

**M. S. Ramaiah University of Applied Sciences
School of Social Sciences**

Department of Data Sciences and Analytics



**RAMAIAH
UNIVERSITY**
OF APPLIED SCIENCES

DATA SCIENCES AND ANALYTICS



AUTOMATED FOOD PRICE AND INGREDIENTS COMPARISON

Submitted by: PRIYANKA

Reg. No.: 21SSDS415019

Programme: B.SC. (HONS) DATA SCIENCES AND ANALYTICS

Supervisor: PROF. DR. SUDHARSANA V IYENGAR AND PROF. SURESH DARLA

May-2024

SCHOOL OF SOCIAL SCIENCES

M. S. RAMAIAH UNIVERSITY OF APPLIED SCIENCES

Bengaluru - 560054

DECLARATION

“AUTOMATED FOOD PRICE AND INGREDIENTS COMPARISON”

The dissertation is submitted in partial fulfilment of academic requirements for the Bachelors’ Degree of M. S. Ramaiah University of Applied Sciences in the Department of Data Sciences and Analytics. This dissertation is a result of my own investigation. All sections of the text and results, which have been obtained from other sources, are fully referenced. I understand that cheating and plagiarism constitute a breach of university regulations, hence this dissertation has been passed through plagiarism check.

Signature:

Name of the student: Priyanka

Reg. No: 21SSDS415019

Date:

**M. S. Ramaiah University of Applied Sciences
School of Social Sciences**



CERTIFICATE

This is to certify that the Dissertation titled **“AUTOMATED FOOD PRICE AND INGREDIENTS COMPARISON”** is a bona fide record of the work carried out by PRIYANKA, Reg. No. 21SSDS415019 in partial fulfilment of requirement for the award of Bachelor’s Degree of M. S. Ramaiah University of Applied Sciences in the Department of Data Sciences and Analytics.

May-2024

Dr. Sudarshana V Iyengar

Professor

Academic Supervisor

Data Sciences and Analytics

School of Social Sciences

Dr. Tapasi Ghosh

HOD

Data Sciences and Analytics

School of Social Sciences

Sri. Gurucharan Gollerkeri

Director

Data Sciences and Analytics

School of Social Sciences

ACKNOWLEDGEMENT

I would like to express my gratitude to Professor Dr. Sudharsana V Iyengar and Professor Suresh Darla, Department of Data Science and Analytics, School of Social Sciences for the indefatigable patience, expertise, understating support they have provided as supervisor throughout this project. I would also like to extend special thanks to all the professors in the Data Science and Analytics department who took time out of their busy days offering us directions and advice throughout this project and being helpful through every step that improved the quality of the final work.

A huge thankyou to my teammates and their effort in helping me push this project over the finishing line. Every member did their job diligently and the team executed well. Their collaborative spirit and willingness to go the extra mile made this project a truly rewarding experience.

Priyanka

UNDER THE GUIDANCE OF:

Dr. Sudharsana V Iyengar

Dr. Sudharsana a physicist with a Ph.D. and a passion for unraveling the universe's mysteries. With expertise in theoretical physics, coding, and ecological modeling, she combines scientific knowledge with data analysis skills (machine learning, deep learning) to explain complex scientific concepts to a wide audience.

Mr. Suresh Darla

Mr. Suresh Darla, is an Electronics and Communication Engineer from Rajiv Gandhi University of Knowledge Technologies, Nuzvid, have lots of experience and is currently working at Ramaiah University of Applied Sciences, he has been instrumental in mentoring our project, providing invaluable assistance in web development and data analysis which includes HTML, CSS, JavaScript, Web Scraping, Python and Python Flask.

CONTENT

Chapter No.	Chapter Name	Page Number
1	Abstract	6
2	Introduction	7
3	Project Overview	8-10
4	Methodology	11-19
5	Individual Progress	20-22
6	Conclusion	23
7	References	24

Chapter 1

Abstract

Convenience and a wider selection are provided by the growing online grocery sector, but evaluating costs and nutritional value across platforms is still difficult. Customers find it difficult to choose the most economical choice when weighing the potential health effects of different food items. In order to close this gap, this study suggests a data science-driven system that examines food items from several e-commerce platforms.

