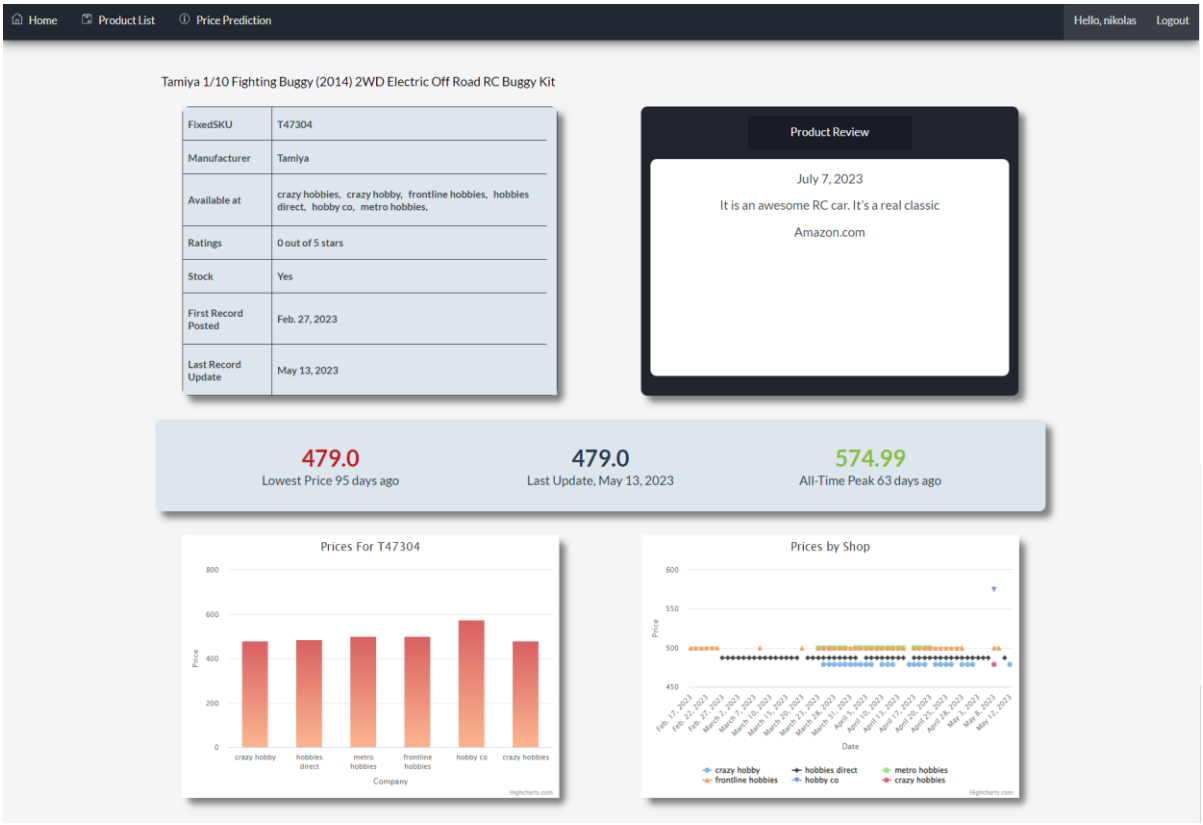

Figure 20: Product Details Page with More Data

Figures 19 and 20 show a product details page that displays information about a specific product, including its details, review, price data visualized through bar and line charts, and spare parts available for the product using ConnectedSKU, which is changed to FixedSKU. The charts are updated in real-time through a Web socket to reflect the latest price information.

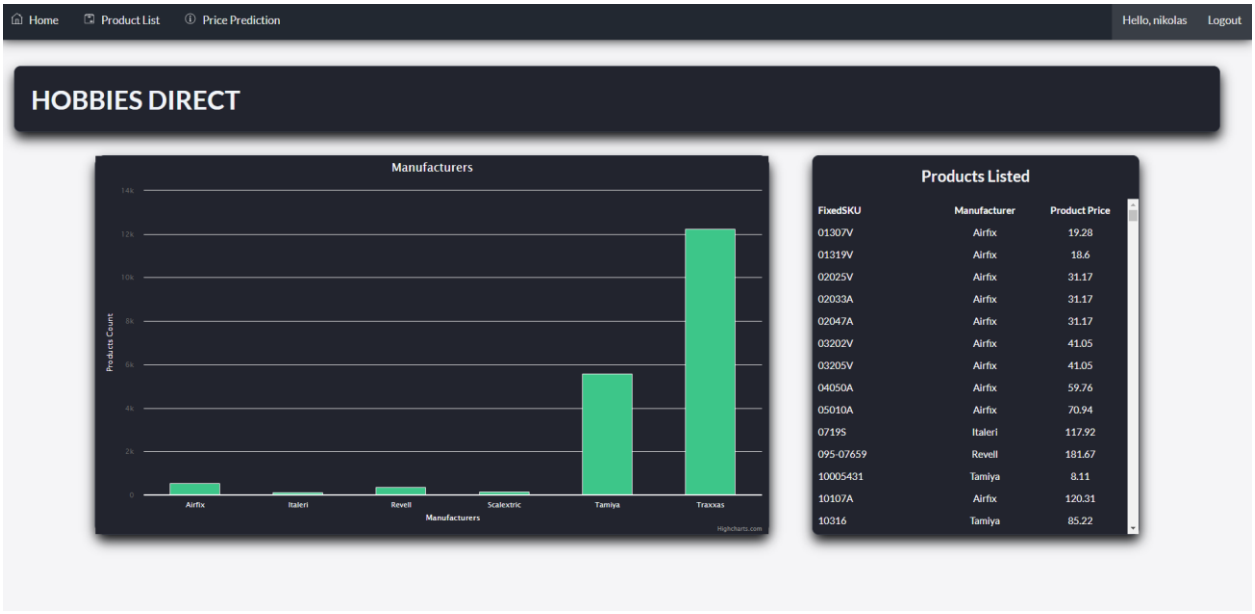

Figure 21: Store Detail Page

Figure 21 shows a shop details page that displays the shop name, a bar chart showing the number of products by manufacturer, and a table listing the available products with their details. The bar chart is dynamically updated based on the number of products listed by each manufacturer.

Presentation of Objective No. 4:

Shows the Price Prediction Page and Explains the Machine Learning used.

