AI PRODUCT/SERVICE PROTOTYPE DEVELOPMENT AND BUSINESS/FINANCIAL MODELLING

Project 03
Report

SUSTAINABLE PALETTE

A SMART FOODPRINT TRACKER FOR CONSCIOUS LIVING

Sona Teresa Jossy

February 2024

Machine Learning Internship Program
Feynn Labs Services

STEP 1: PROTOTYPE SELECTION

ABSTRACT

The Sustainable Palette app, also known as the Foodprint Tracker, is a groundbreaking solution designed to bridge the gap between consumer choices and environmental impact in the realm of food consumption. This report presents a comprehensive exploration of the app's concept, development, and potential impact on individuals' dietary habits. The app leverages artificial intelligence, utilizing machine learning algorithms for carbon footprint analysis, personalized eco-profile creation, and recommendations of sustainable alternatives. The user journey, exemplified through the scenario of assessing different types of milk, showcases the app's seamless integration of environmental impact and nutritional insights, coupled with features like real-time feedback, community engagement, and continuous learning. The report delves into the data sources, algorithms, frameworks, and team requirements critical for the app's development, emphasizing a user-centric design.

The business model, based on a freemium model with premium subscription options, partnerships, and advertising, ensures both sustainability and profitability. A detailed cost analysis, including development, data acquisition, maintenance, and marketing costs, provides a holistic understanding of the financial aspects. In conclusion, the Sustainable Palette app stands as an innovative, AI-driven platform poised to empower users to make informed, sustainable, and healthy dietary choices, fostering a collective commitment to a greener and nourished future.

PROBLEM STATEMENT

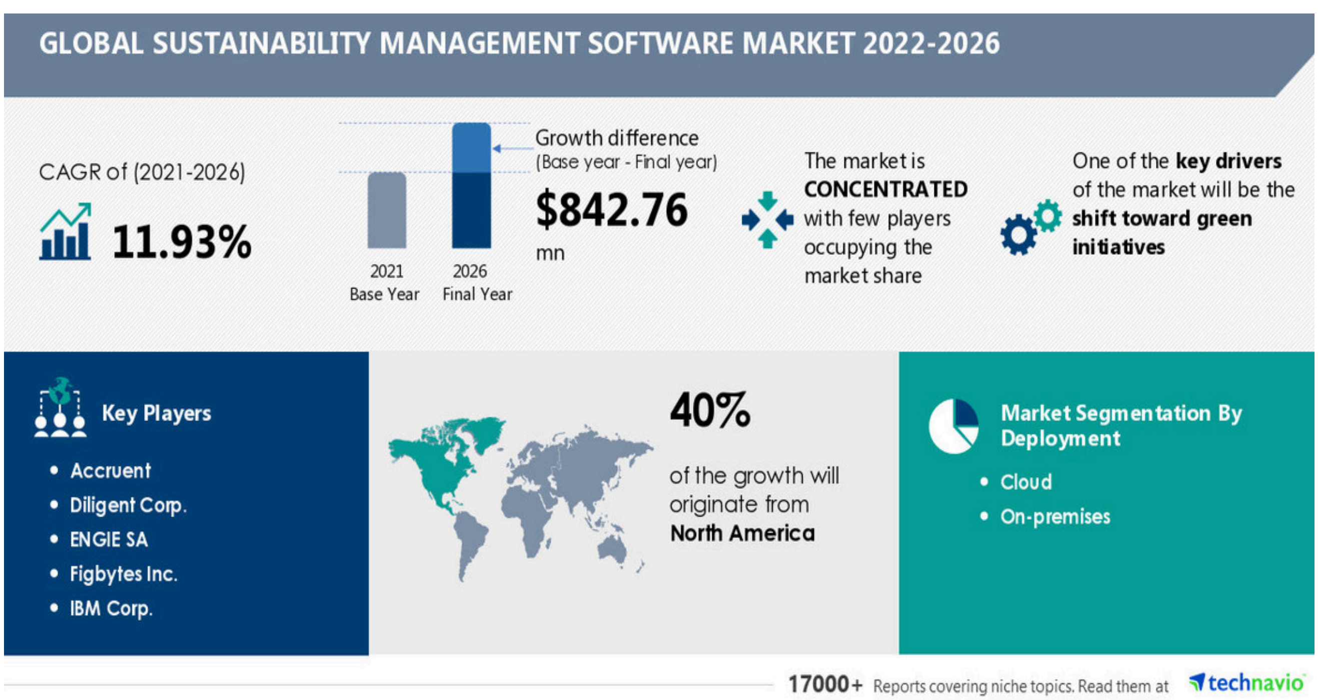
The current global food consumption patterns, characterized by a lack of awareness and tools for making environmentally conscious choices, pose a significant threat to our planet's sustainability. This problem is exacerbated by a general lack of understanding regarding the environmental impact of individual dietary decisions. As populations continue to grow and urbanize, there is a pressing need for innovative solutions that empower individuals to make informed, sustainable choices about the food they consume.

The environmental consequences of conventional food production and consumption are multifaceted, encompassing issues such as excessive water usage, deforestation, greenhouse gas emissions, and the depletion of natural resources. Many consumers are unaware of the intricate web of factors contributing to the carbon footprint of the products they consume, including the production methods, transportation, and processing involved. This knowledge gap creates a disconnect between individuals and the broader environmental implications of their dietary habits.

In light of these challenges, the problem statement revolves around the urgency to bridge the gap between consumer choices and environmental impact, fostering a deeper understanding of the consequences of individual dietary decisions. The lack of a unified, user-friendly platform that integrates artificial intelligence to analyze and personalize sustainability recommendations hinders progress toward more sustainable living. The Foodprint Tracker aims to address this critical problem by providing users with a holistic understanding of the environmental implications of their food choices, thereby empowering them to contribute to a more sustainable and ecologically responsible future.

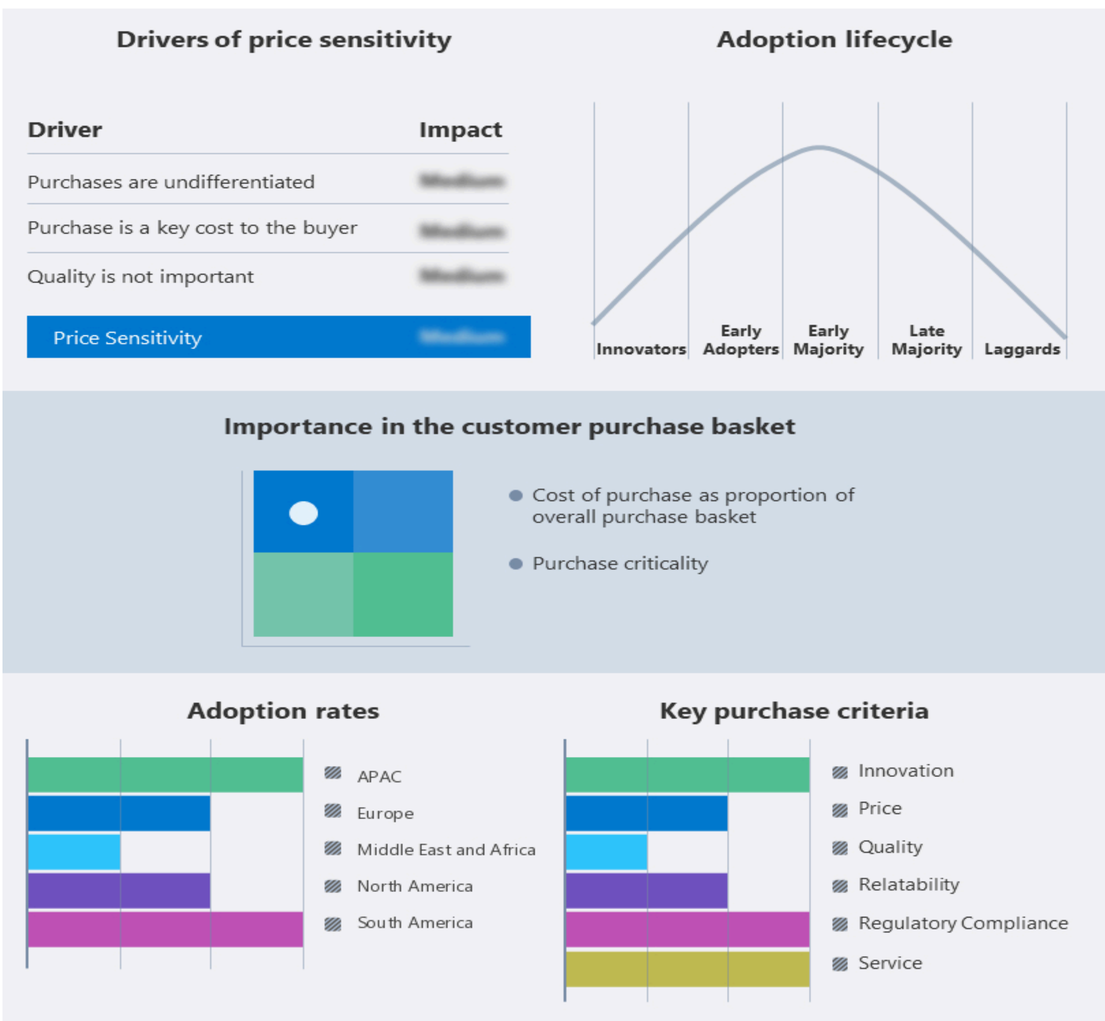
MARKET/CUSTOMER/BUSINESS NEED ASSESSMENT

The market for sustainable living solutions has experienced significant growth in recent years, driven by an increasing global awareness of environmental issues. Consumers are actively seeking products and services that align with their values and contribute to a more sustainable lifestyle. The Foodprint Tracker taps into this expanding market segment, offering a unique proposition by combining artificial intelligence with nutrition insights to guide users towards more eco-friendly and health-conscious dietary choices.



The target customers for the Foodprint Tracker are individuals who prioritize sustainability in their daily lives, encompassing environmentally conscious consumers, health enthusiasts, and those seeking a comprehensive tool to understand and reduce their ecological footprint. This diverse customer base reflects a growing trend where people are not only concerned about personal health but are also eager to contribute positively to environmental conservation efforts through their lifestyle choices.

The need for the Foodprint Tracker is underscored by the existing gap in available tools for consumers to make sustainable food choices. Current solutions often lack personalization and fail to provide holistic insights into the environmental impact of dietary decisions. This creates a prime opportunity for an innovative, AI-powered platform that seamlessly integrates nutritional information with sustainability metrics. Businesses operating in the food industry also stand to benefit from participating in such a platform, as it allows them to showcase their commitment to sustainability and appeal to an increasingly conscientious consumer base.



Global Sustainability Management Software Market Customer Landscape

TARGET SPECIFICATIONS AND CHARACTERIZATION

The Foodprint Tracker is specifically designed for a diverse set of customers who share a common goal of making environmentally conscious and health-oriented dietary choices. The primary target audience comprises individuals who are environmentally aware, seeking to reduce their carbon footprint, and eager to contribute to sustainable living practices. These users are characterized by their interest in understanding the broader impact of their food choices beyond personal health considerations.

Additionally, the application caters to health-conscious consumers who prioritize nutritional well-being. This segment includes individuals following specific diets, such as vegetarian, vegan, or those with dietary restrictions, who are looking for personalized insights that align with their nutritional preferences.

The product is tailored to appeal to a tech-savvy demographic comfortable with mobile applications and seeking user-friendly interfaces. This includes individuals of varying age groups, from young adults to middle-aged consumers, who value the convenience of digital tools in their daily routines.

EXTERNAL SEARCH

References:

1. Health and Wellness Food Market by Product, Distribution Channel, and Geography - Forecast and Analysis 2023-2027
   https://www.technavio.com/report/health-and-wellness-food-market-industry-analysis
2. Sustainability Management Software Market by Application, Deployment, and Geography - Forecast and Analysis 2023-2027
   https://www.technavio.com/report/sustainability-management-software-market-industry-analysis

Dataset:

1. Environment Impact of Food Production
   https://www.kaggle.com/datasets/selfvivek/environment-impact-of-food-production

2. Nutritional Information
   https://www.kaggle.com/datasets/meghanachebolu/nutritional-info

BENCH MARKING ALTERNATE PRODUCTS

Product	Key Features	Differentiators
MyFitnessPal https://www.myfitnesspal.com/	- Nutrition tracking, exercise logging, and weight management.	- Extensive food database for tracking dietary intake.
Yazio https://www.yazio.com/en	- Personalized nutrition plans and food consumption tracking.	- Integration with wearables and devices for comprehensive health tracking.

APPLICABLE PATENTS

Dynamic recommendation platform with artificial intelligence by Gregory M. BELLOWE.

