**A Project Report**

*on*

**Fast-food Nutrition Statistics, Analysis and Prediction**

*carried out as part of the Foundations of Data Science [IT3101]*

by

**Ananda Chaturvedi**

**229302457**

**Vidushi Gupta**

**229302455**

*in partial fulfilment for the award of the degree* *of*

**Bachelor of Technology**

in

**Information Technology**



**School of Information, Security and Data Science**

**Department of Information Technology**

**MANIPAL UNIVERSITY JAIPUR**

**RAJASTHAN, INDIA**

**Nov 2024**

**CERTIFICATE**

Date 20/11/24

This is to certify that the project titled **Fast-food Nutrition Statistics, Analysis and Prediction** is a record of the bonafide work done by **Ananda Chaturvedi (229302457)** and **Vidushi Gupta** **(229302455)** submitted in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology (B. Tech) in **Information Technology** of Manipal University Jaipur, during the academic year 2024-25.

**Dr. Narendra Singh Yadav**

*Project Guide, Dept of Information Technology*

*Manipal University Jaipur*

**Dr. Pratistha Mathur**

*HOD, Dept of Information Technology*

*Manipal University Jaipur*

**ABSTRACT**

The main intention of this project would be to assess the nutritional data of some of the most popular fast-food chains and bring it to the purview of current trends with practices, along with deriving practical insights. The project will use practical and statistical methods and machine learning techniques to assess the nutritional content of fast foods, describe consumer preferences, and model the changes in their offerings over time. The idea behind this research would be to promote good practices in eating behaviors and in doing so be able to foster the health of the public by making the fast-food chains provide them with better choices.

**Objectives**

The project wants to analyze fast-food nutrition data and extract interesting patterns from it. Here are the main objectives:

**1.Nutritional Analysis**:

Understand the details of fast-food products concerning calories, fats, proteins, carbohydrates, sodium, or any other important aspect that would characterize what customers consume.

**2.Understanding Customer Preferences**

Discover patterns and trends revealed in the choices by customers of popular fast-food items whether these differ demographic or regional-wise.

**3.Predicting Future Trends**

Changing tastes and preferences because of social, cultural, and economic factors will soon be predicted with the help of machine learning tools-altering fast-food menus with healthier options or changes brought about by health regulations.

**4.Dietary Guidance**

Make available practical tips for healthier choices, or personalized menu recommendations for each dietary need or goal.

**5.Improving Data Access**

Strengthen the availability of nutritional data by making it more comprehensible and easier for consumers to use in making informed decisions about food.

**6.Impact on Public Health**

Support efforts at making fast food healthier in something stakeholders and policymakers can understand regarding trends and spur improved options for the customers.

By marrying statistical analysis and technology, the aim of this project would be to close the gap of a very different situation between what customers would like and what is available for grabbing.

**LIST OF TABLES**

|  |  |  |  |
| --- | --- | --- | --- |
| **Table No** | **Table Title** | | **Page No** |
|  |  | |  |
| 1. | |  | | --- | |  |  |  | | --- | | Timeline Chart | | | 15 |
|  |  | |  |
|  |  | |  |
|  |  |
|  |  |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No** | **Figure Title** | **Page No** |
|  |  |  |
| 1. | Block Diagram | 11 |
| 2. | Output Data | 15 |
| 3. | Output Graph | 16 |
|  |  |  |
|  |  |  |

|  |  |
| --- | --- |
|  | **Contents** |
|  |  |
|  | Page No |
|  |  |
|  |  |
| Abstract | 3 |
|  |  |
| List Of Figures | 4 |
|  |  |
| List Of Tables | 4 |
|  |  |