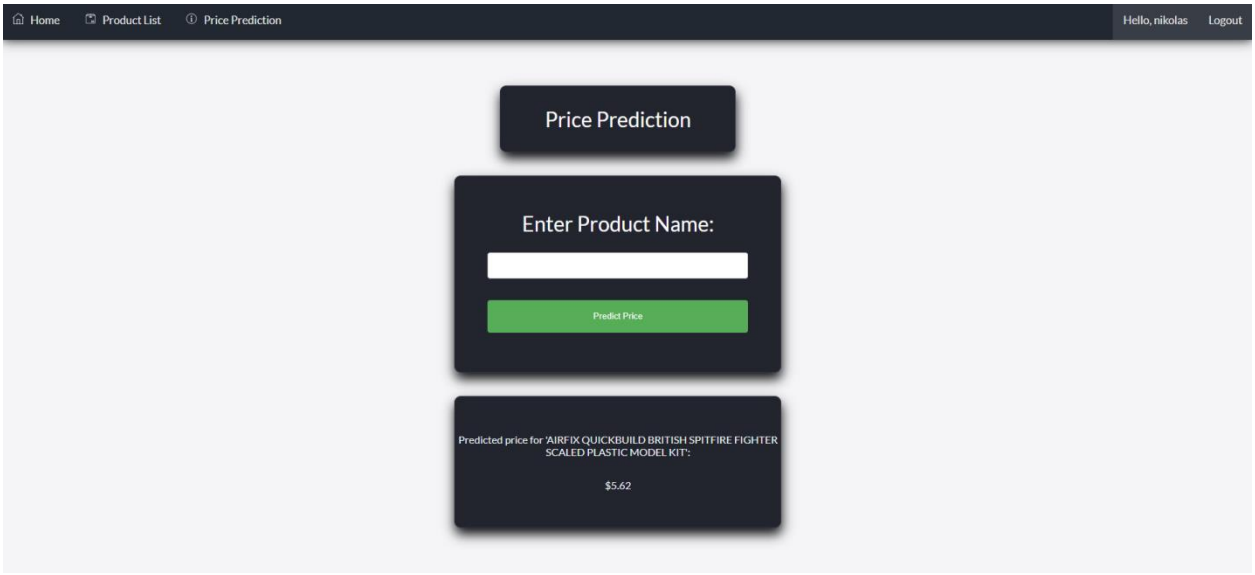

Figure 22: Price Prediction Page

Figure 22 shows a price prediction page with a form where users can enter a product name. Upon submission, the page displays the predicted price for the entered product, if available. The page structure and styling are defined using HTML and CSS.

```python
#Price Prediction Page
@login_required(login_url='login')
def PricePredictionPage(request):
    if request.method == 'POST':
        new_product = request.POST.get('product_name', '')

        # Load the product data from the database
        product_data = Product.objects.all()

        # Train the machine learning model
        vectorizer, model = train_model(product_data)

        # Predict the price for the new product
        predicted_price = predict_price(vectorizer, model, new_product)

        return render(request, 'base/PricePredictionPage.html', {'predicted_price': predicted_price, 'product_name': new_product})
    else:
        return render(request, 'base/PricePredictionPage.html')

def train_model(product_data):
    # Extract the product names and prices into separate lists
    product_names = [x.ProductName for x in product_data]
    prices = [x.Price for x in product_data]

    # Vectorize the product names using bag-of-words
    vectorizer = CountVectorizer()
    X = vectorizer.fit_transform(product_names)

    # Train a linear regression model to predict prices
    model = LinearRegression()
    model.fit(X, prices)

    return vectorizer, model
```

```python
def predict_price(vectorizer, model, new_product):
    # Vectorize the new product name
    new_product_vec = vectorizer.transform([new_product])

    # Predict the price of the new product
    predicted_price = model.predict(new_product_vec)

    return predicted_price


if __name__ == "__main__":
    # Load the product names and prices dataset
    product_data = Product.objects.all()

    # Train the machine learning model
    vectorizer, model = train_model(product_data)

    # Test the model on a new product name
    new_product = ""
    predicted_price = predict_price(vectorizer, model, new_product)

    print(f"Predicted price for '{new_product}': ${predicted_price[0]:.2f}")
```

Figure 23: The code for the Price Prediction

Figure 23 shows a code snippet, which is a Python functioncalled PricePredictionPage that predicts the price of a new product using a machine learning model.

1. The HTTP request that was made by the client is represented by the request object that is passed as a parameter to the function.
2. In order to know whether a user has submitted a form with data, it first determines whether the request method is POST.
3. It obtains the value of the product_name parameter from the POST data of the request inside the POST condition.
4. The code then loads the product data from the database using the Product.objects.all() query, assuming there is a Django model called Product that represents the product data.

5. The loaded product data is passed as an argument when the train_model method is called. The product data will be used to train a machine learning model, which will then be returned together with the feature extraction vectorizer.
6. After training the model, the code calls the predict_price function, passing the trained model, vectorizer, and the new product name as arguments. This function vectorizes the new product name using the same vectorizer and predicts the price using the trained model.
7. After rendering the 'base/PricePredictionPage.html' template, the function renders context variables for the predicted price and the product name.

The train_model function performs the following steps:

1. It receives the product data as an argument.
2. It extracts the product names and prices from the product data, storing them in separate lists.
3. The TF-IDF (Term Frequency-Inverse Document Frequency) algorithm is used to transform textual product names into numerical feature vectors once the function initializes a TfidfVectorizer object.
4. The product names are passed to the vectorizer's fit_transform method, which learns the vocabulary from the product names and transforms them into a sparse matrix representation.
5. A LinearRegression model is initialized.
6. The model is trained by calling its fit method, passing the feature vectors (X) and the corresponding prices.
7. The function returns the trained vectorizer and the trained model.

The predict price function performs the following steps:

1. It receives the vectorizer, model, and new product name as arguments.
2. The function vectorizes the new product name using the transform method of the vectorizer, which converts the name into a feature vector using the learned vocabulary from the training step.
3. The trained model's predict procedure, which forecasts the price, receives the feature vector for the new product name.
4. The predicted price is returned by the function.

Overall, this code demonstrates a straightforward application of machine learning that uses a linear regression model and TF-IDF vectorization to estimate prices based on product names.

PROJECT TIMELINE

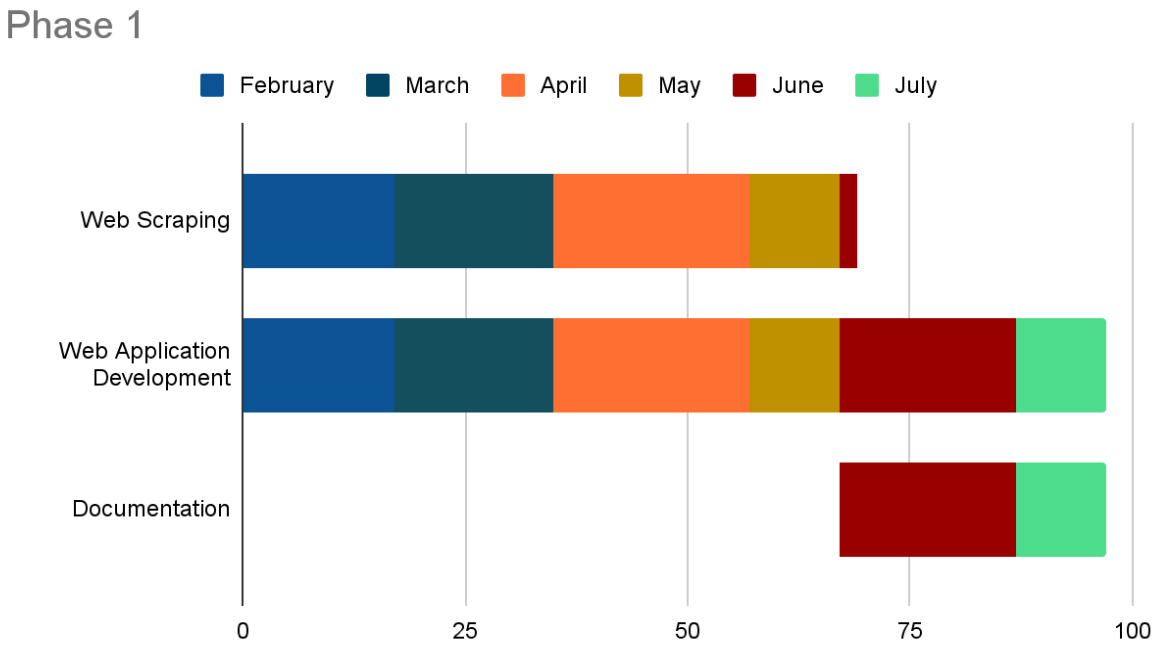

Figure 24: Gantt Chart

Web Scraping

Toy model web scraping is the practice of gathering information from websites that offer toy models and storing it in a database for future use and research. In this case, the objective is to scrape toy model data from multiple internet retailers between February 6, 2023, and May 13, 2023. A Toy Model Database, which has various fields to record pertinent data, will be used to store the data.

The web scraping process will involve writing a script or program that navigates to the toy model shops' websites, retrieves the necessary data such as company name, product name, SKU, price, rating, stock status, and reviews, and saves it into the Toy Model Database using the defined schema.

Once the initial scraping phase is completed, the Toy Model Database will contain a comprehensive collection of toy model data from various shops during the specified period. This data can then be utilized for various purposes, such as analyzing pricing trends, comparing ratings, monitoring stock availability, or generating reports.

Additionally, in June, spare parts information can be incorporated into the Toy Model Database. This can involve extending the existing schema to include fields specific to spare parts, such as SparePartName, SparePartSKU, SparePartPrice, and SparePartStock. The web scraping process can be modified or extended to retrieve spare parts data from the toy model shops' websites and populate the respective fields in the database.

By regularly updating and maintaining the Toy Model Database with the latest toy model and spare parts information, toy enthusiasts, collectors, and sellers can

make informed decisions, track product availability, and efficiently manage their inventory.

Web Application Development

A hard undertaking requiring careful preparation and execution is creating a web application that displays database statistics, product statistics, and replacement parts related to a particular product. With a user-friendly design, this program aims to give users detailed information about the database, product statistics, and the availability of replacement parts.

However, during the development process, several challenges were encountered, resulting in a longer-than-expected development time. One of the primary challenges was the inexperience of the developer. Developing a web application with such intricate features requires a solid understanding of web development concepts, database management, and data visualization techniques. The developer's lack of experience in these areas led to a slower pace of development, as they needed to invest additional time in research and learning.

Another factor that contributed to the extended development time was the developer's frequent illness. Being easily sick caused disruptions in the development schedule, as the developer was unable to work for several days at a time. This further delayed the completion of the web application, as the developer had to take time off to recover and regain productivity.

The development process persisted with tenacity and commitment in spite of these obstacles. A smooth user experience was made possible by the web application's construction employing a combination of front-end and back-end

technologies. The front-end was designed using HTML, CSS, and JavaScript to give users an easy-to-use and responsive interface. To efficiently store and retrieve data, the back-end made use of a server-side programming language like Python or PHP and a database management system like MySQL or PostgreSQL.