APPLICABLE REGULATIONS

The Foodprint Tracker, being a product that involves the analysis and communication of environmental impact data related to food consumption, would need to adhere to various government and environmental regulations. While specific regulations can vary by country, here are some general aspects and considerations:

- Data Privacy and Security Regulations
- Food Labeling Regulations
- Environmental Impact Reporting Standards
- Accuracy and Transparency Requirements
- Consumer Protection Regulations
- Health and Nutrition Labeling Regulations
- Mobile App Regulations
- Local Environmental Regulations

APPLICABLE CONSTRAINTS

The development and implementation of the Foodprint Tracker may encounter various constraints that need to be carefully addressed. Here are some key constraints associated with this product idea:

- Budget Constraints
- Data Availability and Quality
- Expertise in Environmental Impact Assessment
- Integration Challenges
- User Adoption and Behavior Change
- Legal and Regulatory Compliance
- User Interface Design Challenges
- Continuous Learning and Updates
- Scalability

STEP 2: PROTOTYPE DEVELOPMENT

For the prototype development, a small-scale implementation of the carbon footprint analysis and recommendation system can be built using Python programming language and relevant libraries such as TensorFlow and Scikit-learn. This implementation will involve:

- Collecting and preprocessing environmental impact data and nutritional information.
- Training machine learning models to analyze the carbon footprint of various foods and recommend sustainable alternatives.

CONCEPT GENERATION

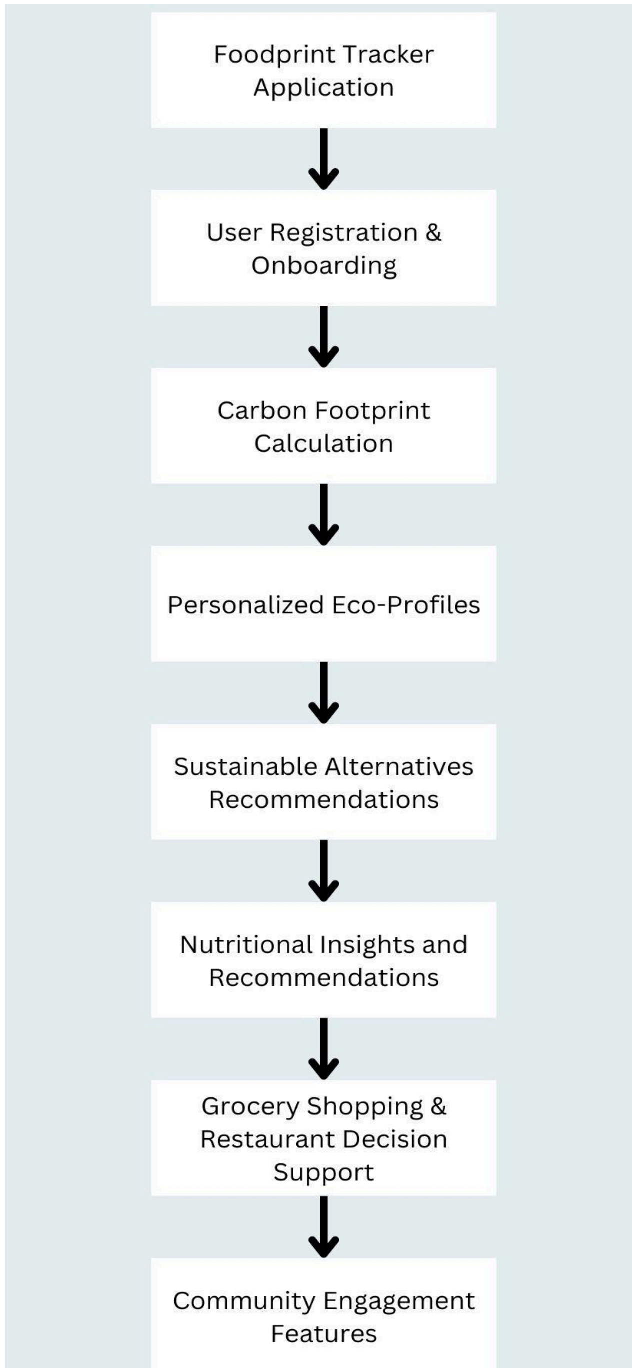
The concept generation for the Foodprint Tracker began with a comprehensive analysis of the problem—lack of awareness regarding the environmental impact of dietary choices. Through research and brainstorming sessions, the idea evolved into an AI-driven application. The core of the product involves leveraging machine learning algorithms for carbon footprint analysis, personalized eco-profile creation, and sustainable alternatives recommendations.

Technologies such as Python, TensorFlow, and Scikit-learn can be employed for the development, while libraries like NumPy and Pandas facilitate data manipulation. Algorithms such as regression models and clustering algorithms can be utilized for the main feature of assessing and recommending sustainable food choices.

CONCEPT DEVELOPMENT

In the concept development phase, the Foodprint Tracker is envisioned as a revolutionary AI-powered application that empowers users to make informed and sustainable dietary choices. The app employs sophisticated machine learning algorithms to analyze the carbon footprint of various foods, creating personalized eco-profiles based on individual preferences. Users receive real-time feedback on the environmental impact of their choices, along with recommendations for eco-friendly alternatives. The concept integrates nutritional insights, encourages community engagement, and envisions partnerships with grocery stores and restaurants. The user-centric design ensures a seamless and intuitive experience, promoting both environmental consciousness and personal well-being through a comprehensive approach to food choices.

Block Diagram:



FINAL PRODUCT PROTOTYPE

The Foodprint Tracker app prototype is an innovative solution designed to revolutionize users' dietary habits by providing real-time insights into the environmental impact of their food choices. Leveraging cutting-edge AI algorithms, the app calculates the carbon footprint of various foods, creating personalized eco-profiles for users. The intuitive user interface seamlessly integrates nutritional data, offering comprehensive information for making sustainable and healthy choices. Community engagement features allow users to share eco-profiles and tips, fostering a collective commitment to sustainable living. The prototype envisions a future where individuals can effortlessly align their dietary decisions with environmental conservation goals.

The schematic diagram given below represents the flow of key features in the Foodprint Tracker app, illustrating how users interact with the application to make sustainable and healthy food choices. The stages include user onboarding, carbon footprint calculation, personalized eco-profiles, recommendations, nutritional insights, and community engagement. The interconnected flow emphasizes a user-centric design, ensuring a holistic and engaging experience.

a) Feasibility

The product idea, Sustainable Palette app, is feasible for short-term development (2-3 years). The technology required for AI-driven analysis and recommendation systems is readily available, and the concept can be implemented within a reasonable timeframe.

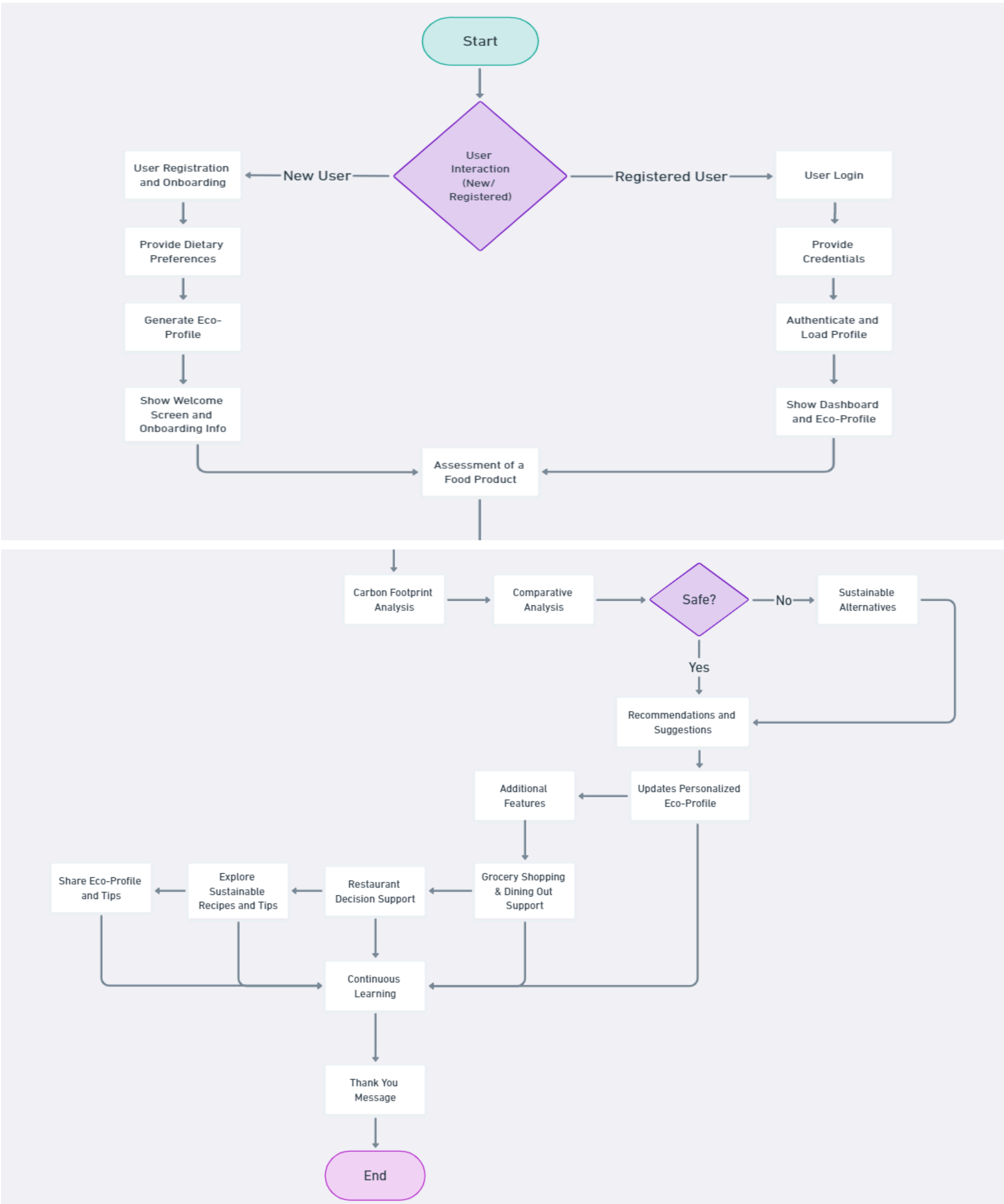
b) Viability

The Sustainable Palette app is highly relevant and has the potential to survive in the long term (20-30 years). As environmental concerns become increasingly prominent and consumers demand more sustainable options, the need for tools like the Foodprint Tracker will only grow over time.

c) Monetization

The Sustainable Palette app is directly monetizable. Revenue can be generated through premium subscription options, partnerships with businesses, and targeted advertising, making it a suitable candidate for this project.

Workflow Diagram:



CODE IMPLEMENTATION/MODEL BUILDING

Small scale code implementation/model building of the Sustainable Palette: Foodprint Tracker prototype is implemented and can be accessed by visiting the below link.

Github Link: ***https://github.com/sonateresa/feynnlabs-sona***

Dataset 01: Environment Impact of Food Production

This dataset contains 43 most common foods grown across the globe and 23 columns as their respective land, water usage and carbon footprints.

1. Land use change - Kg CO2 - equivalents per kg product
2. Animal Feed - Kg CO2 - equivalents per kg product
3. Farm - Kg CO2 - equivalents per kg product
4. Processing - Kg CO2 - equivalents per kg product
5. Transport - Kg CO2 - equivalents per kg product
6. Packaging - Kg CO2 - equivalents per kg product
7. Retail - Kg CO2 - equivalents per kg product

These represent greenhouse gas emissions per kg of food product (Kg CO2 - equivalents per kg product) across different stages in the lifecycle of food production.

Eutrophication – the pollution of water bodies and ecosystems with excess nutrients – is a major environmental problem. The runoff of nitrogen and other nutrients from agricultural production systems is a leading contributor.

Dataset 02: Nutritional Information

This dataset provides nutritional information for 43 common food products, detailing various macronutrients and energy values.

1. Energy Value (kcal): The energy content of the food product
2. Protein (g): The amount of protein present in the food product
3. Carbohydrate(g): The amount of carbohydrates present in the food product
4. Total Sugars (g): The total amount of sugars present in the food product
5. Added Sugars (g): The amount of added sugars present in the food product
6. Total Fat (g): The total amount of fat present in the food product
7. Saturated Fat (g): The amount of saturated fat present in the food product
8. Trans Fat (g): The amount of trans fat present in the food product

9. Cholesterol (mg): The amount of cholesterol present in the food product
10. Sodium (mg): The amount of sodium present in the food product
11. Energy per Kilogram (kcal/kg): The energy content of the food product per kilogram

These nutritional attributes are essential for understanding the dietary composition and health impacts of consuming different food products. They provide insights into the macronutrient composition and energy density of each food item, facilitating informed dietary choices for consumers.

