

# School of Computer Science and Engineering J Component report

**Programme : B.Tech**

# Course Title : Data Visualization

# Course Code : CSE3020

**Slot : D2**

# Title: Nutritional value of fast food and VIT Hostel menu

**Team Members:**

**Adarsh Prasad – 20BCE1519**

**Souradyuti Choudhury – 20BCE1115**

# Faculty: Dr. Pattabirman V Sign: Date:

**Table of contents**

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Title** | **Page number** |
| 1. | Abstract | 3 |
| 2. | Literature review | 4 |
| 3. | Methodology | 5 |
| 4. | Code | 7 |
| 5. | Result | 32 |
| 6. | Conclusion | 34 |
| 7. | References | 35 |

# Abstract

McDonald’s is one of the most renowned food industry giants with a total of 36,000 all around the world and servicing more than 70 million customers daily in a total of 100 countries. One of the main reasons for the success in the growth of the McDonald’s chain is the change in the menu based on the location to please the customers that reside there. Being a student and part of a hostel puts more light on importance of nutrition. A student encounters many tasks such as attending classes, studying, carrying out demanding physical activities. We have considered the hostel menu and calculated the nutritional value of each item. Further in consideration, we have considered to carrying out an analysis of the McDonald’s global and VIT Chennai Hostel Menu. The entire data analysis is carried out with the help of RStudio and Python where libraries like Pandas, NumPy, Seaborn, Matplotlib, etc. have been used. This study can prove to be helpful because the results give a good understating of how the consumption of these food items affects our health.

# LITERATURE REVIEW

McDonald's is a fast-food restaurant chain that was established in the United States in 1940 in San Bernardino, California, by Richard and Maurice McDonald. McDonald's now has 36,900 locations and serves over 69 million individuals each day in more than 100 countries, making it the largest restaurant franchise in the world by sales. In terms of Most Valuable Global Brands in 2014, McDonald's came in fifth. The difficulties McDonald's faces include price sensitivity, competitive pressures, health hazards, and many more, according to internet research. We will concentrate on McDonald's biggest problem, which is the dangers consuming McDonald's poses to your health. Today, the term "fast food" nearly has a bad connotation. McDonald's meals typically include high salt, harmful fat, sugar, and empty carbohydrate content. Actually, the processed fat in McDonald's cuisine encourages endothelial damage, which has a future connection to erectile dysfunction. Numerous studies have suggested that consuming McDonald's meals has significant health hazards, including heart attack, diabetes, and high blood pressure.[1]

The nutritional analysis focuses on assessing the macro and micronutrient content of existing hostel menus, identifying potential gaps or excesses in nutritional provision. Concurrently, surveys are conducted to understand resident dietary preferences, restrictions, and overall satisfaction with the current culinary offerings. The sensory evaluation component employs a panel of trained assessors to objectively measure the taste, aroma, and visual appeal of popular hostel dishes. Findings from this research aim to inform strategic interventions for menu enhancement, incorporating evidence-based nutritional adjustments and culinary innovations. The ultimate goal is to create a balanced and diverse hostel menu that caters to the nutritional needs and taste preferences of a diverse resident population, promoting overall well-being and satisfaction within the hostel community. The insights gained from this study contribute not only to the field of hostel management but also hold broader implications for institutional food services striving to provide optimal nutrition and culinary delight in communal living environments.

# METHODOLOGY

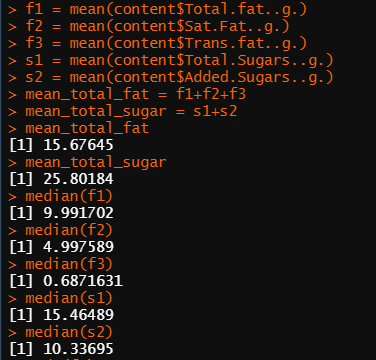
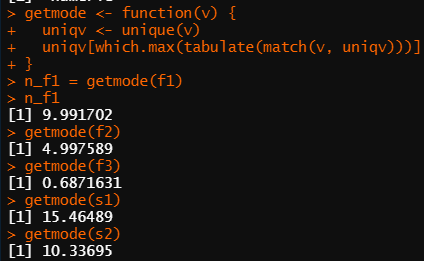
**Cleaning the dataset**

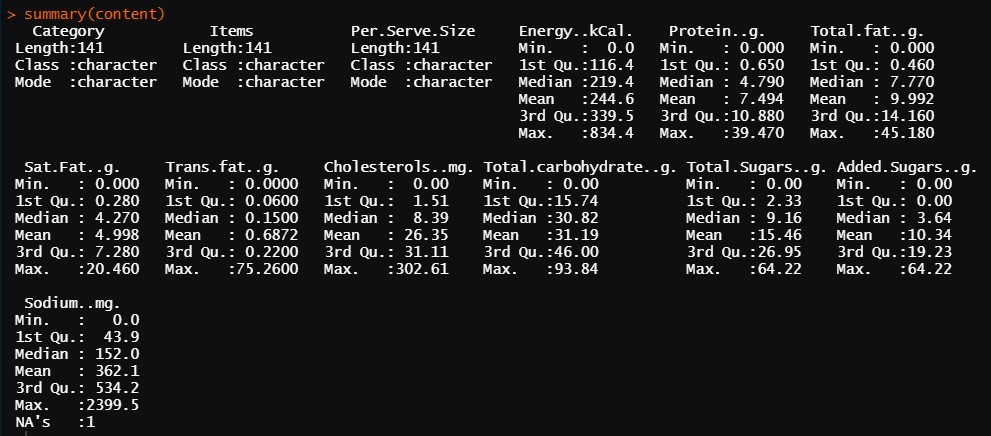
The practice of correcting or deleting inaccurate, damaged, improperly formatted, duplicate, or missing data from a dataset is known as data cleaning. There are several ways for data to be duplicated or incorrectly categorized when merging different data sources. Even though results and algorithms appear to be right, they are unreliable if the data is inaccurate. Because the procedures will differ from dataset to dataset, there is no one definitive way to specify the precise phases in the data cleaning process. The following are the steps followed for data cleaning.[4]

* Step 1: Removing duplicates/nan values from the dataset. – For example, pandas built-in function drop\_duplicates() helps achieve a dataset without duplicates and dropna() helps achieve a dataset without nan values, or the na.omit() in R can be used for removing the nan values.
* Step 2: Filtering the unwanted outliers – removing the rows from the dataset that just don’t fit in well. Sometimes a hypothesis you are working on will be proven by the presence of an outlier. Never forget that an outlier doesn't always indicate that something is wrong. To determine the reliability of the figure, this step is necessary. If an outlier turns out to be incorrect or unimportant for the analysis, it is necessary to remove it.
* Step 3: Handling missing data – Missing data can be worked with in multiple ways, the first case is just removing records that have missing values. The second case being, fill the missing values with the average of the entire attribute. The last case is deriving a method such that it can navigate through the dataset avoiding the missing values
* Step 4: Partitioning the dataset – splitting the dataset into testing, training, and validation. This is an optional step based on the application.

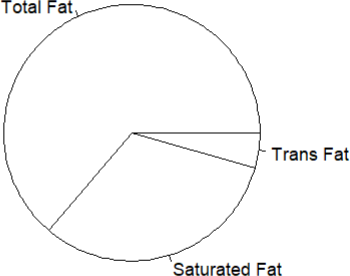
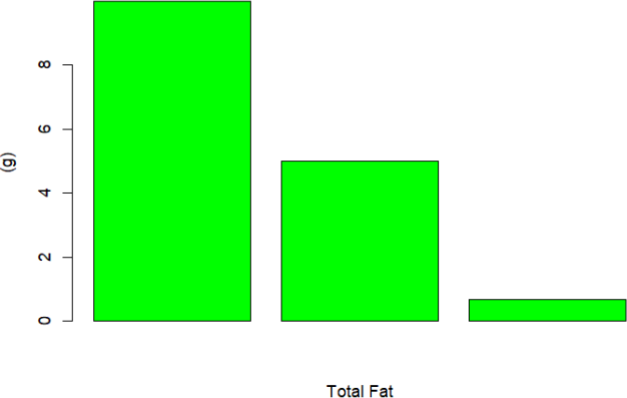
# Basic Statistical analysis

The basic statistical analysis was done on the dataset using RStudios. The analysis included finding the mean, median, and mode of the data in each category of a nutritional category. The results are shown below:



The above figures show a general statistical analysis of the dataset in question.



The above figures show the comparison of the different types of fats.

# CODE

RStudio code for basic statistical analysis

path = "C:/Users/adars/Downloads/India\_Menu.csv" content = read.csv(path)

print(content)

f1 = mean(content$Total.fat..g.) f2 = mean(content$Sat.Fat..g.) f3 = mean(content$Trans.fat..g.)

s1 = mean(content$Total.Sugars..g.) s2 = mean(content$Added.Sugars..g.) mean\_total\_fat = f1+f2+f3 mean\_total\_sugar = s1+s2 mean\_total\_fat

mean\_total\_sugar median(f1) median(f2) median(f3) median(s1) median(s2) mode(f1)

getmode <- function(v) { uniqv <- unique(v)

uniqv[which.max(tabulate(match(v, uniqv)))]

}

n\_f1 = getmode(f1) n\_f1

getmode(f2) getmode(f3) getmode(s1) getmode(s2) summary(content) sd(f1)

var(f1) var(f2)

var(content$Total.fat..g.) sd(content$Total.fat..g.) sd(content$Sat.Fat..g.) sd(content$Trans.fat..g.) sd(content$Total.Sugars..g.) sd(content$Added.Sugars..g.) fat\_graph=c(f1,f2,f3)

fat\_pie\_names = c("Total Fat","Saturated Fat", "Trans Fat") barplot(fat\_graph,xlab="Total Fat",ylab = "(g)",col = "green") pie(fat\_graph,labels=fat\_pie\_names,col = "white")

hist(fat\_graph,xlab = "Fats",ylab = "(g)",col = "yellow") barplot(fat\_graph,content$Per.Serve.Size,xlab = "Per Serving", ylab = "Fat Content (g)",col = "black")

size\_mean = mean(content$Per.Serve.Size) plot(fat\_graph,content$Per.Serve.Size) content$Per.Serve.Size plot(mean\_total\_fat,mean\_total\_sugar) class(getmode)

is.numeric()

Python code for descriptive analysis as well as predictive analysis

# Import Basic Libraries import pandas as pd import numpy as np import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

menu = pd.read\_csv('India\_Menu.csv') menu.head()

menu.shape menu.info() menu.isnull().sum()

#Gives the entire summary of the dataset menu.describe()

menu.Category.unique() menu.Category.value\_counts() sns.countplot(x='Category', data=menu) plt.xticks(rotation=55)

# Make a lis of all nutritions which u want to visualize nutritions = ['Energy (kCal)', 'Protein (g)', 'Total fat (g)', 'Sat Fat (g)',

'Trans fat (g)','Cholesterols (mg)', 'Total carbohydrate (g)',

'Total Sugars (g)', 'Added Sugars (g)', 'Sodium

(mg)']

plt.figure(figsize = (25,55))

for i, ntr in enumerate(nutritions): plt.subplot(7, 2, i+1) sns.barplot(x='Category', y=ntr, data=menu)

plt.xticks(rotation=55) plt.show()

# Texual Description about above visualization for ntr in nutritions:

temp\_list = menu.pivot\_table(ntr, 'Category').sort\_values(by=ntr, ascending=False)

print('{}'.format(temp\_list[:3]))

print('