**Chapter 1 INTRODUCTION** 6

**1.1** Introduction to Fast-food Nutrition Statistics, Analysis, and Prediction 6

**1.2** Project Statement / Objectives of the Project 6

**Chapter 2** **BACKGROUND OVERVIEW** 7

**2.1** Conceptual Overview (*Concepts/ Theory used*) 7

**2.2** Technologies Involved 8

**Chapter 3** **METHODOLOGY** 9

**3.1** Detailed methodology that will be adopted 9

**3.2** Circuit Layouts / block diagram 11

**Chapter 4** **IMPLEMENTATION AND RESULTS** 12

**4.1** Modules 12

**4.2** Prototype and System Architecture 13

**Chapter 5 FUTURE WORK AND CONCLUSION** 14

**5.1** Future Work and Conclusion 14

**5.2** Progress Chart / Timeline Chart 15

**Data and Output** 15

**Reference** 16

**CHAPTER 1: INTRODUCTION**

**1.1 Introduction to Work Done / Motivation.**   
The project Fast-food Nutrition Statistics, Analysis and Prediction comes in response to the alarmingly increased public interest in knowing what fast foods contain and their health impacts. Fast-food chains are often noted to have a huge dearth or unavailability of such nutritional data to the consumer who cannot afford or is largely unable to acquire proper nutrition advice. Using data science and artificial intelligence to analyze and model fast-food nutrition and predict trends therein provides consumers and stakeholders with actionable insights.  
  
The project depends on statistical methods and machine learning techniques, using them to analyze nutrition profiles of several menu items involved, understand their consumption patterns, and forecast how future fast foods will be like. It also promotes the public health strategies necessary for enabling healthy food choices within the contexts of adaptation of stakeholders to changing nutrition demands.  
  
**Some of the Major Challenges are:**   
**1. Data Fragmentation:** Nutritional information is different from chain to chain and makes it very hard to compare or figure out their total dietary impacts.

**2. Lack of Clarity:** Because of the little or often made complicated information available, it is hard for the consumer to know where he really stands in his health.

**3. Personalization:** Very few offers personalized advice based on dietary goals or restrictions.

**4. Trend Predictions:** The change in consumer demand has been difficult to predict by the fast-food industry. For example, they would prefer more health-oriented or plant-based foods.  
  
**1.2 Problem Statement/Objectives of the Project**

**Problem Statement**  
The problem is that fast foods are highly consumed and very easily accessible yet are of generally poor nutrition quality that includes many health problems such as obesity, heart disease, and diabetes. Much nutritionally related data is available, but it is often dispersed, poorly structured, ambiguous, or has access issues for the average consumer.

**Objectives**  
**1. Data Collection and Cleaning:** Nutritional pricing, and consumer behavior data collection from fast-food chains and clear it out.

**2. Exploratory Data Analysis (EDA):** Analyze trends and relationships of nutrition, price and popularity through statistical methods and visualization tools.

**3. Cost-Nutrition Insights:** It deals with the economic connection between the price of food in menus and its nutritional value, leading to healthier yet cheap options.  
**4. Consumer Behavior Analysis:** Understand consumer preferences in-between tradeoffs between price and convenience and nutrition.

**5. Clustering and Segmentation:** Apply machine learning models (K-Means, Decision Trees) to menu items based on their nutritional and economical attributes.

**6. Trend Prediction:** Apply regression models, i.e., Linear regression, Random Forest, and time-series forecasting, such as ARIMA, to model consumer demand changes and alterations in pricing strategy.

**7. Economic Impact Analysis:** The financial effects public health policies (e.g. Calorie labeling) have on pricing and profit margins.

**8. Profitability Analysis:** The profitability and market share impact of healthier menu options in the Fast-food market are evaluated.

**CHAPTER 2: BACKGROUND OVERVIEW**

**2.1 Conceptual Overview (Concepts/Theory Used)**

The Fast-food Nutrition Statistics, Analysis, and Prediction project utilizes various core data science and economics concepts and theories to analyse nutritional data and project industry trends. The approach integrates statistical methods, machine learning algorithms, and economic principles toward insights in fast-food nutrition, consumer behaviour, and market dynamics.

**Data Science Concepts:**

**1. Exploratory Data Analysis (EDA):**

This can be used to explore and summarize key patterns, trends, and outliers in data. Techniques like visualizations-histograms, scatter plots, and statistical tests like correlation, variance help identify the relations among nutritional content, pricing, and popularity.