In addition to the aforementioned features, the web application also incorporates a price prediction feature using linear regression. This feature utilizes historical data and statistical techniques to predict the future prices of products. By leveraging the power of linear regression algorithms, the application can assist users in making informed decisions regarding product pricing and inventory management.

Overall, despite the challenges posed by the developer's inexperience and frequent illness, the web application was eventually completed. It now serves as a valuable tool for users to access and analyze database statistics, product statistics, and spare parts information, while also offering the added benefit of price prediction using linear regression. Despite taking longer than initially planned, the development process produced a strong and feature-rich web application that satisfied the demands of the target audience.

Documentation

By creating comprehensive documentation for the project, both technical and non-technical stakeholders can gain a clear understanding of the web application's purpose, features, and underlying processes. This documentation serves as a valuable resource for developers, users, and future contributors, facilitating seamless project management maintenance, and scalability.

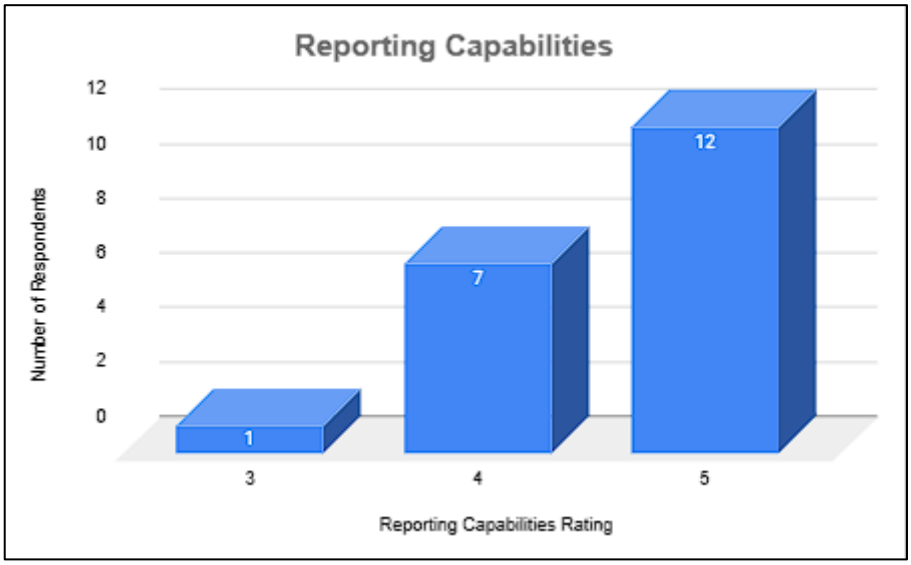

Figure 25. User Acceptance Results showing the Reporting Capabilities rating

The chart displays user ratings on the system's reporting abilities using a scale from 1 to 5, where 1 means in adequate and 5 means highly effective report generation.

Most users rated the system highly, with scores of 4 and 5, indicating a strong belief in its effective report generation. However, some users gave slightly lower ratings at 3 and 4, suggesting differing opinions on its reporting capabilities.

Overall, the chart indicates a mostly positive view of the system's reporting capabilities, despite some users having varying perspectives.

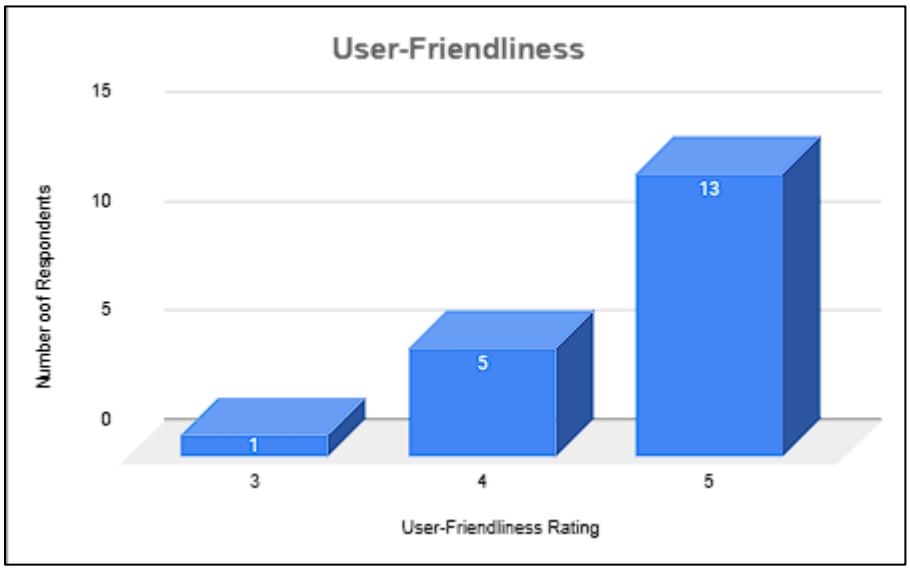

Figure 26. User Acceptance Results showing the User Friendliness Rating

The chart displays user ratings on the system's user-friendliness, using a scale from 1 to 5, where 1 means difficult to use, and 5 means very user-friendly.

Most users rated the system highly, with scores of 4 and 5, suggesting they find it very user-friendly. However, a few users gave slightly lower ratings at 3 and 4, indicating some differences in opinions about its ease of use.

Overall, the chart indicates that most users view the system as user-friendly, although there are varying opinions among some users

Evaluating the Effectiveness of the Data Analytics Interface in Providing Insights and for price monitoring.

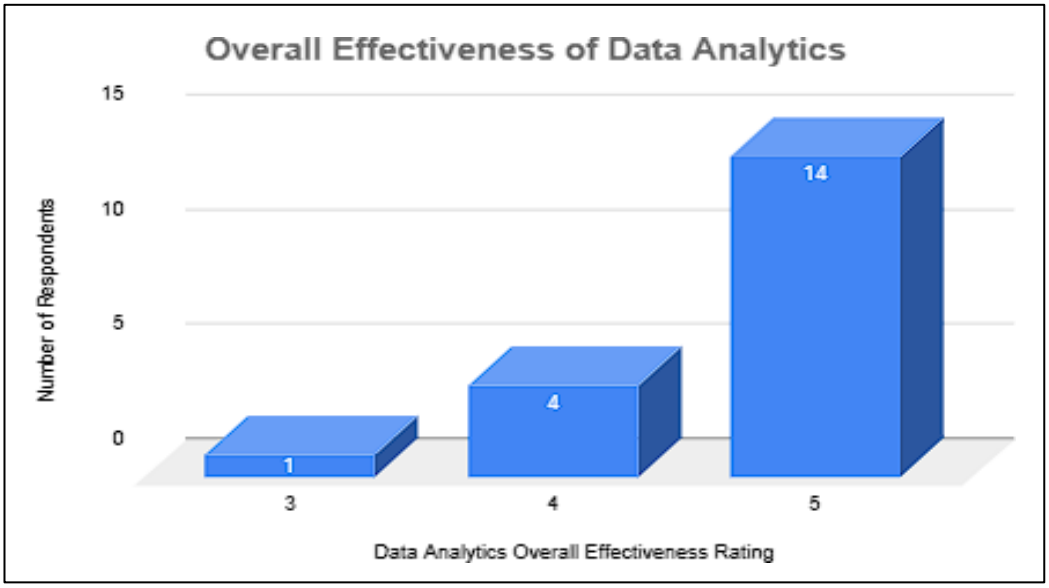

Figure 27. User Acceptance Results showing the Overall Effectiveness of Data Analytics Rating

The chart displays user ratings on how effective the data analytics interface is in providing insights and supporting firearm data monitoring for decision-making .Ratings range from 1 to 5, where 1 means highly ineffective and 5 means highly effective.

Most users rated the interface highly, with scores of 4 and 5, indicating they find it effective in supporting fire arm data monitoring. However, a few users gave slightly lower ratings at 3 and 4, suggesting some differences in opinions.

Overall, the chart suggests that most users perceive the data analytics interface as effective in providing insights for firearm data, although there are varying opinions among some users.