--')

sns.barplot(x='Category', y='Energy (kCal)', data=menu) plt.xticks(rotation=55)

# A function which can plot N Items which contains highest nutrition

# Change value of N as per convenience in visualization

def plot\_item\_vs\_nutrition(nutrition, n): item =

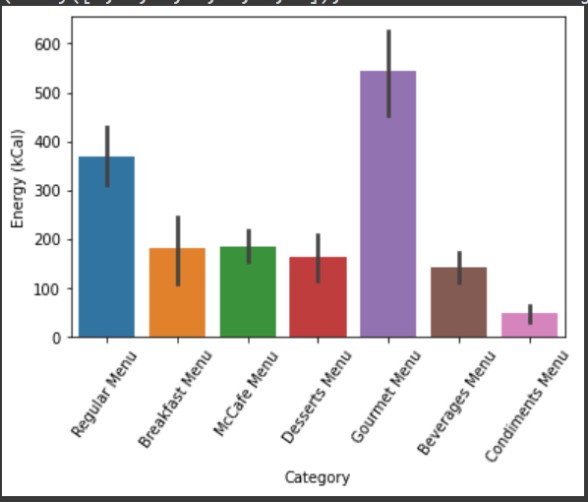
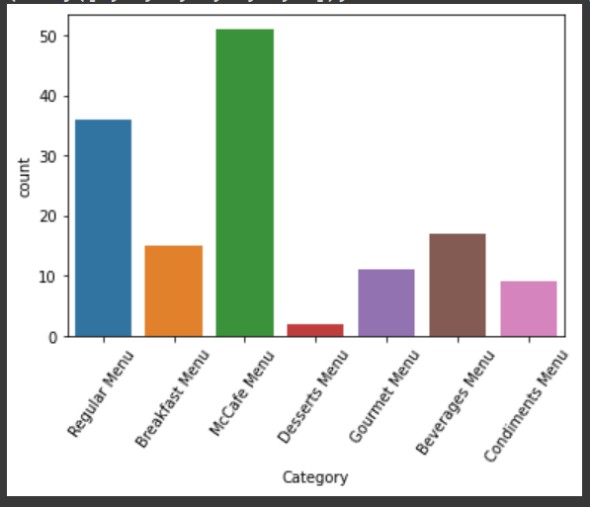
menu.groupby(menu[nutrition])['Items'].sum().sort\_index()[-n:] plt.figure(figsize=(20,20))

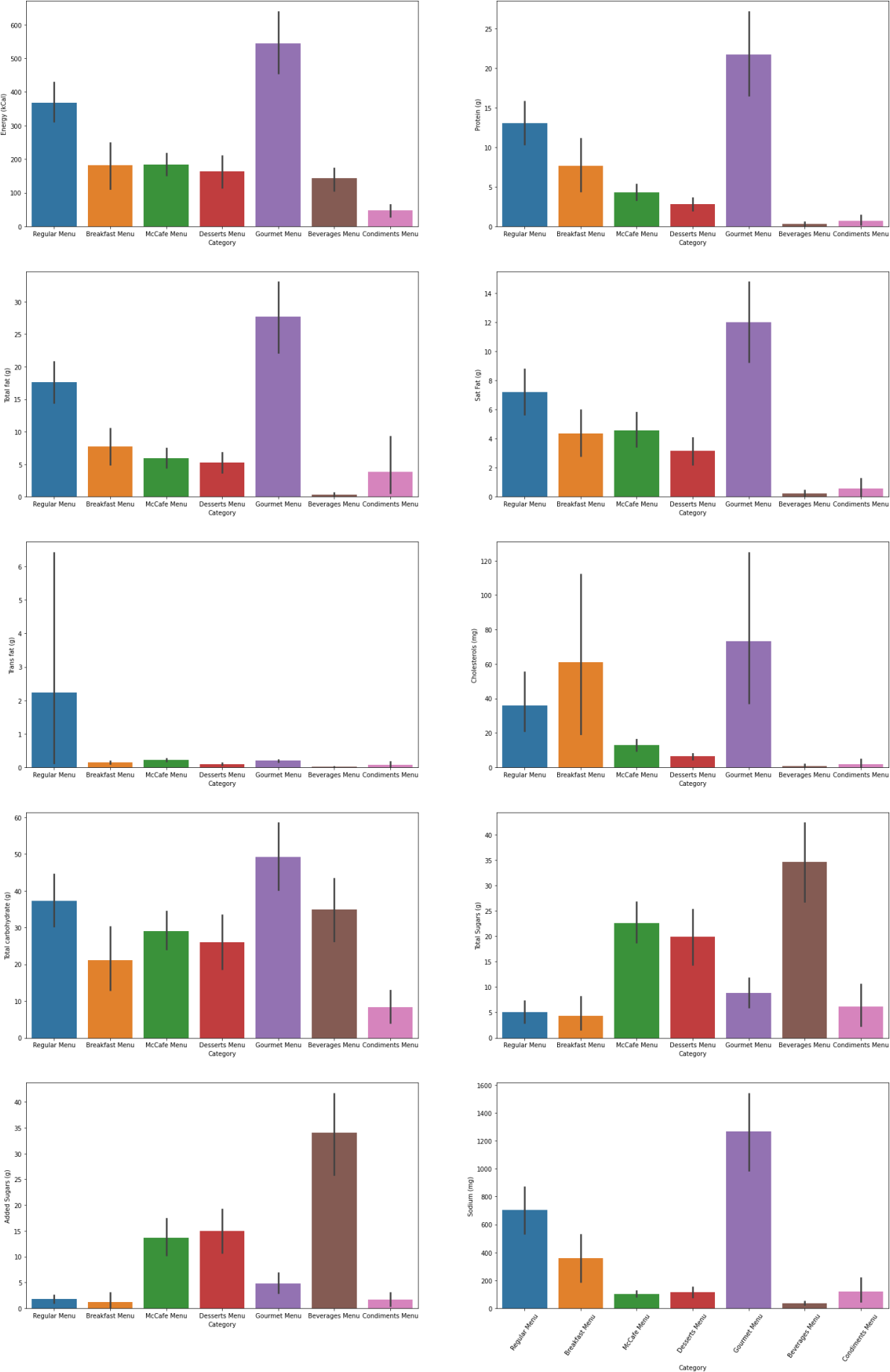
sns.barplot(item.index, item.values) # Let's Start with Energy

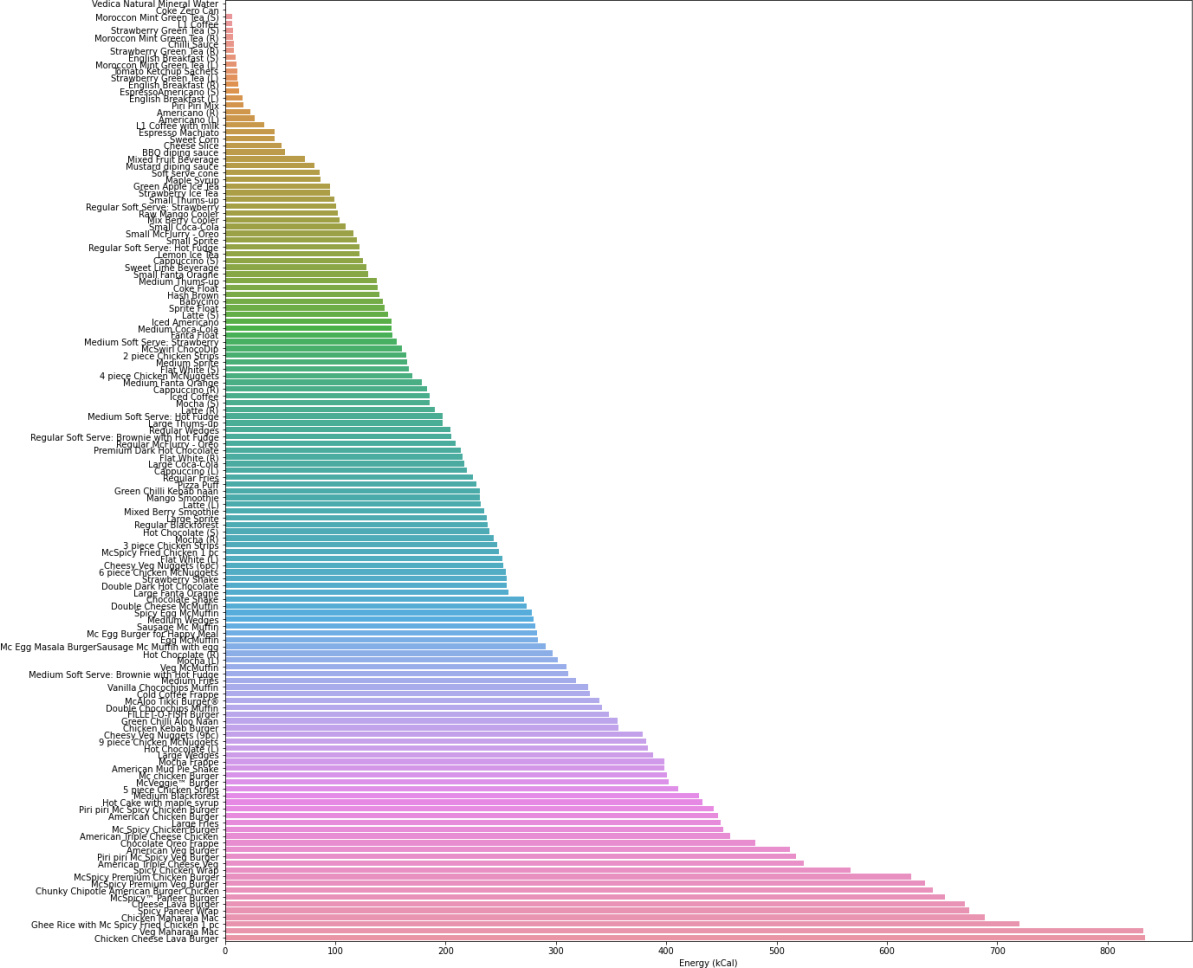
plot\_item\_vs\_nutrition('Energy (kCal)', 5)

plot\_item\_vs\_nutrition('Protein (g)', 5) plot\_item\_vs\_nutrition('Total fat (g)', 5) plot\_item\_vs\_nutrition('Sat Fat (g)', 5) plot\_item\_vs\_nutrition('Trans fat (g)', 5) plot\_item\_vs\_nutrition('Cholesterols (mg)', 5) plot\_item\_vs\_nutrition('Total carbohydrate (g)', 5) plot\_item\_vs\_nutrition('Total Sugars (g)', 5) plot\_item\_vs\_nutrition('Added Sugars (g)', 5) plot\_item\_vs\_nutrition('Sodium (mg)', 5)

