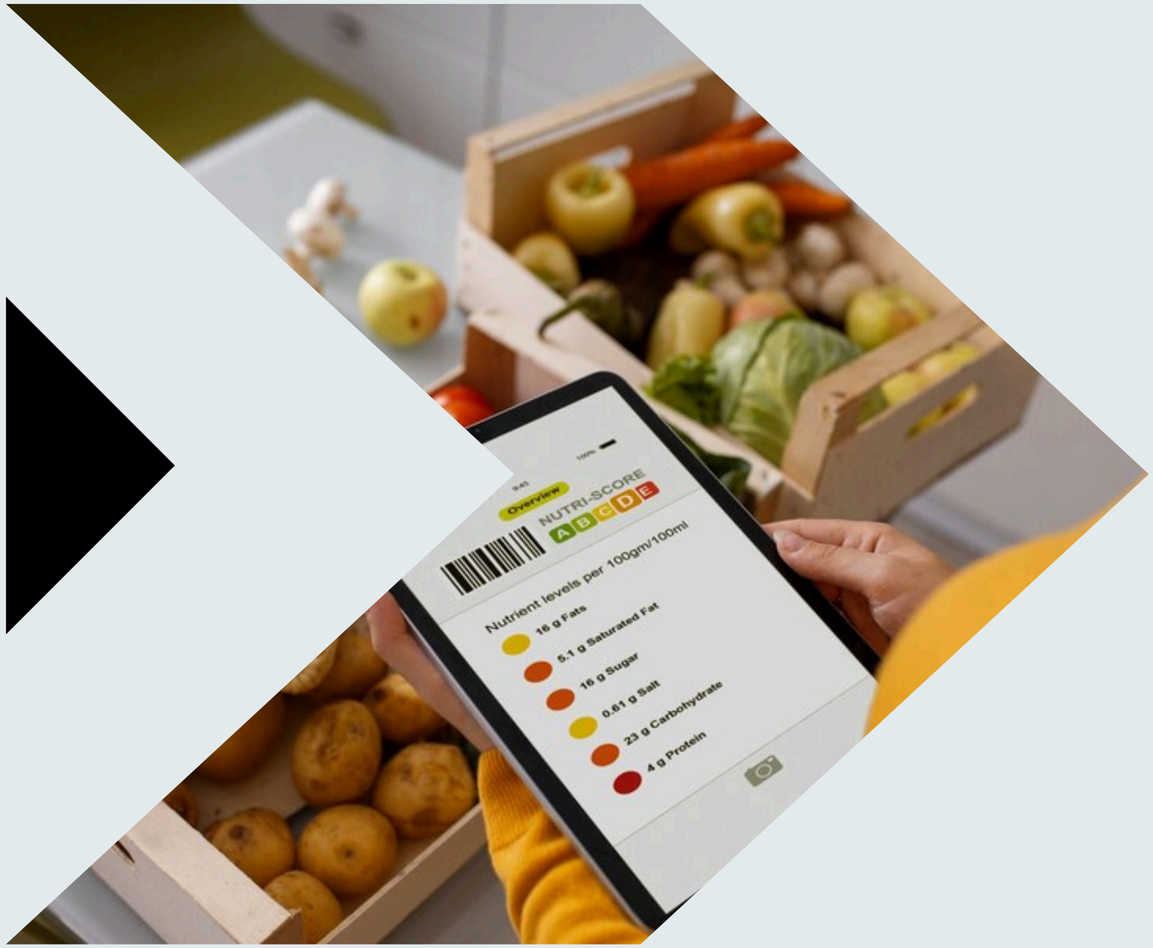
Food Calories Estimation



Overview

- This project focuses on creating a system or application that estimates the caloric content of food items based on their ingredients and macronutrients.
- This project aims to provide accurate caloric information for individuals seeking to monitor their diet, manage health conditions, and achieve wellness goals.

Why Calorie Estimation?