**2. Regression Analysis:**

These models help in the determination of the relationship between variables, such as the effect of the composition of ingredients on calorie content or how price influences the choices of consumers. In this way, nutritional values can be predicted by menu characteristics.

**3. Classification Algorithms**

Supervised machine learning applications, such as Decision Trees, Support Vector Machines, and Random Forests classify menu items into categories using nutritional content and other characteristic features, such as categorizing menu items as healthy or unhealthy.

**4. Clustering Techniques:**

Unsupervised applications of machine learning, like K-Means and Hierarchical Clustering, put menu items with similar nutritional and economical characteristics together, bringing some hidden patterns in the fast-food industry to the open.

**5. Time Series Forecasting:**

Time-series models such as ARIMA (Autoregressive Integrated Moving Average) and Prophet predict future trends on menus of fast food about health options or shifts in demand due to consumer requirements by analysing the past data.

**6.Recommendation Systems:**

The collaborative filtering method along with content-based filtering are applied to generate recommendations customized according to the preferences and past purchase history of the consumers in dietary terms.

**2.2 Technologies Involved**

This Fast-food Nutrition Statistics, Analysis, and Prediction project is implemented in Python, which has a well-stocked ecosystem of libraries, frameworks, and tools that can be used for collecting data, analysis, modelling, and visualization. Below is a brief overview of the technologies applied.

**1. Data Collection and Preparation:**

•Pandas: Used for manipulating data and preprocessing, including cleaning, merging, and feature engineering.

•NumPy: Supports efficient numerical computations for handling large datasets.

•OpenPyXL/CSV: Allows for the import and export of data from spreadsheets and CSV files.

•BeautifulSoup and Requests: Used to perform web scraping to fetch nutritional and pricing data from fast-food chain websites or APIs.

**2. EDA:**

•Matplotlib: It makes static, animated, and even interactive visualizations available for inspecting data.

•Seaborn: Offers a high-level interface for plotting attractive statistical graphics.

•Plotly: Allows creating interactive and dynamic visualizations to find trends or patterns.

•Statsmodels: Used for conducting statistical tests and generating descriptive statistics.

**3. Machine Learning Models:**

•Scikit-learn: Core library for implementing machine learning algorithms, including:

o Linear Regression and Random Forest Regression for predicting nutritional attributes or menu prices.

o Decision Trees and SVM for classification tasks (e.g., identifying healthy/unhealthy menu items).

o K-Means for clustering similar menu items based on nutritional and economic factors.

•XGBoost: Used for efficient and high-performance modelling, especially for regression and classification problems.

•Keras/TensorFlow: Used for building and training deep learning models to analyse complex relationships between nutritional features.

**4. Time Series and Forecasting:**

•ARIMA (from Statsmodels): Used for time-series forecasting to predict trends in menu composition or consumer demand.

•Facebook Prophet: Simplifies time-series forecasting by incorporating seasonality and holiday effects.

**5. Recommendation Systems:**

•Surprise Library: Used for collaborative filtering to build recommendation systems based on user preferences.

• Scikit-learn: Implements content-based filtering using cosine similarity or nearest neighbour algorithms.

**6. Data Visualization and Reporting:**

• Tableau (Optional): Used to create interactive dashboards for effective presentation of results

• Dash/Plotly: Creates web-based interactive visualizations for stakeholders

• Jupyter Notebooks: Allows creating an interactive environment for code, visualizations, and explanations

**7. Deployment and Scalability:**

• Flask/Django: Used as the framework for developing web applications or APIs for disseminating insights or recommendation systems.

• Google Collab/AWS: For computing on the cloud, especially with large datasets or complex model training.

**CHAPTER 3: METHODOLOGY**

**3.1 Methodology**

The approach for the Fast-food Nutrition Statistics, Analysis, and Prediction project has been extended and modified to provide a complete structured approach toward the fulfillment of objectives. The process consists of several key stages:

**1. Data Collection:**

•Collect nutritional, pricing, and consumer behavior data through various sources such as fast-food websites, APIs, public datasets, and web scraping tools.