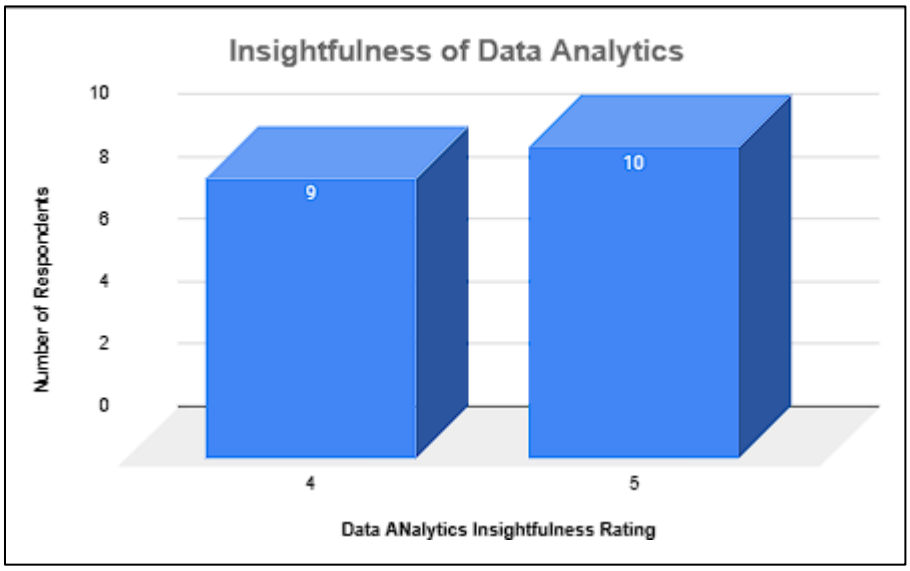

Figure28. User Acceptance Results showing the Insightfulness of Data Analytics Rating

The chart represents user ratings on the insightfulness of the data analysis provided by the interface, rated on a scale from 1 to 5, where 1 signifies not insightful at all and 5 indicate highly insightful.

Analysis of the data shows a consistent trend: most users rated the data analysis highly, with scores primarily at 4 and 5. This suggests that most users find the interface's data analysis quite insightful.

However, a subset of users gave slightly lower ratings at 4, implying some variance in opinions regarding the level of insightfulness provided by the interface's data analysis.

Overall, the chart demonstrates a positive view of the data analysis's insightfulness provided by the interface, although some users hold differing opinions.

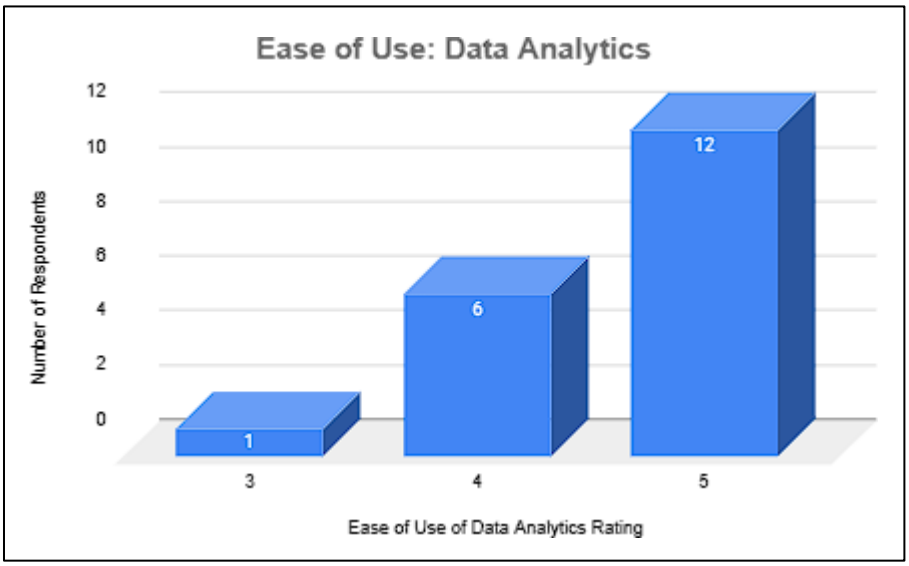

Figure 29. User Acceptance Rating showing the results of Ease of Use of Data Analytics Rating

The chart displays user ratings on how easy it is to use the data analytics interface to extract insights from firearm data, graded on a scale from 1 to 5, where 1 indicates highly challenging and 5 signifies highly user-friendly.

Upon analysis, it's noticeable that the majority of users consistently rated the interface highly, predominantly at 4 and 5. This suggests that most users find the interface quite user-friendly when extracting insights from firearm data.

However, a few users provided slightly lower ratings at 3 and 4, indicating some variation in opinions regarding the ease of use when using the interface for data extraction.

In summary, the chart illustrates a positive perception of the data analytics interface's ease of use for extracting firearm data insights, though some users hold differing opinions.

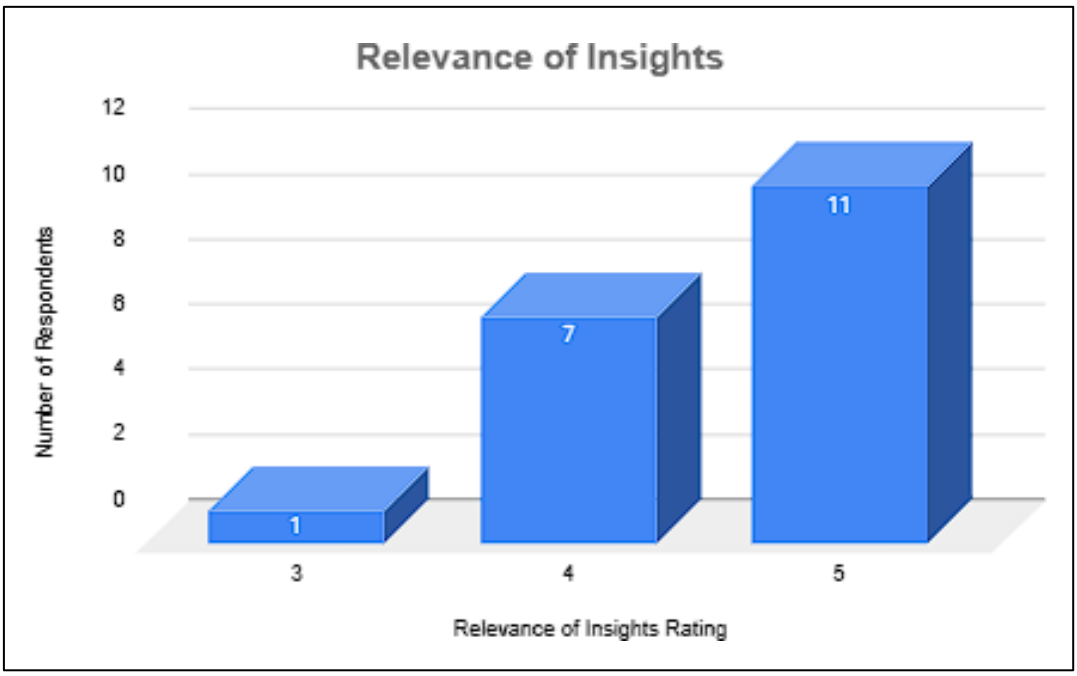

Figure 30. User Acceptance Rating showing the results of the Relevance of Insights Rating

The chart displays user ratings on how well the insights provided by the interface match the needs of the users. Ratings range from 1 to 5, where 1 means not relevant and 5 means highly relevant.

Most users consistently rated the insights highly, at 4and5, suggesting they find the provided insights relevant to their needs. However, a few users gave slightly lower ratings at 3 and 4, indicating some differences in opinions.

Overall, the chart suggests that most users perceive the insights provided by the interface as relevant to their needs, although there are varying opinions among some users.

Chapter 5

SUMMARY, CONCLUSION, AND RECOMMENDATION

Summary

The design and implementation of big data and analytics solutions to supplement dynamic pricing comparison analytics for an Australian e-commerce enterprise company are covered in this chapter. This research explores the critical role of data-driven strategies in improving pricing competitiveness and customer engagement within the Australian e-commerce market. By leveraging cutting-edge data technologies and analytics, this chapter outlines a comprehensive approach to empower the company with real-time, personalized, and competitive pricing strategies, ultimately enhancing its market position and customer satisfaction."

This thesis statement serves as an introduction to the chapter's primary focus on leveraging big data and analytics to enhance dynamic price comparison analytics within an Australian e-commerce company. It highlights the importance of data-driven strategies for improving pricing and customer engagement.

Conclusion

In lieu of, the summary and the findings shown and discussed in the previous section and chapter, the researcher was able to achieve the following:

Design a Web-scrape Script to Gather Data of E-commerce Websites: The research introduced a web-scraping process aimed at collecting data from toy model shops, which involved creating scripts to navigate websites, retrieve necessary information, and save it into a database. This satisfied the objective by demonstrating the capability to harvest pertinent data from e-commerce sites, a foundational step for analyzing e-commerce dynamics.

Design the Model of the Data-Warehouse to House Data from Different Sources:

The SSMS (SQL Server Management Studio) interface plays a crucial role in managing and querying databases, including those in the cloud. When working with a cloud database storing scraped data, SSMS offers a familiar and robust platform for tasks such as connecting to the database, viewing its schema, executing SQL queries, and managing backups. This makes SSMS an essential tool for efficiently handling cloud databases, highlighting its significance in the data scraping and storage processes.

Additionally, the creation of a Model Database with various fields for recording data from e-commerce sites implies a comprehensive approach to designing a data warehouse model capable of integrating diverse data types. This aligns with the objective of building a system that can be customized for different industries, indicating a thoughtful consideration for scalability and adaptability.