- Personal Growth
- Health Awareness
- Technology Integration
- Practical Application

What are calories?

Multiple definitions:

- Mathematically, calorie is the amount of energy that is required to raise the temperature of one mL of water by one degree Celsius.
- A calorie is a measure of energy expenditure and stored energy.

Calories refers to the total number of calories, or “energy” you get from all sources (carbohydrate, fat, protein, and alcohol) in a serving of a food or beverage.

References:
https://www.msdmanuals.com
https://www.livescience.com

More on Calories..

Carbohydrates, proteins, and fats are the main types of macronutrients in food.

- Carbohydrate: 4 calories/gram
- Protein: 4 calories/gram
- Fat: 9 calories/gram
- Fiber: 1.5 calories/gram

References:
https://www.msdmanuals.com
https://www.livescience.com

Data sets

The project contains 2 tabular data sets and 2 image data sets, sourced from Kaggle.

Tabular data set:

https://www.kaggle.com/datasets/utsavdey1410/food-nutrition-dataset

Image data set:

https://www.kaggle.com/datasets/marquis03/fruits-100

https://www.kaggle.com/datasets/kritikseth/fruit-and-vegetable-image-recognition

Tabular data sets

- The Food Calories Dataset captures the essence of various food items, providing insights into their nutritional makeup.
- Each entry details the food name, along with its caloric content and a range of nutrients.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Unnamed:	food	Caloric Value	Fat	Saturated Fats	Monounsaturated Fats	Polyunsaturated Fats	Carbohydrates	Sugars	Protein	Dietary Fiber	Cholesterol
2	0	cream cheese	51	5	2.9	1.3	0.2	0.8	0.5	0.9	0	14.6
3	1	neufchatel cheese	215	19.4	10.9	4.9	0.8	3.1	2.7	7.8	0	62.9
4	2	requeijao cremoso light catupiry	49	3.6	2.3	0.9	0	0.9	3.4	0.8	0.1	0
5	3	ricotta cheese	30	2	1.3	0.5	0.002	1.5	0.091	1.5	0	9.8
6	4	cream cheese low fat	30	2.3	1.4	0.6	0.042	1.2	0.9	1.2	0	8.1
7	5	cream cheese fat free	19	0.2	0.1	0.091	0.075	1.4	1	2.8	0	2.2
8	6	gruyere cheese	116	9.1	5.3	2.8	0.5	0.1	0.1	8.3	0	30.8
9	7	cheddar cheese	113	9.3	5.3	2.6	0.3	0.9	0.1	6.4	0	27.7
10	8	parmesan cheese	71	4.5	2.7	1.4	0.1	0.6	0.046	6.4	0	12.2
11	9	romano cheese	19	1.3	0.9	0.4	0.035	0.2	0.088	1.6	0	5.2
12	10	parmesan cheese grated	21	1.4	0.8	0.4	0.036	0.7	0.075	1.4	0	4.3
13	11	port salut cheese	465	37.2	22	12.3	1	0.8	0.8	31.4	0	162.4
14	12	swiss cheese	98	7.7	4.6	2	0.3	0.4	0	6.7	0	23.3
15	13	goat cheese hard	128	10.1	7	2.3	0.2	0.6	0.6	8.6	0	29.7

