Mahatma Education Society's

Pillai College of Arts, Commerce & Science

(Autonomous)

Affiliated to University of Mumbai

NAAC Accredited 'A' grade (3 cycles)

Best College Award by University of Mumbai

ISO 9001:2015 Certified

CERTIFICATE

This is to certify that Miss AMRITHA RAJ RAJESHof S.Y. B.Sc. IT Semester 4th has completed the project work in the Subject of Data Structures using Python during the academic year 2025-26 under the guidance of Mrs.NIKITA BAHALEY being the partial requirement for the fulfilment of the curriculum of Degree of Bachelor of Science in Information Technology, University of Mumbai.

MAHATMA EDUCATION SOCIETY'S

PILLAI COLLEGE OF ARTS, COMMERCE & SCIENCE
(Autonomous) NEW PANVEL

PROJECT REPORT ON

IN PARTIAL FULFILLMENT OF

BACHELOR OF SCIENCE IN IT

SEMESTER IV– 2025-26

PROJECT GUIDE : Mrs.NISHA YADAV

SUBMITTED BY:

AMRITHA RAJ RAJESH

ROLL NO: 5506

CONTENT

1	**Introduction**
2	**Dataset description**
3	**Data analysis topic**
4	**Methodology**
5	**Source Code**
6	**Output**
7	**Conclusion**

INTRODUCTION

In recent years, fast food consumption has increased significantly due to rapid urbanization, changing lifestyles, and busy schedules. Fast food restaurants such as McDonald's, Burger King, KFC, and Subway offer quick, affordable, and convenient meals, which has made them popular among people of all age groups, especially students and working professionals.

However, fast food is often high in calories, fats, sodium, and sugar, which can lead to serious health issues such as obesity, diabetes, heart diseases, and high blood pressure. Understanding the nutritional content of fast food is important to make healthier dietary choices and to spread awareness about the impact of unhealthy eating habits.

Data Science is a field that combines statistics, mathematics, programming, and data analysis techniques to extract meaningful insights from large datasets. It helps organizations and researchers analyze large amounts of data, identify patterns, and make data-driven decisions. In the food industry, data science can be used to analyze nutritional values, customer preferences, and health risks associated with food consumption.

In this project, a Fast Food Nutrition Dataset is analyzed using Python programming and data science techniques. The dataset contains information about various fast food items, their calorie content, fat content, sodium, sugar, and protein values. The main objective of this project is to explore and analyze the dataset, perform data cleaning and transformation, visualize the data using graphs, and identify trends and patterns related to fast food nutrition.

The project uses exploratory data analysis (EDA), data visualization techniques, and statistical measures such as mean, median, and mode to understand the dataset. Graphical representations such as bar charts, histograms, box plots, scatter plots, pie charts, and heatmaps are used to make the analysis more understandable and visually attractive.

This project demonstrates how data science techniques can be applied to real-world datasets to gain insights into fast food nutrition and its impact on health. The results of this analysis can help consumers make informed food choices and raise awareness about the importance of a balanced and healthy diet.

DATASET DESCRIPTION

The dataset used in this project is the **Fast Food Nutrition Dataset**, which was downloaded from Kaggle, an online platform that provides real-world datasets for data science and machine learning projects. The dataset contains nutritional information about various fast food items from popular fast food restaurants such as McDonald's, Burger King, KFC, Subway, and others.

The main purpose of this dataset is to provide detailed nutritional values of fast food products so that they can be analyzed to understand their impact on human health. The dataset includes information such as calorie content, fat, sodium, sugar, and protein levels for each food item. These attributes help in identifying unhealthy food items and comparing nutritional values across different restaurants.

The dataset consists of more than **300 fast food items** with multiple attributes for each item. It includes both numerical and categorical data. Numerical data includes values such as calories, fat, sodium, and protein, while categorical data includes restaurant name and food item name.

Dataset Source

The dataset was obtained from Kaggle, a popular platform for sharing datasets and data science projects. Kaggle provides free access to datasets for educational and research purposes.

