****

**COURSE CODE: INT353**

**COURSE NAME: EDA PROJECT**

**Continuous Assessment – II**

**NAME: B. MANOHAR**

**REG NO: 12100789**

**ROLL NO: RK21UWB51**

**SECTION: K21UW**

**Table of Contents:**

|  |  |
| --- | --- |
| Name | S. No. |
| Introduction | 03 |
| Brief Overview of Dataset | 04 |
| Reason for selecting Dataset | 05 |
| Libraries Used | 05 |
| Steps of EDA | 06 |
| Dataset Description | 08 |
| Data Exploration | 09 |
| Data Cleaning | 10 |
| Univariate Analysis | 11 |
| Bivariate Analysis | 14 |
| Multivariate Analysis | 16 |
| Distribution | 18 |
| Hypothesis Testing | 18 |
| Finding and Insights | 19 |
| Limitations | 20 |
| Recommendations | 20 |
| Conclusion | 20 |
| References | 21 |
| Acknowledgement | 22 |
| Project Code | 23 |

**Introduction:**

McDonald's is one of the largest and most popular fast-food chains in the world. It is known for its convenient and affordable meals, but there have been concerns about the nutritional quality of its food.

This report presents the results of an exploratory data analysis (EDA) of McDonald's nutrition data. The data was collected from the and includes information on the calories, fat, sugar, and other nutrients in each menu item.

The goals of this EDA were to:

* Understand the nutritional content of McDonald's food.
* Identify any relationships between the different nutrients.
* Make recommendations for how people can make healthier choices when eating at McDonald’s.

**Domain Knowledge:**

The nutritional quality of McDonald's food has been a source of debate for many years. Some people argue that McDonald's food is high in calories, fat, and sugar, and can contribute to obesity and other health problems. Others argue that McDonald's food has improved in recent years and now offers a variety of healthier options, such as salads, wraps, and grilled chicken sandwiches.

The domain knowledge for this EDA is McDonald's nutrition. This includes knowledge of the different nutrients found in McDonald's food, the recommended daily intake for each nutrient, and the potential health implications of consuming too much or too little of each nutrient.

For example,

Saturated fat is a type of fat that can raise LDL cholesterol levels. High LDL cholesterol levels can increase the risk of heart disease. The recommended daily intake for saturated fat is no more than 10% of total daily calories.

Here is a brief overview of the nutritional content of McDonald's food:

|  |  |
| --- | --- |
| **Column** | **Description** |
| **Item** | The name of the menu item |
| **Calories** | The total number of calories in the menu item **Calories From Fat** |
| **Calories From Fat** | The number of calories from fat in the menu item |
| **Total Fat** | The total amount of fat in the menu item |
| **Saturated Fat** | The amount of saturated fat in the menu item |
| **Trans Fat** | The amount of trans fat in the menu item |
| **Cholesterol** | The amount of cholesterol in the menu item |
| **Sodium** | The amount of sodium in the menu item |
| **Carbohydrates** | The total amount of carbohydrates in the menu item |
| **Fiber** | The amount of fibre in the menu item |
| **Sugars** | The amount of sugar in the menu item |
| **Proteins** | The amount of protein in the menu item |
| **Weight** | The weight of the menu item |

**Reason for Selecting the dataset:**

* Relevance:

Understanding the nutritional content of McDonald's food is important for people who make choices about what to eat.

* Availability:

The McDonald's nutrition dataset is publicly available and easy to access. This makes it a good choice for an EDA project, as it is not necessary to spend time and resources collecting the data.

* Completeness:

This allows for a comprehensive analysis of the nutritional content of McDonald's food.

I believe that this dataset is well-suited for an EDA project, as it is relevant, available, complete. I hope that my analysis will provide valuable insights into the nutritional content of McDonald's food and help people make healthier choices when eating at McDonald's.

**Libraries Used:**

**Numpy**:

* Numpy is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays.

**Pandas**:

* Pandas is a Python library that provides high-performance, easy-to-use data structures and data analysis tools. It is built on top of the NumPy library and provides a variety of features for working with data

**Matplotlib and Seaborn:**

* Matplotlib and Seaborn were used for data visualization. Matplotlib provided essential plotting functions, while Seaborn offered a higher-level interface for creating aesthetically pleasing and informative visualizations.

**SciPy**:

* SciPy is a Python library that provides a wide range of scientific and technical computing algorithms and functions. It is built on top of the NumPy library and provides a variety of features.