Image data sets

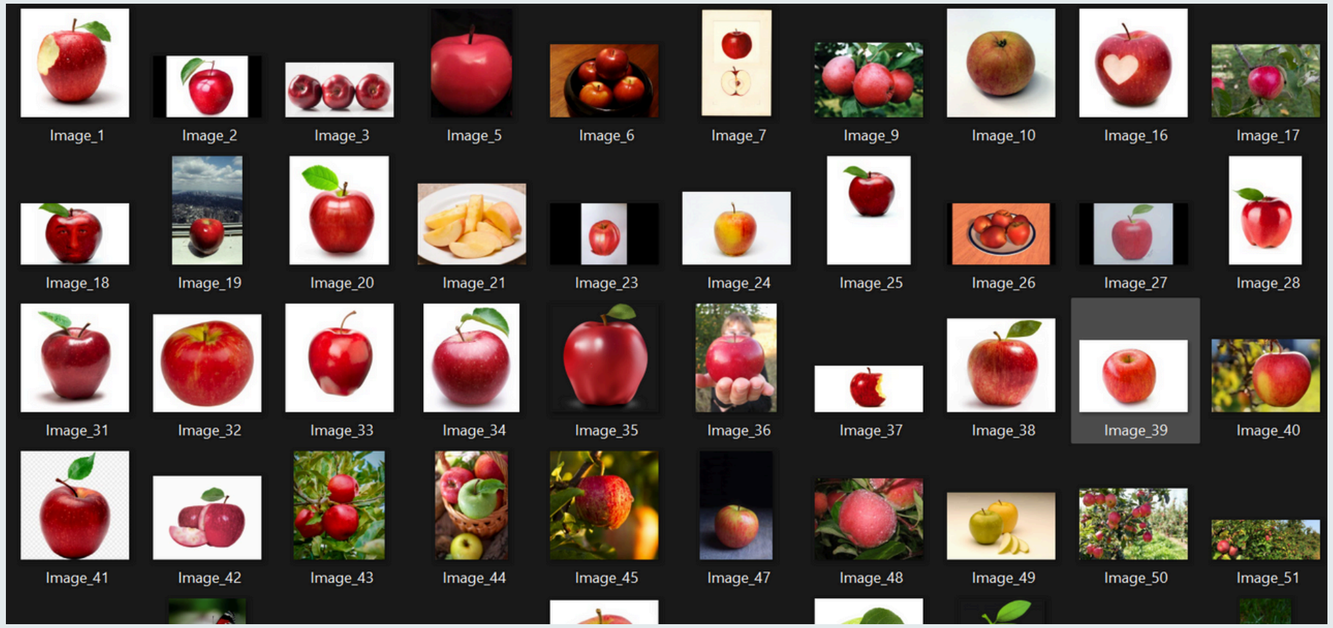


Image_1
Image_2
Image_3
Image_5
Image_6
Image_7
Image_9
Image_10
Image_16
Image_17
Image_18
Image_19
Image_20
Image_21
Image_23
Image_24
Image_25
Image_26
Image_27
Image_28
Image_31
Image_32
Image_33
Image_34
Image_35
Image_36
Image_37
Image_38
Image_39
Image_40
Image_41
Image_42
Image_43
Image_44
Image_45
Image_47
Image_48
Image_49
Image_50
Image_51