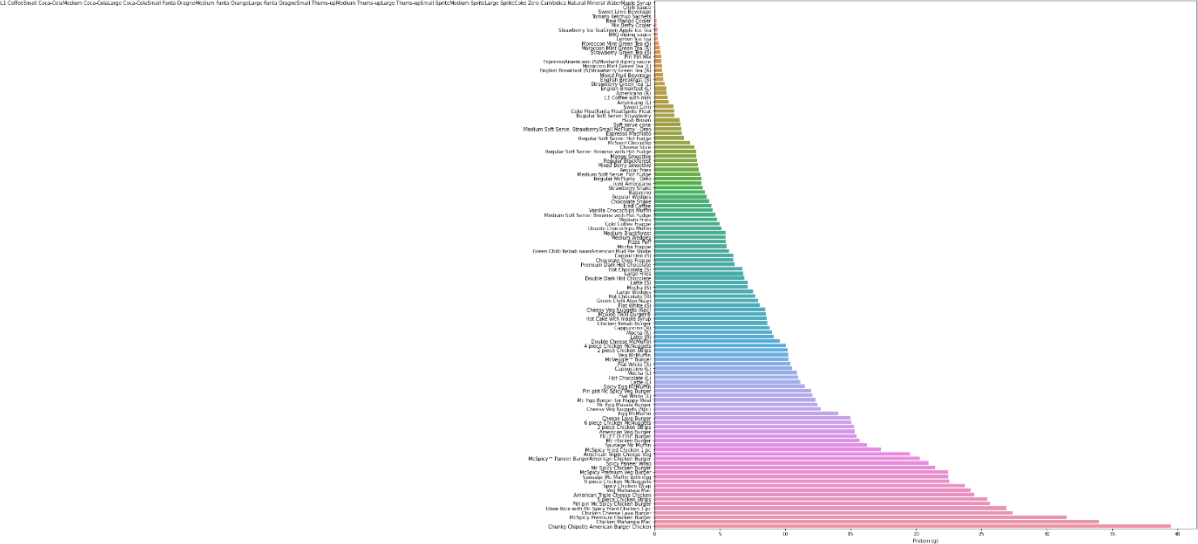
Graphs and outputs



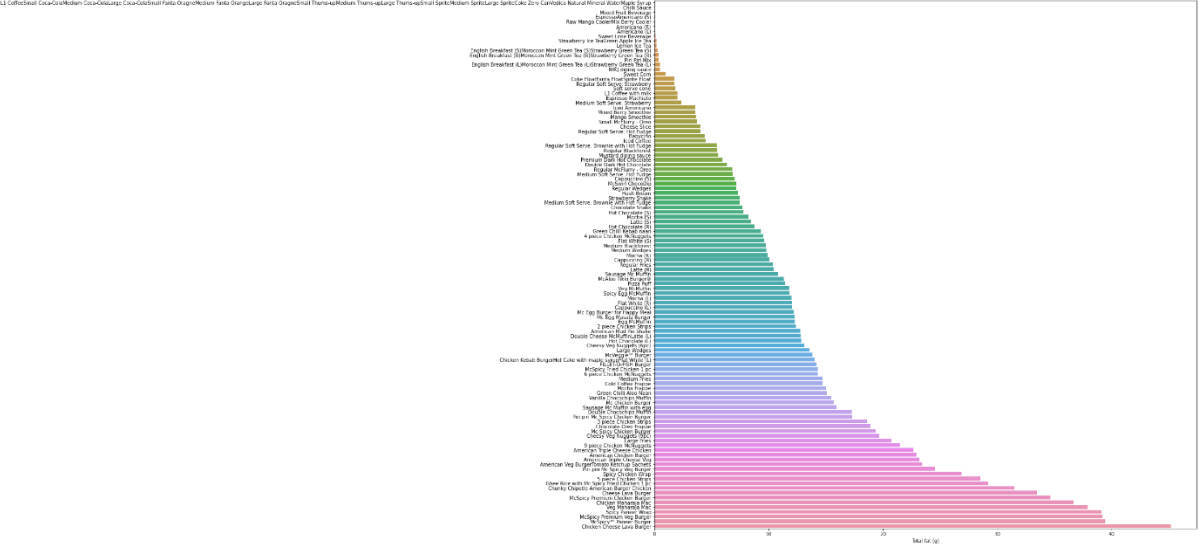




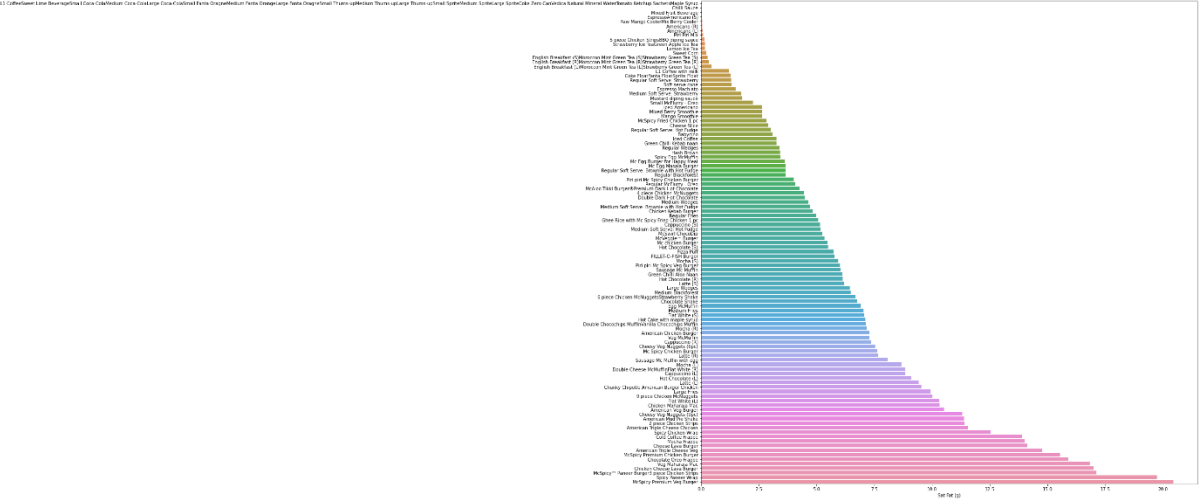
Energy content in each item on the menu from the least to the highest.



Protein content in each item on the menu from the least to the highest.



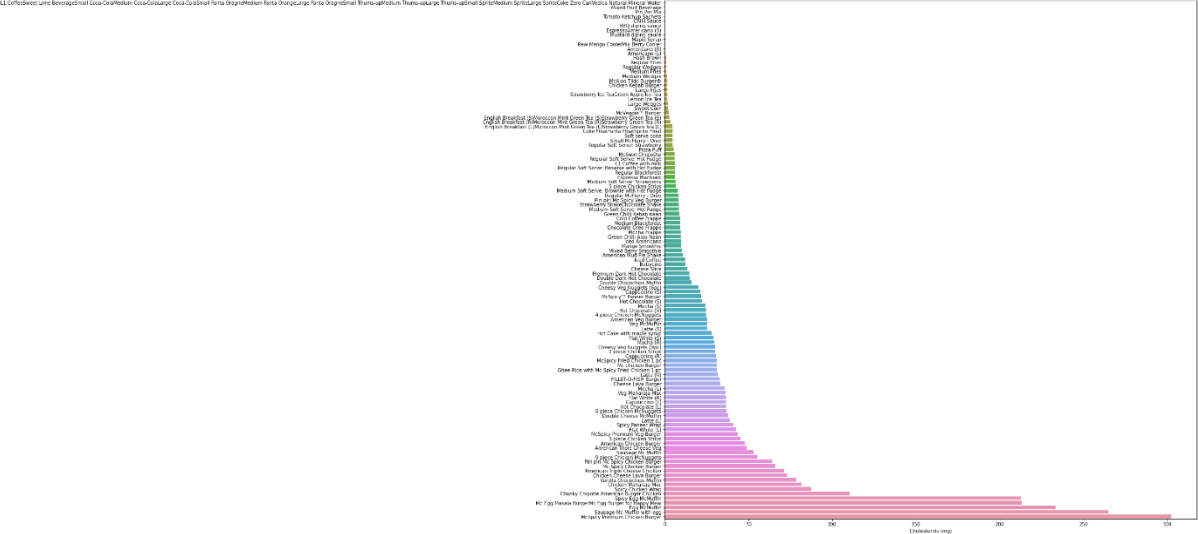
Total fat content in each item on the menu from the least to the highest.



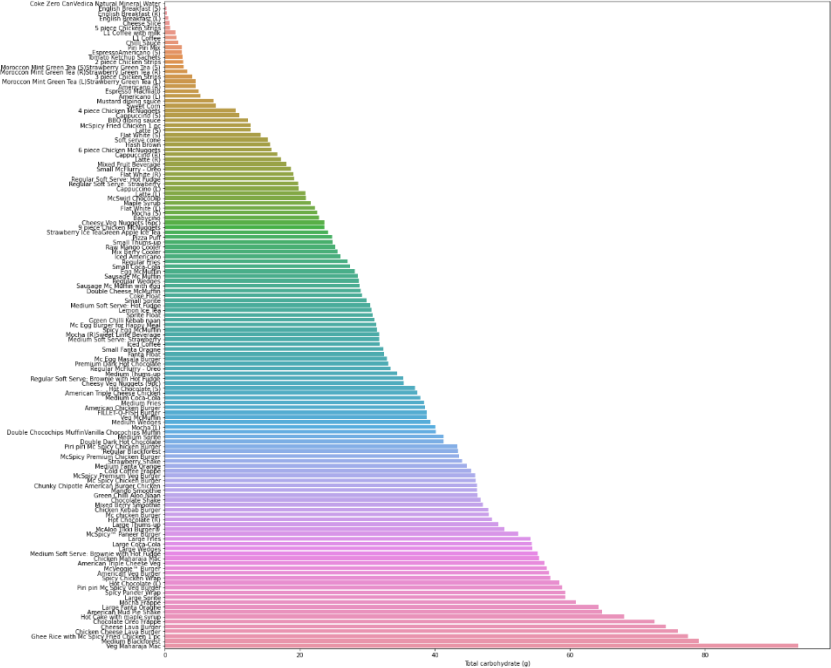
Sat fat content in each item on the menu from the least to the highest.



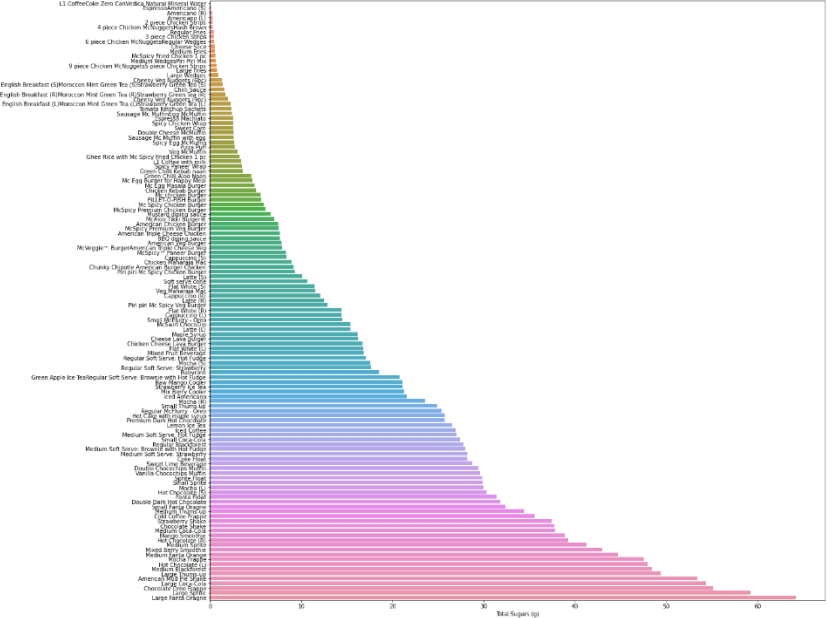
Trans-fat content in each item on the menu from the least to the highest