PROTOTYPE DEVELOPMENT

To develop a prototype for the Sustainable Palette app, which includes carbon footprint analysis and recommendation system, we'll need to merge and preprocess the provided datasets: `Food_Production.csv` and `NUTRITIONAL INFO.csv`. After preprocessing, we can proceed with training machine learning models to analyze the carbon footprint and recommend sustainable alternatives based on nutritional information.

Step 1: Data Preprocessing and Integration

1. Read the datasets into pandas dataframes.
2. Clean and preprocess the data, handling missing values and ensuring data consistency.

Step 2: Feature Engineering

1. Calculate additional features if needed, such as energy value per kilogram or per serving.
2. Normalize numerical features to bring them to a similar scale.

Step 3: Machine Learning Model Training

1. Split the dataset into training and testing sets.
2. Define the target variable (e.g., carbon footprint).
3. Train machine learning models using regression algorithms for carbon footprint analysis.
4. Evaluate the models' performance using appropriate metrics (e.g., mean squared error, R-squared).

Step 4: Recommendation System Development

1. Cluster food products based on their nutritional attributes using clustering algorithms.

2. Develop a recommendation system that suggests sustainable alternatives from the same cluster for a given food product.
3. Evaluate the recommendation system's performance using relevant metrics (e.g., accuracy, precision, recall).

Step 5: Integration and Testing

1. Integrate the trained machine learning models and recommendation system into the Sustainable Palette app prototype.
2. Develop a user interface to allow users to input food items and receive carbon footprint analysis and sustainable alternatives.
3. Test the prototype thoroughly to ensure functionality and accuracy.

Step 6: Iterative Improvement

1. Gather feedback from users and stakeholders to identify areas for improvement.
2. Continuously refine the prototype based on user feedback and performance evaluation.
3. Explore additional features and enhancements to make the app more useful and user-friendly.

DATA ANALYSIS AND FINDINGS

1. Data Loading and Preprocessing:

```
print("\nFood Production dataset:")
food_production_df.head()
```

```
Food Production dataset:
```

	food product	land use change	animal feed	farm	processing	transport	packging	retail	total_emissions	eutrophying emissions per 1000kcal (gPO₄eq per 1000kcal)
0	Wheat & Rye (Bread)	0.1	0.0	0.8	0.2	0.1	0.1	0.1	1.4	0.000000
1	Maize (Meal)	0.3	0.0	0.5	0.1	0.1	0.1	0.0	1.1	0.000000
2	Barley (Beer)	0.0	0.0	0.2	0.1	0.0	0.5	0.3	1.1	0.000000
3	Oatmeal	0.0	0.0	1.4	0.0	0.1	0.1	0.0	1.6	4.281357
4	Rice	0.0	0.0	3.6	0.1	0.1	0.1	0.1	4.0	9.514379

5 rows × 23 columns

	freshwater withdrawals per 100g protein (liters per 100g protein)	freshwater withdrawals per kilogram (liters per kilogram)	greenhouse gas emissions per 1000kcal ($kgco_2eq$ per 1000kcal)	greenhouse gas emissions per 100g protein ($kgco_2eq$ per 100g protein)	land use per 1000kcal (m^2 per 1000kcal)	land use per kilogram (m^2 per kilogram)	land use per 100g protein (m^2 per 100g protein)	scarcity-weighted water use per kilogram (liters per kilogram)	scarcity-weighted water use per 100g protein (liters per 100g protein)	scarcity-weighted water use per 1000kcal (liters per 1000 kilocalories)
.	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.00000	0.000000
.	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.00000	0.000000
.	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.000000	0.0	0.00000	0.000000
.	371.076923	482.4	0.945482	1.907692	2.897446	7.6	5.846154	18786.2	14450.92308	7162.104461
.	3166.760563	2248.4	1.207271	6.267606	0.759631	2.8	3.943662	49576.3	69825.77465	13449.891480

```
[8] print("\nNutritional Info dataset:")
    nutritional_info_df.head()
```

```
Nutritional Info dataset:
```

	food product	energy value(kcal)	protein(g)	carbohydrate(g)	total sugars(g)	added sugars(g)	total fat(g)	saturated fat(g)	trans fat(g)	cholesterol(mg)	sodium(mg)
0	Britannia 50-50 Biscuits	487.0	7.0	72.0	19.5	18.5	19.0	9	0	0.0	466.0
1	Parle Monaco Biscuits	492.0	7.9	68.8	8.4	7.8	20.6	9.7	0	0.0	891.0
2	Parle-G Biscuits	454.0	6.9	77.3	25.5	25.0	13.0	6	0	0.0	296.0
3	Britannia Bourbon Biscuits	493.0	5.1	72.6	38.0	35.1	20.3	10	0	0.0	108.0
4	Britannia Marie GOLD	445.0	8.4	78.9	21.9	20.0	10.7	5.2	0	0.0	0.0

2. Feature Engineering:

Normalized Food Production Dataset:

```scss
         food product  land use change  animal feed      farm  processing  \
0  Wheat & Rye (Bread)         0.119565          0.0  0.017812    0.153846
1         Maize (Meal)         0.130435          0.0  0.010178    0.076923
2        Barley (Beer)         0.114130          0.0  0.002545    0.076923
3              Oatmeal         0.114130          0.0  0.033079    0.000000
4                 Rice         0.114130          0.0  0.089059    0.076923

   transport  packging    retail  total_emissions  \
0      0.125    0.0625  0.333333         0.020202
1      0.125    0.0625  0.000000         0.015152
2      0.000    0.3125  1.000000         0.015152
3      0.125    0.0625  0.000000         0.023569
4      0.125    0.0625  0.333333         0.063973

   eutrophying emissions per 1000kcal (gpo₄eq per 1000kcal)  ...
0                                           0.000000          ...
1                                           0.000000          ...
2                                           0.000000          ...
3                                           4.281357          ...
4                                           9.514379          ...

[5 rows x 23 columns]
```

Normalized Nutritional Info Dataset:

```scss
               food product  energy value(kcal)  protein(g)  \
0    Britannia 50-50 Biscuits            0.510083    0.328484
1       Parle Monaco Biscuits            0.516014    0.370718
2            Parle-G Biscuits            0.470937    0.323792
3  Britannia Bourbon Biscuits            0.517200    0.239324
4        Britannia Marie GOLD            0.460261    0.394181

   carbohydrate(g)  total sugars(g)  added sugars(g)  total fat(g)  \
0         0.736196         0.241935         0.230386         0.190
1         0.703476         0.104218         0.097136         0.206
2         0.790389         0.316377         0.311333         0.130
3         0.742331         0.471464         0.437111         0.203
4         0.806748         0.271712         0.249066         0.107

   saturated fat(g)  trans fat(g)  cholesterol(mg)  sodium(mg)  energy_per_kg
0          0.145161           0.0              0.0    0.078070       0.510083
1          0.156452           0.0              0.0    0.149271       0.516014
2          0.096774           0.0              0.0    0.049590       0.470937
3          0.161290           0.0              0.0    0.018093       0.517200
4          0.083871           0.0              0.0    0.000000       0.460261
```

3. Model Training:

The data has been split into training and testing sets and trained in three regression models: *Linear Regression, Random Forest Regression,* and *Gradient Boosting Regression*. The models have been calculated using Mean Squared Error (MSE) and R-squared values.

```
Linear Regression - MSE: 29.40853495674892 R-squared: 0.5185943713882497
Random Forest Regression - MSE: 46.93430599999991 R-squared: 0.23170470352855732
Gradient Boosting Regression - MSE: 57.752895455565685 R-squared: 0.054608841214821346
```

These metrics represent the performance of different regression models in predicting the target variable ('total_emissions'):

1. Linear Regression:
   Mean Squared Error (MSE): 29.4085
   R-squared (R2): 0.5186
2. Random Forest Regression:
   Mean Squared Error (MSE): 46.9343
   R-squared (R2): 0.2317
3. Gradient Boosting Regression:
   Mean Squared Error (MSE): 57.7529
   R-squared (R2): 0.0546

Mean Squared Error (MSE): This measures the average squared difference between the actual and predicted values. A lower MSE indicates a better fit of the model to the data.
R-squared (R2): This is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, where 1 indicates a perfect fit. A higher R-squared value indicates that more variance is explained by the model.

Observation:

The *Linear Regression model* has the lowest MSE and the highest R-squared value among the three models, indicating that it performs the best in terms of accuracy and explaining the variance in the target variable.

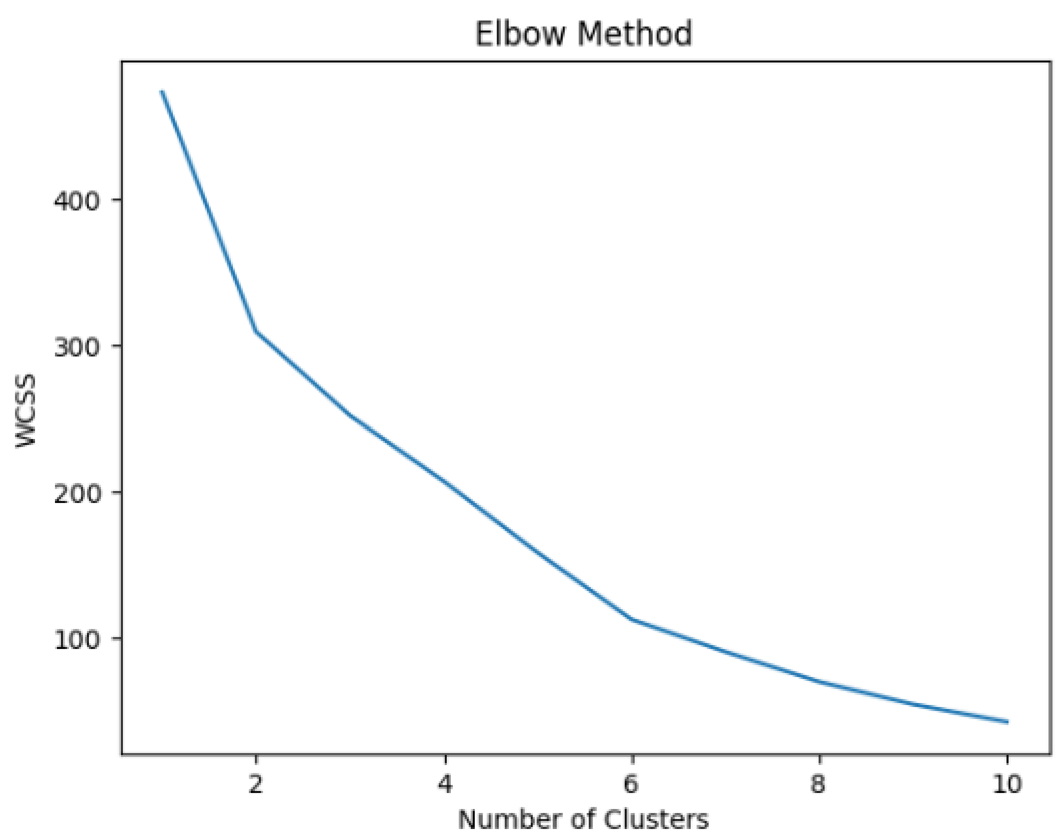
Random Forest Regression performs better than Gradient Boosting Regression but worse than Linear Regression in terms of both MSE and R-squared.

4. Cluster Analysis:

Within-Cluster-Sum-of-Squares (WCSS): WCSS measures the compactness of the clusters. It calculates the sum of squared distances between each data point and its centroid within a cluster. Lower WCSS indicates that the data points within each cluster are closer to their centroids, implying better clustering.

Elbow Method: The Elbow Method is a heuristic used to determine the optimal number of clusters in a dataset. It involves plotting the number of clusters against the corresponding WCSS values and identifying the "elbow" point in the plot. The elbow point represents the number of clusters where the rate of decrease in WCSS starts to slow down, suggesting diminishing returns in terms of clustering improvement beyond that point.

By plotting the number of clusters against the WCSS values, you can visually inspect the plot to determine the optimal number of clusters for your dataset. Typically, the optimal number of clusters is chosen at the point where adding more clusters does not significantly reduce the WCSS.