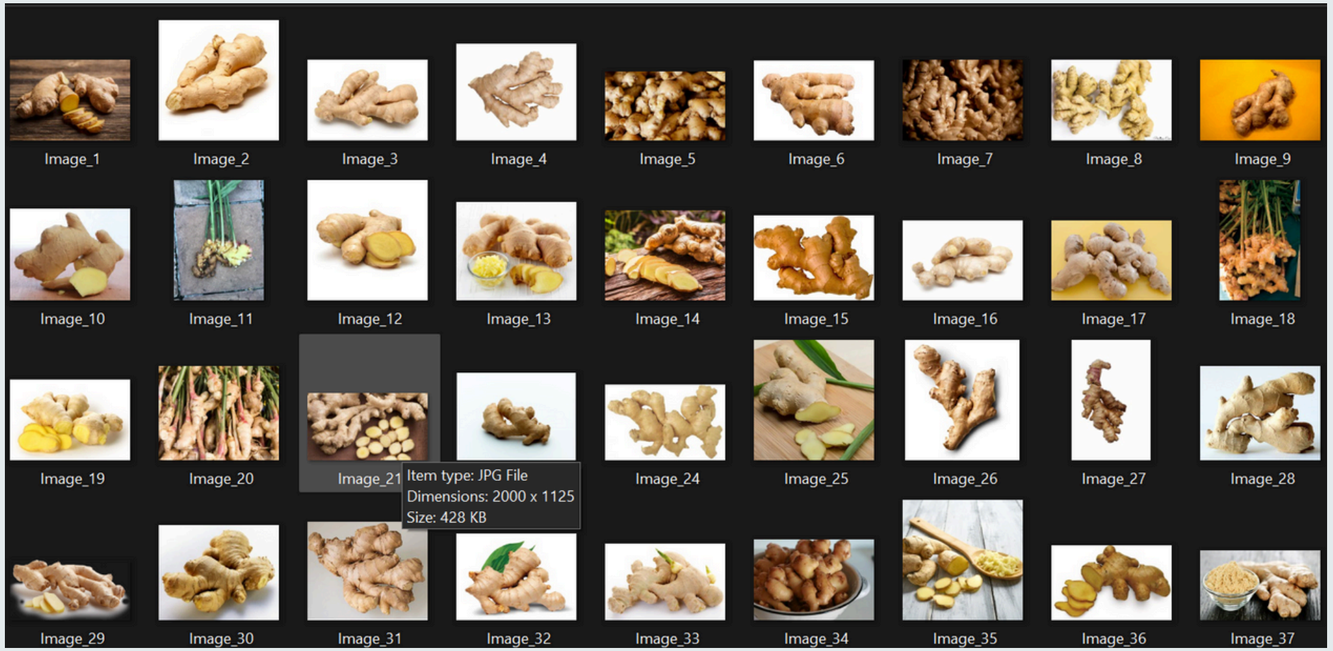


Image_1
Image_2
Image_3
Image_4
Image_5
Image_6
Image_7
Image_8
Image_9
Image_10
Image_11
Image_12
Image_13
Image_14
Image_15
Image_16
Image_17
Image_18
Image_19
Image_20
Image_21
Item type: JPG File
Dimensions: 2000 x 1125
Size: 428 KB
Image_24
Image_25
Image_26
Image_27
Image_28
Image_29
Image_30
Image_31
Image_32
Image_33
Image_34
Image_35
Image_36
Image_37

More on data sets...

Instance: The data set contains about 2300+ instances, i.e. food items.
Attributes: The data set contains about 30+ attributes, i.e. the micronutrients and macronutrients.
Image data sets: The image data set contains about 12,000+ images.

- Overall, the dataset offers a comprehensive look at the nutritional profiles of foods, enabling analysis of dietary habits and health implications.
- By exploring this data, one can appreciate how different foods contribute to our overall health and well-being.

Statistical summary

```
data.describe()
```

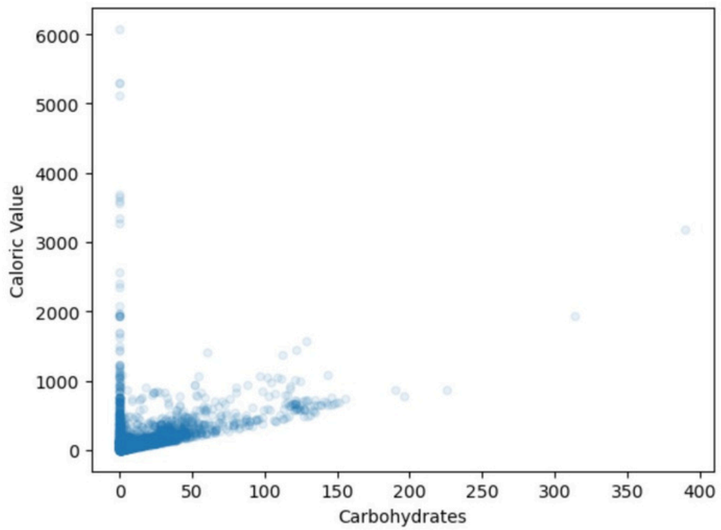
	No	Caloric Value	Fat	Saturated Fats	Monounsaturated Fats	Polyunsaturated Fats	Carbohydrates	Sugars	Protein	Dietary Fiber	...	Calcium	2
count	2395.00000	2395.000000	2395.000000	2395.000000	2395.000000	2395.000000	2395.000000	2395.000000	2395.000000	2395.000000	...	2395.000000	
mean	1197.00000	223.769520	10.176276	3.924917	4.133622	2.152844	18.589021	4.457459	13.400777	2.235790	...	52.047728	
std	691.52127	384.728244	29.008915	19.502262	12.939587	7.145738	29.406134	13.339929	32.294246	5.404483	...	115.933379	
min	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	
25%	598.50000	44.500000	0.300000	0.064000	0.058000	0.071000	0.500000	0.000000	0.800000	0.000000	...	0.600000	
50%	1197.00000	117.000000	2.100000	0.500000	0.500000	0.400000	6.800000	0.086000	3.500000	0.200000	...	13.900000	
75%	1795.50000	258.000000	9.400000	2.700000	3.400000	1.700000	25.050000	3.200000	13.300000	2.200000	...	48.250000	
max	2394.00000	6077.000000	550.700000	672.000000	291.100000	188.000000	390.200000	291.500000	560.300000	76.500000	...	1283.500000	1