**Steps of EDA:**

Exploratory Data Analysis (EDA) is a crucial step in the data analysis process that helps you understand your dataset, identify patterns, and prepare the data for further analysis or modelling. Here are the steps typically involved in EDA:

* **Data Collection**: Gather your dataset from reliable sources. Ensure that you have all the necessary data to perform your analysis.
* **Data Loading**: Load the dataset into your analysis environment. Common tools for this step include Pandas in Python or read functions in R
* **Initial Data Inspection**: Use Pandas to check the first few rows of the dataset with df.head() to get an overview of the data's structure. Use df.shape to determine the number of rows and columns in the dataset. Use df.info() to get information about the data types of each column and check for missing values.
* **Data Cleaning and Handling Missing Values**: Identify missing values using functions like df.isnull().sum(). Decide on a strategy for handling missing values (e.g., imputation or removal) based on the nature of the data and the analysis goals. Use Pandas functions to perform data cleaning, such as dropping unnecessary columns or duplicates.
* **Data Visualization**: Create visualizations like scatter plots, box plots, and violin plots to explore relationships between variables. Visualize trends and patterns in the data. Use Seaborn's pair plots or heatmaps for a comprehensive view of variable relationships.
* **Outlier Detection**: Identify potential outliers by visualizing data using box plots or other relevant techniques. Decide on an approach for handling outliers, which may include removal or transformation.
* **Feature Engineering**: Create new features, if necessary, based on domain knowledge or relationships observed during EDA. Perform feature transformations, such as scaling or normalization, if it improves the quality of the data.
* **Hypothesis Testing** (Optional): If relevant, conduct statistical tests or hypothesis tests to make data-driven conclusions about specific relationship or hypotheses.
* **Documentation and Reporting**: Document your findings, insights, and any actions taken during EDA. Create visualizations and reports that communicate your results effectively.
* **Iterate**: EDA is an iterative process. You may need to revisit previous steps as you gain more insights or encounter issues during the analysis.

Reading CSV File:



Renaming Columns:

A screenshot of a computer

Description automatically generated

A list of food items

Description automatically generatedDescribe Dataset:

A white rectangular object with a white background

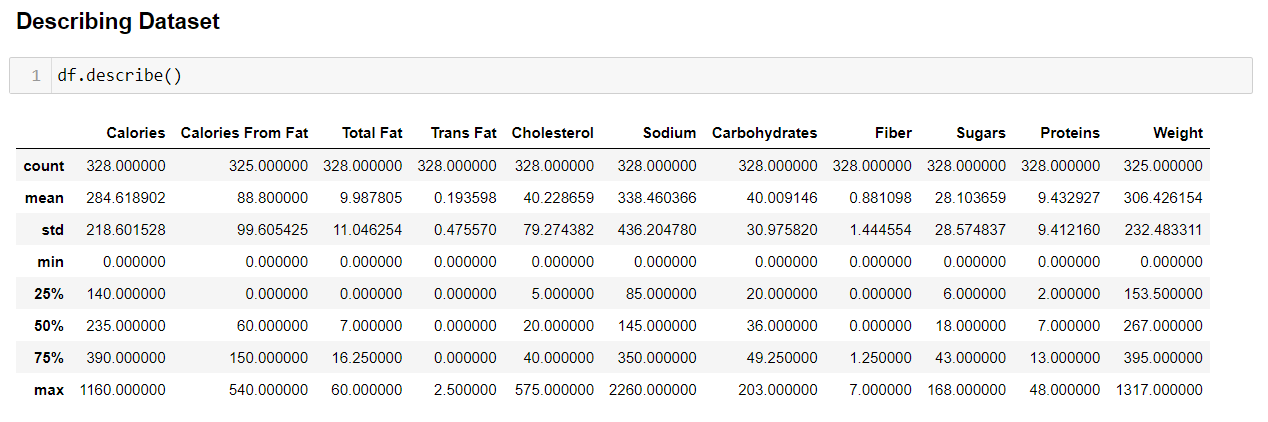
Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