The cluster centers represent the average values of each feature within each cluster.

```
Cluster Centers:
   energy value(kcal)  protein(g)  carbohydrate(g)  total sugars(g)  \
0          389.376471    4.876471        65.105882        32.247059
1          462.526316    9.371579        61.835263         3.898947
2          897.000000    0.000000         0.000000         0.000000
3          900.000000    0.000000         0.000000         0.000000
4          245.000000    2.900000        15.150000         4.200000

   added sugars(g)  total fat(g)  saturated fat(g)  trans fat(g)  \
0        22.094118     12.229412          6.505882      0.033529
1         2.790526     18.780526          8.894737      0.043684
2         0.000000     99.700000         62.000000      1.500000
3         0.000000     99.997500         10.000000      1.000000
4         0.000000     19.200000          3.650000      0.000000

   cholesterol(mg)   sodium(mg)  energy_per_kg
0     2.976471e+00   184.782353       0.432641
1    -1.776357e-15   651.715789       0.513918
2     2.000000e+02     0.000000       0.996667
3     0.000000e+00     0.000000       1.000000
4     0.000000e+00  5315.000000       0.272222
```

Observation:

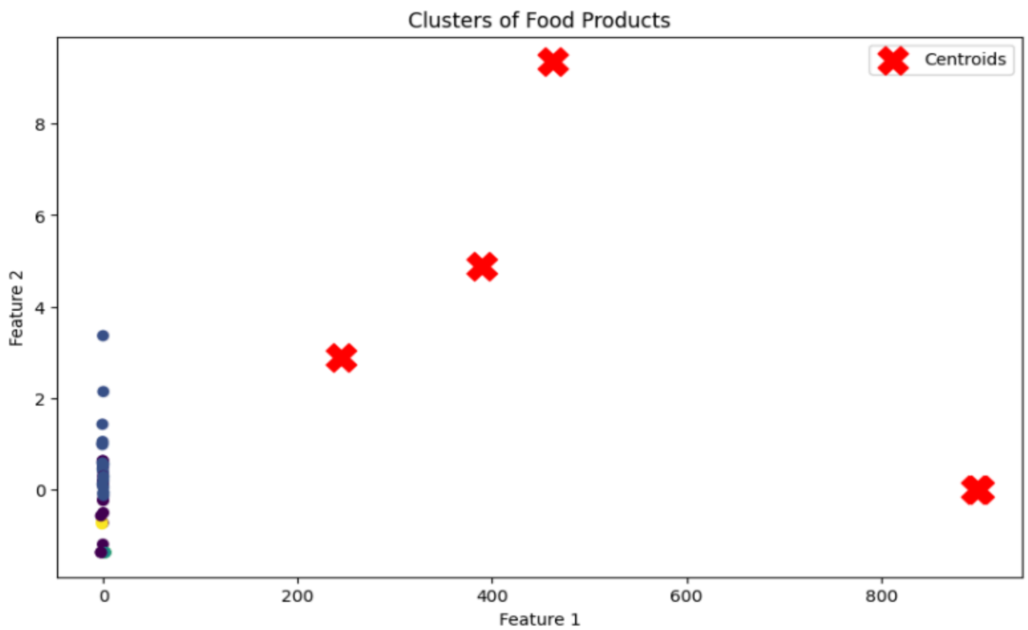
Cluster 0: This cluster has moderate energy value, moderate protein content, high carbohydrate content, high total sugars, and moderate total fat content. It also has moderate levels of saturated fat, low trans fat, low cholesterol, and moderate sodium content.

Cluster 1: This cluster has relatively high energy value, high protein content, moderate carbohydrate content, low total sugars, and low added sugars. It has high total fat content, high saturated fat, low trans fat, and moderate cholesterol. It also contains high sodium content.

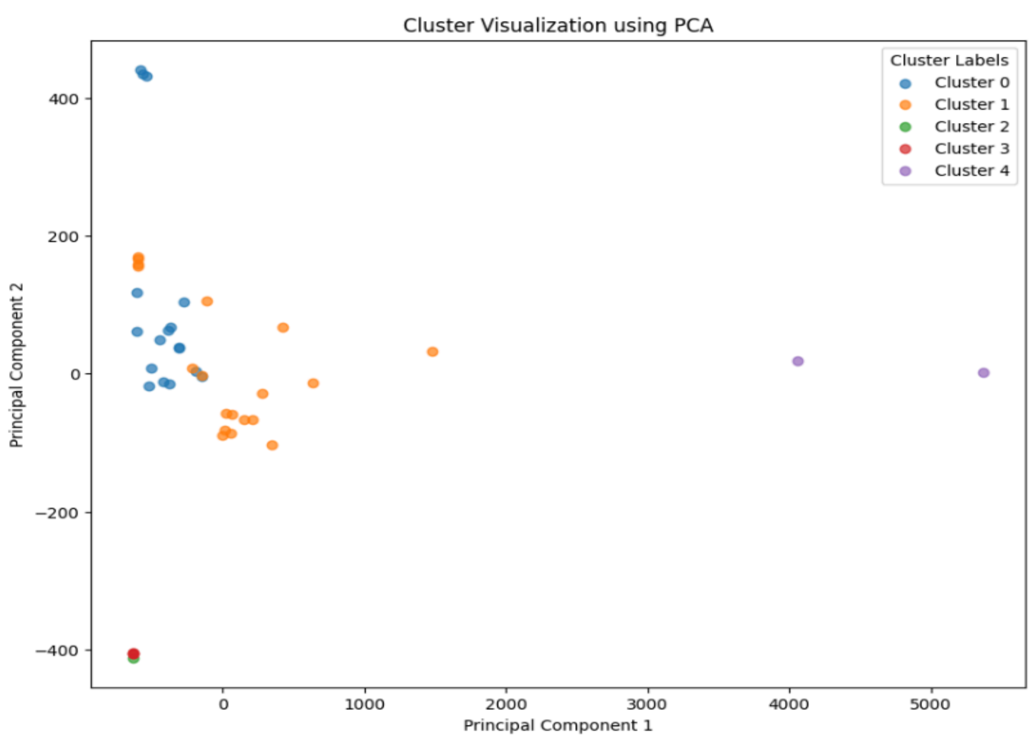
Cluster 2: This cluster has a very high energy value and zero content for all other nutritional attributes. This could represent oils or fats which are primarily composed of energy and have negligible amounts of other nutrients.

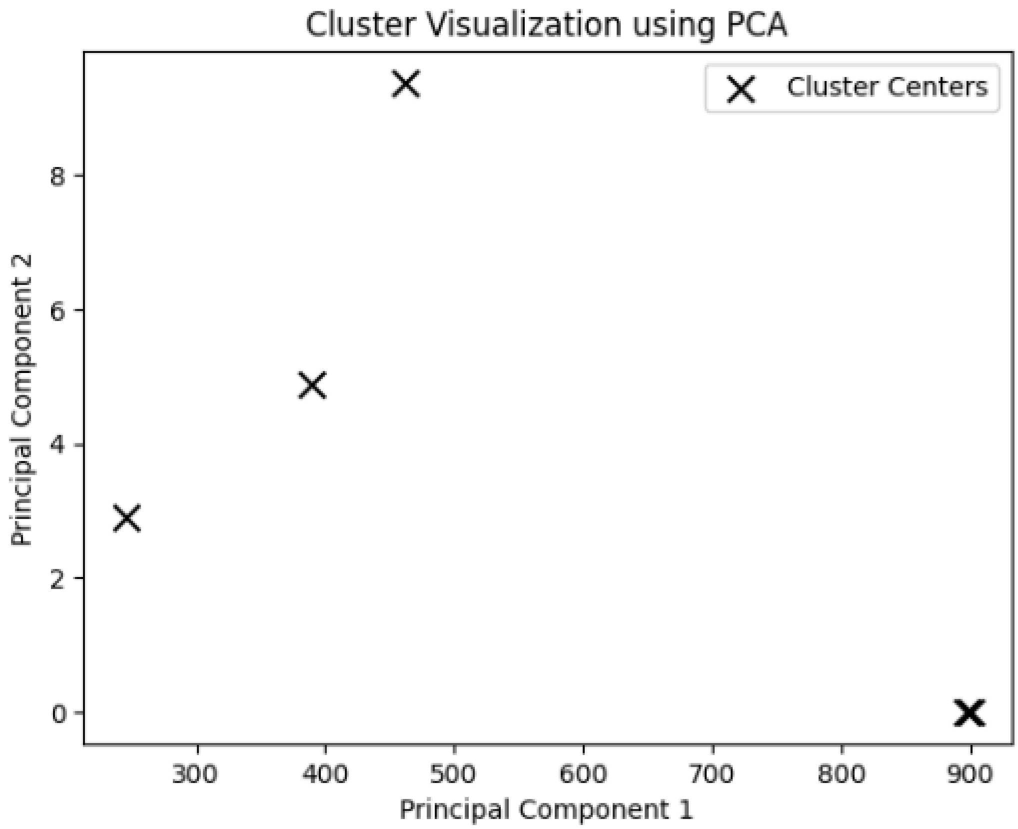
Cluster 3: This cluster has very high energy value, zero content for all other nutritional attributes, except for a high level of saturated fat and cholesterol. This could represent oils or fats with extremely high levels of saturated fat and cholesterol.

Cluster 4: This cluster has relatively low energy value, low protein content, low carbohydrate content, and low total sugars. It also has low total fat content, low saturated fat, and no trans fat. It contains minimal cholesterol and extremely high sodium content, indicating a potentially processed or snack food.



Principal Component Analysis (PCA): PCA is a technique used for dimensionality reduction. It transforms high-dimensional data into a lower-dimensional space while preserving as much of the variance in the data as possible. After reducing the dimensionality of the data using PCA, the next step is to visualize the clusters in the reduced two-dimensional space. Each data point is represented by its coordinates in the new two-dimensional space obtained from PCA.





5. Recommendation System:

A recommendation system has been implemented that recommends products from the same cluster as the given product. Two functions have been designed to work with the clustering model that has been trained. Here's how each function works:

1. assign_cluster_labels(data, kmeans_model): This function takes two parameters - `data` and `kmeans_model`. Inside the function, it uses the `predict` method of the `kmeans_model` to assign cluster labels to each data point in the dataset `data`. It returns an array of cluster labels corresponding to each data point.

2. recommend_products(product_name, cluster_labels, products_df): This function takes three parameters - `product_name`, `cluster_labels` and `products_df. It first identifies the cluster to which the given product belongs by searching for its index in the `products_df`, extracting the cluster label from `cluster_labels`. Then, it filters the `products_df` to find other products belonging to the same cluster and creates a list of recommended products. Finally, it excludes the given product from the recommended list and returns the list of recommended products.

Example:

```
Recommended products for 'Parle Monaco Biscuits':
0     Britannia 50-50 Biscuits
5      Britannia Nutri Choice
11                OREO Biscuits
13               Yippee Noodles
14                Maggi Noodles
24           BINGO Potato Chips
25                  LAYS TOMATO
26               KURKURE TOMATO
27        KURKURE GREEN CHUTNEY
28          HALDIRAM SOYA STICKS
30          HALDIRAM ALOO BHUJIA
31             HALDIRAM MIXTURE
33              BINGO MAD ANGLES
38         KELLOGG'S CORN FLAKES
```

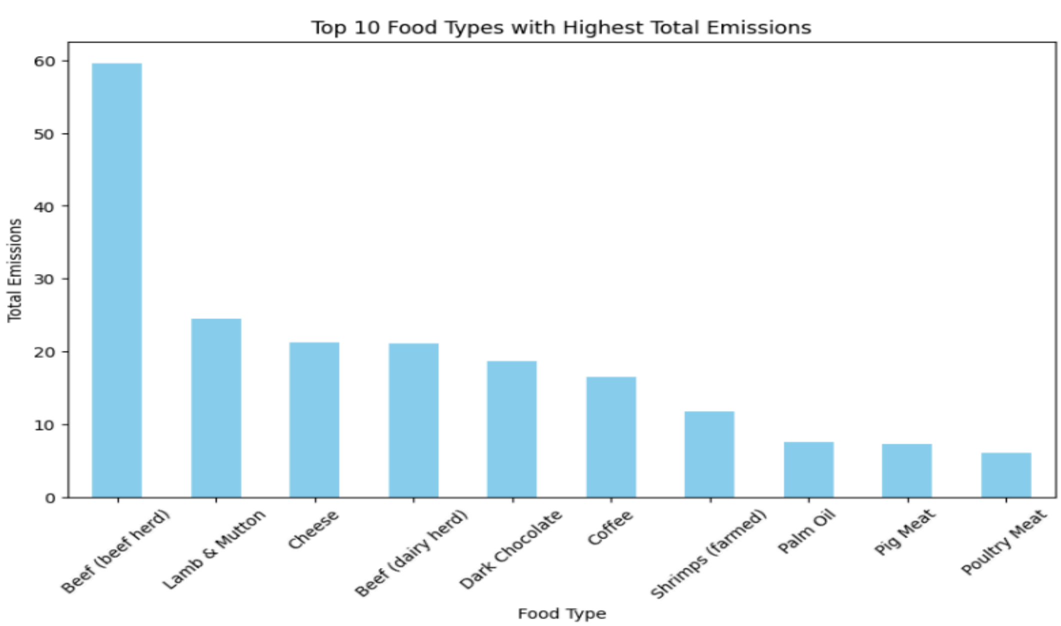
```
Recommended products for 'Parle-G Biscuits':
3      Britannia Bourbon Biscuits
4            Britannia Marie GOLD
6            Britannia Milk Bikis
7             Britannia Good Day
15                 Aashirvaad ATTA
16             Sunfeast MOM'S MAGIC
17                 PERK Chocolate
18               KOPIKO Chocolate
19             MANGO BITE Chocolate
20                  Annapurna ATTA
21              Melody Chcocolate
22                  BRITANNIA RUSK
29              HALDIRAM MOONG DAL
32                          CHOCOS
36         Visakha Dairy Ganga Milk
37                    Saffola Oats
40              Daawat Basmati Rice
41               Maaza Mango Drink
42           Real Fruit Apple Juice
```