For a given product, the system will gather information on prices from different e-commerce websites, ingredient lists for nutritional facts. A price comparison will show you which option is the most economical. By highlighting potentially harmful ingredients like high sugar content or preservatives, ingredient analysis will enable consumers to make decisions that are in line with their dietary requirements. Lastly, from product descriptions, the system will collect and compare important nutritional data (calories, protein, and vitamins). Customers will be able to quickly grasp the nutritional profile of various options and make wise decisions thanks to the data's user-friendly, visual presentation on the dashboard.

This research provides a data-driven tool for knowledgeable food shopping, which advances the field of consumer decision support systems. Customers may optimize their grocery purchases based on budget and health goals through the dashboard, which provides a comprehensive picture of price, ingredients, and nutrition across e-commerce platforms.

Chapter 2

Introduction

With the ability to browse aisles from the comfort of your sofa, access a broader range of products, and avoid congested stores, online grocery shopping has become increasingly popular. But this ease has a price: it can be difficult to compare food products across a variety of e-commerce sites. Two main issues that consumers frequently have to balance are cost and health. Finding the best deal while also assessing the nutritional content of different items can be a laborious and time-consuming process. The overwhelming amount of information included in individual product listings frequently makes it difficult to make well-informed decisions.

This project proposes a data science-driven solution that enables customers to make knowledgeable decisions about their online grocery purchases, taking on this major obstacle head-on. We acknowledge the growing movement toward healthy lives, in which dietary decisions are quite important. It's difficult to analyse product costs and ingredients while remaining within a budget, and the current state of internet shopping leaves a lot to be desired. By offering a simple solution that demystifies the world of online grocery buying, this project seeks to close this gap.

Many customers struggle with the question of how to eat healthily while still giving in to personal tastes and urges. It's common to believe that affordable selections are less nutritious or that healthy options are intrinsically more expensive. This breeds tension and can be a significant impediment to leading a healthy lifestyle. When it comes to the food that they place on their plates, consumers demand transparency and information. They require a tool that makes it easier for them to analyse ingredients and gives them the power to make wise decisions without sacrificing their financial constraints or taste preferences.

The foundation of our suggested remedy is the idea that "knowledge is power." Our goal is to enable customers to make knowledgeable and health-conscious decisions by offering a thorough and user-friendly platform for comparing ingredients. Our system will examine product details from different e-commerce websites by utilizing data science.

Chapter – 3

Project Overview

AUTOMATED FOOD PRICE & INGREDIENTS COMPARISON Website:

<https://yumyield.000webhostapp.com/https://drive.google.com/file/d/1YklbPE5z3hekVLzObgKXHGYcjWqB8EmM/view>

Project Idea

We aim to build a system that compares the prices and ingredients of food products from different eCommerce websites and prepares a dashboard containing insights on the nutritional information list. A buyer can then analyze the list to decide which website they should purchase the product from and suggest healthier alternatives.

With the growth of e-commerce platforms, consumers face difficulties comparing prices, ingredients, and nutritional value across different websites. This project proposes a data science-driven system that serves as a "Food Assistant" for online shoppers. The project addresses the growing need for informed decision-making in online grocery shopping.

Objective

Many of us are now drifting to a healthy lifestyle which is directly influenced by the food choices. One needs to constantly analyze product ingredients and prices to come across the perfect buying options, which can be sometimes misleading as well.

Problem that we are planning to tackle through this project: - How to maintain a healthy diet without cutting out the cravings and preferences of the individual? Solution: Knowledge is power, especially when it comes to food. Let's get empowered with ingredient comparisons.

The principal aim is enabling customers to make cost-effective and healthy grocery purchases by providing them with an intuitive dashboard. Three main categories will be covered by this dashboard's insights:

Price Comparison: Examine costs on several e-commerce sites to determine which food product selection is the most economical.

Analyze ingredient lists to identify ingredients that may be harmful, such as high sugar or preservatives. Customers are now more equipped to make decisions depending on their dietary requirements.

Nutritional Information Visualization: Take essential nutritional data (vitamins, protein, and calories) and compare it with other data from product listings. Users will be able to quickly comprehend the nutritional profile of various options thanks to the visual presentation of this data.