Develop a Business Intelligence Tool to Provide Insights on Price Comparisons of Different E-commerce Websites: The development of a web application that displays database statistics, product statistics, and information on spare parts indirectly addresses this objective. By presenting various statistics and visualizations related to the scraped data, including the number of products per shop and price details, the research outputs function as a form of Business Intelligence tool. This tool enables stakeholders to compare prices and other key metrics across different e-commerce websites, fulfilling the objective.

Implement Machine Learning to Predict Possible Price of Product: The research successfully addressed this objective through the development of a price

prediction feature that utilizes a machine learning model (linear regression) to predict the price of a new product based on historical data. This feature, demonstrated through a user interface and underlying code, shows the effective application of machine learning techniques for price prediction, aligning closely with the stated objective.

Recommendation

To enhance dynamic price comparison analytics for an e-commerce enterprise in Australia, it is crucial to establish a comprehensive strategy that encompasses data management, advanced analytics, and customer-centric solutions. Firstly, defining clear objectives is fundamental; understanding the specific business goals, such as improving pricing strategies, optimizing product recommendations, and increasing sales, will serve as a guiding principle. The foundation of this enhancement lies in robust data strategies and a reliable big data infrastructure capable of handling the volume, velocity, and variety of data sources, including competitor pricing, historical sales data, and customer behavior. Implementing real-time data integration, advanced analytics techniques, and dynamic pricing models allows the enterprise to stay competitive by adjusting prices in real-time and making data-driven decisions. A user-friendly interface with data visualization tools provides access to insights, and personalization based on customer preferences should be integrated. Scalability, compliance, and security measures must be prioritized to handle data growth and ensure data privacy. Collaboration with competitors to share non-sensitive pricing data and implementing a feedback loop for customer input fosters transparency and customer-centric pricing strategies. Regular monitoring, skill development, and continuous improvement, along with legal and ethical considerations, are essential components of a successful implementation. Lastly, a well-defined set of KPIs

ensures that the enterprise can measure the success of its dynamic pricing analytics and make data-driven decisions to gain a competitive edge in the Australian e-commerce market

References

Ahmad Tasnim Siddiqui and Sultan Aljahdali. Web mining techniques e-commerce applications. ArXiv preprint arXiv:1311.7388, 2013.

Ahmad Tasnim Siddiqui and Sultan Aljahdali. Web mining techniques in e-commerce applications. arXiv preprint arXiv:1311.7388, 2013

Andrea Horch, Holger Kett, and Anette Weisbecker. Mining e-commerce data from e- shop websites. In 2015 IEEE Trustcom/BigDataSE/ISPA, volume 2, pages 153–160. IEEE,2015.

Antal, M., “A ‘Parasite Market:’ A Competitive Market of Energy Price Comparison Websites Reduces Consumer Welfare,” Energy Policy 138 (2020), 111228.

Amna Shifia Nisafani, Rully Agus Hendrawan, and Arif Wibisono. Eliciting data from website using scrapy: An example. SEMNASTE-KNOMEDIA ONLINE, 5(1):2–1, 2017.

Armstrong, M., “Competition in Two-Sided Markets,” RAND Journal of Economics 37 (2006), 66891. “Search and Ripoff Externalities,” Review of Industrial Organization 47 (2015),
273–302.

Arnold, M. A., C. Li, C. Saliba, and L. Zhang, “Asymmetric Market Shares, Advertising and Pricing: Equilibrium with an Information Gatekeeper,” Journal of Industrial Economics 59 (2011), 63–84.

A. Alam, A. A. Anjum, F. S. Tasin, M. R. Reyad, S. A. Sinthee, and N. Hossain, “Upoma: ADynamic Online Price Comparison Tool for Bangladeshi E-commerce Websites.” 2020IEEE Region 10 Symposium (TENSYMP), 2020, doi: 10.1109/tensymp50017.2020.9230862.

A. Cavallo, B. Neiman and R. Rigobon, Currency unions, product introductions, and the real exchange rate, Quarterly Journal of Economics, vol. 129, no. 2, pp. 529-595, 2014.

A. Chandra and M. Tappata, Consumer search and dynamic price dispersion: An application togasoline markets, The RAND Journal of Economics, vol. 42, no. 4, pp. 681-704, 2011.

A. Hortaçsu and C. Syverson, Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds, Quarterly Journal of Economics, vol. 119, no. 2, pp. 403-456, 2004.

A. Wolitzky, “A Search Cost Model of Obfuscation,” RAND Journal of Economics 43(2012), 417–41.

Bakos,J.Y. (1997), “Reducing buyer search costs: implications for electronic marketplaces”,Management Science, Vol. 43 No. 12, pp. 1676-1692.

Bakos, Y. (1998), “The emerging role of electronic marketplaces on the internet” , Communications of theACM, Vol. 41 No. 8, pp. 35-42.

Baye, M. R., X. Gao, and J. Morgan, “On the Optimality of Clickthrough Fees in Online Markets,” Economic Journal 121 (2011), F340–F367.

Baylis, K. and Perloff, J.M. (2002), “Price dispersion on the internet: good firms and bad firms”, Review of Industrial Organization, Vol. 21 No. 3, pp. 305-324.

Belleflamme, P., and M. Peitz, “Platform Competition and Seller Investment Incentives,” European Economic Review 54 (2010), 1059–76.

Boik, A., and K. Corts, “The Effects of Platform MFNs on Competition and Entry,” Journal of Lawand Economics 59 (2016), 105–34.

Brynjolfsson, E. and Smith, M.D. (2000), “Frictionless commerce? A comparison of internet andconventional retailers”, Management Science, Vol. 46 No. 4, pp. 563-585.

Brand and Price Advertising in Online Markets,” Management Science 55 (2009), 1139–51.

Brynjolfsson, E., and M. D. Smith, “Frictionless Commerce? A Comparison of Internet and Conventional Retailers,” Management Science 46 (2000), 563–85.

Brown, J. R., and A. Goolsbee, “Does the Internet Make Markets More Competitive? Evidencefrom the Life Insurance Industry,” Journal of Political Economy 110 (2002), 481–507.

Burdett, K., and K. L. Judd, “Equilibrium Price Dispersion,” Econometrica 51 (1983), 955–69.

Byrne, D. P., N. de Roos, and D. Tiong, “The Internet, Search and Asymmetric Pricing:

A Natural Experiment in Retail Gasoline,” Working Paper University of Melbourne, 2014.

B. D. L. Santos, A. Hortaçsu and M. R. Wildenbeest, Testing models of consumer search using data on web browsing and purchasing behavior, American Economic Review, vol. 102,no. 6, pp. 2955-2980, 2012.

B. K. Pathak, Comparison shopping agents and online price dispersion: A search cost based explanation, Journalof Theoretical and Applied Electronic Commerce Research, vol. 7, no. 1, pp. 64-76, 2012.

Camacho, A. and Conover, E. (2011), “The impact of receiving price and climate information in the agricultural sector”, IDB Working Paper Series No. IDB-WP- 220, Washington, DC.

Campbell, J.R. and Eden, B. (2014), “Rigid prices: evidence from U.S. scanner data”,

International Economic Review, Vol. 55 No. 2, pp. 423-442.

Caillaud, B., and B. Jullien, “Chicken & Egg: Competition Among Intermediation Service Providers,”RAND Journal of Economics 34 (2003), 309–28.

Cavallo, A. (2013), “Online and official price indexes: measuring Argentina’s inflation”, Journal of Monetary Economics, Vol. 60 No. 2, pp. 152-165.

Cavallo, A. (2017), “Are online and offline prices similar? Evidence from large multi-channel retailers”, American Economic Review, Vol. 107 No. 1, pp. 283-303.

Cavallo, A. (2018), “Scraped data and sticky prices”, Review of Economics and Statistics, Vol.100 No. 1, pp. 105-119.

Cavallo, A. and Rigobon, R. (2016), “The billion prices project: using online prices for measurement and research”, Journal of Economic Perspectives, Vol. 30 No. 2, pp. 151-178.

Chevalier, J. and Goolsbee, A. (2003), “Measuring prices and price competition online: Amazon.comand BarnesandNoble.com”, Quantitative Marketing and Economics, Vol. 1No. 2, pp. 203-222.

Chioveanu, I., “Advertising, Brand Loyalty and Pricing,” Games and Economic Behavior 64 (2008),68–80.

Chunmei Zheng, Guomei He, and Zuojie Peng. A study of web information extraction technology based on beautiful soup. JCP, 10(6):381–387, 2015.

CMA, “Digital Comparison Tools: Consumer Research,” (2017a), Report Retrieved from https://assets.publishing.service.gov.uk/media/58e224f5e5274a06b3000099/dcts-consumer-research-final-report.pdf, accessed March 29, 2019.