8 rows × 35 columns

Scatter Plots

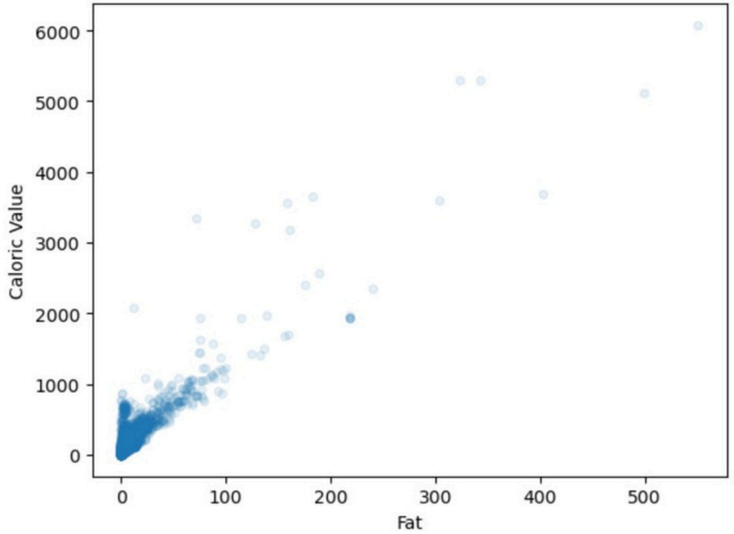
```
data.plot(kind="scatter", x="Carbohydrates", y="Caloric Value", alpha=0.1)
```

```
<Axes: xlabel='Carbohydrates', ylabel='Caloric Value'>
```



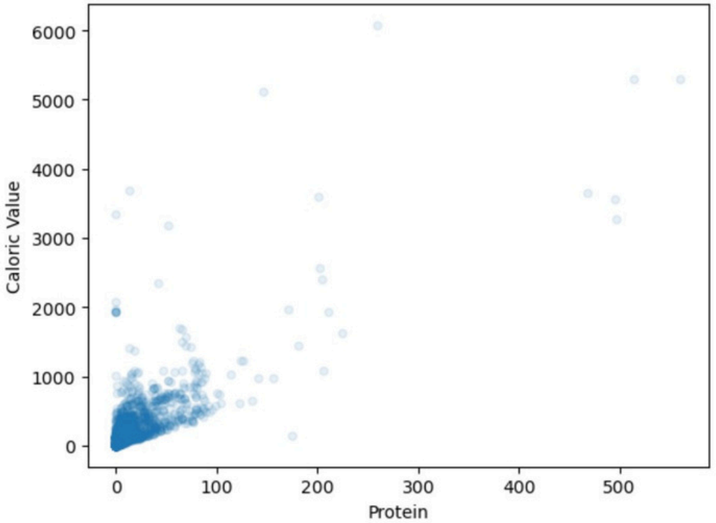

```
data.plot(kind="scatter", x="Fat", y="Caloric Value", alpha=0.1)
```

```
<Axes: xlabel='Fat', ylabel='Caloric Value'>
```




```
data.plot(kind="scatter", x="Protein", y="Caloric Value", alpha=0.1)
```

```
<Axes: xlabel='Protein', ylabel='Caloric Value'>
```