Dataset Link:

https://www.kaggle.com/datasets/ulrikthygepedersen/fastfood-nutrition

DATA ANALYSIS TOPICS USED

Data Cleaning

- Handling missing values
- Removing duplicates

Data Transformation

- Converting data types
- Normalization

Exploratory Data Analysis (EDA)

- Mean, Median, Mode
- Graphs and charts

Visualization

- Bar graph
- Histogram
- Pie chart

METHODOLOGY

Step 1: Download Dataset

Download CSV file from Kaggle.

Step 2: Open Google colab

Step 3: Load Dataset in Python

Step 4: Clean Data

Step 5: Analyze Data

Step 6: Create Graphs

Step 7: Upload project to GitHub

SOURCE CODE

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler, LabelEncoder

plt.style.use("ggplot")

data = pd.read_csv("fastfood.csv")
print("First 5 Rows of Dataset:")
print(data.head())

print("\nDataset Info:")
print(data.info())

print("\nDataset Description:")
print(data.describe())



# Missing values
print("\nMissing Values:")
print(data.isnull().sum())
data.fillna(data.mean(numeric_only=True), inplace=True)

# Remove duplicates
data.drop_duplicates(inplace=True)

# Rename column (if exists)
if "Total Fat (g)" in data.columns:
    data.rename(columns={"Total Fat (g)": "Total_Fat"}, inplace=True)

# Handle inconsistent text
data["Restaurant"] = data["Restaurant"].str.title()



# Data type conversion
```

```
data["Calories"] = data["Calories"].astype(int)

# Scaling Calories
scaler = MinMaxScaler()
data["Calories_Scaled"] = scaler.fit_transform(data[["Calories"]])

# Binning Calories
data["Calorie_Level"] = pd.cut(data["Calories"], bins=3, labels=["Low", "Medium",
"High"])

# Aggregation
avg_calories = data.groupby("Restaurant")["Calories"].mean()
print("\nAverage Calories by Restaurant:")
print(avg_calories)

print("\nMean Calories:", data["Calories"].mean())
print("Median Calories:", data["Calories"].median())
print("Mode Calories:", data["Calories"].mode())

# Label Encoding Restaurant
le = LabelEncoder()
data["Restaurant_Label"] = le.fit_transform(data["Restaurant"])

# Feature Extraction (Length of item name)
data["Item_Length"] = data["Item"].apply(len)

data["Item"] = data["Item"].str.lower()   # lowercase
data["Item"] = data["Item"].str.replace(r"[^\w\s]", "", regex=True)  # remove punctuation