“Digital Comparison Tools Market Study: Final Report,” (2017b), Report Retrieved from https://assets.publishing.service.gov.uk/media/59c93546e5274a77468120d6/digital- comparison-tools-market-study-final-report.pdf, accessed March 29, 2019. ,

"Digital Comparison Tools Market Study: Final Report. Paper A: Consumer Views, Behaviour and Experiences," (2017c), Report Retrieved from https://assets.publishing.service.gov.uk/media/59c9356bed915d7bd5d75dda/pa per-a- consumer-experiences.pdf, accessed March 29, 2019.

"Digital Comparison Tools Market Study: Final Report. Paper E: Competitive Landscape and Effectiveness of Competition," (2017d), Report Retrieved from https://assets.publishing.service.gov.uk/media/59e093f5e5274a11ac1c4970/pa per-e- competitive-landscape.pdf, accessed March 29, 2019.

Cotterill, R.W. (1994), "Scanner data: new opportunities for demand and competitive strategyanalysis", Agricultural and Resource Economics Review, Vol. 23 No. 2, pp. 125-139.

Comparison shopping website - Wikipedia. https://en.wikipedia.org/wiki/Comparison_shopping_website (accessed:Sep. 28, 2021).

Degeratu, A.M., Rangaswamy, A. and Wu, J. (2000), "Consumer choice behavior in online andtraditional supermarkets: the effects of brand name, price, and other search attributes",International Journal of Research in Marketing, Vol. 17 No. 1, pp. 55-78.

Dingledine, R., Mathewson, N. and Syverson, P. (2004), Tor: The Second-Generation Onion Router,US Naval Research Laboratory, Washington DC.

Edelman, B. (2012), "Using internet data for economic research", Journal of Economic Perspectives, Vol. 26 No. 2, pp. 189-206.

Diamond, P. A., "A Model of Price Adjustment," Journal of Economic Theory 3 (1971), 156–68. 1971

De Cornière, A., and G. Taylor, "A Model of Biased Intermediation," The RAND Journal ofEconomics 50 (2019), 854–82.

De los, Santos, B., A. Hortaçsu, and M. R. Wildenbeest, “Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior,” American Economic Review 102 (2012),2955–80.

Dwijen Rudrapal, Amitava Das, and Baby Bhattacharya. Measuring semantic similarity for bengali tweets using wordnet. In Proceedings of the International ConferenceRecent Advances in Natural Language Processing, pages 537–544, 2015.

D. Fudenberg, “Knife-Edge or Plateau: When Do Market Models Tip?,” Quarterly Journal ofEconomics 118 (2003), 1249–78.

Edelman, B. G., and J. Wright, “Price Coherence and Excessive Intermediation,” Quarterly Journal of Economics 130 (2015), 1283-1328.

Eloisa Vargiu and Mirko Urru. Exploiting web scraping in a collabora-tive filtering-basedapproach to web advertising. Artif. Intell. Research, 2(1):44–54, 2013.

Ehrenberg, A.S.C., Goodhardt, G.J. and Barwise, T.P. (1990), “Double jeopardy revisited”, Journal of Marketing, Vol. 54 No. 3, pp. 82-91.

Einav, L., Farronato, C., Levin, J. and Sundaresan, N. (2018), “Auctions versus posted prices in online markets”, Journal of Political Economy, Vol. 126 No. 1, pp. 178- 215.

Ellison, G. and Ellison, S.F. (2009), “Search, obfuscation, and price elasticities on the internet”, Econometrica, Vol. 77 No. 2, pp. 427-452.

Esther Shaulova and Lodovica Biagi. ecommerce report 2019

E. Brynjolfsson and M. D. Smith, Frictionless commerce? A comparison of internet and conventional retailers, Management Science, vol. 46, no. 4, pp. 563-585, 2000.

E. K. Clemons, I. H. Hann and L.M. Hitt, Price dispersion and differentiation in online travel: Anempirical investigation, Management Science, vol. 48, no. 4, pp. 534- 549, 2002.

E. Honka, Quantifying search and switching costs in the US auto insurance industry, The RAND Journal of Economics, vol. 45, no. 4, pp. 847-884, 2014.

E. Orlov, How does the internet influence price dispersion? Evidence from the airline industry,Journal of Industrial Economics, vol. 59, no. 1, pp. 21-37, 2011.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, VincentMichel, Bertrand Thirion, OlivierGrisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit- learn: Machine learning in python. Journal of machine learning research, 12(Oct):2825–2830, 2011.

Fankar Armash Aslam, Hawa Nabeel Mohammed, Jummal MusabMohd, Murade Aaraf Gulamgaus, and PS Lok. Efficient way of web development using python and flask.International Journal of Advanced Research in Computer Science, 6(2), 2015.

Fedoseeva, S., Grein, T. and Herrmann, R. (2017), “How German online retailers price foods: an empirical analysis for chocolate products”, International Journal on Food System Dynamics, Vol. 8 No. 1, pp. 32-44.

FTC, “The Effect of Mortgage Broker Compensation Disclosures on Consumers and Competition: A Controlled Experiment,” (2004), Report Retrieved from https://www.ftc.gov/sites/default/files/documents/reports/effect-mortgage-brokercompensation-disclosures-consumers-and-competition-controlled-experiment/030123mortgagefullrpt.pdf, accessed April 6, 2019.

F. Ancarani and V. Shankar, Price levels and price dispersion within and across multiple retailer types: Further evidence and extension, Journal of the Academy of Marketing Science,vol. 32, no. 2, pp. 176 187, 2004.

F.S. Morton, F. Zettelmeyer and J. Silva-Risso, Consumer information and discrimination: Does the internet affect the pricing of new cars to women and

minorities? Quantitative Marketing and Economics, vol. 1, no. 1, pp. 65-92,2003.

Gorodnichenko, Y. and Talavera, O. (2017), “Price setting in online markets: basic facts, international comparisons, and cross-border integration”, American Economic Review, Vol. 107 No. 1, pp. 249-282.

Gorodnichenko, Y., Sheremirov, V. and Talavera, O. (2018), “Price setting in online : doesIT click?”, Journal of the European Economic Association, Vol. 16 No. 6, pp. 1764-1811.

Gorodnichenko, Y., V. Sheremirov, and O. Talavera, “Price Setting in Online Markets: Does IT Click?,” Journal of the European Economic Association 16 (2018), 1764–811.

Gomes, R., and A. Mantovani, “Regulating Platform Fees under Price Parity,” Working Paper University of Toulouse, 2020.

Grewal, D., Ailawadi, K.L., Gauri, D., Hall, K., Kopalle, P. and Robertson, J.R. (2011), “Innovations in retail pricing and promotions”, Journal of Retailing, Vol. 87 S1, pp. S43-S52.G. Ellison and S. Ellison, Search, obfuscation, and price elasticities on the internet, Econometrica, vol. 77, no. 2, pp. 427-452, 2009.

G. Lewis, Asymmetric information, adverse selection and online disclosure: The case of eBaymotors, American Economic Review, vol. 101, no. 4, pp. 1535-1546, 2011.

Hagiu, A., “Two-Sided Platforms: Product Variety and Pricing Structures,” Journal of Economics & Management Strategy 18 (2009), 1011–43.

Hemenway, K. and Calishain, T. (2004), Spidering Hacks, O’Reilly Publishing, Sebastopol, CA, ISBN 0596005776.

Herrmann, R., Moeser, A. and Weber, S.A. (2005), “Price rigidity in the German grocery-retailing sector: scanner-data evidence on magnitude and causes”, Journal of Agricultural & Food Industrial Organization, Vol. 3 No. 1, pp. 1-37.

Hirschey, J.K. (2014), “Symbiotic relationships: pragmatic acceptance of data scraping”Berkeley Technology Law Journal, Vol. 29 No. 4, pp. 897-927.

Honka, E., and P. Chintagunta, “Simultaneous or Sequential? Search Strategies in the US AutoInsurance Industry,” Marketing Science 36 (2016), 21–42.

Inderst, R., and M. Obradovits, “Loss Leading with Salient Thinkers,” RAND Journal of Economics 51 (2020), 260–78.

Ireland, N. J., “The Provision of Information in a Bertrand Oligopoly,” Journal of Industrial Economics 41 (1993), 61–76.

Janssen, M., and S. Shelegia, “Consumer Search and Double Marginalization,” American Economic Review 105 (2015), 1683–710.

Janssen, M. C., and J. L. Moraga-González, “Strategic Pricing, Consumer Search and the Number of Firms,” Review of Economic Studies 71 (2004), 1089–118.

Jawahire Nakash, Shaikh Anas, Siddiqi Muzammil Ahmad, Ansari Mohd. Azam, and Tabrez Khan. Real time product analysis using data mining. International Journal of Advanced Research in Computer Engineering Technology, Mar 2015.

Joeran Beel, Stefan Langer, Marcel Genzmehr, Bela Gipp, Corinna Breitinger, and Andreas Nürnberger. Research paper recommender system evaluation: a quantitative literature survey. In Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation, pages 15–22, 2013.