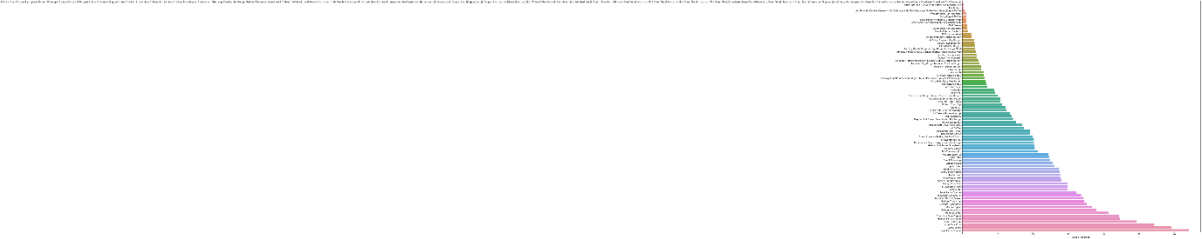
Cholesterol content in each item on the menu from the least to the highest



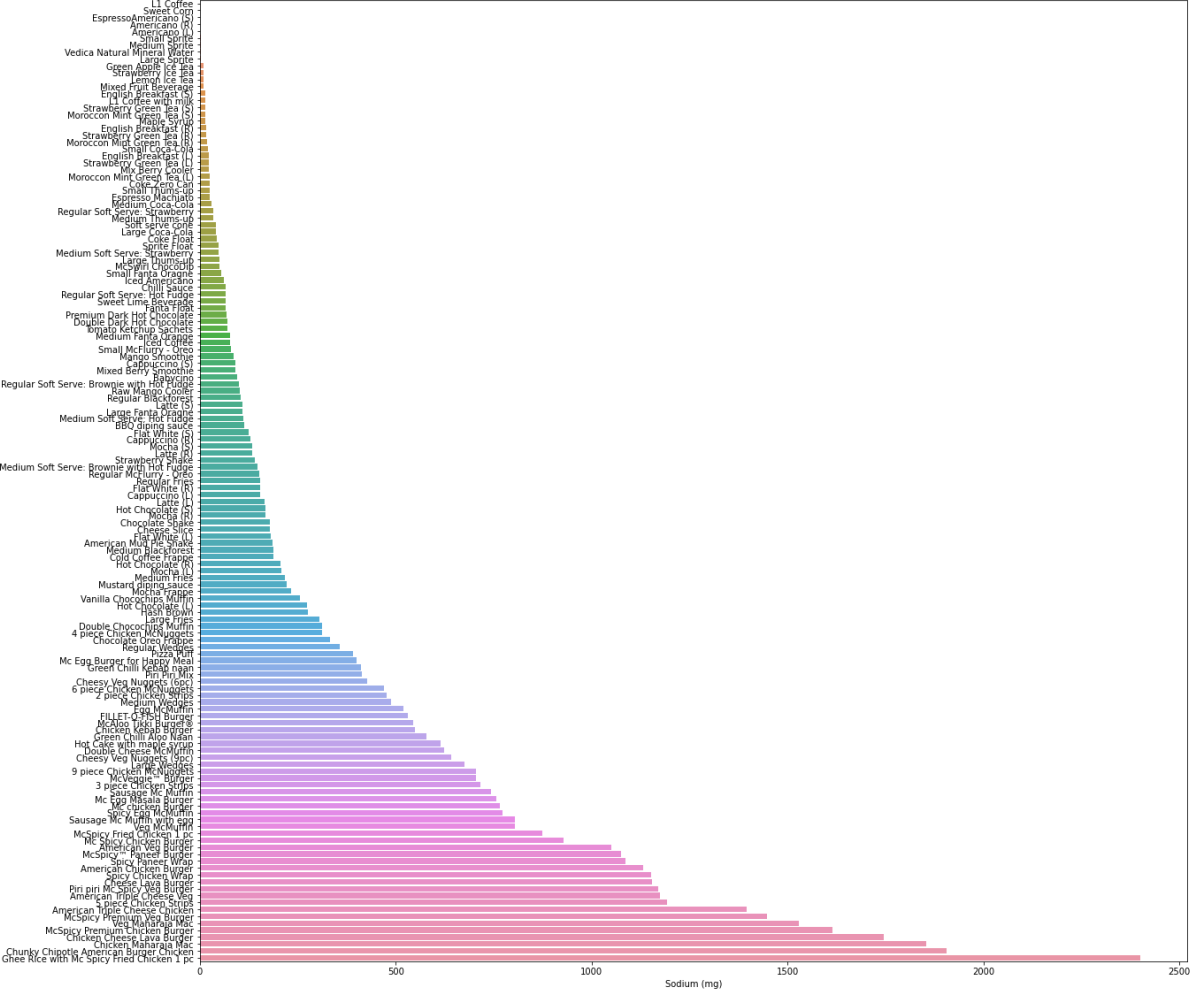
Total carbohydrate content in each item on the menu from the least to the highest



Total sugars content in each item on the menu from the least to the highest



Added sugars content in each item on the menu from the least to the highest



Sodium content in each item on the menu from the least to the highest

For HOSTEL

Veg Menu

# -\*- coding: utf-8 -\*-

"""VegMenu.ipynb

Automatically generated by Colaboratory.

Original file is located at

https://colab.research.google.com/drive/1MbbrSkjlqgOempYYCrLBeEc7LYXQe7yD

#Importing the libraries

"""

# Commented out IPython magic to ensure Python compatibility.

# Import Basic Libraries

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

# %matplotlib inline

from google.colab import drive

drive.mount('/content/drive')

file\_veg="veg\_menu\_nov.csv"

file\_path\_veg = f"/content/drive/My Drive/dv\_dataset/{file\_veg}"

file\_nonveg="nonveg\_menu\_nov.csv"

file\_path\_nonveg = f"/content/drive/My Drive/dv\_dataset/{file\_nonveg}"

"""#Loading the dataset for VEG MENU"""

menu = pd.read\_csv(file\_path\_veg)

menu.head()

"""#Shape of the dataset"""

menu.shape

menu.info()

menu.isnull().sum()

"""Analysis:

1. The dataset has 67 records (menu items)

2. 5 out of 7 attributes have the datatype float64

3. 2 out of 7 attributes have the object datatype

4. Total memory usage of the dataset = 3.8+ KB

#Effective Data Analysis

"""

#Gives the entire summary of the dataset

menu.describe()

"""#Unique values in Category Column"""

unique\_menu\_categories = menu['Category'].unique()

print(unique\_menu\_categories)

category\_counts = menu['Category'].value\_counts()

print(category\_counts)

sns.countplot(x='Category', data=menu)

plt.xticks(rotation=55)

"""Analysis:

1. The different types of menus are - breakfast(23 items), lunch(27 items),dinner (18 items)

#Visualize Average Nutritions Value per Category

"""

# Make a lis of all nutritions which u want to visualize

nutritions = ['protein(g)', 'carbohydrates(g)',

'sodium(mg)', 'iron(mg)', 'calcium(mg)']

plt.figure(figsize = (25,55))

for i, ntr in enumerate(nutritions):

plt.subplot(7, 2, i+1)

sns.barplot(x='Category', y=ntr, data=menu)

plt.xticks(rotation=55)

plt.show()

# Texual Description about above visualization

for ntr in nutritions:

temp\_list = menu.pivot\_table(ntr, 'Category').sort\_values(by=ntr, ascending=False)

print('{}'.format(temp\_list[:7]))

print('-----------------------------------------------------')

"""Analysis:

1. The above code cell prints top 3 categorical source for each nutrition listed

2. Also by onserving graphs we can get rich source of each nutrition

#Plot Average Carbohydrates each Category contains

"""

sns.barplot(x='Category', y='carbohydrates(g)', data=menu)

plt.xticks(rotation=55)

"""Analysis:

1. The above graph shows the carbohydrates that each type of menu has on an average.

\* breakfast - around 25g

\* lunch - around 24g

\* dinner - around 23g

#Plot which items contains which nutritions

"""

# A function which can plot N Items which contains highest nutrition

# Change value of N as per convenience in visualization

def plot\_item\_vs\_nutrition(menu, nutrition, n):

top\_n\_items = menu.groupby('item')[nutrition].sum().nlargest(n)

# Print maximum and minimum values along with corresponding item names

max\_item = top\_n\_items.idxmax()

min\_item = top\_n\_items.idxmin()

max\_value = top\_n\_items.max()

min\_value = top\_n\_items.min()

print(f"Item with Maximum {nutrition}: {max\_item} ({max\_value})")

print(f"Item with Minimum {nutrition}: {min\_item} ({min\_value})")

plt.figure(figsize=(12, 8))

sns.barplot(x=top\_n\_items.values, y=top\_n\_items.index, orient='h')

plt.xlabel(nutrition)

plt.ylabel('Menu Item')

plt.title(f'Top {n} Items based on {nutrition}')

plt.show()

# Let's Start with protein

plot\_item\_vs\_nutrition(menu,'protein(g)', len(menu) - 1)

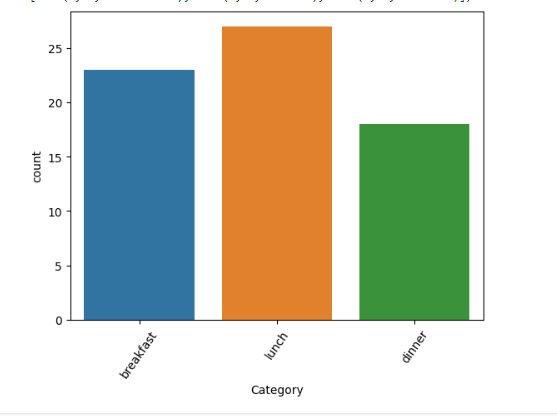
plot\_item\_vs\_nutrition(menu,'carbohydrates(g)', len(menu) - 1)

plot\_item\_vs\_nutrition(menu,'sodium(mg)', len(menu) - 1)