Data Type:

Encountered errors while trying to retrieve the data through web scraping tools like beautifulsoup and selenium from the required website like bigbasket. While, we could successfully scrape few of the required data from Amazon and zepto website. Our current system appears to be relying on static data, the project initially aimed to leverage real-time data. However, challenges were encountered during the real-time data scraping process.

Static data offers a stable foundation for analysis, unaffected by scraping bugs. Static data used for the website include: Zepto, Bigbasket and Amazon. These data were retrieved from Kaggle and other reliable platforms.

The dataset includes:

Name of the products, price of the products from each e-commerce platforms, ratings, and the ingredients included in the products. The products only include packed edible items.

Tools/Language/Platforms Used:

- Python
- Google Colab
- HTML
- CSS
- JavaScript
- Python Flask
- VS Code
- BeautifulSoup (BS4)
- Selenium
- WordPress
- Kissflow
- Open Food Facts
- Nutrionix
- Kaggle
- SQL Workbench
- MySQL

Application Logo:



Chapter – 4

Methodology

Data Acquisition:

First step in the methodology includes gathering information or data from several e-commerce websites about food products, including prices, ingredients, and nutritional values. Data collecting methods include online scraping and collecting the static data from Kaggle or any related community.

The primary goal was to collect comprehensive product data from these platforms using web scraping techniques. Selenium and Scrapy were utilized for this purpose, and the process is divided into sections for each website.

Zepto

Objective: To scrape product data from various categories such as biscuits, beverages, etc.

Tools Used: Selenium

Initial Challenges: Initially, the scraper could only retrieve approximately 20 products per category.

Improvements: By consulting online tutorials and blogs, the script was refined to scrape over 100 products per category.

Outcome: Successfully increased the data collection efficiency and volume, as evidenced by the attached screenshots.

Challenges Overcome:

Pagination Handling: Ensured the script navigated through multiple pages to gather more data.

Dynamic Content Loading: Managed the loading of dynamic content using appropriate wait conditions.

Big Basket

Objective: To scrape extensive product data from Big Basket.

Tools Used: Selenium

Initial Approach: Applied a similar code used for Zepto, assuming a uniform approach would suffice.

Challenges:

Sold Out Products: The scraper halted whenever it encountered a sold-out product, even if more products were available on the page.

Attempts to Resolve:

Error Handling: Implemented try-except blocks to skip sold-out products and continue scraping.

Failure to Overcome: Despite these efforts, the issue persisted, leading to incomplete data retrieval.

Alternative Approach:

Static Data Source: Due to persistent issues, data was sourced from Kaggle.

Data Preprocessing: Applied various preprocessing techniques to clean the dataset, removing irrelevant and duplicate entries.

Outcome: Achieved a clean, processed dataset from a static source, ensuring data integrity and usability.

Amazon

Objective: To scrape product data similar to Zepto and Big Basket.

Tools Used: Selenium and Scrapy (methodology and challenges similar to those encountered with Zepto and Big Basket).

Outcome: Not explicitly mentioned, but assumed to follow a similar pattern of iterative improvement and overcoming challenges.

This project encountered and addressed several challenges, particularly with dynamic content and error handling. While Zepto's data scraping was successfully optimized, Big Basket posed significant challenges that necessitated the use of static data from Kaggle. This report highlights the iterative nature of web scraping, requiring continuous learning and adaptation to overcome obstacles.

Data analysis:

To guarantee correctness and consistency, the gathered data is cleaned and pre-processed using jupyter notebook and google colab. We used data science methods including statistical analysis for pricing and nutritional data, and text analysis for ingredient lists and to ensure the quality and consistency of the datasets.

Key Findings from the analysis:

Product Variety: A wide range of products across various categories like biscuits, beverages and snacks.

Nutritional Information: Detailed nutritional content for each product, including calories, protein, fat, and carbohydrate content.

Pricing Trends: Price variations across different brands and product sizes.

Big Basket

Despite initial difficulties with scraping data from Big Basket, static datasets from Kaggle were used to supplement the analysis.

Key Findings:

Diverse Product Categories: Included fresh produce, packaged goods, and beverages.

Nutritional Data: Available for most products, enabling a detailed nutritional analysis.

Price Comparison: Static data allowed for comparison across different time periods and product categories.

Challenges:

Scraping Issues: The scraper stopped at sold-out products, which necessitated the use of static datasets.