Johansen, B. O., and T. Vergé, “Platform Price Parity Clauses with Direct Sales,” Working Paper University of Bergen, 2016.

Johnen, J., and D. Ronayne, “The Only Dance in Town: Unique Equilibrium in a Generalized Model of Price Competition,” forthcoming, Journal of Industrial Economics (2020).

Johnson, J., “The Agency Model and MFN Clauses,” Review of Economic Studies 84 (2017),1151–85.

Jensen, R. (2007), “The digital provide: information (technology), market performance, and welfare in the South Indian fisheries sector”, The Quarterly Journal of Economics, Vol.122 No. 3, pp. 879-924.

Joydeep Bhattacharjee. Working with data. In Practical Machine Learning with Rust, pages 141–186. Springer, 2020.

Jun Ye. Cosine similarity measures for intuitionistic fuzzy sets and their applications. Mathematical and computer modelling, 53(1-2):91–97, 2011.

J. Alba, J. Lynch, B. Weitz, C. Janiszewski, R. Lutz, A. Sawyer, and S. Wood, Interactive homeshopping:Consumer, retailer, and manufacturer incentives to participate in electronic marketplaces, Journal of Marketing, vol. 61, no. 3, pp. 38-53, 1997.

J. Berger, A. T. Sorensen and S. J. Rasmussen, Positive effects of negative publicity: When negative reviews increase sales, Marketing Science, vol. 29, no. 5, pp. 815-827, 2010.

J. Boivin, R. Clark and N. Vincent, Virtual borders, Journal of International Economics, vol. 86,no. 2, pp. 327-335, 2012.

J. Chevalier and A. Goolsbee, Measuring prices and price competition online: Amazon.com and BarnesandNoble.com, Quantitative Marketing and Economics, vol. 1, no. 2, pp. 203-222,2003.

J. G. Lynch and D. Ariely, Wine online: Search costs affect competition on price, quality, and distribution, Marketing Science, vol. 19, no. 1, pp. 83-103, 2000.

J.E. Stiglitz, Equilibrium in Product Markets with Imperfect Information, American Economic Review, vol. 69, no.2, pp. 339-345, 1979.

J.E. Stiglitz, Competition and the number of firms in a market: Are duopolies more competitive than atomistic markets? Journal of Political Economy, vol. 96, no. 5, pp. 1041-1061, 1987.

J. Lindsey-Mullikin and D. Grewal, Imperfect information: The persistence of price dispersion on the web, Journal of the Academy of Marketing Science, vol. 34, no. 2, pp. 236 243, 2006.

J. Perloff, V. Suslow and P. Seguin, Higher Prices from Entry: Pricing of Brand-Name Drugs . Mimeo/Universityof California, Berkeley, 2006.

J.P. Bailey, Electronic commerce: Prices and consumer issues for three products: Books, compact discs andsoftware, OECD Digital Economy Papers 32, OECD Publishing, Paris,1998.

J. Morgan, “Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets,” American Economic Review 91 (2001), 454– 74.

J. R. Brown and A. Goolsbee, Does the internet make markets more competitive? Evidencefrom the life insurance industry, Journal of Political Economy, vol. 110, no. 3, pp. 481-507. 2002.

J. Vickers, “Patterns of Competitive Interaction,” Working Paper University of Oxford,2020.

J. Waldfogel and L. Chen, Does information undermine brand? Information intermediary useand preference for branded web retailers, Journal of Industrial Economics, vol. 54,no.4, pp. 425-449, 2006.

J.Y. Bakos, Reducing buyer search costs: Implications for electronic marketplaces, Management Science, vol.43, no. 12, pp. 1676-1692, 1997.

J. Zhou, “Paying for Prominence,” Economic Journal 121 (2011), F368–F395.

Kali Pradeep, I Bhagyasri, and P Praneetha. E-commerce with backbone of data mining. International Journal of Engineering and Technical Research, 2:460, 10 2018.

Karle, H., M. Peitz, and M. Reisinger, “Segmentation versus Agglomeration: Competitionbetween Platforms with Competitive Sellers,” Journal of Political Economy 128 (2020),2329–74.

Kienle, H.M., German, D. and Müller, H.A. (2004), “Legal concerns of web site reverse engineering”, Web Site Evolution, Conference Proceedings of the 6th International Workshop on Web Site Evolution(WSE 2004) – Testing, Chicago, IL, 11 September, pp. 41-50, doi: 10.1109/WSE.2004.10000.

Kultti, K., T. Takalo, and O. Vähämaa, “Intermediation in a Directed Search Model,” WorkingPaper Bank of Finland, 2019.

K. Clay, R. Krishnan and E. Wolff, Prices and price dispersion on the web: Evidence from the online book industry, Journal of Industrial Economics, vol. 49, no. 4, pp. 521-539, 2001.

K. Clay, R. Krishnan, E. Wolff, and D. Fernandes, Retail strategies on the web: Price and non-price competition in the online book industry, Journal of Industrial Economics, vol. 50,no.3, pp. 351-367, 2002.

K. L. Wakefield and J. J. Inman, Who are the price vigilantes? An investigation of differentiatingcharacteristics influencing price information processing, Journal of Retailing, vol. 69, no.2, pp. 216-233, 1993.

L. Kutlu, Limited memory consumers and price dispersion, Review of Industry Organization, vol.46, no. 4, pp.349-357, 2015.

L. Zhang, “The Unique Equilibrium in a Model of Sales with Costly Advertising,” Eco- nomicsLetters 124 (2014), 457–60.

K. Zhao, X. Zhao and J. Deng, Online price dispersion revisited: How do transaction prices differfrom listing prices? Journal of Management Information Systems, vol. 32, no. 1, pp. 261290, 2015.

Massimino, B. (2016), “Accessing online data: web-crawling and information-scraping

techniques to automate the assembly of research data”, Journal of Business

Logistics, Vol. 37 No. 1, pp. 34-42.

McAfee, R. P., “Endogenous Availability, Cartels, and Merger in an Equilibrium Price Dispersion,”Journal of Economic Theory 62 (1994), 24–47.

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest, “Prices and Heterogeneous Search Costs,” The RAND Journal of Economics 48 (2017), 125–46.

Morgan, J., H. Orzen, and M. Sefton, “An Experimental Study of Price Dispersion,” Games and Economic Behavior 54 (2006), 134–58.

Myatt, D., and D. Ronayne, “A Theory of Stable Price Dispersion,” Working Paper University of Ox-ford, 2019.

M. A. Kwarteng, A. B. Jibril, E. Botha, and C. N. Osakwe, “The Influence of Price Comparison Websites on Online Switching Behavior: A Consumer Empowerment Perspective.” Lecture Notes in Computer Science, pp. 216-227, 2020, doi: 10.1007/978-3-030-44999- 5_18.

M. C. W. Janssen and José, Luis Moraga-González, Strategic pricing, consumer search and thenumber of firms, Review of Economic Studies, vol. 71, no. 4, pp. 1089-1118, 2004.

M. Dinerstein, L. Einav, J. Levin, and N. Sundaresan, Consumer price search and platform design in internet commerce, American Economic Review, vol. 108, no. 7 pp.1820-1859, 2018.

M. R. Baye, J. Morgan and P. Scholten, Price dispersion in the small and in the large: Evidencefrom an internet price comparison site, Journal of Industrial Economics, vol. 52, no. 4, pp. 463-496, 2004.

M. R. Wildenbeest, “Comparison Sites,” in M. Peitz and J. Waldfogel, eds., The Oxford Handbookof the Digital Economy (Oxford, England: Oxford University Press, 2012),224–253.

Nakash, J., Anas, S., Ahmad, S.M., Azam, A.M. and Khan, T. (2015), “Real time product analysis using data mining”, International Journal of Advanced Research in ComputerEngineering& Technology (IJARCET), Vol. 4 No. 3, pp. 815-820.

Nielsen (2015), “The future of grocery: E-commerce, digital technology and changing shoppingpreferences around the world”, available at: www.nielsen.com/wp- content/uploads/sites/3/2019/04/nielsen-global-e-commerce-new-retail-report- april-2015.pdf (accessed 24 September 2019).

Nielsen (2017), “What’s in-store for online grocery”, available at: www.nielsen.com/wp- content/uploads/sites/3/2019/04/nielsen-global-connected-commerce-report- january-2017.pdf (accessed24 September 2019)

N. UzZaman and M. Khan. A double metaphone encoding for bangla and its application in spellingchecker. In 2005 International Conference on Natural Language Processing and Knowledge Engineering, pages 705–710, Oct 2005.

Oliver Wyman (2014), “AmazonFresh in the U.S.”, available at: www.oliverwyman.de/content/dam/oliver-wyman/global/en/2014/aug/OW_AmazonFresh_ENG.pdf (accessed 24 September 2019).