•Convert this data into a structured format: CSV, JSON, databases, etc. for later analysis.

•Validate the data to ensure completeness, accuracy, and consistency.

**2. Data Preprocessing and Cleaning:**

•Handle missing values, outliers, and inconsistencies using techniques like imputation and filtering.

•Normalize and standardize numerical features to ensure uniformity.

•Perform feature engineering to derive new attributes, such as cost per calorie or macro ratios, relevant for analysis.

**3. Exploratory Data Analysis (EDA):**

•Use statistical methods and visualizations to understand the relationships between nutritional attributes, pricing, and popularity.

•Trends such as calorie-dense items being cheaper or consumer demand for certain macronutrient compositions.

•Demographic and regional segmentations to reveal regional and local trends.

**4. Modeling and Analysis:**

•Regression Analysis: Regression models (Linear Regression, Random Forest Regression) that could be used to forecast nutritional values or prices using characteristics of the menu.

•Classification Models: Decision Trees and Logistic Regression models could be trained to classify the menu items as healthy or unhealthy.

• Clustering: Apply K-Means or DBSCAN to cluster menu items with similar nutritional and economic profiles.

• Time Series Forecasting: Implement ARIMA or Prophet to predict trends in fast-food consumption and menu innovations.

• Recommendation Systems: Build collaborative or content-based filtering models to suggest healthier alternatives or personalized menu options.

**5. Economic Analysis:**

• Price elasticity studies to assess how price changes influence consumer demand.

•Conduct cost-benefit analysis on healthier menu options, weighing costs of ingredients against consumer demand and market share.

•Analyze the profitability implications of regulations such as calorie labeling on price.

**6. Validation and Optimization:**

• Divide the data into training, validation, and test set for proper model evaluation

• Use cross-validation for hyperparameter optimization on the machine learning models

•Back test your forecasting models to validate the accuracy.

**7. Results Visualization and Interpretation:**

•Visualization of findings using interactive dashboards like Tableau or Plotly.

•Key insights such as nutritional breakdown of menu items or monetary savings from healthier options presented in a way that will be easily interpreted by key stakeholders.

**8. Deployment:**

•Web applications built using Flask or Django to deploy models and recommendation systems.

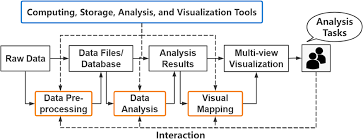
•Findings made available over APIs for integration with other external tools or platforms.