Data Cleaning: Extensive preprocessing was required to remove irrelevant and duplicate entries.

Amazon

Data was collected from Amazon using similar methodologies as Zepto and Big Basket.

Key Findings:

Wide Range of Products: Included both branded and generic food items.

Customer Reviews:

Provided insights into consumer preferences and satisfaction levels.

Datasets

The two datasets used for this analysis are:

1. Food Products Dataset

2. Food Ingredients Dataset

Dataset 1: Food Products

Description: Contains information about various food products, including product names, categories, prices, ratings, and availability.

Attributes:

- Product Name
- Category
- Price
- Rating
- Availability

Dataset 2: Food Ingredients

Description: Lists ingredients for various food products, including their nutritional information and potential allergens.

Attributes:

- Product Name
- Ingredients
- Nutritional Information (calories, fat, protein, etc.)
- Allergens

Data Preprocessing

Cleaning and Preparation

1. Missing Values: Identified and handled missing values through imputation and removal of irrelevant records.
2. Duplicate Entries: Removed duplicate entries to ensure data integrity.
3. Standardization: Standardized categorical values for consistency.

Analysis

1. Product Distribution

Category-Wise Distribution

- Objective: To analyze the distribution of products across different categories.
- Method: Used frequency distribution and visualization techniques (bar charts).
- Findings: Certain categories like snacks and beverages had higher product counts, indicating market trends and consumer preferences.

Price Distribution

- Objective: To analyze the price range of products within each category.
- Method: Used box plots to visualize price spread and identify outliers.
- Findings: Significant price variation was observed within categories, with premium products skewing the average prices.

2. Ingredient Analysis

Nutritional Information

- Objective: To assess the nutritional profile of products.
- Method: Statistical analysis of nutritional attributes (calories, fat, protein, etc.).

3. Health Implications

Allergen Analysis

- Objective: To identify potential allergens, present in food products.
- Method: Extracted allergen information and analyzed their presence across different categories.
- Findings: Common allergens like nuts, gluten, and dairy were frequently present, highlighting the need for clear labeling and consumer awareness.

Visualizations

- 1.Bar Charts: Displaying the distribution of products across categories.
- 2.Box Plots: Illustrating price distribution within categories.
- 3.Word Clouds: Visualizing the frequency of common ingredients.
- 4.Histograms: Showing the distribution of nutritional values across products.

The analysis provided valuable insights into the distribution, pricing, and nutritional composition of food products.

Key findings include:

- Market Trends: Categories like snacks and beverages dominate the market.
- Price Variation: Significant price differences within categories, influenced by product quality and brand positioning.
- Nutritional Concerns: High prevalence of calorie-dense, nutrient-poor products, raising health concerns.
- Allergens: Frequent presence of common allergens, emphasizing the need for proper labeling.

Attachments:

- Datasets:

<https://drive.google.com/file/d/1Z7k21vuBTPDIHE4QOfueZuZ2c5LMtGO/view?usp=drivesdk>

- <https://drive.google.com/file/d/1Z85lh-2GV-GtaB4SfdjdOufO-RUXjIQF/view?usp=drivesdk>

- Code:

https://drive.google.com/drive/folders/13Rb0u4f5oMaRczsPTdh7PKiwowAmAJf?usp=drive_link

Dashboard Design:

The project involved building a dynamic, user-friendly web application with a robust back-end and interactive front-end components created using HTML, CSS, JavaScript, and Python Flask.

Technologies Used:

- Python
- Front-end: HTML, CSS, JavaScript
- Back-end: Python Flask
- Database: SQLite, MySQL
- VS Code
- WordPress
- Kissflow

Development Process

1. Planning and Design

- Requirements Gathering: Defined the website's objectives, target audience, and core functionalities.
- Wireframing: Created wireframes to layout the structure and design of the web pages.
- Design Tools: Utilized tools such as Adobe XD for detailed design mockups.
- Team Assignment: Assigned roles and responsibilities to team members based on their expertise.

2. Front-end Development

- **HTML:** Structured the content of the web pages using semantic HTML elements.
- **CSS:** Styled the web pages with CSS to ensure a cohesive and attractive visual presentation. CSS frameworks like Bootstrap were used for responsive design.
- **JavaScript:** Added interactivity to the web pages with JavaScript. Implemented features such as user feedback mechanisms, search bars, filters and comparison tables.