Data Correlation

```
corr_matrix=num_data.corr()
corr_matrix["Caloric Value"].sort_values(ascending=False)
```

```
Caloric Value           1.000000
Fat                     0.901783
Monounsaturated Fats    0.845348
Protein                 0.748770
Phosphorus              0.735810
Vitamin B3              0.693851
Potassium               0.681601
Vitamin B6              0.614840
Saturated Fats          0.606614
Polyunsaturated Fats    0.603871
Nutrition Density       0.535323
Water                   0.534724
Zinc                    0.534415
Magnesium               0.474511
Vitamin B5              0.467535
Vitamin B1              0.391420
Iron                    0.373881
Vitamin B2              0.305870
Carbohydrates           0.297667
Vitamin E               0.270989
Cholesterol             0.269212
Calcium                 0.265974
Dietary Fiber           0.152123
Sodium                  0.144128
Sugars                  0.118609
Selenium                0.067144
Manganese               0.057497
Copper                  0.025887
```

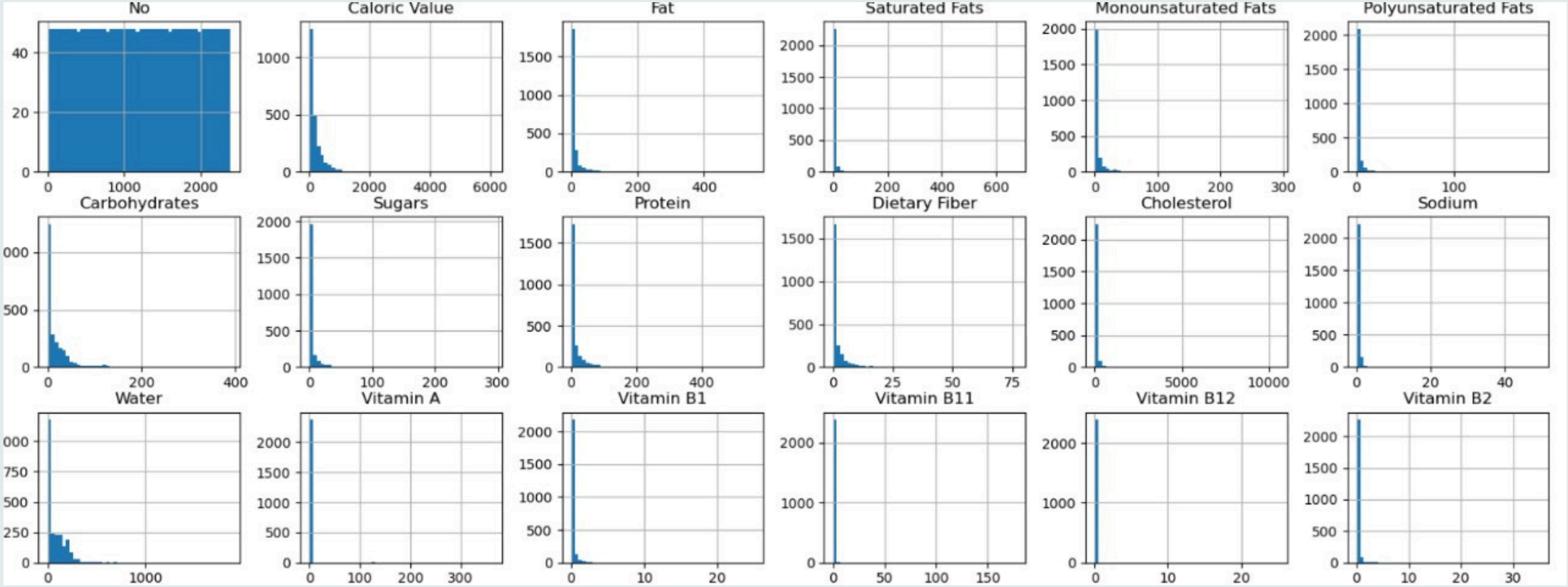
```
Carbohydrates           0.29766?
Vitamin E               0.270989
Cholesterol             0.269212
Calcium                 0.265974
Dietary Fiber           0.152123
Sodium                  0.144128
Sugars                  0.118609
Selenium                0.067144
Manganese               0.057497
Copper                  0.025887
Vitamin A               0.012179
Vitamin B11             0.008006
Vitamin C              -0.002313
Vitamin B12            -0.002386
Vitamin K              -0.006541
Vitamin D              -0.059747
No                     -0.106808
Name: Caloric Value, dtype: float64
```