# 1. Bar Chart
plt.figure(figsize=(10,5))
avg_calories.sort_values(ascending=False).plot(kind="bar")
plt.title("Average Calories by Restaurant")
plt.xlabel("Restaurant")
plt.ylabel("Calories")
plt.show()
```

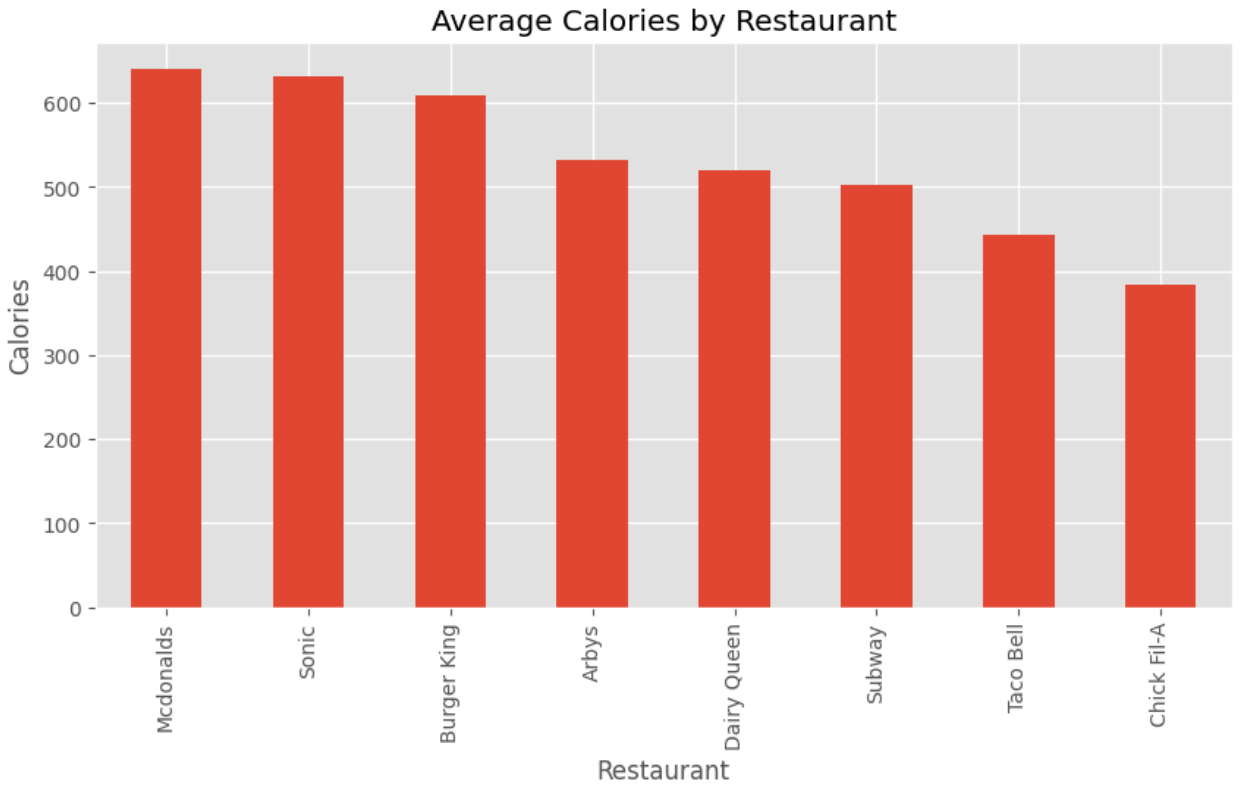
```
# 2. Histogram
plt.figure(figsize=(8,5))
plt.hist(data["Calories"], bins=20)
plt.title("Calories Distribution")
plt.xlabel("Calories")
plt.ylabel("Number of Food Items")
plt.show()

# 3. Box Plot
plt.figure(figsize=(8,5))
data.boxplot(column="Calories")
plt.title("Outliers in Calories")
plt.show()

# 4. Scatter Plot
if "Total_Fat" in data.columns:
    plt.figure(figsize=(8,5))
    plt.scatter(data["Total_Fat"], data["Calories"])
    plt.xlabel("Total Fat (g)")
    plt.ylabel("Calories")
    plt.title("Fat vs Calories Relationship")
    plt.show()

# 5. Pie Chart
restaurant_counts = data["Restaurant"].value_counts()
plt.figure(figsize=(7,7))
restaurant_counts.plot(kind="pie", autopct="%1.1f%%")
plt.title("Food Items Distribution by Restaurant")
plt.ylabel("")
plt.show()