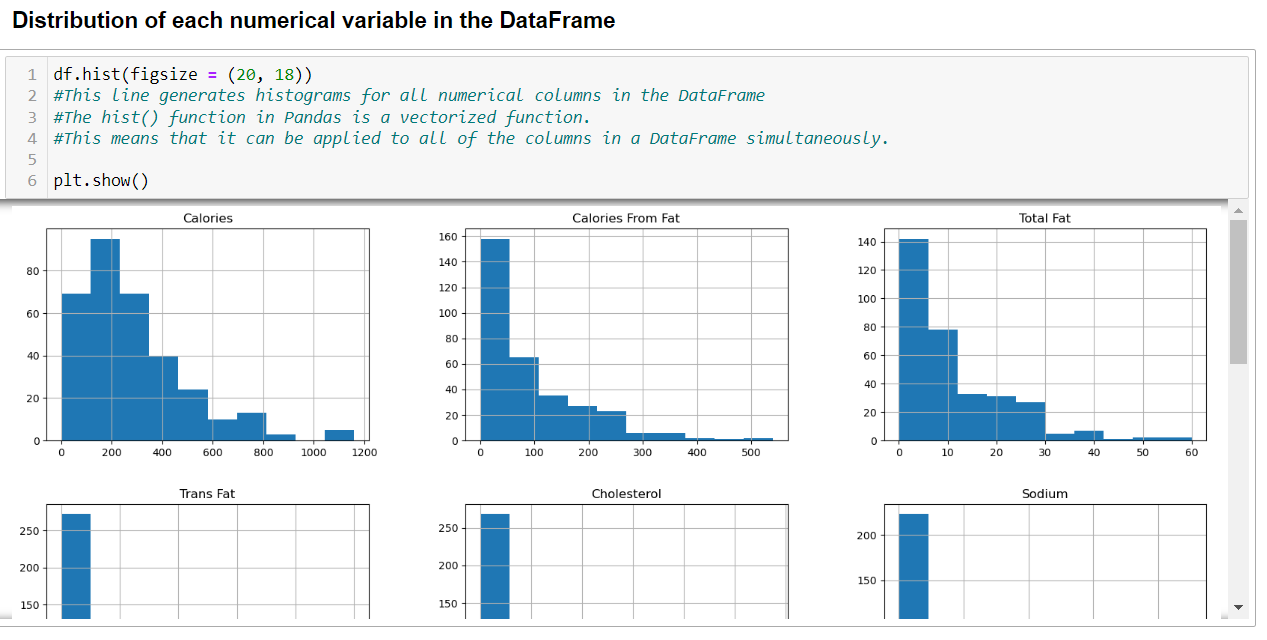
**A white rectangular box with blue text

Description automatically generated1.Data Cleaning:**

**A screenshot of a computer

Description automatically generated**

**2.Univariate Analysis**

****

**A screenshot of a computer

Description automatically generated**

**A graph of weight loss

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A graph with blue rectangular objects

