SJIF Impact Factor 2021: 8.013| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 2 | February 2022 - Peer Reviewed Journal

# FOOD NUTRITIONAL ANALYSIS AND EDA

**Atishya Mahesh Jain*[1], KrishnaPriya B[2] , Seno Sunil[3] , Helen Grace Jikku[4] , Sheethal Kj[5]**
[1,2,3,4,5]*Karunya Institute of Technology and Sciences, Coimbatore*

**ABSTRACT**

*Food and nutrition are the way that we get fuel, providing energy for our bodies. We need to replace nutrients in our bodies with a new supply every day. Water is an important component of nutrition. Fats, proteins, and carbohydrates are all required. Nutrition is the science that interprets the nutrients and other substances in food in relation to maintenance, growth, reproduction, health and disease of an organism. It includes ingestion, absorption, assimilation, biosynthesis, catabolism and excretion. Knowing and eating mindfully is not only essential for a healthy gut but also for peace of mind. Also, A diet filled with vegetables, fruits and whole grains could help prevent major conditions such as stroke, diabetes and heart disease. More often than not, we like to gorge on our favourite foods which are not exactly the best for our bodies. While it is okay for such binges to occur occasionally, such diets can be extremely harmful if the person does not strike a balance with healthy foods. This article analyses the most common available foods and the nutritional facts in them.*

**INDEX TERMS** - *Analysis , Food , EDA , Nutrition*

## 1 - INTRODUCTION

Nutritional analysis is the process of determining the nutritional content of food. It is a vital part of analytical chemistry that provides information about the chemical composition, processing, quality control and contamination of food. It ensures compliance with trade and food laws. There are a variety of certified methods used for performing nutritional analysis. Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset. EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. In this article, we will understand EDA with the help of dataset and do the nutritional analysis . We will use Python language (Pandas library) for this purpose.

Everybody nowadays is mindful of what they eat. Counting calories and reducing fat intake is the number one advice given by all dieticians and nutritionists. Therefore, we need to know what foods are rich in what nutrients, don't we? The dataset contains a csv file with more than 300 foods each with the amount of Calories, Fats, Proteins, Saturated Fats, Carbohydrates, Fibers labelled for each food. Also, the foods are also categorized into various groups like Desserts, Vegetables, Fruits etc.

## 2 - PROCEDURE

### 2.1 Cleaning Data

Data cleaning is always the first step in any data science project. Although the data here seems clean, some minor alterations are required. Data cleaning is the process that removes data that does not belong in your dataset. Data transformation is the process of converting data from one format or structure into another. Transformation processes can also be referred to as data wrangling, or data munging, transforming and mapping data from one "raw" data form into another format for warehousing and analyzing.

```
import pandas as pd
import numpy as np
import plotly.express as px
import seaborn as sns
import plotly.offline as py
import plotly.graph_objects as go
```

+ Code + Markdown

```
nutrients=pd.read_csv('../kaggle/input/nutrition-details-for-most-common-foods/nutrients_csvfile.csv')
nutrients.head()
```

[2]:

| | Food | Measure | Grams | Calories | Protein | Fat | Sat.Fat | Fiber | Carbs | Category |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Cows' milk | 1 qt. | 976 | 660 | 32 | 40 | 36 | 0 | 48 | Dairy products |
| 1 | Milk skim | 1 qt. | 984 | 360 | 36 | 1 | 1 | 0 | 52 | Dairy products |
| 2 | Buttermilk | 1 cup | 240 | 127 | 9 | 5 | 4 | 0 | 13 | Dairy products |
| 3 | Evaporated, undiluted | 1 cup | 252 | 345 | 16 | 20 | 18 | 0 | 24 | Dairy products |
| 4 | Fortified milk | 6 cups | 1,419 | 1,371 | 89 | 42 | 23 | 14 | 119 | Dairy products |

SJIF Impact Factor 2021: 8.013| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 2 | February 2022 - Peer Reviewed Journal

First things first, the t's in the data denote miniscule amounts so we might as well replace them by 0.

```
nutrients=nutrients.replace('t',0)
nutrients=nutrients.replace('t',0)

nutrients.head()
```

| | Food | Measure | Grams | Calories | Protein | Fat | Sat.Fat | Fiber | Carbs | Category |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Cows milk | 1 cup | 972 | 660 | 33 | 40 | 30 | 0 | 48 | Dairy products |
| 1 | Milk skim | 1 cup | 984 | 360 | 36 | 0 | 0 | 0 | 52 | Dairy products |
| 2 | Buttermilk | 1 cup | 245 | 127 | 9 | 5 | 4 | 0 | 13 | Dairy products |
| 3 | Evaporated, undiluted | 1 cup | 252 | 345 | 16 | 20 | 16 | 0 | 24 | Dairy products |
| 4 | Fortified milk | 8 ounce | 1470 | 1373 | 80 | 42 | 23 | 14 | 178 | Dairy products |