# 6. Heatmap
plt.figure(figsize=(10,6))
sns.heatmap(data.corr(numeric_only=True), annot=True)
plt.title("Correlation Between Nutrition Values")
plt.show()
```

OUTPUT

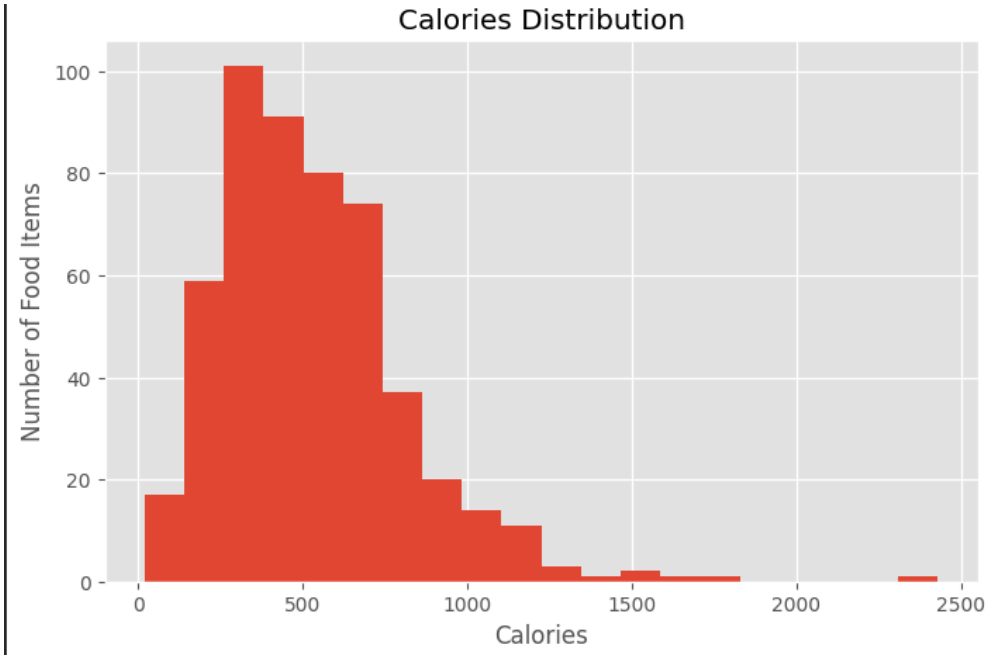
```
Mean: 530.9126213592233
Median: 490.0
Mode: 0    350
1    380
Name: calories, dtype: int64
```

	calories	Calorie_Level
0	380	Low
1	840	Medium
2	1130	Medium
3	750	Low
4	920	Medium

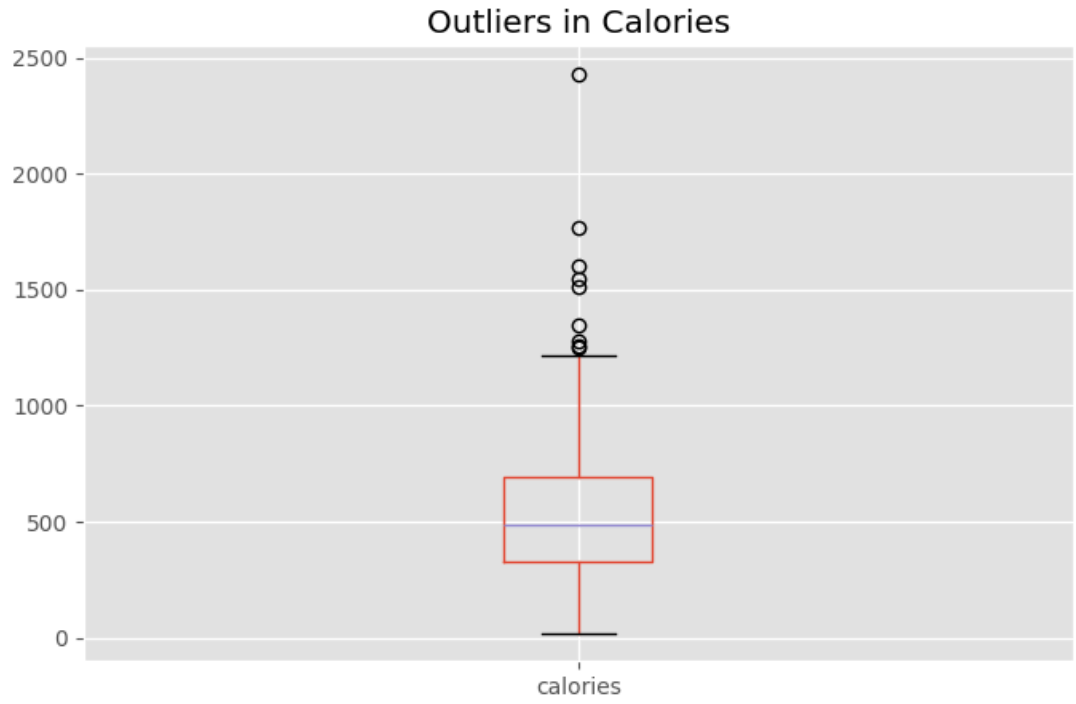
Bar chart:



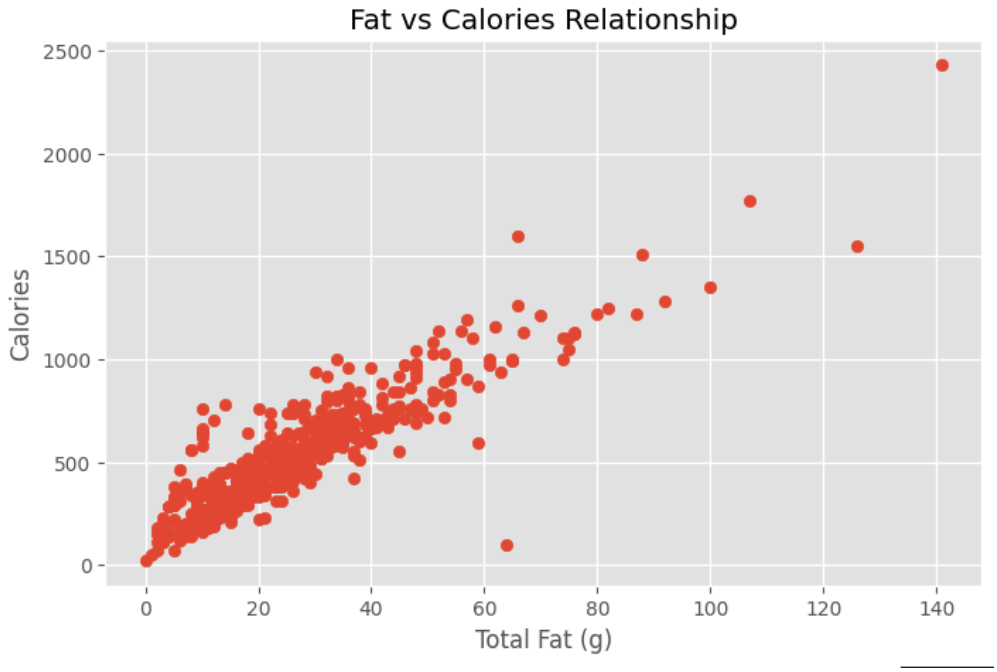
Histogram:



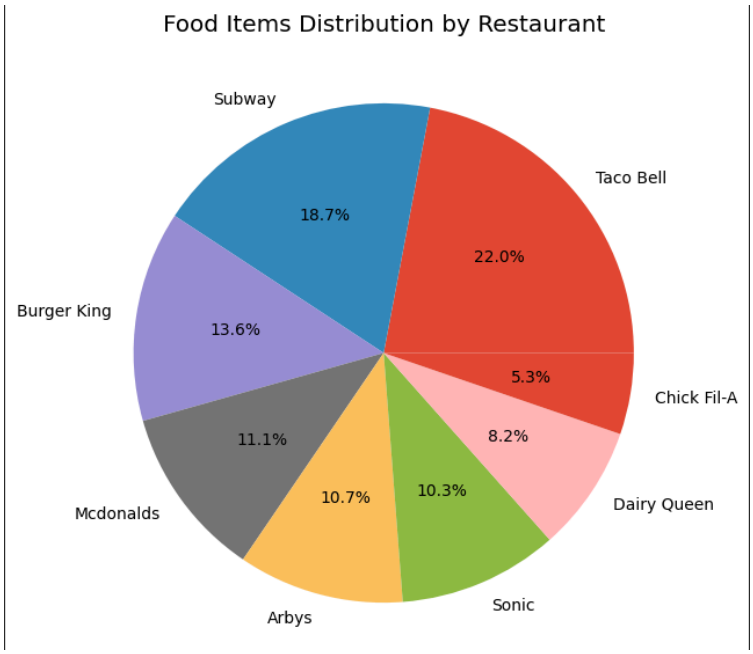
Box plot:



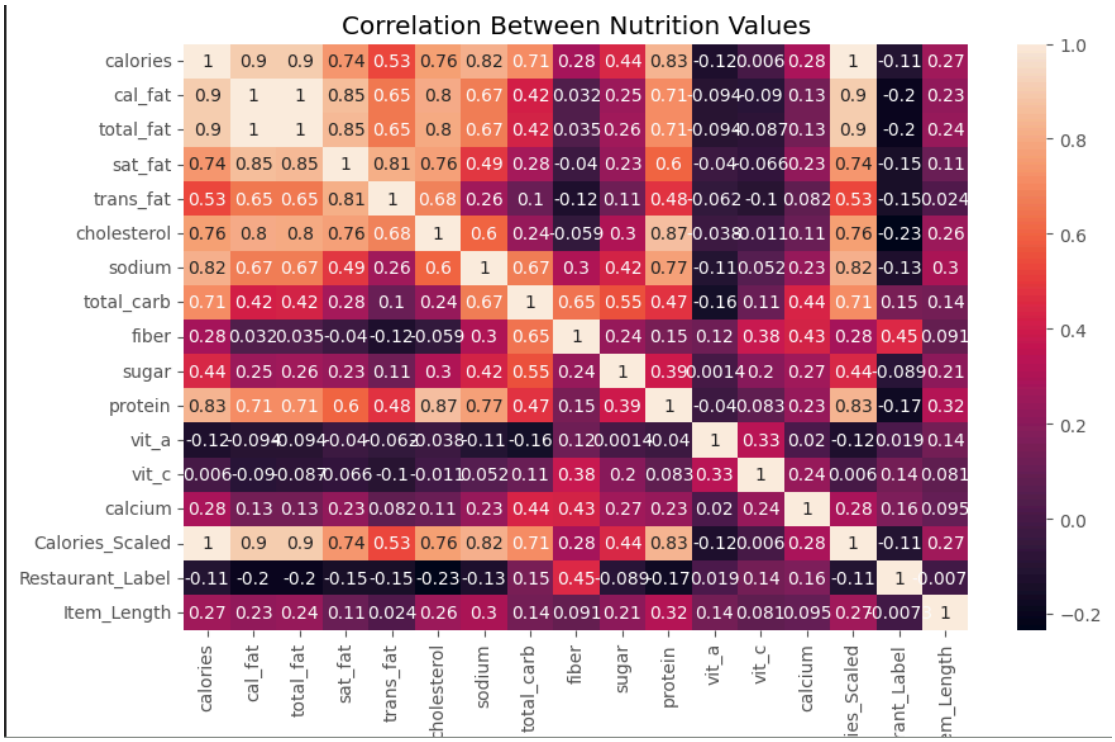
Scatter plot:



Pie Chart:



Heat Map:


Correlation Between Nutrition Values
calories 1 0.9 0.9 0.74 0.53 0.76 0.82 0.71 0.28 0.44 0.83 -0.12 0.006 0.28 1 -0.11 0.27
cal_fat 0.9 1 1 0.85 0.65 0.8 0.67 0.42 0.032 0.25 0.71 -0.094 -0.09 0.13 0.9 -0.2 0.23
total_fat 0.9 1 1 0.85 0.65 0.8 0.67 0.42 0.035 0.26 0.71 -0.094 -0.087 0.13 0.9 -0.2 0.24
sat_fat 0.74 0.85 0.85 1 0.81 0.76 0.49 0.28 -0.04 0.23 0.6 -0.04 -0.066 0.23 0.74 -0.15 0.11