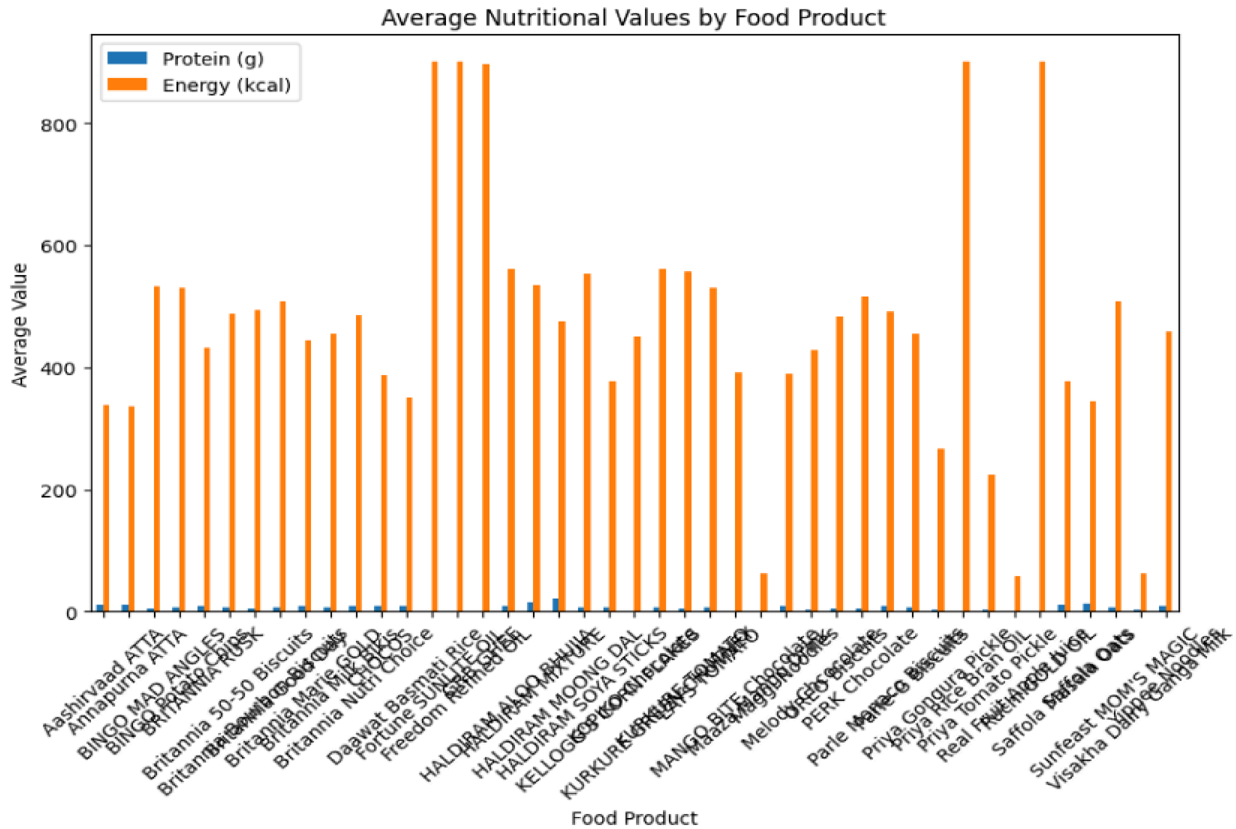
Observation:

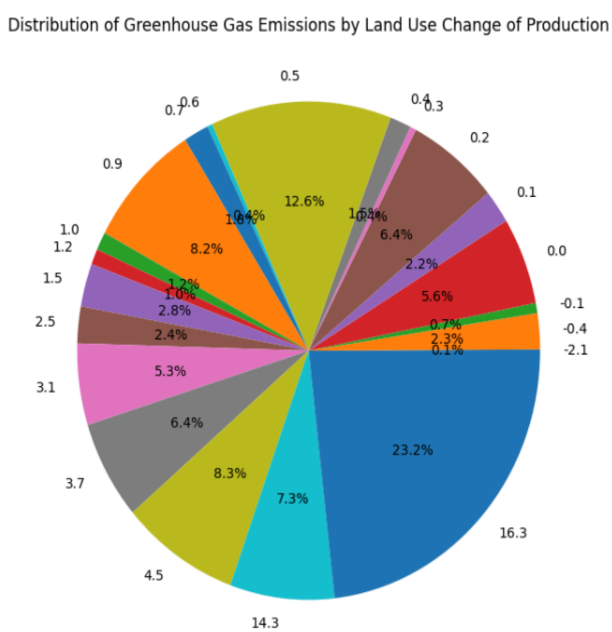
To develop a recommendation system, we use a content-based filtering approach since we have nutritional information available for each food product. After clustering the food products, we assign each product its corresponding cluster label. We define a function that takes a food product as input and recommends other products from the same cluster.

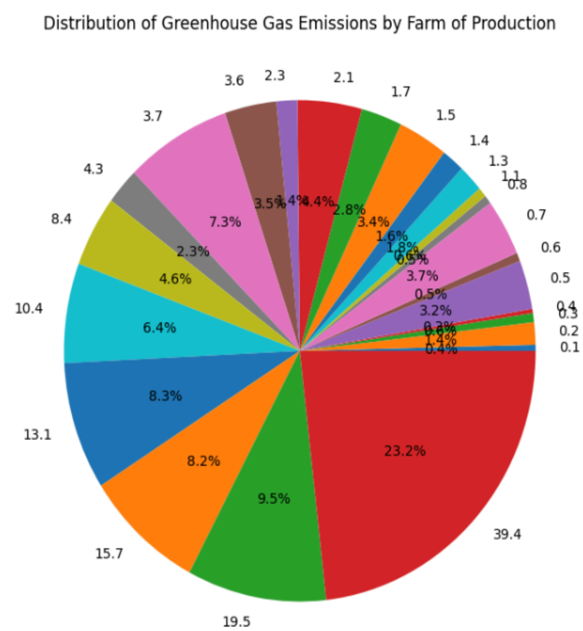
This output indicates the recommended products that belong to the same cluster as 'Parle Monaco Biscuits'. In other words, these are food products that have similar nutritional attributes to 'Parle Monaco Biscuits' and are therefore considered sustainable alternatives or potential substitutes. These recommendations are based on the clustering of food products according to their nutritional content.

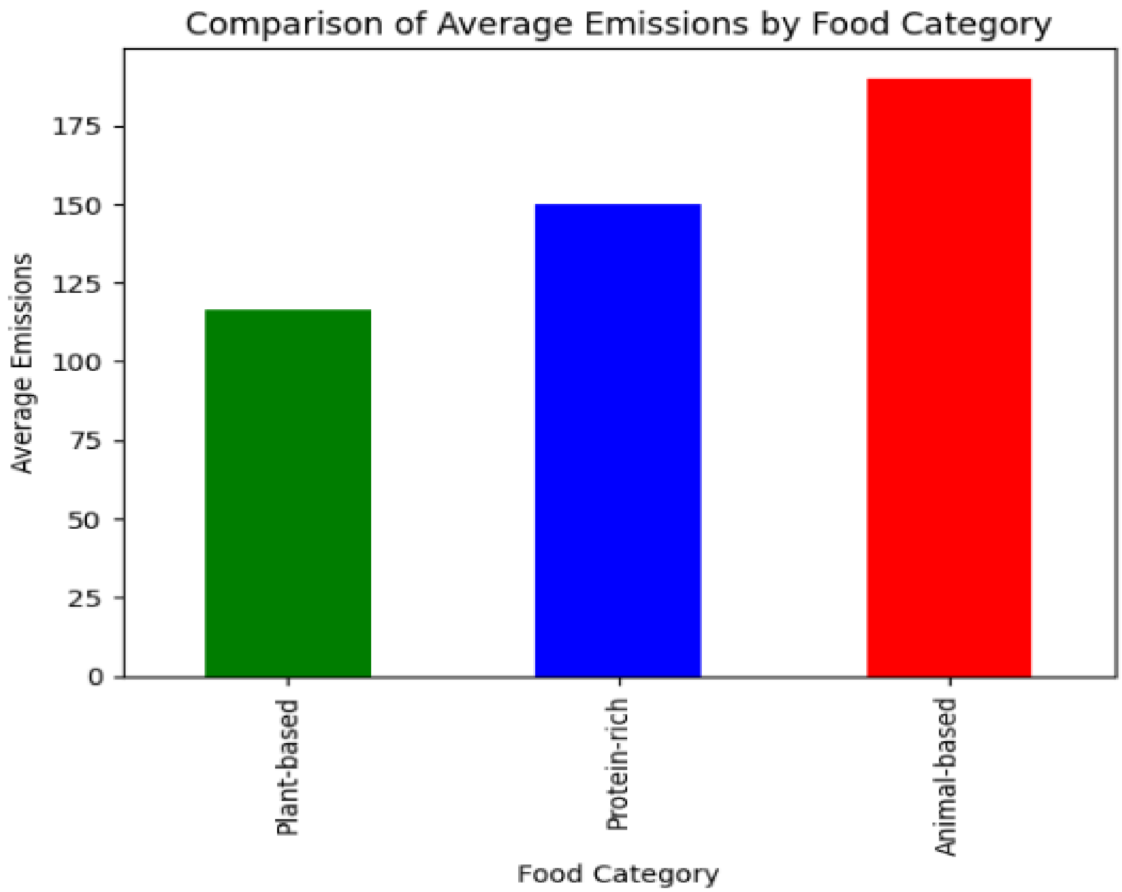
6. Visualization and Analysis:

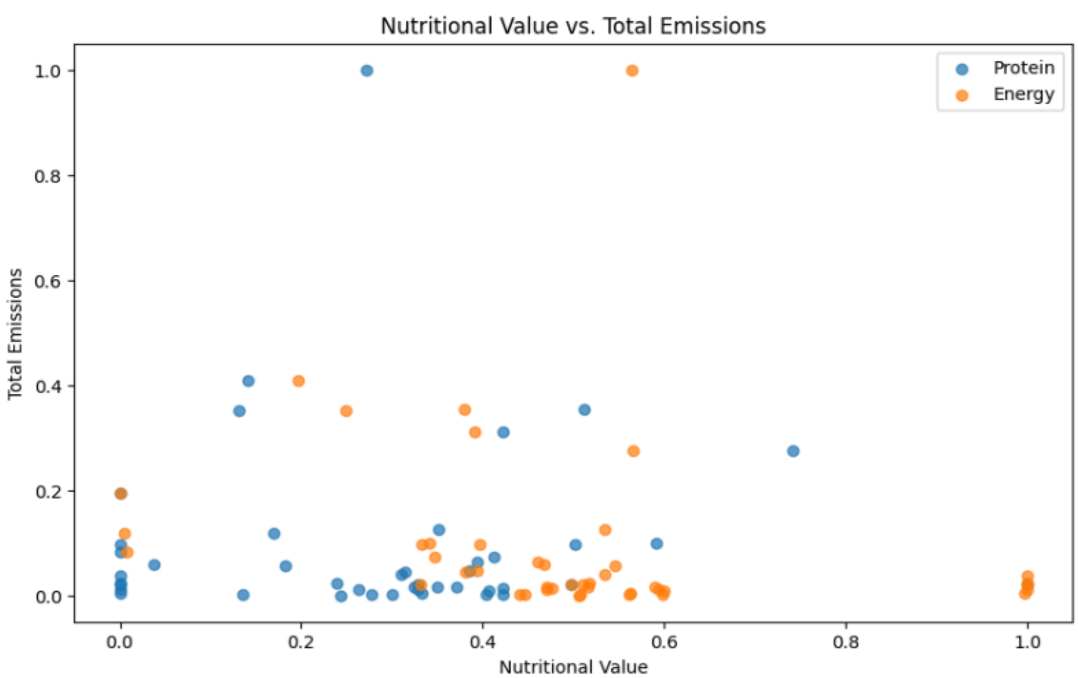


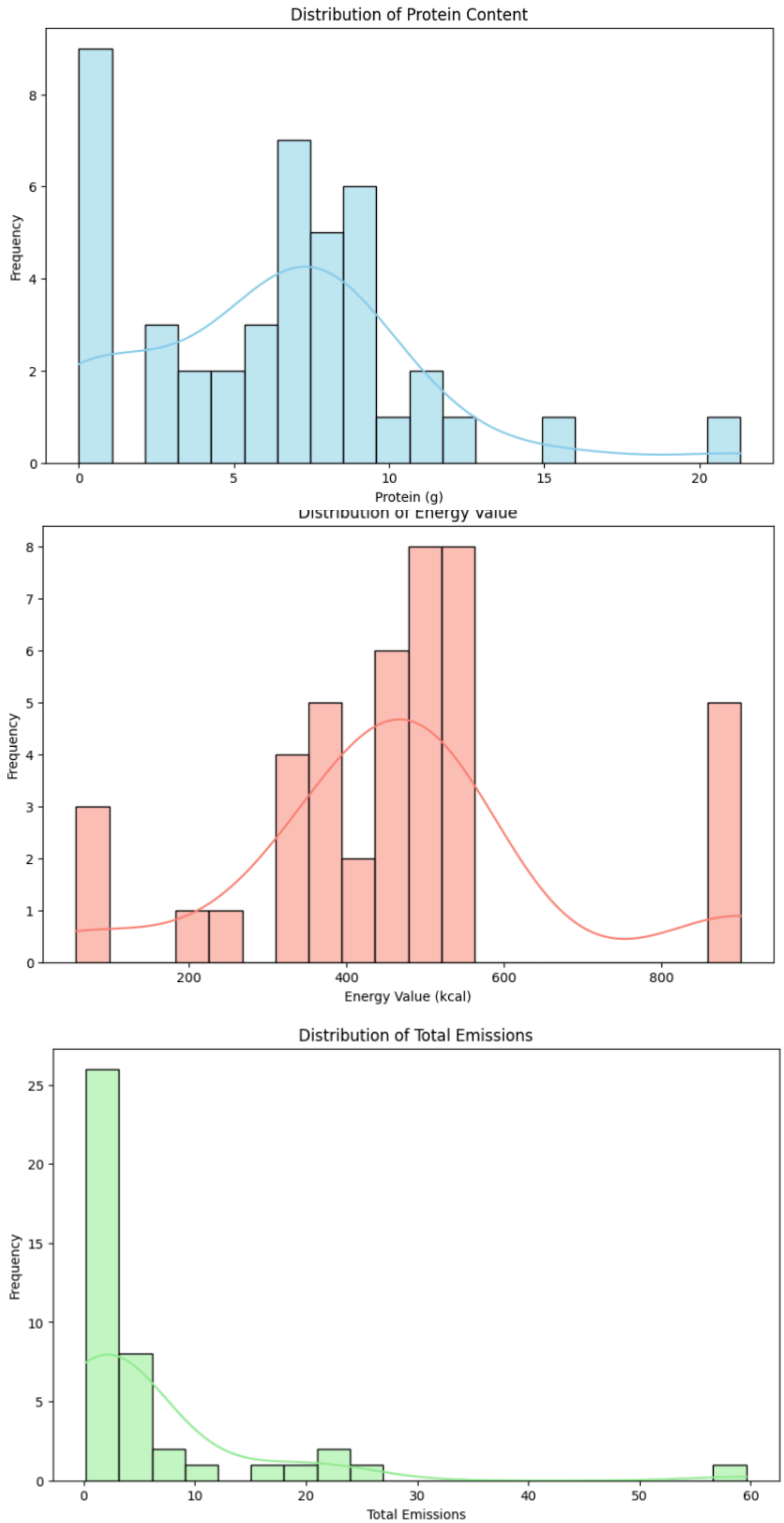

Average Nutritional Values by Food Product
Protein (g)
Energy (kcal)
Average Value
Food Product


Distribution of Greenhouse Gas Emissions by Land Use Change of Production

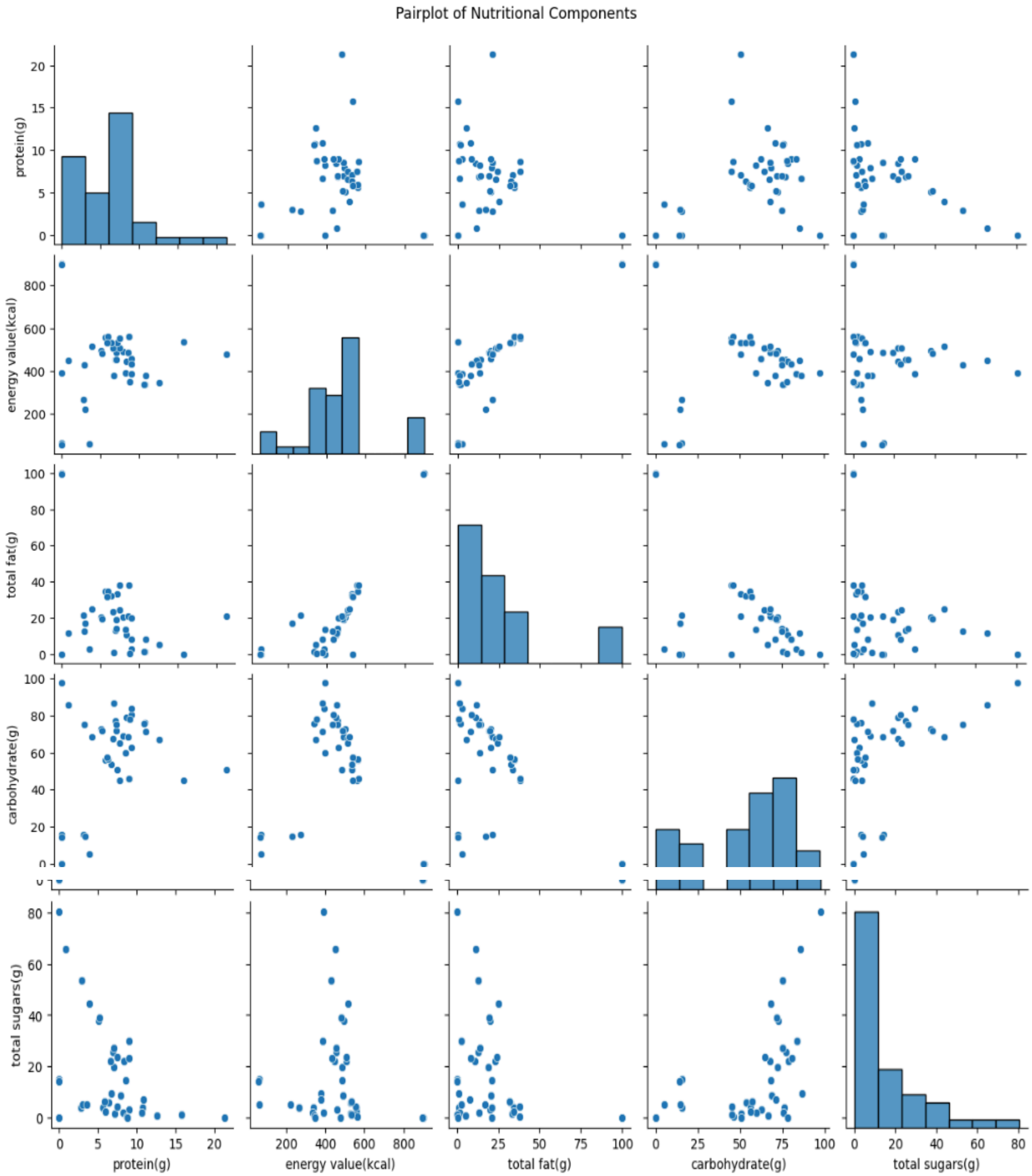

Distribution of Greenhouse Gas Emissions by Farm of Production


Comparison of Average Emissions by Food Category
Average Emissions
0
25
50
75
100
125
150
175
Plant-based
Protein-rich
Animal-based
Food Category

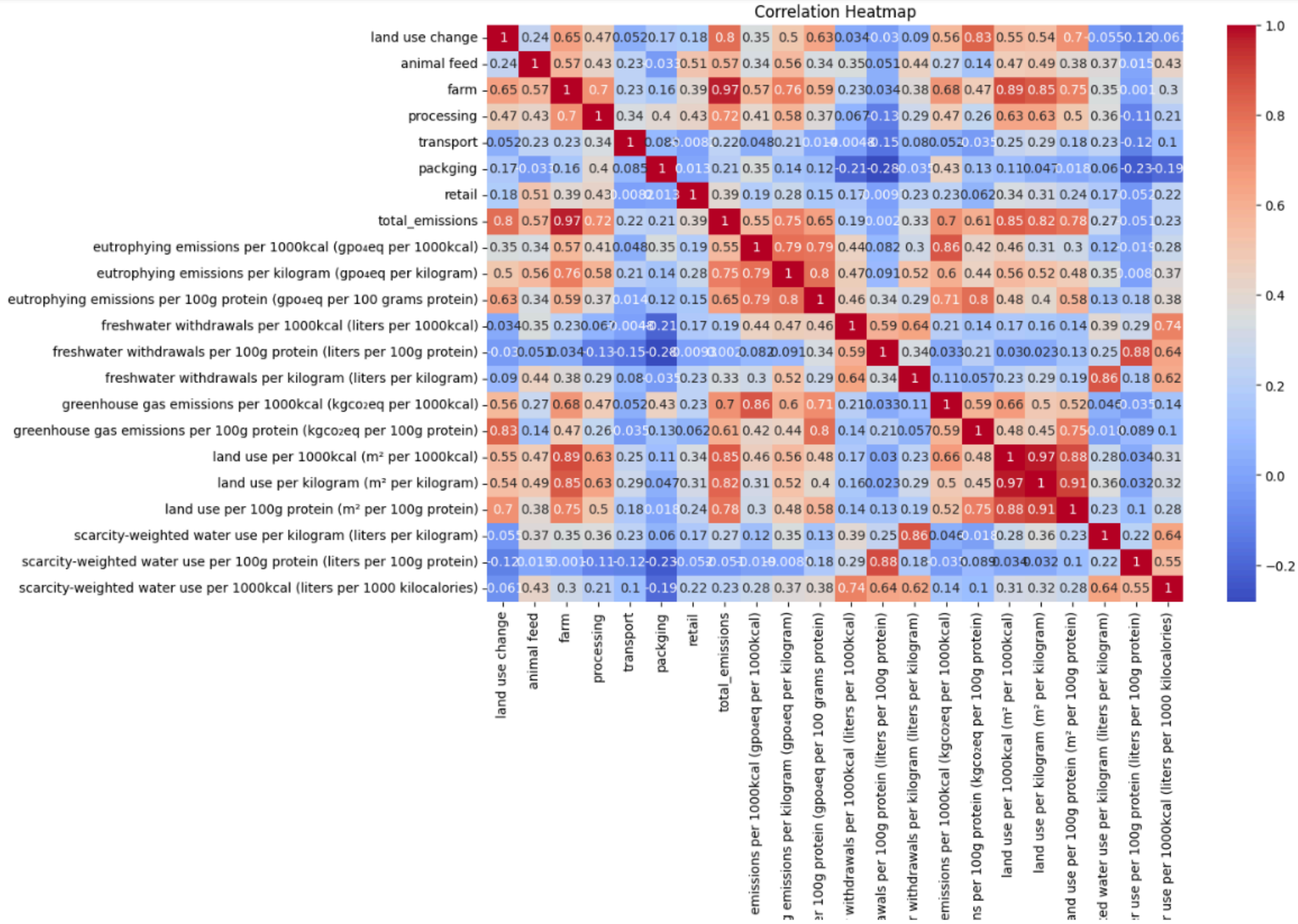

Nutritional Value vs. Total Emissions
Protein
Energy
Total Emissions
0.0
0.2
0.4
0.6
0.8
1.0
Nutritional Value