Now, we need to remove all the expressions like commas from the dataset so as to convert the numerical data to the respective integer or float variables

```
nutrients=nutrients.replace(',',"", regex=True)
nutrients['Fiber']=nutrients['Fiber'].replace("a","", regex=True)
nutrients['Calories'][91]=(8+44)/2
```

Now, let us convert grams, calories, protein, fat, saturated fat, fiber and carbs datatypes to int.

```
nutrients['Grams']=pd.to_numeric(nutrients['Grams'])
nutrients['Calories']=pd.to_numeric(nutrients['Calories'])
nutrients['Protein']=pd.to_numeric(nutrients['Protein'])
nutrients['Fat']=pd.to_numeric(nutrients['Fat'])
nutrients['Sat.Fat']=pd.to_numeric(nutrients['Sat.Fat'])
nutrients['Fiber']=pd.to_numeric(nutrients['Fiber'])
nutrients['Carbs']=pd.to_numeric(nutrients['Carbs'])
```

```
nutrients.dtypes
```

```
Food          object
Measure       object
Grams          int64
Calories     float64
Protein        int64
Fat            int64
Sat.Fat      float64
Fiber        float64
Carbs        float64
Category      object
dtype: object
```

Nice, all our data is in desired datatypes.

Lets have a quick data quality check

```
print(nutrients.isnull().any())
print('-'*245)
print(nutrients.describe())
print('-'*245)
```

```
Food        False
Measure     False
Grams       False
Calories    False
Protein     False
Fat         False
Sat.Fat     False
Fiber        True
Carbs       False
Category    False
dtype: bool
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
            Grams     Calories      Protein          Fat      Sat.Fat  \
count  355.000000   334.000000  335.000000   335.000000   333.000000
mean   143.211940   188.802395    8.573134     8.540299     6.438438
std    138.668626   184.453018   17.733722    19.797871    18.517656
min     11.000000     0.000000   -1.000000     0.000000     0.000000
25%     60.000000    75.000000    1.000000     0.000000     0.000000
50%    108.000000   131.000000    3.000000     1.000000     0.000000
75%    200.000000   250.000000   12.000000    10.000000     8.000000
max   1419.000000  1373.000000  232.000000   233.000000   234.000000

            Fiber       Carbs
count  334.000000  335.000000
mean     2.376078   24.982388
std     16.078272   35.833106
min      0.000000    0.000000
25%      0.000000    3.000000
50%      0.200000   14.000000
75%      1.000000   38.500000
max    235.000000  236.000000
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
```

There's a null value in the fiber column, lets drop that row entirely.

```
nutrients=nutrients.dropna()
nutrients.shape
```

```
(331, 10)
```