3. Back-end Development

- **Flask Framework:** Set up the Flask framework to handle server-side logic. Flask was chosen for its simplicity and flexibility.
- **Templates:** Used Flask's Jinja2 templating engine to dynamically generate HTML pages based on user inputs and database queries.

4. Database Integration

- **Database Setup:** Configured the database to store user data, content, and other relevant information. The choice of database (e.g., SQLite for development, PostgreSQL for production) depended on the project's requirements.
- **CRUD Operations:** Implemented Create, Read, Update and Delete (CRUD) functionalities to manage database records.

5. Testing and Debugging

- **Unit Testing:** Wrote unit tests for both front-end and back-end components to ensure individual parts function correctly.
- **Integration Testing:** Tested the integration of different components to ensure they work together seamlessly.

Challenges Encountered:

Transitioned our project from using WordPress to a custom-built solution using UI components and Python Flask for enhanced flexibility and functionality.

Responsive Design: Ensuring the website looked and functioned well on various devices and screen sizes required careful planning and testing.

Cross-Browser Compatibility: Addressing inconsistencies in how different browsers render HTML/CSS.

Database Performance: Optimizing database queries to handle large volumes of data efficiently.

Collaboration

Maintaining an open communication with the team in development process to ensure design alignment with project goals and technical feasibility.

Providing technical guidance and support to other team members.

Attachments:

Link of the website we were working on: <https://yumyield.000webhostapp.com/>

Current website:

<https://drive.google.com/file/d/1YklbPE5z3hekVLzObgKXHGYcjWqB8EmM/view>

YouTube playlist:

<https://youtube.com/playlist?list=PLfqMhTWNBTte0PY9xunOzsP5kmYIz2Hu7i&si=OOzNwqP174r2buNa> <https://roadmap.sh/full-stack>

Code:

<https://drive.google.com/drive/folders/18Xj7LX3TErWA43nD2k1u60Xc9OXwgVT0>

Chapter – 5

Individual Progress

Identified the core objectives and functionalities of the website, focusing on user needs and business goals. Utilized wireframes to establish the basic layout and structure of the web pages, ensuring a clear vision before moving to development. Adobe XD was employed for creating detailed design mockups, providing a visual guide for the development team. Roles were assigned based on expertise, ensuring efficient task management and collaboration.



Used semantic HTML to structure content effectively, ensuring accessibility and SEO friendliness. Applied CSS for styling, with a focus on consistency and visual appeal. Bootstrap was integrated for responsive design, ensuring the site adapts well to different screen sizes. Implemented dynamic features like search bars, filters, and interactive elements to enhance user experience.

Chose Flask for its simplicity and flexibility, setting up the server-side logic to handle requests and responses. Leveraged Flask's Jinja2 templating engine for generating dynamic HTML content based on user inputs and data retrieved from the database.

Configured databases (SQLite for development, MySQL for production) to store and manage user data, content, and other relevant information. Implemented standard CRUD functionalities to manage database records, allowing for data manipulation through the web interface.



Developed unit tests for both front-end and back-end components, ensuring that individual parts worked correctly. Ensured seamless integration of various components, verifying that they worked together as intended.

Provided technical guidance and support to team members, fostering a collaborative and productive environment.


[Home](#)
[Product](#)
[About](#)
[Review](#)
[Blog](#)
[Contact](#)


BUY YUMYIELD RECOMENDED PRODUCTS ONLINE


Get Yours Now


[Home](#)
[Product](#)
[About](#)
[Review](#)
[Blog](#)
[Contact](#)


Limited Sales

Vegetables


Shop Now



Limited Sales

Fruits


Shop Now





Limited Sales


Assorted

Shop Now



OUR PRODUCTS


[Home](#)
[Product](#)
[About](#)
[Review](#)
[Blog](#)
[Contact](#)







Kisan Fresh Tomato Ketchup


200g

BigBasket Price- Rs.65 Zepto Price- Rs.65

★★★★★

Energy (Kcal) 295, Sugar(G) 8.18, Fiber(G) 21.52, Total Fat(G) 0.86, Saturated Fat(G) 0.02, Carbohydrate(G) 57.8, Protein(G) 3.48, Sodium(G) 2.32