Oliver Wyman (2018), “AmazonFresh undercuts supermarkets by 10–20 percent”,4 January,availableat: www.oliverwyman.com/our-expertise/insights/2018/jan/amazon-fresh-undercuts-supermarkets-by-1020-percent.html (accessed 24 September 2019).

Onur Bodur, Noreen M. Klein, Neeraj Arora " Online Price Search: Impact of Price ComparisonSites on Offline Price Evaluations

Pan, X., Ratchford, B.T. and Shankar, V. (2002), “Can price dispersion in online markets be explained by differences in e-tailer service quality?”, Journal of the Academy of Marketing Science, Vol. 30No. 4, pp. 433-445.

Papin, J., Andrès, F. and d’Orazio, L. (2018), “A method to build a geolocalized food price time series knowledge base analyzable by everyone”, First Latin America Data ScienceWorkshop, 27 August, Rio de Janeiro, available at: http://ceur- ws.org/Vol- 2170/paper13.pdf (accessed 3 September 2019).

Pigatto, G., Machado, J.G.D.C.F., Negreti, A.D.S. and Machado, L.M. (2017), “Have you chosenyour request? Analysis of online food delivery companies in Brazil”, British Food Journal, Vol. 119 No. 3, pp. 639-657.

P. Lünnermann and L. Wintr, Price stickiness in the U.S. and Europe revisited: Evidence from internet prices, Oxford Bulletin of Economics and Statistics, vol. 73, no. 5, pp. 593-621, 2011.

P. Pereira, Do lower search costs reduce prices and price dispersion? Information Economicsand Policy, vol. 17,no. 1, pp. 61-72, 2005.

P. Scholten, “Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site,” Journal of Industrial Economics 52 (2004), 463–96. Information, Search, and Price Dispersion,” in T. Hendershott, ed., Handbook on Economics andInformation Systems, Volume 1 (Amsterdam: Elsevier, 2006) 323–375.

Reisinger, M., “Two-Part Tariff Competition between Two-Sided Platforms,” European Economic Review 68 (2014), 168–80.

Rigby, N. (2018), “Amazon grocery 2017 review: one click retail study”, 16 January, available at:http:// oneclickretail.com/amazon-grocery-2017-review/ (accessed 24 January 2019).

Rochet, J.-C., and J. Tirole, “Platform Competition in Two-Sided Markets,” Journal of the European Economic Association 1 (2003), 990–1029.

Ronayne, D., and G. Taylor, “Competing Sales Channels,” Working Paper University of Oxford, 2020. Rosenthal, R. W., “A Model in Which an Increase in the Number of Sellers Leadsto a Higher Price,” Econometrica 48 (1980), 1575–79.

Rubinstein, A., and A. Wolinsky, “Middlemen,” Quarterly Journal of Economics 102 (1987)58193.

R. Thomadsen, Product positioning and competition: the role of location in the fast food industry, Marketing Science, vol. 26, no. 6, pp. 792-804, 2007.

Salop, S., and J. Stiglitz, “Bargains and Ripoffs: A Model of Monopolistically Competitive PriceDispersion,” Review of Economic Studies 44 (1977), 493–510.

Shilony, Y., “Mixed Pricing in Oligopoly,” Journal of Economic Theory 14 (1977), 373– 88.

Shpanya, A. (2013), “Five trends to anticipate in dynamic pricing [Blog post]”, Retail Touchpoints, 14May, available at: www.retailtouchpoints.com/features/executive- viewpoints/5-trends-to-anticipate-in-dynamic-pricing (accessed 24 January 2019).

Silver, M. and Heravi, S. (2001), “Scanner data and the measurement of inflation”, The

Economic Journal, Vol. 111 No. 472, pp. 383-404.

Smith, M.D., Bailey, J. and Brynjolfsson, E. (2001), “Understanding digital markets: review andassessment”, MIT Sloan School of Management Working Paper No. 4211-01,Cambridge, MA.

Stahl, D. O., “Oligopolistic Pricing with Sequential Consumer Search,” American EconomicReview 79 (1989), 700–12.

“Oligopolistic Pricing with Heterogeneous Consumer Search,” International Journal of Industrial Organization 14 (1996), 243–68.

Stigler, G. J., “The Economics of Information,” Journal of Political Economy 69 (1961), 213–25.

S. C. Salop, Information and monopolistic competition, American Economic Review, vol. 66, pp.240-245, 1976.

Tang, F.-F. and Xing, X. (2001), “Will the growth of multi-channel retailing diminish the pricingefficiency of the web?”, Journal of Retailing, Vol. 77 No. 3, pp. 319-333

The HubSpot Marketing Blog.” https://blog.hubspot.com/marketing (accessed: Sep. 28, 2021).

Tiago Pessoa, Raul Medeiros, Thiago Nepomuceno, Gui-Bin Bian, V.H.C. Albuquerque, and Pedro Pedrosa Filho. Performance analysis of google colaboratory as a tool for accelerating deep learning applications. IEEE Access, PP:1–1, 10 2018.

T. Blake, S. Moshary, K. Sweeney, and S. Tadelis, Price salience and product choice, National Bureau of Economic Research, Cambridge, MA, Working Paper 25186, 2018.

T. Hossain and J. Morgan, Plus shipping and handling: Revenue (non) equivalence in field experiments on eBay, Advances in Economic Analysis and Policy, vol. 5, no. 2, 2006.

T. Scitovsky, Ignorance as a source of oligopoly power, American Economic Review, vol. 40, pp. 48-53, 1950.

Riya Shah, Karishma Pathan, Anand Masurkar, and Shweta Rewatkar. Comparison of ecommerce products using web mining.

Shakra Mehak, Rabia Zafar, Sharaz Aslam, and Sohail Masood Bhatti. Exploiting filtering approach with web scrapping for smart online shopping penny wise a wise tool for online shopping. In 2019 2nd International Conference on Computing, Mathematics and EngineeringTechnologies (iCoMET), pages 1–5. IEEE, 2019.

S. Ba, J. Stallaert and Z. Zhang, Research note-Online price dispersion: A game-theoreticperspective and empirical evidence, Information Systems Research, vol. 23, no. 2, pp. 575-592, 2012.

Teh, T.-H., and J. Wright, "Intermediation and Steering: Competition in Prices and Commissions,"forthcoming, American Economic Journal: Microeconomics (2020).

Toto, C.S. and Buffington, K. (2016), "How binding is your browsewrap agreement? [Blog post]",Pillsbury Winthrop Shaw Pittman LLP, 6 June, available at: www.lexology.com/library/detail.aspx?g=4b4b93da-c40a-4724-916c3bdd9011698d (accessed 24 September 2019).

Vanden Broucke, S. and Baesens, B. (2018), Practical Web Scraping for Data Science, Apress, New York NY, ISBN-13: 978-1-4842-3582-9.

Wang, C., and ———, "Search Platforms: Showrooming and Price Parity Clauses," RAND Journal of Economics 51 (2020), 32–58.

Wang, O. and Somogyi, S. (2018), "Consumer adoption of online food shopping in China",British Food Journal, Vol. 120 No. 12, pp. 2868-2884.

X. Pan, B. T. Ratchford and V. Shankar, Can price dispersion in online markets be explained by differences in e-tailer service quality? Journal of the Academy of Marketing Science, vol. 30, no. 4, pp. 433-445, 2002.

X. Pan, B. T. Ratchford and V. Shankar, Price dispersion on the internet: A review and directions for future research, Journal of Interactive Marketing, vol. 18, no. 4, pp.116-135, 2004.

Yang, Y., Wilson, L.T. and Wang, J. (2010), “Development of an automated climatic data scraping, filtering and display system”, Computers and Electronics in Agriculture, Vol71 No. 1, pp. 77-87.

Ying Ming-Hsiung and Hsu Yeh-Yen. A commodity search system for online shopping based on ontology and web mining. In IET Conference Proceedings. The Institution of Engineering & Technology, 2014.

Y. Chen, G. Iyer and A. Pazgal, Limited memory, categorization, and competition, Marketing Science, vol. 29, no.4, pp. 650-670, 2010.

Y. Gorodnichenko, V. Sheremirov and O. Talavera, Price setting in online markets: Does it click? Journal of the European Economic Association, vol.16, no.6, pp. 1764-1811, 2018.

Zhu, H. and Madnick, S.E. (2010), “Legal challenges and strategies for comparison shoppingand data reuse”, Journal of Electronic Commerce Research, Vol. 11 No. 3, pp. 231-239.

Varian, H. R., “A Model of Sales,” American Economic Review 70 (1980), 651–59.6 w ways to strengthen your online presence to benefit your business. https://www.experian.co.uk/blogs/latest-thinking/small-business/6-benefits-to-an online-presence/ (accessed: Sep. 28, 2021).