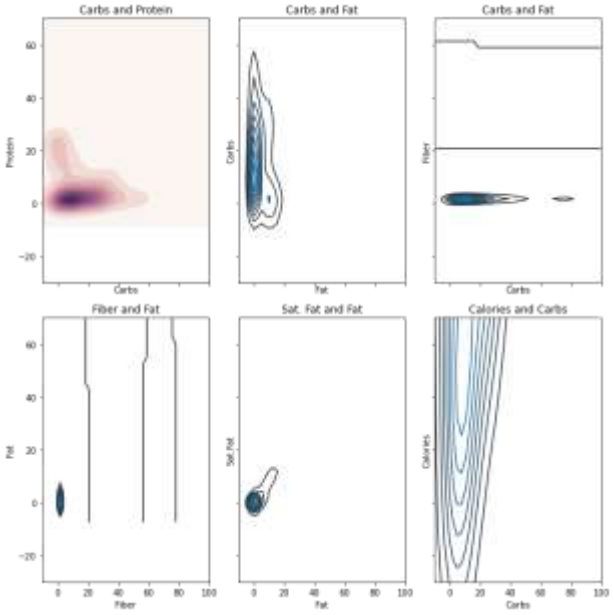
### 2.2 Data Visualization and Analysis

Let's start the analysis by plotting the features with one another. This will not only provide us the distribution of features with one another but also give a quick quantitative feel of the data

```
# Plotting the KDEplots

import matplotlib.pyplot as plt


f, axes = plt.subplots(2, 3, figsize=(10, 10), sharex=True, sharey=True)

s = np.linspace(0, 3, 10)
cmap = sns.cubehelix_palette(start=0.0, light=1, as_cmap=True)

sns.kdeplot(nutrients['Carbs'],nutrients['Protein'],cmap=cmap,shade=True, ax=axes[0,0])
axes[0,0].set(xlim=(-10, 50), ylim=(-30, 70), title = 'Carbs and Protein')

cmap = sns.cubehelix_palette(start=0.25, light=1, as_cmap=True)

sns.kdeplot(nutrients['Fat'],nutrients['Carbs'], ax=axes[0,1])
axes[0,1].set(xlim=(-10, 50), ylim=(-30, 70), title = 'Carbs and Fat')

cmap = sns.cubehelix_palette(start=0.33, light=1, as_cmap=True)

sns.kdeplot(nutrients['Carbs'],nutrients['Fiber'], ax=axes[0,2])
axes[0,2].set(xlim=(-10, 50), ylim=(-30, 70), title = 'Carbs and Fat')

cmap = sns.cubehelix_palette(start=0.45, light=1, as_cmap=True)

sns.kdeplot(nutrients['Fiber'],nutrients['Fat'], ax=axes[1,0])
axes[1,0].set(xlim=(-10, 50), ylim=(-30, 70), title = 'Fiber and Fat')
```

SJIF Impact Factor 2021: 8.013| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 2 | February 2022 - Peer Reviewed Journal

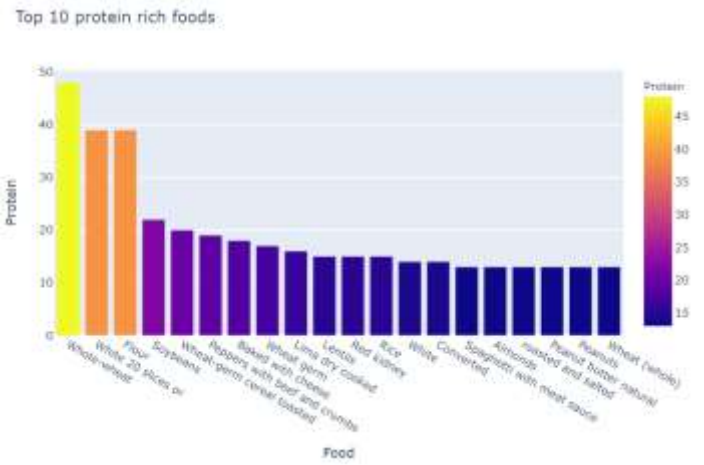
```
cmap = sns.cubehelix_palette(start=0.56, light=1, as_cmap=True)

sns.kdeplot(nutrients['Fat'],nutrients['Sat.Fat'], ax=axes[1,1])
axes[1,1].set(xlim=(-10, 50), ylim=(-30, 70), title = 'Sat. Fat and Fat')

cmap = sns.cubehelix_palette(start=0.68, light=1, as_cmap=True)

sns.kdeplot(nutrients['Carbs'],nutrients['Calories'], ax=axes[1,2])
axes[1,2].set(xlim=(-10, 100), ylim=(-30, 70), title = 'Calories and Carbs')

f.tight_layout()
```




Let's dive into individual metrics

What is the most protein rich food in the category of vegetables and grains?

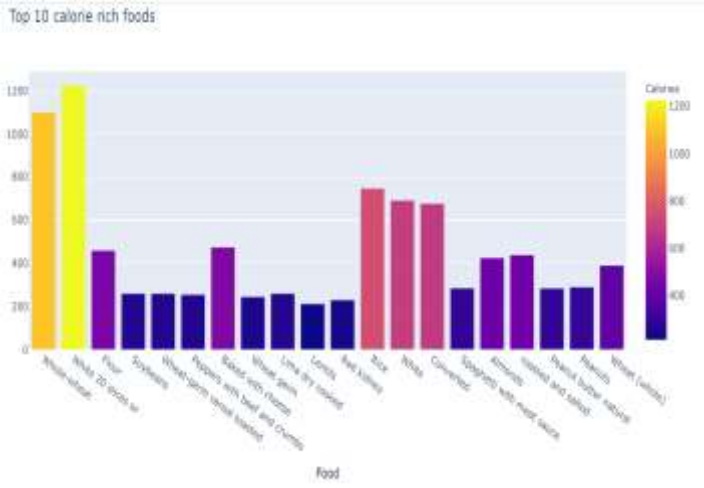
```
alls=['Vegetables A-E',
 'Vegetables F-P',
 'Vegetables R-Z','Breads cereals fastfoodgrains', 'Seeds and Nuts']

prot= nutrients[nutrients['Category'].isin(alls)]

protein_richs prot.sort_values(by='Protein', ascending= False)
top_20=protein_rich.head(20)
fig = px.bar(top_20, x='Food', y='Protein', color='Protein', title=' Top 10 protein rich foods')
fig.show()
```




Therefore, from the category of Grains, Vegetables and Seeds, whole wheat has the most protein content followed by white bread. Soybeans are also in the top 20s. Also, Almonds rank no. 1 in the Seeds category.