**9. Feedback and Iteration:**

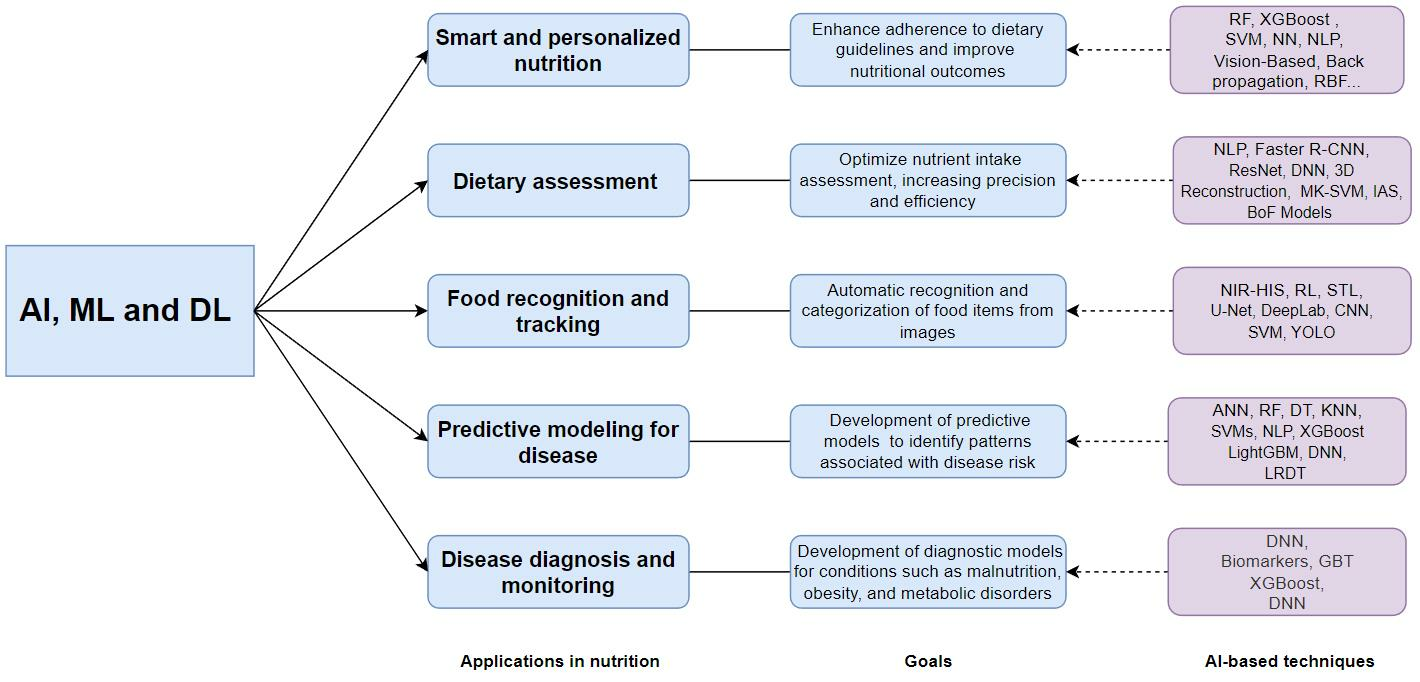
•Feedback from all stakeholders to refine the model and insights.

•The system can be updated with new data to reflect changes in consumer behavior or industry trends.

This methodology ensures that there is a comprehensive exploration of fast-food nutrition data, combining advanced analysis and predictive modeling to generate actionable insights for consumers, businesses, and policymakers.



**3.2 Circuit Layout/Block Diagram**



**CHAPTER 4: IMPLEMENTATION AND RESULTS**

The Fast-food Nutrition Statistics, Analysis, and Prediction project is implemented into key modules that eventually result in the final development of a prototype including all functionalities.

**4.1 Modules**

**1. Data Collection and Preprocessing Module:**

Implementation:

Obtained data from the websites of fast-food chains, APIs, and public datasets.

Pandas and NumPy were used for cleaning and organizing the data.

Missing data is imputed, and cost per calorie and macronutrient ratios are engineered.

Results:

 A clean and structured dataset formatted consistently and with features enhanced for analysis.

**2. EDA Module:**

Implementation:

 Visualizations were developed using Matplotlib and Seaborn to determine trends and correlations.

 Statistical insights on the relationship between price vs. nutritional value and calorie density on items.

Results:

 Clusters of high-calorie, low-priced items and trends in consumer preference for specific food categories.

**3. Modeling and Prediction Module:**

Implementation:

 Regression models used: Linear Regression, Random Forest to predict nutritional values and prices.

 Classification models (Decision Trees, Logistic Regression) classified healthy vs. unhealthy items.

 Clustering algorithms (K-Means) clustered menu items by nutritional profiles.

 Time-series models (ARIMA, Prophet) forecasted future trends in consumer demand and menu innovation.

Results:

 Attained high prediction accuracy and meaningful segmentation of menu items, thus supporting trend predictions and decision-making.

**4. Recommendation System Module:**

Implementation:

 Developed Content-based filtering and collaborative filtering algorithms with Scikit-learn and Surprise library.

 Personalized recommendations based on nutritional preferences and past choices.

Results:

Generated user-specific recommendations for healthier or more economical menu items.

**5.Visualization and Reporting Module:**

Implementation:

Created Dashboards using Plotly and Tableau for interactive visualization

Provided insights through web-based visual reports integrated via Flask

Results:

Stakeholders were able to dynamically explore nutritional patterns, pricing insights, and model predictions

**4.2 Prototype and System Architecture**

**1.Design and Functionality:**

A web-based interface was created using Flask. It enabled users to upload data, visualize insights, and interact with models.