trans_fat 0.53 0.65 0.65 0.81 1 0.68 0.26 0.1 -0.12 0.11 0.48 -0.062 -0.1 0.082 0.53 -0.15 0.024
cholesterol 0.76 0.8 0.8 0.76 0.68 1 0.6 0.24 -0.059 0.3 0.87 0.038 0.011 0.11 0.76 -0.23 0.26
sodium 0.82 0.67 0.67 0.49 0.26 0.6 1 0.67 0.3 0.42 0.77 -0.11 0.052 0.23 0.82 -0.13 0.3
total_carb 0.71 0.42 0.42 0.28 0.1 0.24 0.67 1 0.65 0.55 0.47 -0.16 0.11 0.44 0.71 0.15 0.14
fiber 0.28 0.032 0.035 -0.04 -0.12 -0.059 0.3 0.65 1 0.24 0.15 0.12 0.38 0.43 0.28 0.45 0.091
sugar 0.44 0.25 0.26 0.23 0.11 0.3 0.42 0.55 0.24 1 0.39 0.0014 0.2 0.27 0.44 -0.089 0.21
protein 0.83 0.71 0.71 0.6 0.48 0.87 0.77 0.47 0.15 0.39 1 -0.04 0.083 0.23 0.83 -0.17 0.32
vit_a -0.12 -0.094 -0.094 -0.04 -0.062 -0.038 -0.11 -0.16 0.12 0.0014 -0.04 1 0.33 0.02 -0.12 0.019 0.14
vit_c 0.006 -0.09 -0.087 0.066 -0.1 -0.011 0.052 0.11 0.38 0.2 0.083 0.33 1 0.24 0.006 0.14 0.081
calcium 0.28 0.13 0.13 0.23 0.082 0.11 0.23 0.44 0.43 0.27 0.23 0.02 0.24 1 0.28 0.16 0.095
Calories_Scaled 1 0.9 0.9 0.74 0.53 0.76 0.82 0.71 0.28 0.44 0.83 -0.12 0.006 0.28 1 -0.11 0.27
Restaurant_Label -0.11 -0.2 -0.2 -0.15 -0.15 -0.23 -0.13 0.15 0.45 -0.089 -0.17 0.019 0.14 0.16 -0.11 1 0.007