Aavin Badam Milkshake

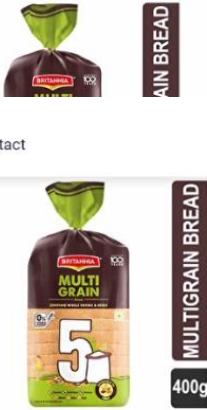
170ml

BigBasket Price- Rs.50 Zepto Price- Rs.54

★★★★★

Energy (Kcal) 93.3, Total Carbohydrates (G) 13.7, Lactose (G) 5.0, Sucrose (G) 8.7, Proteins (G) 2.9, Total Fat (G) 3.0, Minerals (Calcium) (Mg) 65






Britannia Multigrain Bread

400g

BigBasket Price- Rs.50 Zepto Price- Rs.60

★★★★★

Energy (Kcal) 265, Carbohydrate (G) 47, Sugars (G) 1.4, Dietary Fibre (G) 6.5, Protein (G) 9.6, Fat (G) 4.3

Challenges Encountered:

Migrated from a WordPress-based solution to a custom-built application, which provided greater flexibility but required careful planning and execution. Ensured the website was fully responsive, functioning well on various devices and screen sizes. Addressed inconsistencies in how different browsers render HTML/CSS, ensuring a consistent user experience across platforms. Optimized database queries to efficiently handle large volumes of data, improving the overall performance of the application.

Technologies Used:

- Programming Languages: Python
- Front-end Technologies: HTML, CSS, JavaScript (with frameworks like Bootstrap)
- Back-end Technologies: Python Flask, Jinja2
- Databases: SQLite (for development), PostgreSQL/MySQL (for production)
- Development Tools: VS Code, Adobe XD, WordPress, Kissflow

The project covered the entire development lifecycle, from requirements gathering to deployment, focusing on creating a robust and scalable web solution.

Chapter - 6

Conclusion

In conclusion, the development of a content comparison website like ours holds immense potential. It empowers shoppers with informed choices, drives sales for businesses, and creates a valuable profit tool for its owners. While building such a platform from scratch is a complex and time-consuming endeavor, our team's dedication has resulted in significant progress. We are incredibly grateful to our mentor, Prof. Sudharsana V Iyengar and Prof. Suresh Darla, for their invaluable guidance, support, and encouragement throughout this journey. Additionally, Prof. Darla's expertise in web development and web scraping using BeautifulSoup proved instrumental in navigating the initial data acquisition challenges. Their expertise and belief in our abilities helped us push beyond perceived limitations and solidify our commitment to this project.

As development continues, we remain focused on creating a user-friendly and comprehensive food price and ingredient comparison website. We are confident that this platform will ultimately benefit both consumers and businesses, making grocery shopping a more efficient and cost-effective experience for all.

References

1. <https://www.kaggle.com/datasets/devshahoff/zepto-dataset>
2. <https://www.kaggle.com/datasets/surajjha101/bigbasket-entire-product-list-28k-datapoints>
3. <https://drive.google.com/file/d/1YklbPE5z3hekVLzObgKXHGYcjWqB8EmM/view>
4. <https://youtube.com/playlist?list=PLfqMhTWNBT0PY9xunOzsP5kmYlz2Hu7i&si=OOzNwqP174r2buNa>
5. <https://roadmap.sh/full-stack>
6. <https://drive.google.com/drive/folders/18Xj7LX3TErWA43nD2k1u60Xc9OXwgVT0>
7. <https://archive.ics.uci.edu/ml/datasets/Online+Retail>
8. <https://drive.google.com/drive/folders/10VmxnBrocd49SDQTOX8gdVYQNJpwqbE>
9. <https://drive.google.com/file/d/1Z7k21vuBTPDIHE4QOfueZuZ2c5LMtGO/view?usp=drivesdk>
10. <https://drive.google.com/file/d/1Z85Ih-2GV-GtaB4SfdjdOufO-RUXjlQF/view?usp=drivesdk>
11. [https://drive.google.com/drive/folders/13Rb-0u4f5oMaRczsPTdh7PKiwowAmAJf?usp=drive link](https://drive.google.com/drive/folders/13Rb-0u4f5oMaRczsPTdh7PKiwowAmAJf?usp=drive_link)