Description automatically generated with medium confidence**

**A white rectangular object with black text

Description automatically generated**

**A diagram of a box plot

Description automatically generated**

**A graph of calories and weight

Description automatically generatedA white rectangular object with blue text

Description automatically generated**

**A screenshot of a computer code

Description automatically generated**

**A pie chart with text

Description automatically generated**

**4.Bivariate Analysis**

**A screen shot of a computer code

Description automatically generated**

**A chart of food items

Description automatically generated**

**A white rectangular object with black text

Description automatically generated**

**A diagram of a weight scale

Description automatically generated**

**A graph with text overlay

Description automatically generatedA white background with text

Description automatically generated with medium confidence**

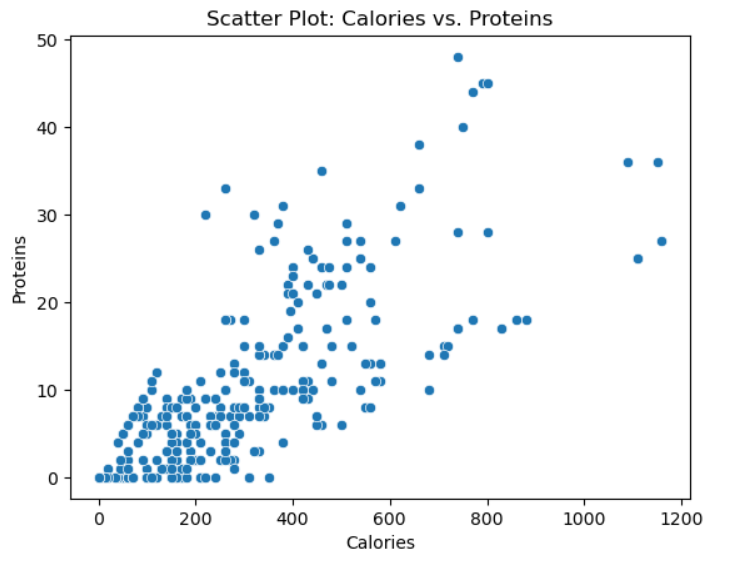
**A screenshot of a computer

Description automatically generated**



**A white rectangular object with a blue and black design

Description automatically generated**

****

A black and white text

Description automatically generated

**A graph showing a number of calories

Description automatically generated**

**5.Multivariate Analysis**

A black text on a white background

Description automatically generated

A chart of food content

Description automatically generated with medium confidence

A black text on a white background

Description automatically generated

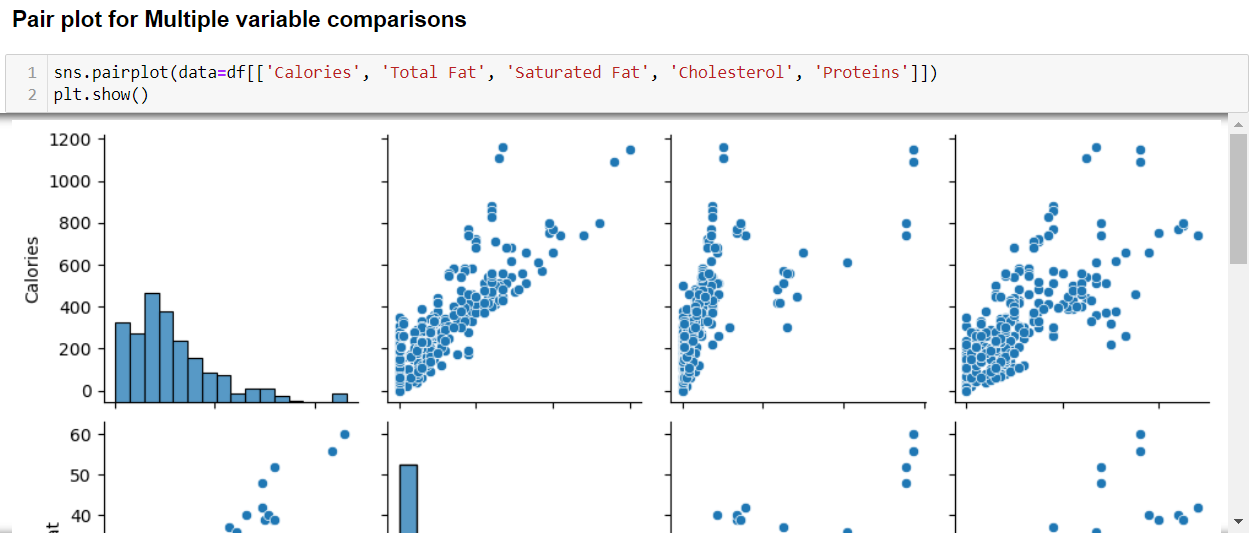
A white graph with green and black lines