Foods to stay away from:
What food has the most calories?

```
cals= nutrients.sort_values(by='Calories', ascending= False)
top_20_cals=cals.head(20)
fig = px.bar(top_20, x='Food', y='Calories' , color='Calories',title=' Top 10 calorie rich foods')
fig.show()
```



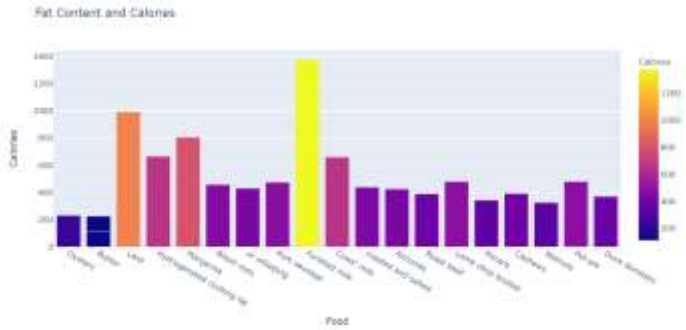

Fortified milk has the most calories, followed by white bread. Also, notice how whole wheat has the most proteins but has almost equal amount of calories. Lard is fat source with most calories and 1/2 cup of ice-creams tops the charts in the dessert category.

SJIF Impact Factor 2021: 8.013| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 2 | February 2022 - Peer Reviewed Journal

### Fat Content

Normally, fat sources are often looked down upon. But, a certain amount of fat is required for a healthy gut. Let's look at some fatty foods.

```
fats= nutrients.sort_values(by='Fat', ascending= False)
top_20_fat=fats.head(20)
fig = px.bar(top_20_fat, x='Food', y='Calories', color='Calories', title= 'Fat Content and Calories')
fig.show()
```




Therefore, Oysters and Butter have the largest combination of calories and fats, followed by lard.

## Analysing categories

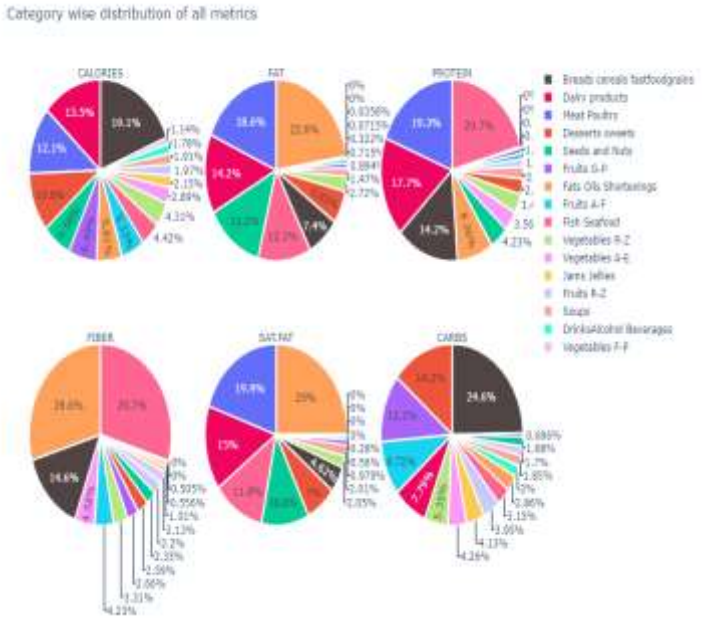
Grouping the data into categories can give us the total count of all metrics and thus we can analyse the categories.

```
category_dist=nutrients.groupby(['Category']).sum()
category_dist
```