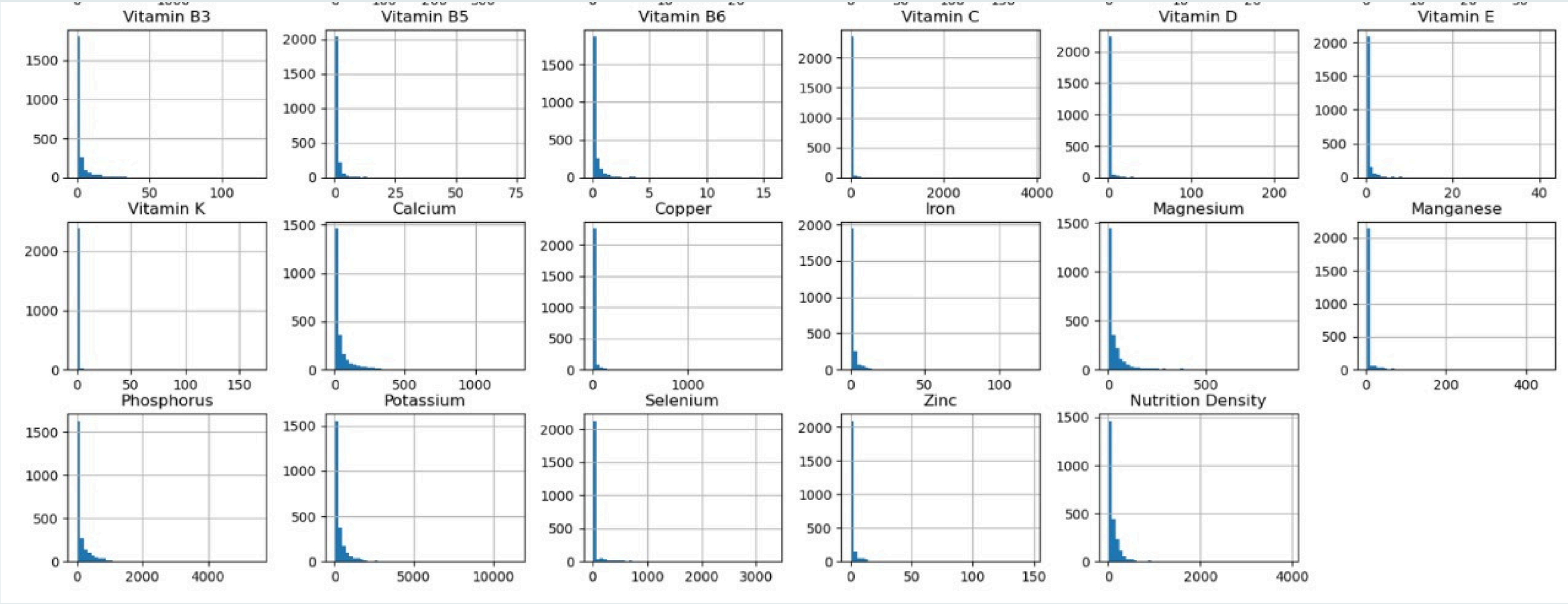
Data Information

```
[5]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2395 entries, 0 to 2394
Data columns (total 36 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   No                     2395 non-null   int64
 1   food                   2395 non-null   object
 2   Caloric Value          2395 non-null   int64
 3   Fat                    2395 non-null   float64
 4   Saturated Fats         2395 non-null   float64
 5   Monounsaturated Fats   2395 non-null   float64
 6   Polyunsaturated Fats   2395 non-null   float64
 7   Carbohydrates          2395 non-null   float64
 8   Sugars                 2395 non-null   float64
 9   Protein                2395 non-null   float64
 10  Dietary Fiber          2395 non-null   float64
```