Item_Length 0.27 0.23 0.24 0.11 0.024 0.26 0.3 0.14 0.091 0.21 0.32 0.14 0.081 0.095 0.27 0.007 1
calories
cal_fat
total_fat
sat_fat
trans_fat
cholesterol
sodium
total_carb
fiber
sugar
protein
vit_a
vit_c
calcium
ies_Scaled
rant_Label
m_Length
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

CONCLUSION

In this project, the Fast Food Nutrition Dataset was analyzed using various data science techniques. The dataset was collected from Kaggle and contained nutritional information such as calories, fat, sodium, sugar, and protein for different fast food items from popular restaurants. The project involved data collection, data cleaning, data transformation, exploratory data analysis (EDA), and data visualization using Python.

Data cleaning techniques were applied to handle missing values, remove duplicate records, and correct inconsistent data. Data transformation methods such as scaling and binning were used to prepare the data for analysis. Exploratory data analysis was performed using statistical measures like mean, median, and mode to understand the distribution of calorie values. Various graphs such as bar charts, histograms, box plots, scatter plots, pie charts, and heatmaps were generated to visualize the data and identify patterns.

The results of the analysis showed that many fast food items contain high calories and fat content, which can contribute to health problems if consumed frequently. The correlation analysis also showed a strong relationship between calories and fat content, indicating that foods with higher fat tend to have higher calories. This project highlights the importance of analyzing nutritional data to make healthier food choices.

Overall, this project demonstrates how data science can be used to analyze real-world datasets and extract meaningful insights. The techniques used in this project can be applied to other datasets in the future to support data-driven decision-making. This study also helps in creating awareness about the health risks associated with fast food consumption and encourages people to adopt healthier eating habits.