| Category | Grams | Calories | Protein | Fat | Sat.Fat | Fiber | Carbs |
|---|---|---|---|---|---|---|---|
| Breads cereals fastfoodgrains | 5253 | 11921.0 | 403 | 207 | 99.0 | 115.91 | 2059.0 |
| Dairy products | 7412 | 8434.0 | 503 | 396 | 322.0 | 4.40 | 651.0 |
| Desserts sweets | 2958 | 6608.0 | 78 | 163 | 150.0 | 20.30 | 1184.0 |
| DrinksAlcohol Beverages | 3284 | 1112.0 | 0 | 0 | 0.0 | 0.00 | 167.0 |
| Fats Oils Shortenings | 695 | 3629.0 | 234 | 631 | 536.0 | 234.00 | 239.0 |
| Fish Seafood | 1807 | 2757.0 | 588 | 338 | 252.0 | 235.00 | 263.0 |
| Fruits A-F | 3844 | 3328.0 | 29 | 20 | 12.0 | 83.50 | 812.0 |
| Fruits G-P | 5412 | 4054.0 | 28 | 25 | 21.0 | 21.10 | 1009.0 |
| Fruits R-Z | 1973 | 1228.0 | 7 | 1 | 0.0 | 17.40 | 330.0 |
| Jams Jellies | 432 | 1345.0 | 0 | 0 | 0.0 | 0.00 | 345.0 |
| Meat Poultry | 2724 | 7529.0 | 546 | 520 | 427.0 | 0.00 | 57.3 |
| Seeds and Nuts | 682 | 4089.0 | 120 | 368 | 232.0 | 18.60 | 140.0 |
| Soups | 2495 | 1191.0 | 59 | 41 | 43.0 | 4.00 | 153.0 |
| Vegetables A-E | 3520 | 1804.0 | 101 | 9 | 6.0 | 26.30 | 356.0 |
| Vegetables F-P | 1725 | 711.0 | 40 | 2 | 0.0 | 16.90 | 142.0 |
| Vegetables R-Z | 3360 | 2684.0 | 98 | 76 | 44.0 | 26.20 | 447.0 |

```
category_dist=nutrients.groupby(['Category']).sum()
from plotly.subplots import make_subplots
import plotly.graph_objects as go

fig = make_subplots(
    rows=2, cols=3,
    specs=[[{'type': 'domain'},{'type': 'domain'},{'type': 'domain'}],[{'type': 'domain'},
{'type': 'domain'},{'type': 'domain'}]])

fig.add_trace(go.Pie(values=category_dist['Calories'].values, title='CALORIES', labels=categ
ory_dist.index,marker=dict(colors=['#100b','#f00560'], line=dict(color='#FFFFFF', width=2.
5)),
              row=1, col=1)

fig.add_trace(go.Pie(values=category_dist['Fat'].values,title='FAT', labels=category_dist.in
dex,marker=dict(colors=['#100b','#f00560'], line=dict(color='#FFFFFF', width=2.5)),
              row=1, col=2)

fig.add_trace(go.Pie(values=category_dist['Protein'].values,title='PROTEIN', labels=category
_dist.index,marker=dict(colors=['#100b','#f00560'], line=dict(color='#FFFFFF', width=2.5)),
              row=1, col=3)

fig.add_trace(go.Pie(values=category_dist['Fiber'].values,title='FIBER', labels=category_dis
t.index,marker=dict(colors=['#100b','#f00560'], line=dict(color='#FFFFFF', width=2.5)),
              row=2, col=1)

fig.add_trace(go.Pie(values=category_dist['Sat.Fat'].values,title='SAT.FAT', labels=category
_dist.index,marker=dict(colors=['#100b','#f00560'], line=dict(color='#FFFFFF', width=2.5)),
              row=2, col=2)

fig.add_trace(go.Pie(values=category_dist['Carbs'].values,title='CARBS', labels=category_dis
t.index,marker=dict(colors=['#100b','#f00560'], line=dict(color='#FFFFFF', width=2.5)),
              row=2, col=3)
fig.update_layout(title_text='Category wise distribution of all metrics',height=700, width=1
000)

fig.show()
```

## 3- RESULT



SJIF Impact Factor 2021: 8.013| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 2 | February 2022 - Peer Reviewed Journal

## 4 - CONCLUSION

Some inferences from the above pie charts :-

- It is clear that breads, grains and cereals have the highest amount of Carbs and Calories.
- Largest percentage of protein is in seafood (God bless the vegetarians!)
- Surprisingly, same amount of fiber content is present in Fats and Seafood.
- Seeds and nuts have about 14% fat content.
- Fruits do not have a large percentage in any of the categories except carbs, they have about 10% carbohydrates.
- Dairy products (15%) have more saturated fat content than seafood (11.8%).
- We can expand this project and analyse the other aspects our diet as well using similar methods.

## 4 - ACKNOWLEDGEMENT

We would like to thank our teachers , family and friends.

.