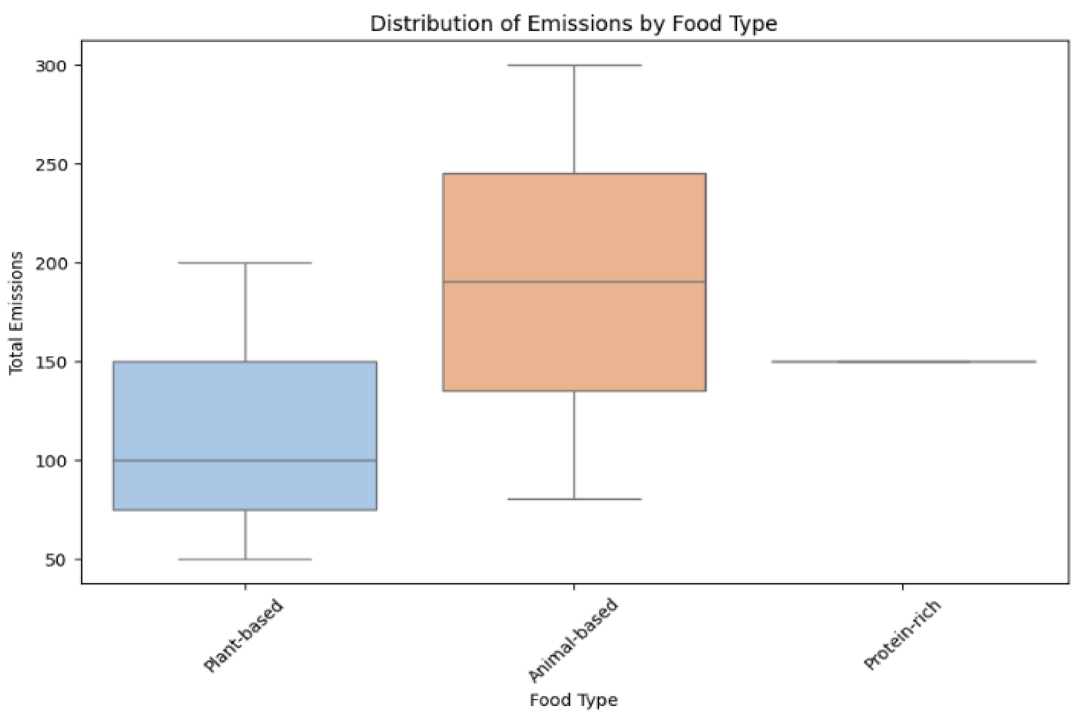


Distribution of Protein Content
Frequency
Protein (g)
Distribution of Energy Value
Frequency
Energy Value (kcal)
Distribution of Total Emissions
Frequency
Total Emissions



Pairplot of Nutritional Components
protein(g)
energy value(kcal)
total fat(g)
carbohydrate(g)
total sugars(g)

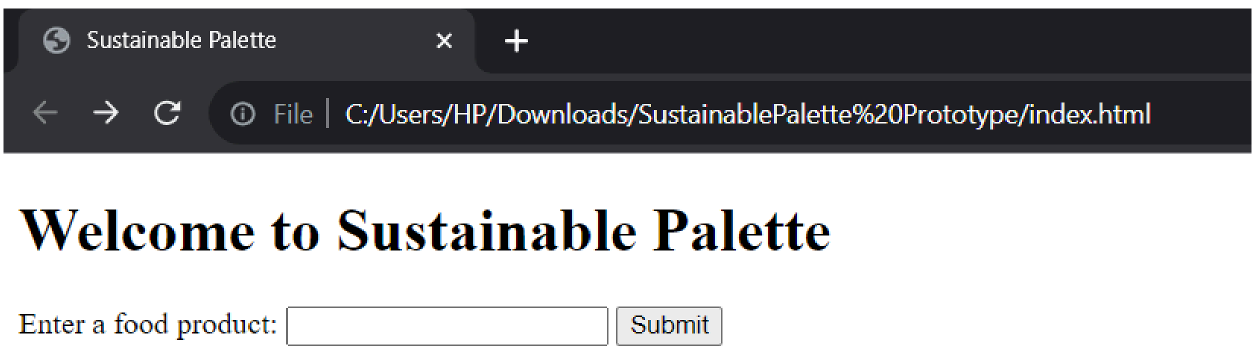

Correlation Heatmap
land use change
animal feed
farm
processing
transport
packging
retail
total_emissions
eutrophying emissions per 1000kcal (gpo₄eq per 1000kcal)
eutrophying emissions per kilogram (gpo₄eq per kilogram)
eutrophying emissions per 100g protein (gpo₄eq per 100 grams protein)
freshwater withdrawals per 1000kcal (liters per 1000kcal)
freshwater withdrawals per 100g protein (liters per 100g protein)
freshwater withdrawals per kilogram (liters per kilogram)
greenhouse gas emissions per 1000kcal (kgco₂eq per 1000kcal)
greenhouse gas emissions per 100g protein (kgco₂eq per 100g protein)
land use per 1000kcal (m² per 1000kcal)
land use per kilogram (m² per kilogram)
land use per 100g protein (m² per 100g protein)
scarcity-weighted water use per kilogram (liters per kilogram)
scarcity-weighted water use per 100g protein (liters per 100g protein)
scarcity-weighted water use per 1000kcal (liters per 1000 kilocalories)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2


Distribution of Emissions by Food Type
Total Emissions
300
250
200
150
100
50
Plant-based
Animal-based
Protein-rich
Food Type

7. Integration:

To integrate the trained machine learning models and recommendation system into the Sustainable Palette app prototype, we perform the following steps:

1. Incorporate the trained machine learning models and recommendation system into the backend of the Sustainable Palette app. This involves writing the necessary code to load the models, preprocess user input, make predictions, and generate recommendations.
2. Create a user interface (UI) for the app that allows users to input food items and receive carbon footprint analysis and sustainable alternatives.
3. Develop a backend API to handle requests from the UI and communicate with the machine learning models and recommendation system. This API should handle data validation, process user input, and return relevant results to the frontend.



STEP 3: BUSINESS MODELLING

BUSINESS OPPORTUNITY

The Sustainable Palette app presents numerous opportunities in the burgeoning market of sustainable living solutions. By addressing the growing demand for tools that empower consumers to make environmentally conscious choices, the app can capitalize on the rising tide of eco-consciousness. Partnerships with grocery stores and restaurants offer avenues for revenue generation while showcasing the app's value to businesses committed to sustainability. Additionally, targeted advertising from eco-friendly brands allows for further monetization. With the global shift towards sustainable practices gaining momentum, the Sustainable Palette app is well-positioned to capitalize on this trend and establish itself as a leading platform for informed, eco-friendly living.

MONETIZATION IDEA

The business model for the Sustainable Palette app includes multiple revenue streams:

- Freemium Model: Basic features are offered for free, with premium subscription options providing access to advanced features.
- Partnerships: Collaborations with grocery stores and restaurants for highlighted visibility and revenue sharing.
- Advertising: Non-intrusive, targeted advertisements from eco-friendly brands.

The business model for the Foodprint Tracker is designed to ensure sustainability and profitability while offering value to users. The monetization strategy is based on a freemium model, offering both free and premium subscription options, and exploring partnerships for additional revenue streams. Here's a detailed elaboration of the Business Model:

Freemium Model:

1. Basic (Free) Features: The core features of the Foodprint Tracker, including carbon footprint analysis, personalized eco-profile creation, and sustainable

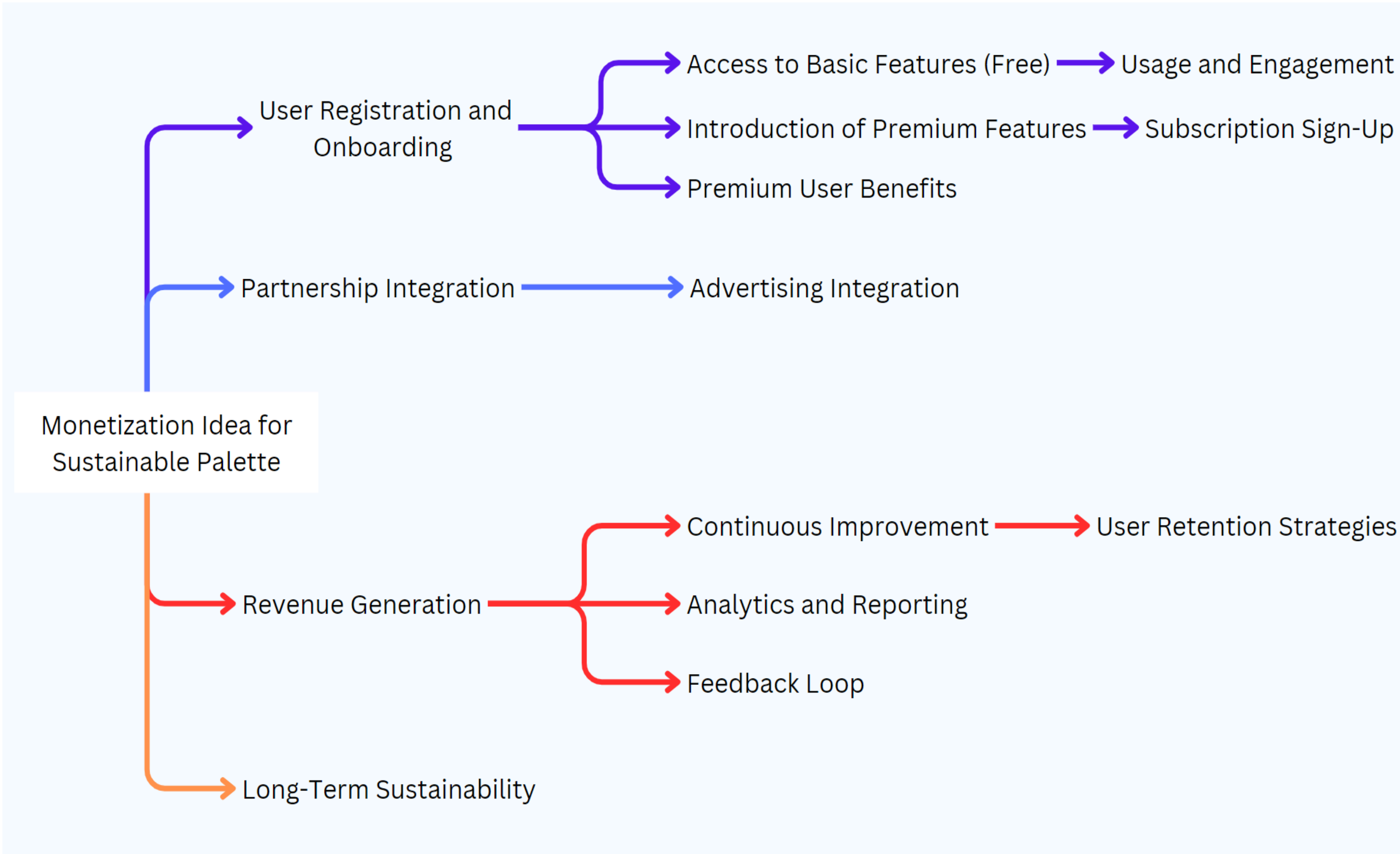
alternatives recommendations, will be accessible to users for free. This ensures a wide user base and encourages initial adoption.

2. Premium Subscription: A premium subscription tier will offer advanced features and enhanced user experiences. Premium subscribers may benefit from additional services such as:
    - ➢ Personalized Meal Plans: Tailored meal plans based on individual preferences and sustainability goals.
    - ➢ Detailed Nutritional Insights: In-depth nutritional analysis of food choices, including micronutrient breakdowns.
    - ➢ Access to a Broader Database: Premium users may have access to an expanded database of sustainable recipes and food products.
3. Partnerships with Grocery Stores and Restaurants: The Foodprint Tracker can collaborate with grocery stores and restaurants to feature their eco-friendly products or menu items. These businesses could pay for highlighted visibility, showcasing their commitment to sustainability within the app. This partnership model can contribute to additional revenue.
4. Advertising Opportunities: Non-intrusive, targeted advertisements from eco-friendly brands, sustainable products, or health-conscious services can be included in the app. Advertisers can pay for space within the app to reach a highly relevant and engaged audience.

Revenue Projections and Sustainability:

1. Subscription Plans: The premium subscription model provides a recurring revenue stream, while the free version ensures a large user base for potential upsells.
2. Partnerships and Advertising: Revenue from partnerships with grocery stores, restaurants, and advertising can be diversified, contributing to overall financial sustainability.
3. Continuous Value Addition: Regular updates, new features, and expanded databases for both free and premium users will enhance the app's value, driving customer retention and attracting new users.