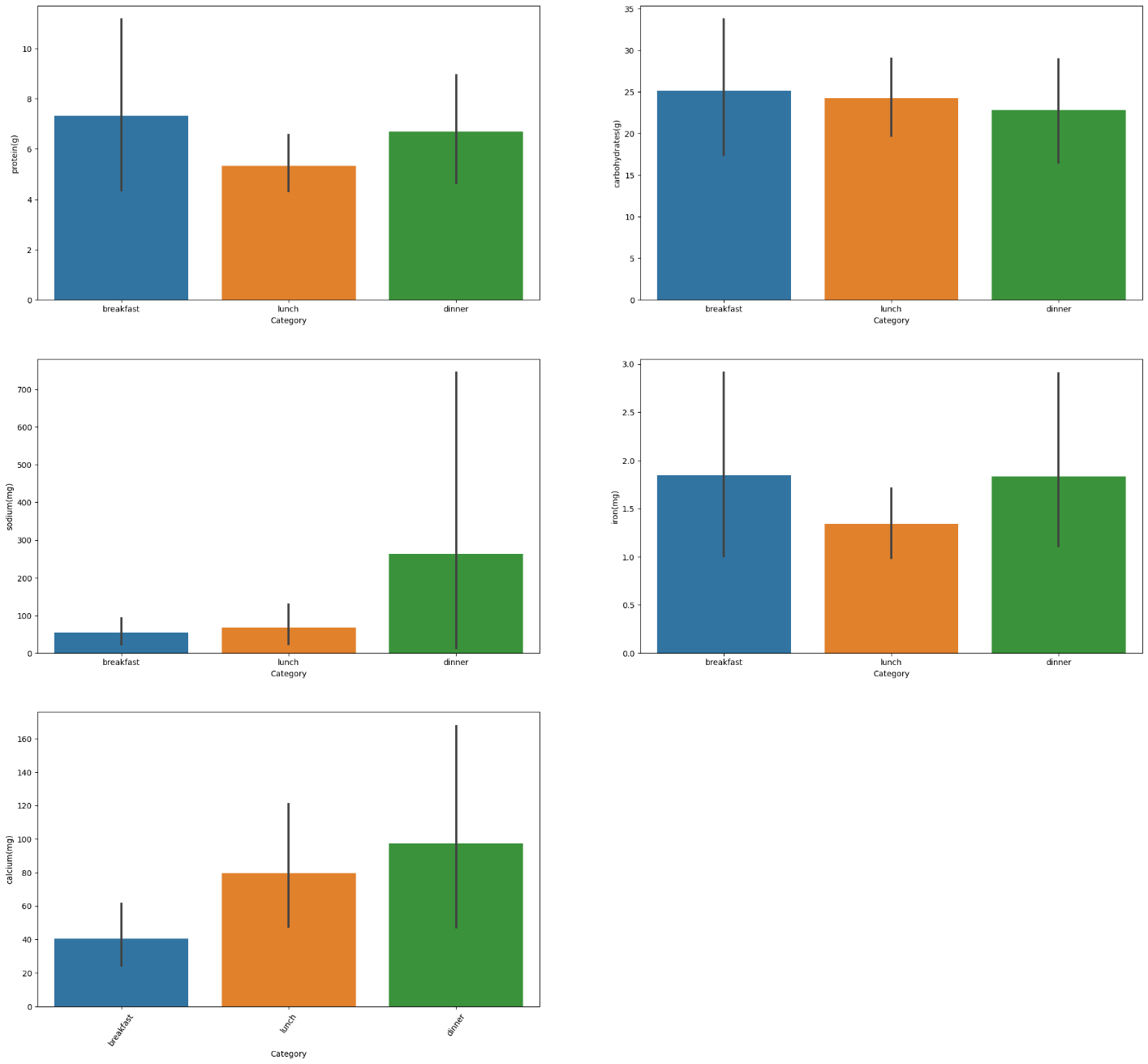
plot\_item\_vs\_nutrition(menu,'iron(mg)', len(menu) - 1)

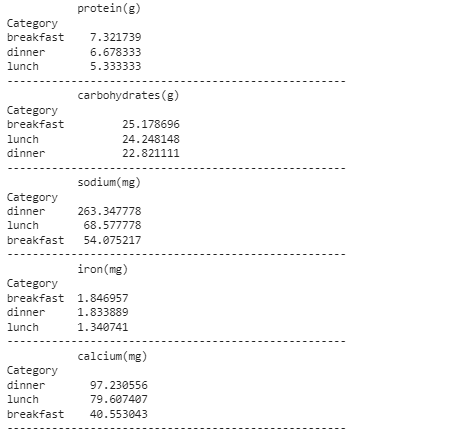
plot\_item\_vs\_nutrition(menu,'calcium(mg)', len(menu) - 1)