Features:

1.Exploratory visualization of fast-food nutritional data.

2.Calorie and pricing trend forecasting tools.

3.Recommendation system for healthy choices.

**2.Key Findings:**

1.The prototype was able to show real-time predictions and personalized insights.

2.Actionable results included flags on unhealthy menu items or demand forecasting for plant- based foods.

3.Dashboards enabled intuitive exploration of trends to inform business and consumer decisions.

4.Modular design and functional prototype of the concept proved feasibility and practical influence of the project toward better, healthier decision-making, thus helping the fast-food industry evolve with consumer needs.

**Prototype Technologies**

The prototype of Fast-food Nutrition Statistics, Analysis, and Prediction project is using the following technologies:

**1.Backend:**

1.Flask: to develop RESTful API for handling data processing, model predictions, and recommendations.

2.SQLite/PostgreSQL: to store datasets, user inputs, and model outputs.

**2.Frontend:**

1.HTML/CSS/Bootstrap: To create a responsive and user-friendly interface.

2.JavaScript: To add interactivity and dynamic elements.

**3.Machine Learning Models:**

1.Scikit-learn and XGBoost: For regression, classification, and clustering models.

2.Prophet/ARIMA: For demand forecasting and trend prediction.

3.Surprise Library: For the recommendation system.

**4.Visualization**

1.Plotly/Dash: To create interactive charts and dashboards embedded into the web app.

**5.Deployment**

1.Heroku/AWS: For hosting the web application.

2.Docker: For ensuring consistent deployment across environments.

This stack ensures a functional, scalable, and user-friendly prototype for fast-food analysis and insights.

**CHAPTER 5: FUTURE WORK AND CONCLUSION**

**5.1 Future Work**

The Fast-food Nutrition Statistics, Analysis, and Prediction project provides a solid foundation to analyse and improve nutritional trends within the fast-food industry. Improvements for the future include:

**1.Integrate Real-Time Data Streams:** Incorporate dynamic streams of data from APIs and social media to provide real-time trend analysis.

**2.Advanced Recommendation Systems:** Develop hybrid models combining collaborative and content-based filtering for more personalized recommendations.

**3. Mobile Application Development:** Extend the project to mobile platforms, so that users can have on-the-go insights and recommendations.

**4. Regional and Cultural Analysis:** Adapt models to reflect cultural and regional dietary preferences for better applicability across geographies.

**5. Sustainability Metrics:** Incorporate environmental data, such as carbon footprint of menu items, to make sustainable choices.

**6. Policy Impact Simulation:** Model the effects of health regulations, such as taxes or mandatory calorie disclosures, on consumer and business behaviour.

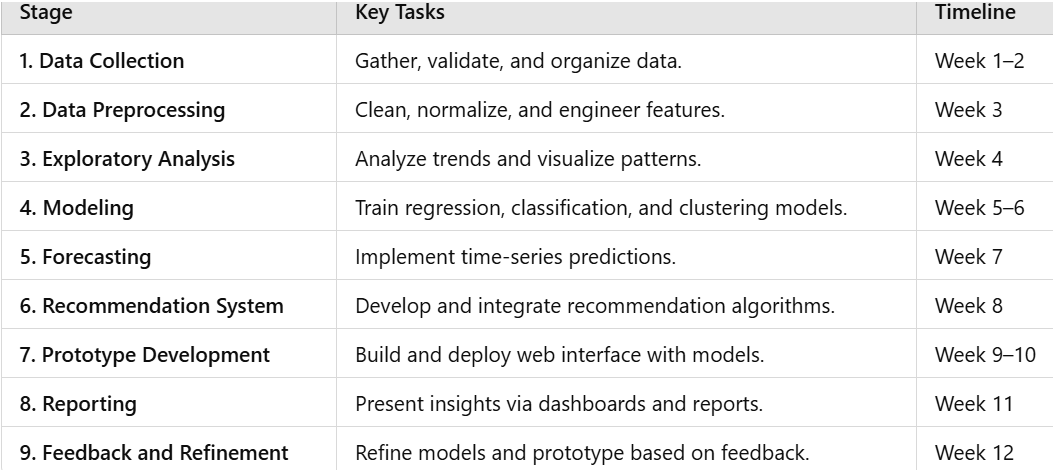
**7. Improved Visualization Tools:** Create more advanced dashboards with predictive capabilities for business users and policymakers.