Description automatically generated



A chart with different colored boxes

Description automatically generated



**6.Distribution**

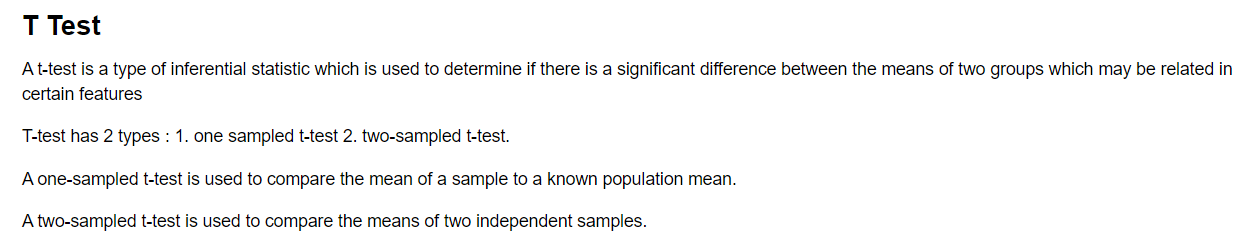




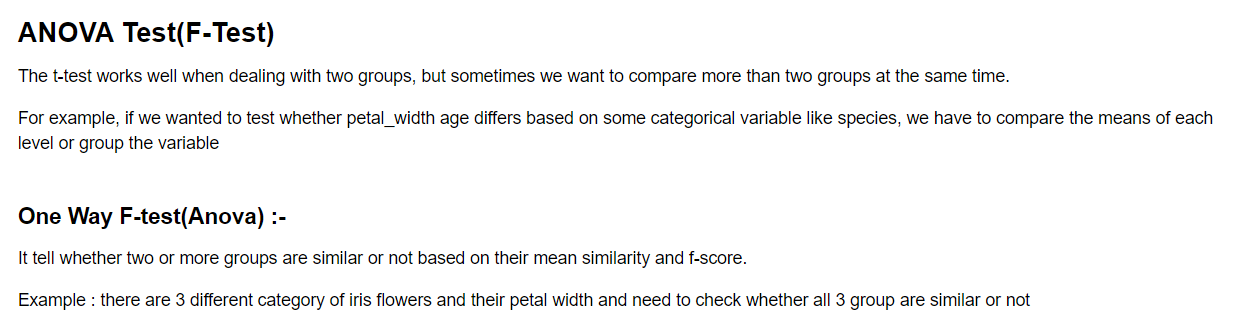




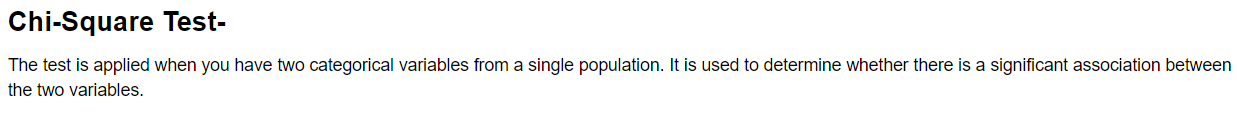


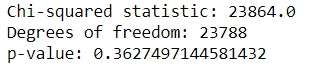
**7.Hypothesis Testing**

****

****







**Finding and Insights:**

* Overall, McDonald's food is high in calories, fat, and sugar.
* The type of fat in McDonald's food varies depending on the menu item.

Some items, such as cheeseburgers and French fries, are high in saturated fat, while others, such as grilled chicken sandwiches and salads, are low in saturated fat.

* Sodium intake is a concern for many people who eat at McDonald's.
* There are some healthier options available on the McDonald's menu. Salads, grilled chicken sandwiches, and fruit smoothies are all healthier choices than burgers, fries, and milkshakes.

**Limitations:**

* The data is only from one source, the McDonald's website. It is possible that the data on the website is not accurate or complete.
* The data does not include information about the cooking methods used to prepare the food. This information can affect the nutritional content of the food.
* The main limitation in the data set was that the data was in continuous form due to which the sample would be large , and if we want to Perform plots like Boxplot the plots would be shown small, so we need to limit the usage of columns or slice the values to get a more perfect plot.

**Recommendations:**

From using the data, I had performed many types of plots but for more deep analysis we can use “Machine Learning models” like

Decision Tree , K-means , Natural language processing , etc

By using theses ML models, we can have a Deep understanding on the data and perform multiple types of analysis.

**Conclusion:**

McDonald's nutrition dataset provides valuable information about the nutritional content of McDonald's food. The data shows that there is a wide range of nutritional content across the McDonald's menu, with some items being relatively healthy and others being very unhealthy. The data also shows that sodium intake is a concern for many people who eat at McDonald's.

Based on my findings, I recommend that McDonald's take steps to educate consumers about the nutritional content of its food, offer healthier menu options, and work with suppliers to improve the quality of its food. McDonald's could also conduct further research on the nutritional impact of its food and partner with public health organizations to promote healthy eating and lifestyle habits.

By taking these actions, McDonald's could help to make its food healthier and more appealing to consumers.

**Reference:**

<https://www.kaggle.com/>

<https://www.youtube.com/>

<https://www.statisticshowto.com/>

<https://study.com/>

**Acknowledgement:**

I hereby declare that the project work entitled(“McDonald’s Nutrition”) is an authentic record of my own work carried out as

requirements of Project for the award of B. Tech degree in EDA Project from

Lovely Professional University, Phagwara, under the guidance of Shivangini

Gupta(Mam) ,during August to November 2023. All the information furnished

in this capstone project report is based on my own intensive work and is genuine.

**Name**: B. MANOHAR  
**Reg No**: 12100789

**Project Code:**

* **Dataset:**

[**https://drive.google.com/file/d/19TvteRkxhzIf-e7kJAOjvKzADjnlejh8/view?usp=sharing**](https://drive.google.com/file/d/19TvteRkxhzIf-e7kJAOjvKzADjnlejh8/view?usp=sharing)

* **Project Code:**

[**https://drive.google.com/file/d/1DCzlON0RLXjf9RYmDRxXhCrZRne0ouzj/view?usp=sharing**](https://drive.google.com/file/d/1DCzlON0RLXjf9RYmDRxXhCrZRne0ouzj/view?usp=sharing)

* **PPT:**

[**https://docs.google.com/presentation/d/1SW-0D\_AjQHMgdqY7nCPAfeoO\_Tq7cmKl/edit?usp=sharing&ouid=114995003587537249276&rtpof=true&sd=true**](https://docs.google.com/presentation/d/1SW-0D_AjQHMgdqY7nCPAfeoO_Tq7cmKl/edit?usp=sharing&ouid=114995003587537249276&rtpof=true&sd=true)