Count of menu category wise

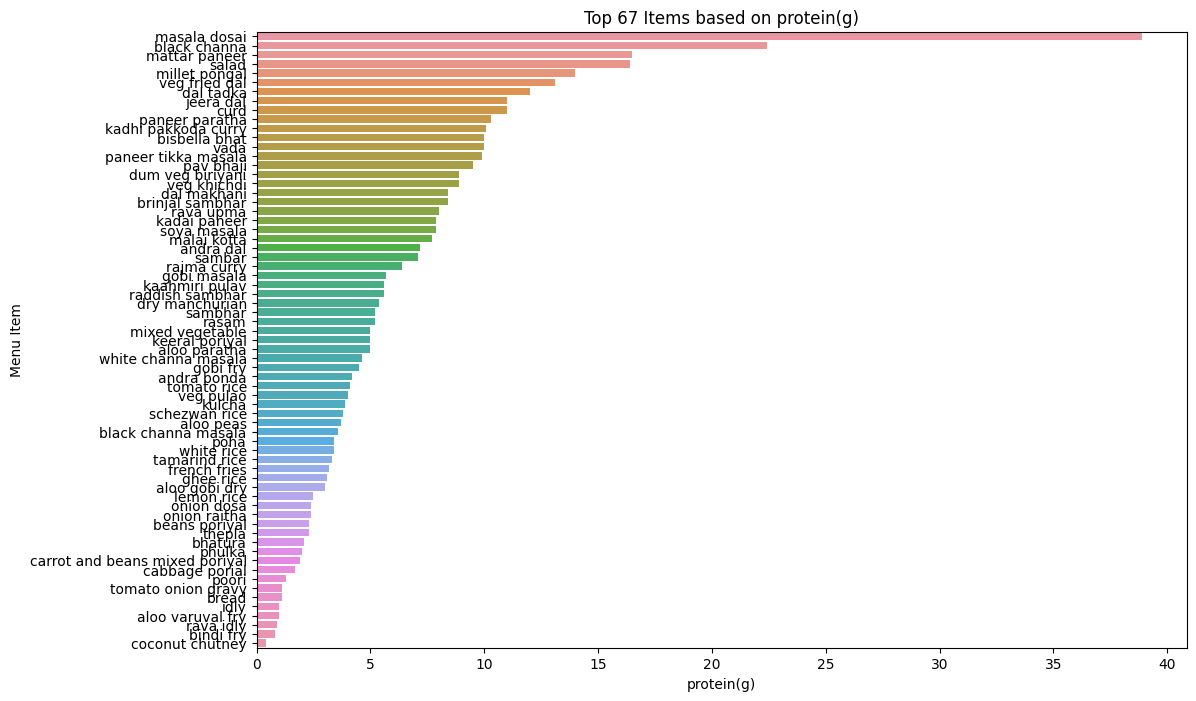


Avg nutrition per category

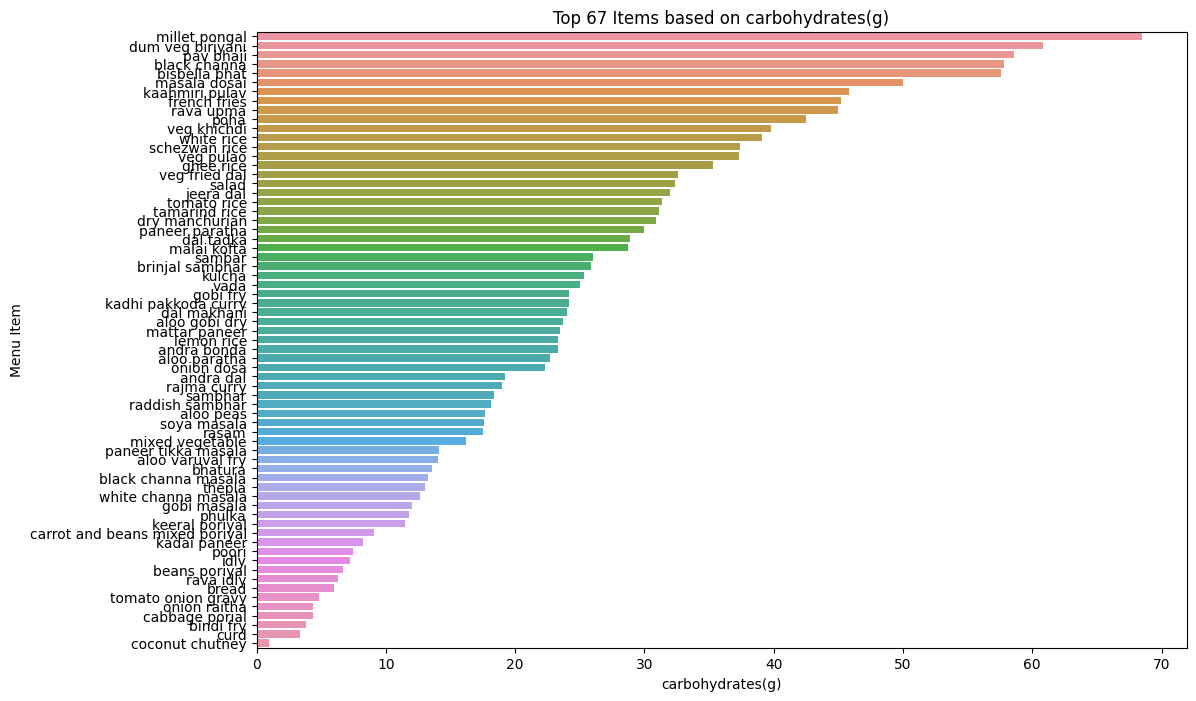




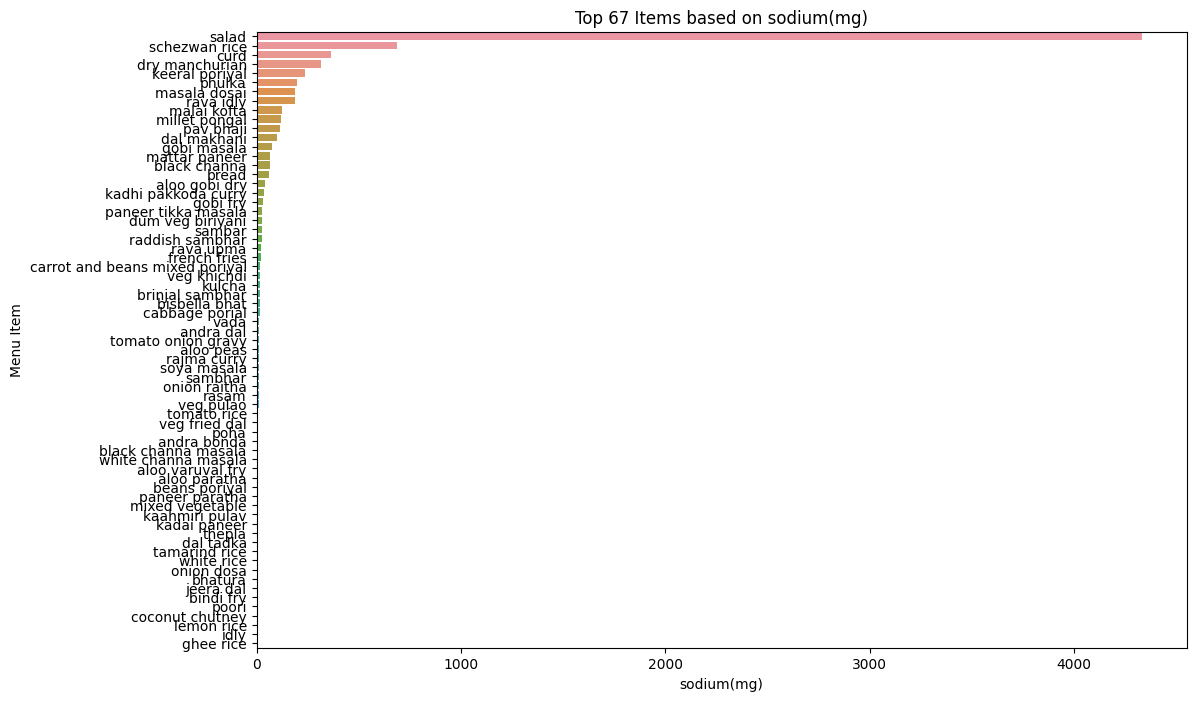
Protein in each Item



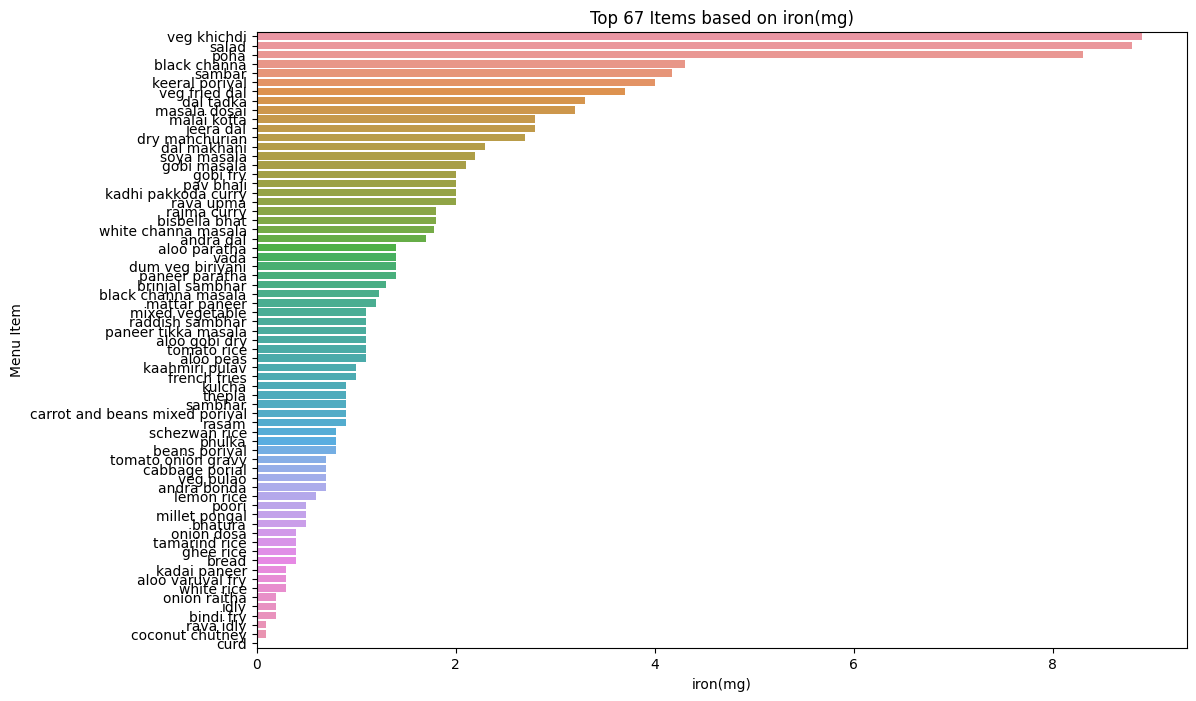
Carbohydrate in each item



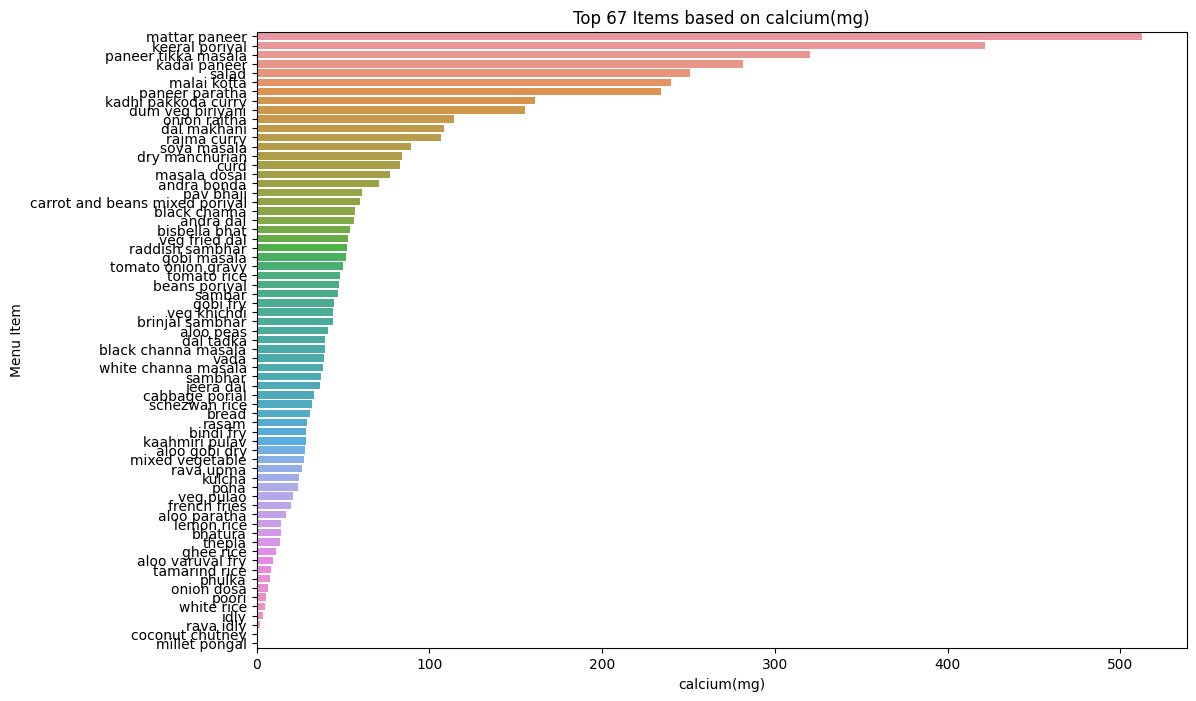
Sodium in each item



Iron in each item



Calcium in each item



For Non Veg

# -\*- coding: utf-8 -\*-

"""NonvegMenu.ipynb

Automatically generated by Colaboratory.

Original file is located at

https://colab.research.google.com/drive/179PHzFlFuJhSJin3zMVk6GdcrYnsx1wG

#Importing the libraries

"""

# Commented out IPython magic to ensure Python compatibility.

# Import Basic Libraries

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

# %matplotlib inline

from google.colab import drive

drive.mount('/content/drive')

file\_veg="veg\_menu\_nov.csv"

file\_path\_veg = f"/content/drive/My Drive/dv\_dataset/{file\_veg}"

file\_nonveg="nonveg\_menu\_nov.csv"

file\_path\_nonveg = f"/content/drive/My Drive/dv\_dataset/{file\_nonveg}"

"""#Loading the dataset for VEG MENU"""

menu = pd.read\_csv(file\_path\_nonveg)

menu.head()

"""#Shape of the dataset"""

menu.shape

menu.info()

menu.isnull().sum()

"""Analysis:

1. The dataset has 37 records (menu items)

2. 5 out of 7 attributes have the datatype float64

3. 2 out of 7 attributes have the object datatype

4. Total memory usage of the dataset = 2.1+ KB

#Effective Data Analysis

"""

#Gives the entire summary of the dataset

menu.describe()

"""#Unique values in Category Column"""

unique\_menu\_categories = menu['Category'].unique()

print(unique\_menu\_categories)

category\_counts = menu['Category'].value\_counts()

print(category\_counts)

sns.countplot(x='Category', data=menu)

plt.xticks(rotation=55)

"""Analysis:

1. The different types of menus are - breakfast(12 items), lunch(13 items),dinner (12 items)

#Visualize Average Nutritions Value per Category

"""

# Make a lis of all nutritions which u want to visualize

nutritions = ['protein(g)', 'carbohydrates(g)',

'sodium(mg)', 'iron(mg)', 'calcium(mg)']

plt.figure(figsize = (25,55))

for i, ntr in enumerate(nutritions):

plt.subplot(7, 2, i+1)

sns.barplot(x='Category', y=ntr, data=menu)

plt.xticks(rotation=55)

plt.show()

# Texual Description about above visualization

for ntr in nutritions:

temp\_list = menu.pivot\_table(ntr, 'Category').sort\_values(by=ntr, ascending=False)

print('{}'.format(temp\_list[:7]))

print('-----------------------------------------------------')

"""Analysis:

1. The above code cell prints top 3 categorical source for each nutrition listed

2. Also by onserving graphs we can get rich source of each nutrition

#Plot Average Carbohydrates each Category contains

"""

sns.barplot(x='Category', y='carbohydrates(g)', data=menu)

plt.xticks(rotation=55)

"""Analysis:

1. The above graph shows the carbohydrates that each type of menu has on an average.

\* breakfast - around 42g

\* lunch - around 27g

\* dinner - around 37g

#Plot which items contains which nutritions

"""

# A function which can plot N Items which contains highest nutrition

# Change value of N as per convenience in visualization

def plot\_item\_vs\_nutrition(menu, nutrition, n):

top\_n\_items = menu.groupby('item')[nutrition].sum().nlargest(n)

# Print maximum and minimum values along with corresponding item names

max\_item = top\_n\_items.idxmax()

min\_item = top\_n\_items.idxmin()

max\_value = top\_n\_items.max()

min\_value = top\_n\_items.min()

print(f"Item with Maximum {nutrition}: {max\_item} ({max\_value})")

print(f"Item with Minimum {nutrition}: {min\_item} ({min\_value})")

plt.figure(figsize=(12, 8))

sns.barplot(x=top\_n\_items.values, y=top\_n\_items.index, orient='h')

plt.xlabel(nutrition)

plt.ylabel('Menu Item')

plt.title(f'Top {n} Items based on {nutrition}')

plt.show()

# Let's Start with protein

plot\_item\_vs\_nutrition(menu,'protein(g)', len(menu) - 1)

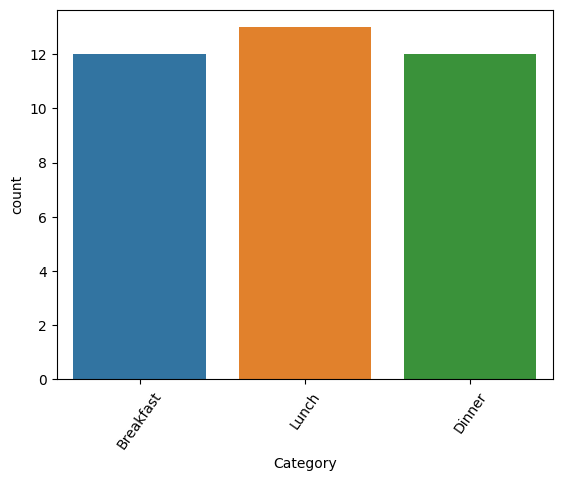
plot\_item\_vs\_nutrition(menu,'carbohydrates(g)', len(menu) - 1)

plot\_item\_vs\_nutrition(menu,'sodium(mg)', len(menu) - 1)