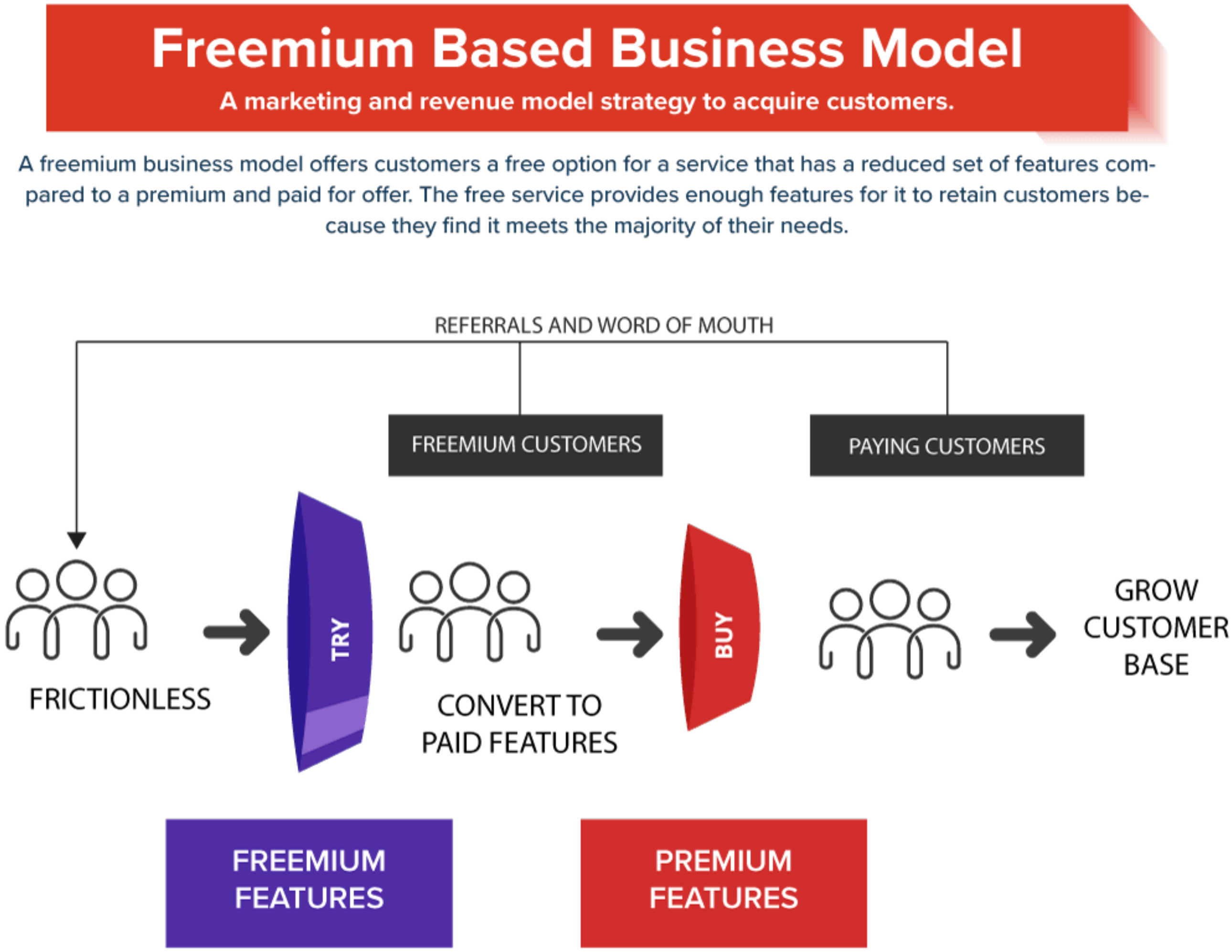
Detailed Workflow:



BUSINESS MODEL DEVELOPMENT

Based on the product idea and requirements of the Sustainable Palette app, as well as the target market of environmentally conscious consumers and health enthusiasts, a hybrid business model combining elements of the freemium model and partnerships/advertising model seems to be the most suitable.

1. Freemium Model: Offering basic features of the app for free allows for wide user adoption and engagement. Basic features could include carbon footprint analysis, personalized eco-profile creation, and basic sustainable alternatives recommendations. This ensures that the app is accessible to a broad audience and encourages initial user engagement.



2. Partnerships and Advertising: Collaborating with grocery stores, restaurants, and eco-friendly brands for highlighted visibility and revenue sharing can be a significant source of income. Partnerships can involve featuring eco-friendly products or menu items within the app, promoting sustainability initiatives, and offering exclusive deals to app users. Additionally, targeted advertising from eco-friendly brands can provide additional revenue streams without compromising user experience.

By combining these models, the Sustainable Palette app can generate revenue from both premium subscriptions and partnerships/advertising while maintaining a user-friendly and accessible platform for consumers. This approach aligns with the app's mission of promoting sustainable living and empowers users to make informed dietary choices while also ensuring long-term financial sustainability.

PRODUCT WORKING

Consider a scenario where a user is evaluating the environmental impact of different types of milk. This scenario illustrates how the app assists users in making environmentally conscious and personalized dietary choices, fostering a sense of community and continuous improvement in sustainability practices.

Scenario Flow:

1. User Input:

- Sarah logs into the Foodprint Tracker app and enters her dietary preferences and sustainability goals.
- She navigates to the "Dairy" section to assess the environmental impact of different types of milk.

2. Carbon Footprint Analysis:

- The app employs AI algorithms to analyze the carbon footprint of various milk options, considering factors like production methods, transportation, and packaging.
- Sarah sees a comparative analysis of regular cow's milk, almond milk, and oat milk.

3. Personalized Eco-Profile:

- Based on Sarah's preferences and previous consumption patterns, the app generates a personalized eco-profile.
- The profile highlights her current carbon footprint from milk consumption and suggests goals for reduction.

4. Sustainable Alternatives:

- The app recommends sustainable alternatives based on Sarah's profile, considering her vegetarian diet.
- It suggests almond milk as a plant-based alternative with a lower carbon footprint compared to cow's milk.

5. Grocery Shopping Assistance:

- Sarah decides to purchase almond milk and adds it to her virtual shopping cart within the app.
- The app provides real-time feedback, suggesting other eco-friendly products and offering information on sustainable brands.

6. Community Engagement:

- Sarah shares her eco-profile and the decision to switch to almond milk on the app's community forum.
- Other users comment, offering additional suggestions and sharing their experiences with sustainable milk choices.

7. Nutritional Guidance:

- Alongside the environmental impact, the app provides nutritional insights, ensuring that Sarah makes an informed decision considering both health and sustainability.

8. Continuous Learning:

- The app, through continuous learning mechanisms, updates Sarah's eco-profile based on her evolving preferences and consumption patterns.
- Sarah receives periodic recommendations for new sustainable products and practices.

STEP 4: FINANCIAL MODELLING WITH MACHINE LEARNING & DATA ANALYSIS

FINANCIAL MODELLING

(a) Market Identification

The Sustainable Palette app targets the market of sustainable living solutions, appealing to environmentally conscious consumers and health enthusiasts alike. With a focus on promoting eco-friendly and nutritious food choices, the app aims to tap into the growing demand for sustainable products and lifestyles. By catering to individuals who prioritize health and environmental sustainability, the app positions itself as a valuable tool for those seeking to make informed dietary decisions that align with their values.

(b) Data Collection

To support its market identification efforts, the Sustainable Palette app relies on comprehensive data collection from various online sources. Market research reports, industry publications, and government databases provide valuable insights into the sustainable living market. These sources offer data on growth rates, consumer preferences, spending patterns, and emerging trends within the industry. By leveraging this information, the app can better understand the needs and preferences of its target audience, ensuring its offerings align with market demand.

(c) Performance and Forecasting

Performance and forecasting play crucial roles in guiding the Sustainable Palette app's strategic decisions. Time series forecasting techniques, such as regression models or exponential smoothing, are employed to analyze historical data and predict future market trends. By analyzing patterns and trends in consumer behavior, the app can anticipate shifts in demand for sustainable food products and adjust its strategies accordingly. This proactive approach enables the app to stay ahead of market fluctuations and capitalize on emerging opportunities within the sustainable living market.

(d) Designing

The design of the Sustainable Palette app's financial model hinges on the growth trajectory of the sustainable living market. Based on insights gathered from market research and forecasting analyses, a suitable financial equation is formulated to guide the app's financial planning and decision-making processes. This equation takes into account factors such as revenue projections, cost structures, and investment requirements. By designing a robust financial model, the app can establish clear financial goals and strategies to drive its long-term sustainability and success in the competitive marketplace.

FINANCIAL EQUATION

The financial equation for the development of the Sustainable Palette app depends on the growth pattern of the target market. Two primary models are considered: linear and exponential.

(a) Linear Financial Model: If the market is growing linearly, a linear financial model can be employed. In this model, the total profit *(y)* is represented as a function of total sales *(x(t))* over time, incorporating pricing *(m)* and production/maintenance costs *(c)*. The equation for the linear financial model is given by:

$$y = mx(t) + c$$

In this equation:

- *(y)* represents the total profit generated by the app.
- *(m)* represents the pricing strategy adopted by the app, influencing the revenue generated per unit sold.
- *(x(t))* represents the total sales volume, which is considered as a function of time to account for market growth.
- *(c)* represents the production and maintenance costs associated with developing and maintaining the app.

(b) Exponential Financial Model: Alternatively, if the market exhibits exponential growth, an exponential financial model is applied to reflect the rapid expansion of the market. In this model, the total profit *(y)* is calculated exponentially based on the growth of the market over time. The equation for the exponential financial model is given by:

$$y = ae^{bt}$$

In this equation:

- *(y)* represents the total profit generated by the app.
- *a* represents a constant multiplier that influences the initial profit level.
- *e* is the base of the natural logarithm.
- *b* represents the growth rate of the market.
- *t* represents time, indicating the duration of market growth.

Selection of Financial Model:

The choice between the linear and exponential financial models depends on the observed growth trend of the sustainable living market. Market data analysis and forecasting techniques are employed to determine whether the market is growing linearly or exponentially. Once the growth pattern is identified, the corresponding financial model is applied to guide the app's financial planning and decision-making processes, ensuring alignment with market dynamics and maximizing profitability.

Given the nature of the Sustainable Palette app and its potential market, a *linear financial model* can be considered and employed initially.

- ➔ While the market for sustainable living solutions is growing rapidly, it may not yet be exhibiting the kind of exponential growth that would warrant an exponential financial model. Linear growth is often more common in markets that are still developing and haven't reached their full potential.

- ➔ A linear financial model provides a more stable and predictable framework for financial planning, especially in markets where growth rates may fluctuate or plateau over time. This allows for more accurate revenue projections and cost estimations, which are essential for sustainable business operations.

- ➔ Linear financial models are generally easier to implement and understand, making them well-suited for startups or new ventures like the Sustainable Palette app. They require fewer assumptions about future growth rates and are therefore less prone to errors in forecasting.

A linear financial model provides a solid foundation for the Sustainable Palette app to plan and manage its finances effectively in the initial stages. As the app gains traction and the market matures, it can reassess its growth trajectory and consider transitioning to an exponential financial model if warranted by the market conditions at that time.

Financial Equation:

To derive the financial equation for the Sustainable Palette app based on a linear financial model, we can use the formula for a simple linear regression:

$$y = mx + b$$

Where:

- (y) represents the dependent variable (total profit)
- (x) represents the independent variable (time, representing market growth)

- *(m)* represents the slope of the line (pricing)
- *(b)* represents the y-intercept (production and maintenance costs)

In the context of the Sustainable Palette app:

- *Total Profit (y)*: This represents the overall revenue generated by the app, which includes revenue from premium subscriptions, partnerships, and advertising.
- *Time (x)*: This represents the passage of time, reflecting the growth of the sustainable living market. As the market grows, the app's revenue is expected to increase.
- *Slope (m)*: This represents the pricing strategy of the app. It reflects how changes in market growth (time) affect total profit. A higher pricing coefficient indicates that the app's revenue grows faster with market growth.
- *Y-intercept (b)*: This represents the fixed costs associated with production and maintenance of the app. It includes expenses such as development costs, data acquisition, maintenance, and marketing costs.

Therefore, the financial equation for the Sustainable Palette app can be expressed as:

Total Profit = (Pricing x Time) + Production and Maintenance Costs

This equation captures the relationship between total profit and market growth, incorporating the app's pricing strategy and fixed costs. By monitoring changes in market growth over time and adjusting pricing and cost structures accordingly, the app can optimize its profitability and sustainability in the long run.

CONCLUSION

The Sustainable Palette app offers a timely solution to the global challenge of unsustainable food consumption. By leveraging artificial intelligence and nutritional insights, it empowers users to make informed, eco-conscious dietary choices, fostering a collective commitment to environmental conservation. The business model, with its freemium approach, partnerships, and targeted advertising, ensures both sustainability and profitability. Financial modeling using a linear financial model provides a stable framework for financial planning, enabling accurate revenue projections and cost estimations. The derived financial equation captures the app's potential for profitability over time, guiding its long-term success. Overall, the Sustainable Palette app stands poised to become a leading platform for sustainable living, driving positive change in individuals' dietary habits and contributing to a greener, healthier future for our planet.