**5.2 Conclusion**

This project demonstrates how data science and machine learning can provide actionable insights into the fast-food industry's nutritional trends and pricing strategies. By analysing large datasets and building predictive models, the project supports healthier decision-making for consumers and more informed strategies for businesses.

The interactive prototype demonstrates practical applications, such as predicting trends, offering personalized recommendations, and enabling economic analysis of healthier options. With further enhancements, the project has the potential to influence public health policies, foster sustainable practices, and reshape consumer behaviour toward healthier, more affordable fast-food choices.

**5.3 Progress Chart / Time Line Chart**



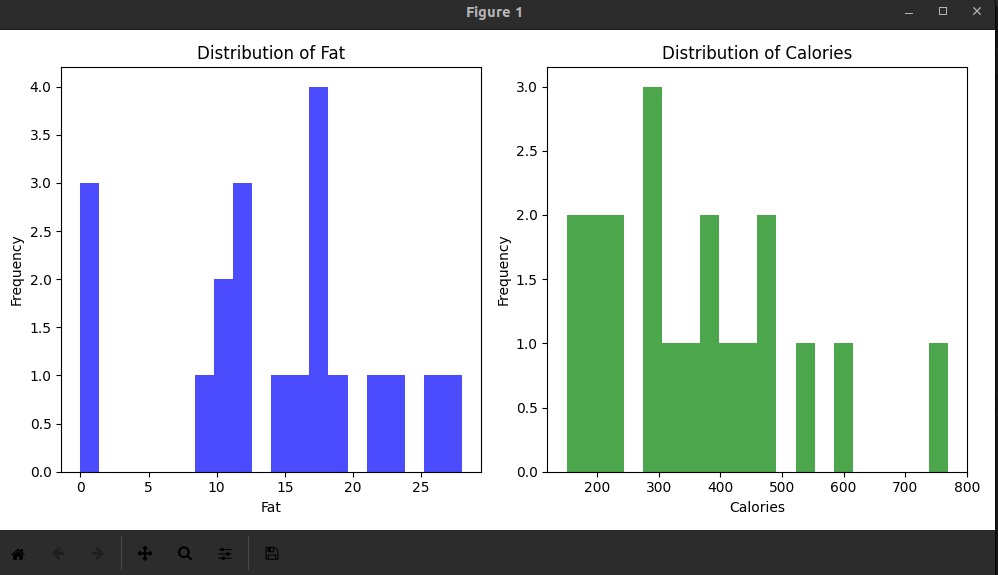
**Data and Output**



A screen shot of a computer

Description automatically generated

**Graph:**



A screenshot of a computer screen

Description automatically generated

**References**

**Books and Papers:**

1. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research, 2011.
2. McKinney, W., *Data Structures for Statistical Computing in Python*, Proceedings of the 9th Python in Science Conference, 2010.
3. Hunter, J. D., *Matplotlib: A 2D Graphics Environment*, Computing in Science & Engineering, 2007.
4. Hyndman, R.J., & Athanasopoulos, G., *Forecasting: Principles and Practice*, 2nd edition, OTexts, 2018.

**Websites and Online Resources:**

1. Surprise Library Documentation: <https://surprise.readthedocs.io>.
2. USDA FoodData Central: https://fdc.nal.usda.gov/.
3. Open Food Facts: https://world.openfoodfacts.org/.
4. Flask Documentation: https://flask.palletsprojects.com/.
5. Plotly Documentation: <https://plotly.com/>.