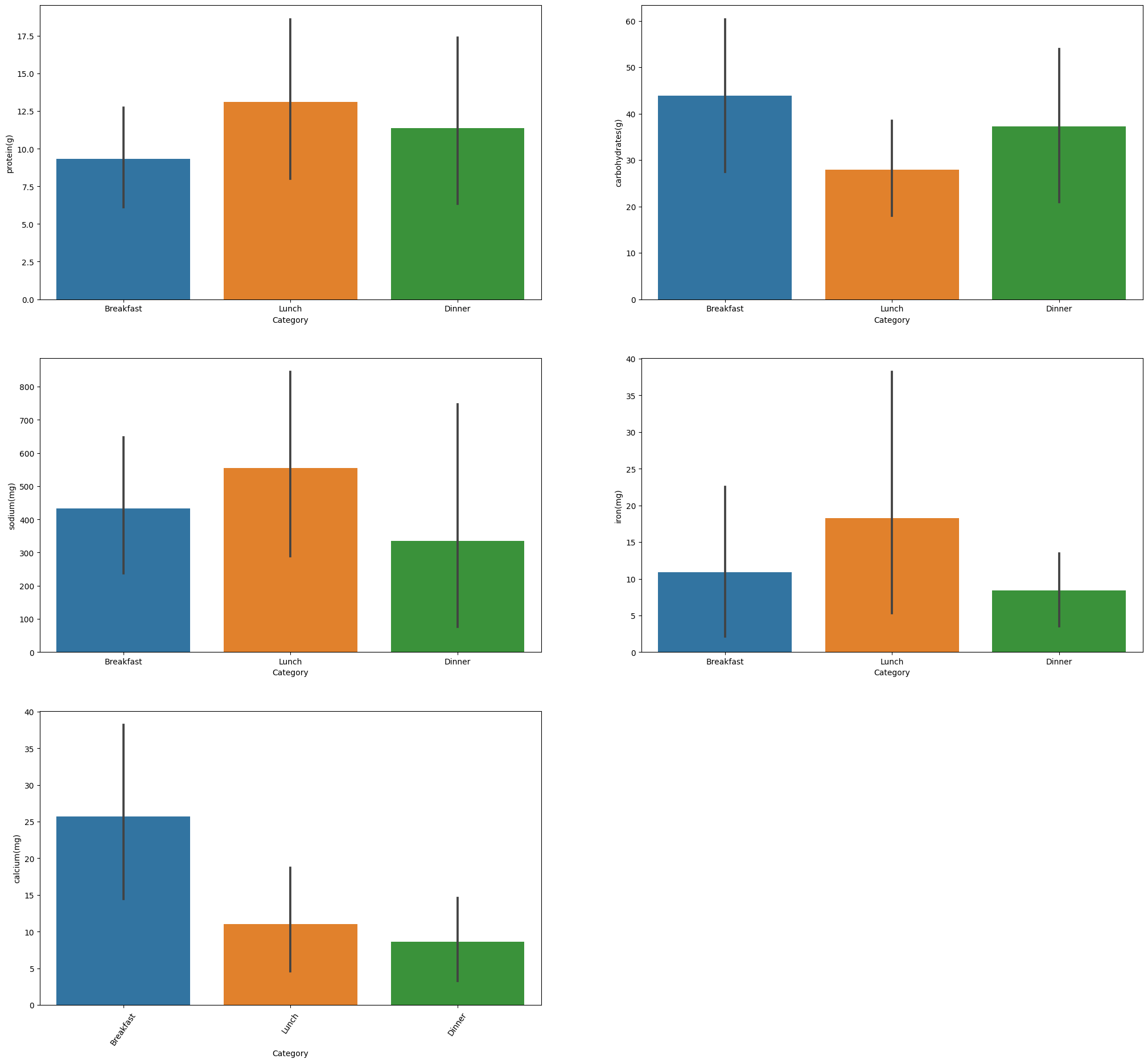
plot\_item\_vs\_nutrition(menu,'iron(mg)', len(menu) - 1)

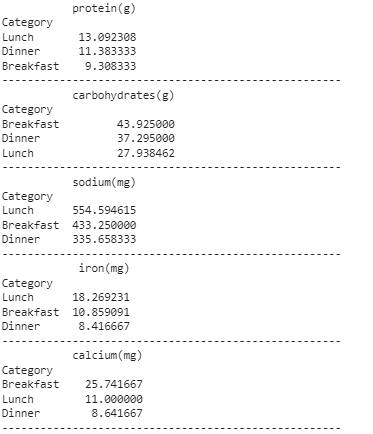
plot\_item\_vs\_nutrition(menu,'calcium(mg)', len(menu) - 1)