Histograms


No
Caloric Value
Fat
Saturated Fats
Monounsaturated Fats
Polyunsaturated Fats
Carbohydrates
Sugars
Protein
Dietary Fiber
Cholesterol
Sodium
Water
Vitamin A
Vitamin B1
Vitamin B11
Vitamin B12
Vitamin B2


Vitamin B3
Vitamin B5
Vitamin B6
Vitamin C
Vitamin D
Vitamin E
Vitamin K
Calcium
Copper
Iron
Magnesium
Manganese
Phosphorus
Potassium
Selenium
Zinc
Nutrition Density

Research Questions:

1. To predict the caloric value of a given food item.
2. To identify a certain food item from the image and predict its calorific value.

Hypothesis:

1. The model/project accurately outputs the calories of a food item.
2. The model/project will accurately identify the image of the food item and output its calories.

Previous Uses in the ML Community

The dataset has been utilized in several ways within the machine learning community:

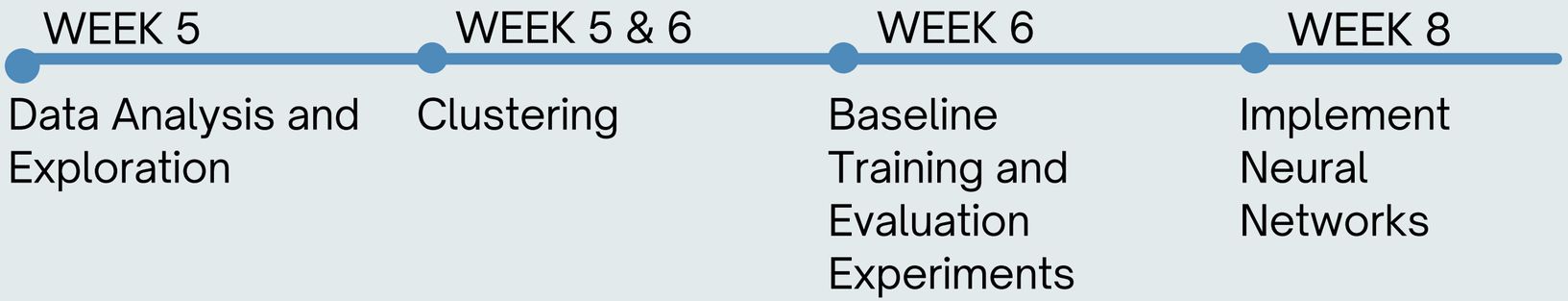
- Nutritional Analysis
- Predictive Modeling
- Clustering and Classification
- Recommendation Systems
- Health and Dietary Studies

Contributions

Topic for this model was researched and decided collectively by all group members.

- Data set research and Literature Survey - Yashica
- Description of Data - Alfiya T.
- General research and ppt - Tanishka
- Pre-processing of data - Rukhsana and Alfia

Workplan


WEEK 5
Data Analysis and Exploration
WEEK 5 & 6
Clustering
WEEK 6
Baseline Training and Evaluation Experiments
WEEK 8
Implement Neural Networks