Count of items category wise

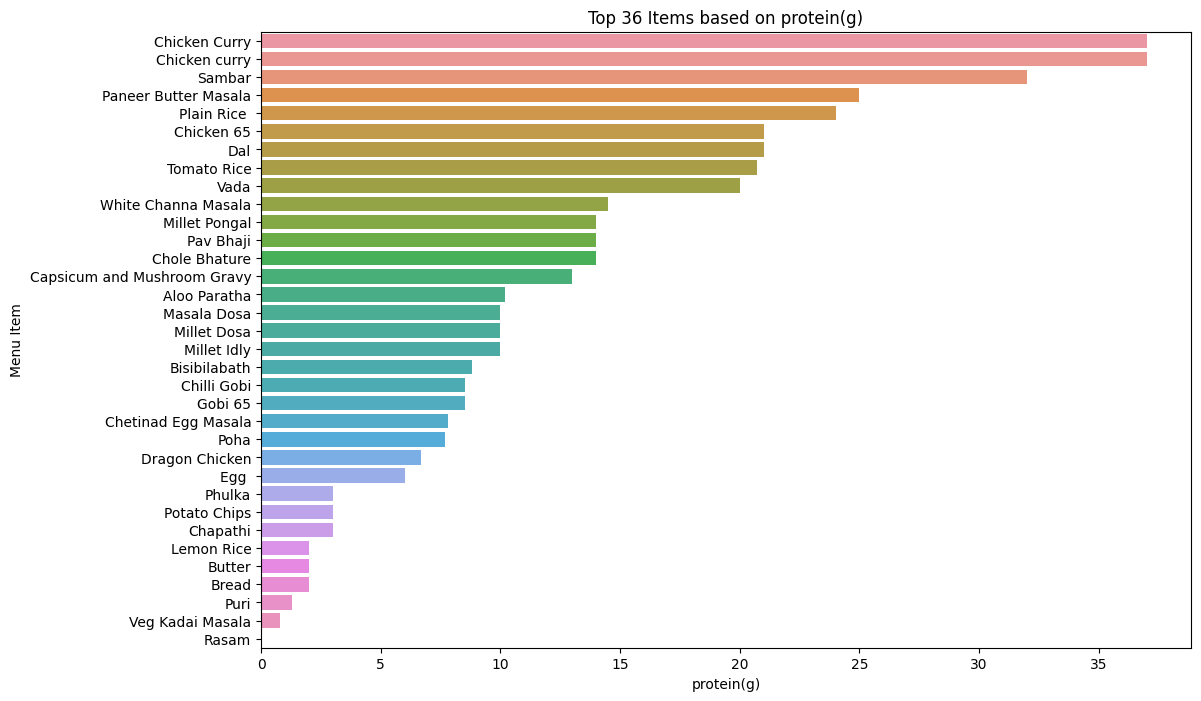


Avg Nutrition in each item

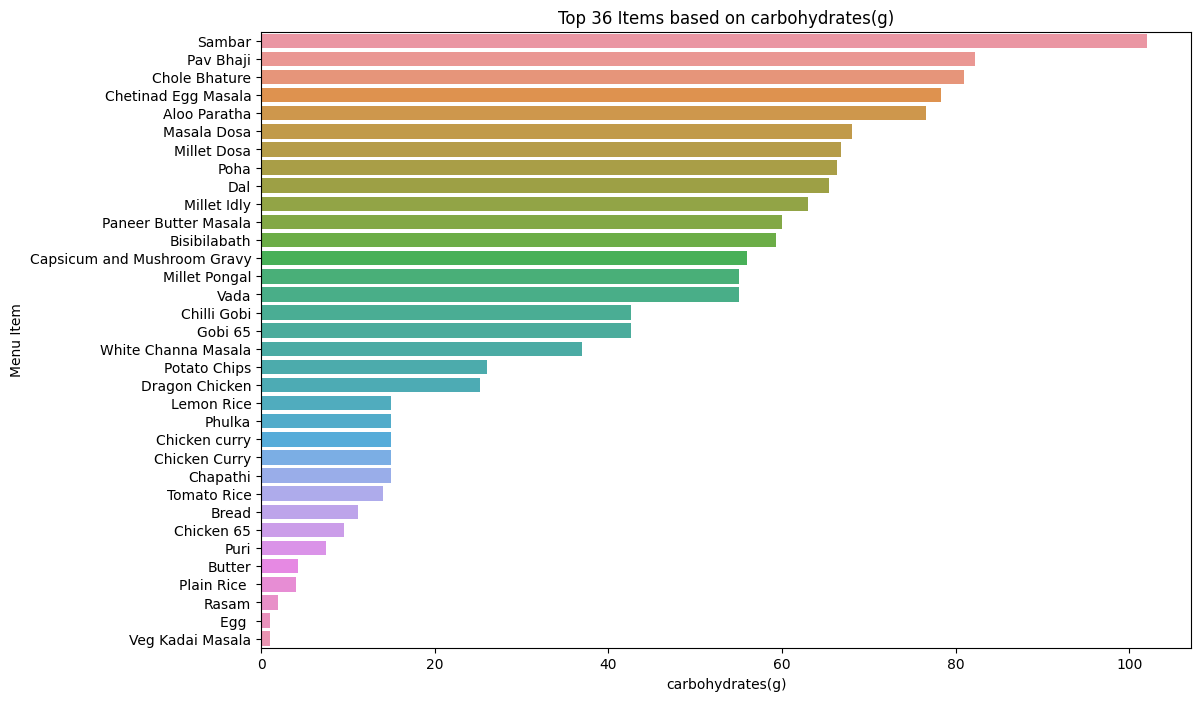




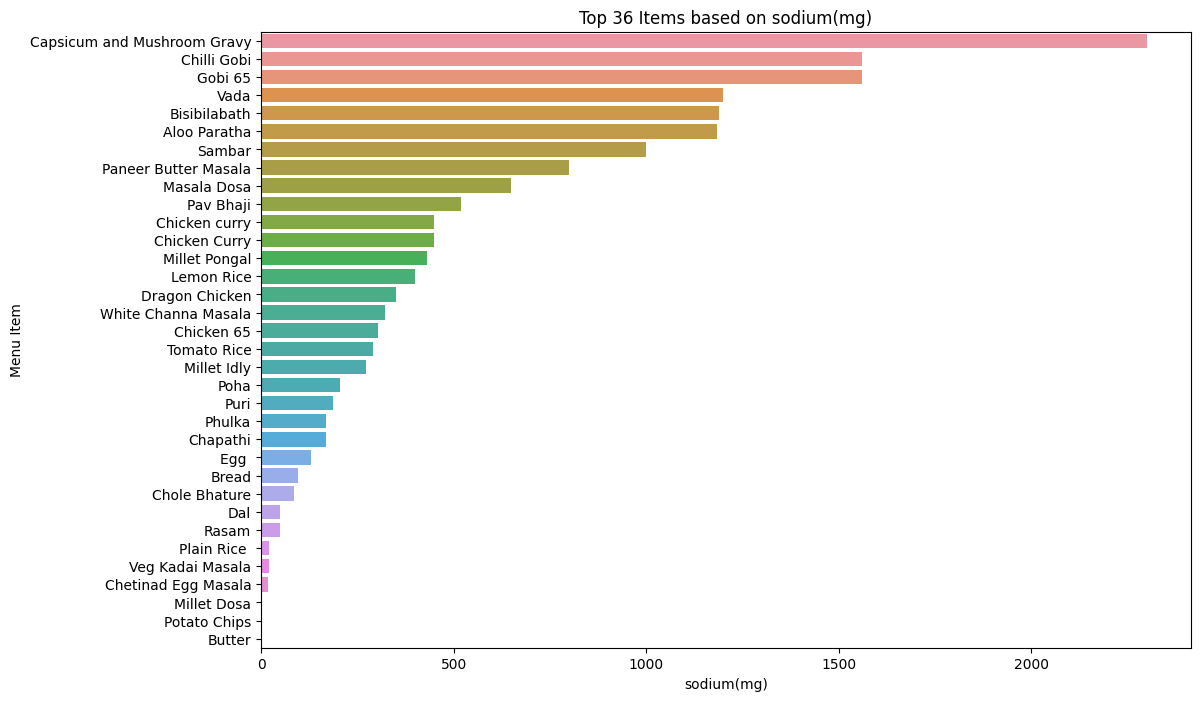
Protein in each item



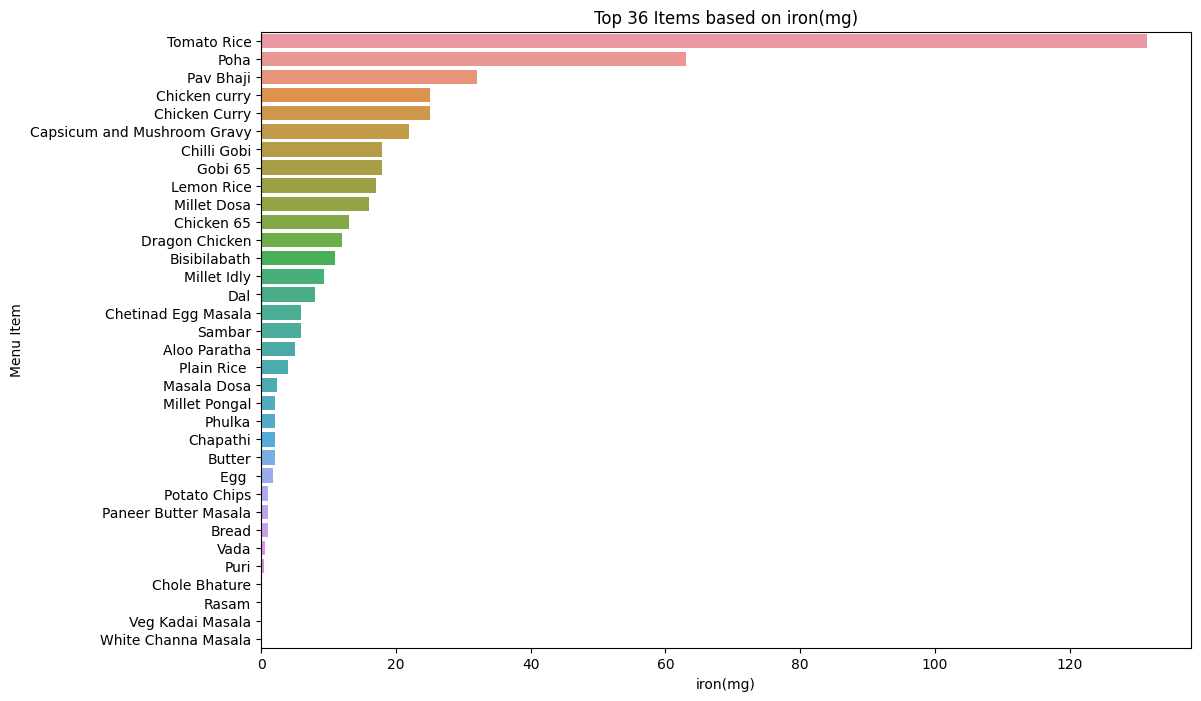
Carbohydrate in each item



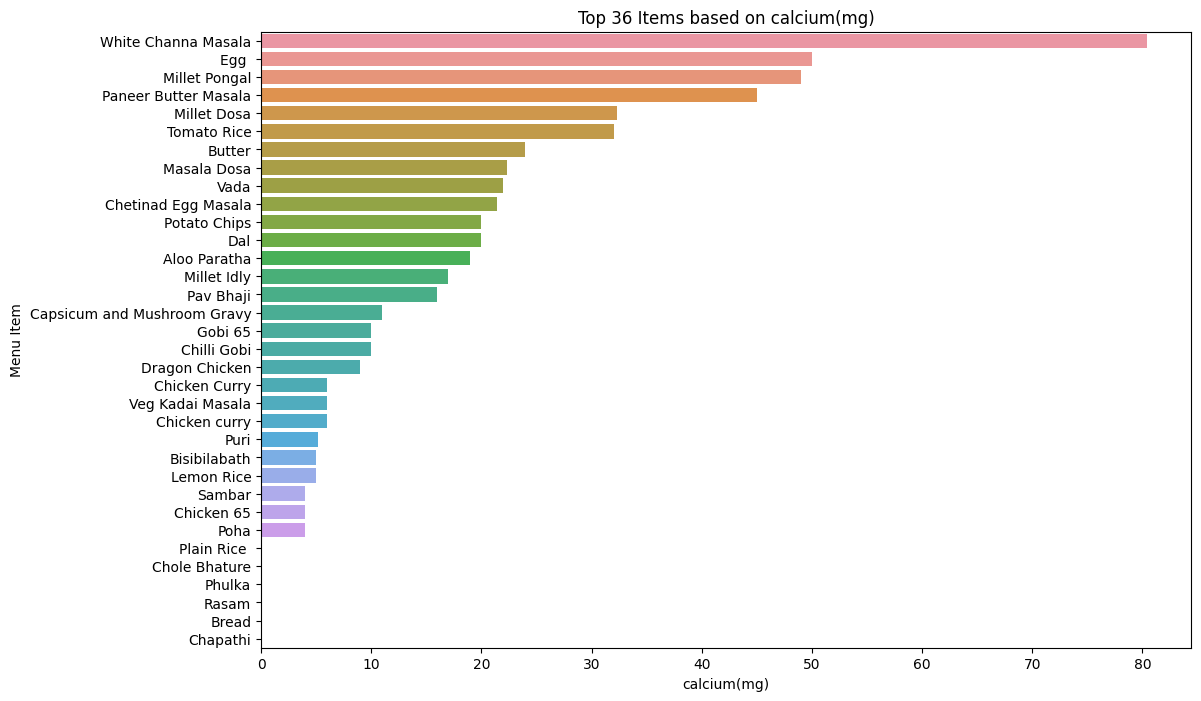
Sodium in each item



Iron in each item

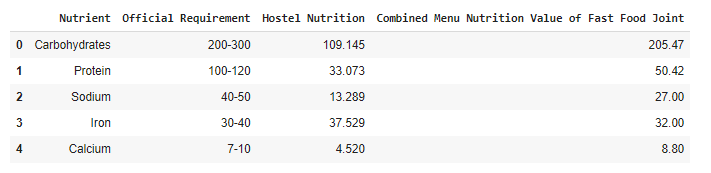


Calcium in each item



# RESULT

In terms of the nutrients being considered, the veg Maharaja mac and Chicken Cheese Lava Burger were the most appeared items after the analysis. It was observed that beverages appeared much lesser in number than compared to the food items. The highest source of protein, fat and cholesterol were all Chicken products which were cooked in a deep-fried manner. Method of cooking being a major contributor to the overall nutritive values needs recognition as well.



As you can see, the nutritional value provided by the hostel mess is below the requirements whereas the nutritional value provided by the fast food menu is over the required limit in some categories.

Hence there can be improvements made to allow the Hostel Authorities in creating a safe and more nutritious food menu for all the students.

The differences in nutritional values for hostel menu are mentioned below: -

1. Protein - ~80g

2. Sodium - ~40mg

3. Carbohydrates - ~100g

4. Iron - within range

5. Calcium - ~4mg

The differences in nutritional values for fast-food menu are mentioned below: -

1. Protein - ~50g

2. Sodium - ~20mg

3. Carbohydrates - within range

4. Iron - within range

5. Calcium - within range

Another interesting observation was that the unhealthiest food item on the menu was the adaptation of American recipes of food stuff. This shows how the western culture has influenced the food we eat and how it is also causing damage to our health. As mentioned earlier, the age group of the customers is a very wide range hence the damage caused will also be on a similarly larger group of people.

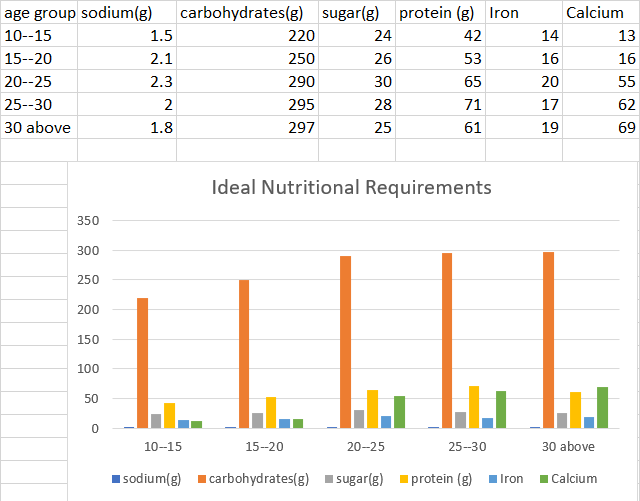
Along with the analysis of food items on the menu, the menu itself showed some interesting facts. The Gourmet menu contains the highest number of calories and contains the most expensive items. Results meaning that the majority of customers are willing to spend large sums of money on item which are available at a more economical tone as well due to the quality and variety of the menu itself.

# CONCLUSION

Through the various computations carried out, the team concurred by finding out the mentioned points: -

1. Foreign influence of food and food habits have caused more harm to the health of people in India than good.
2. Fast food joints need to be more and more responsible for the type of food they serve.
3. There is an urgent need for such joints to adapt to the food habits of the place of their venture.
4. A committee should be put in place to manage and keep check of the actual nutrient values being served by the big shots in the industry.
5. Consumers need to be more aware of the type of food they are actively consuming so as to increase their overall health and functioning of the body.

Age Wise Nutrients Requirement



# REFERENCES

[1] Petimar, J., Ramirez, M., Rifas-Shiman, S.L. *et al.* Evaluation of the impact of calorie labeling on McDonald’s restaurant menus: a natural

experiment. *Int J Behav Nutr Phys Act* **16**, 99 (2019). https://doi.org/10.1186/s12966-019-0865-7

[2] Kaggle dataset: https://[www.kaggle.com/datasets/deepcontractor/mcdonalds-](http://www.kaggle.com/datasets/deepcontractor/mcdonalds-) india-menu-nutrition-facts?resource=download

[3] Xu Chu, Ihab F. Ilyas, Sanjay Krishnan, and Jiannan Wang. 2016. Data Cleaning: Overview and Emerging Challenges. In Proceedings of the 2016 International Conference on Management of Data (SIGMOD '16). Association for Computing Machinery, New York, NY, USA, 2201–2206. https://doi.org/10.1145/2882903.2912574K.

[4] https://[www.tableau.com/learn/articles/what-is-data-](http://www.tableau.com/learn/articles/what-is-data-) cleaning#:~:text=Having%20clean%20data%20will%20ultimately,clients%2 0and%20less%2Dfrustrated%20employees..

[5] https://colab.research.google.com/drive/1\_TGvePvU2Vg2ypQPwesBRlsE--

L9-\_cG?usp=sharing

[6] https://pandas.pydata.org/docs/

[7] https://numpy.org/doc/

[8] https://seaborn.pydata.org/

[9] https://matplotlib.org/stable/index.html