**McDonald's Menu Nutritional Analysis – Project**

**Instruction**

McDonald's is a global fast-food chain known for its diverse menu offerings. As a data analyst, your task is to analyze the nutritional content of the menu items available at McDonald's outlets. This analysis will provide valuable insights into the calorie count and nutrition facts of various menu items.

**Objectives:**

1. Extract meaningful information from the McDonald's menu nutritional dataset.
2. Perform exploratory data analysis to understand the nutritional distribution and trends.
3. Create visualizations to present the calorie count and nutrition facts of different menu items.
4. Identify healthy and less healthy menu options based on nutritional content.

**Deliverables:**

A report (PDF) containing:

Description of data analysis approach and methodology.

Exploratory data analysis findings and insights.

Visualizations depicting nutritional information.

Source code used for data preprocessing, analysis, and visualization.

Tasks/Activities List:

Data Collection: Download the McDonald's menu nutritional dataset from this link.

Data Preprocessing:

Load and inspect the dataset.

Handle missing values and data cleaning if necessary.

Exploratory Data Analysis (EDA):

Analyze the distribution of calorie counts across menu items.

Explore the nutritional content (e.g., fat, protein, carbohydrates) of different items.

Identify trends and patterns in the dataset.

Data Visualization:

Create bar charts, histograms, and box plots to visualize calorie distribution and nutritional content.

Compare nutritional characteristics of different food categories (e.g., burgers, salads, desserts).

Nutrition-Based Insights:

Identify menu items with the highest and lowest calorie counts.

Determine the average nutritional content of popular menu categories.

Documentation and Reporting:

Summarize the findings and insights from the analysis.

Explain how the nutritional analysis could benefit McDonald's customers and the organization.

**Task/ Activities List:**

**Data Collection:**

* The data set was downloaded, observed the trends on nutrition contents.

**Data Pre-processing:**

1. Load and inspect the dataset.
2. Handle missing values and data cleaning if necessary.

* The dataset is uploaded to the Jupyter notebook, imported pandas to with the dataset, as depicted below:

Syntax:

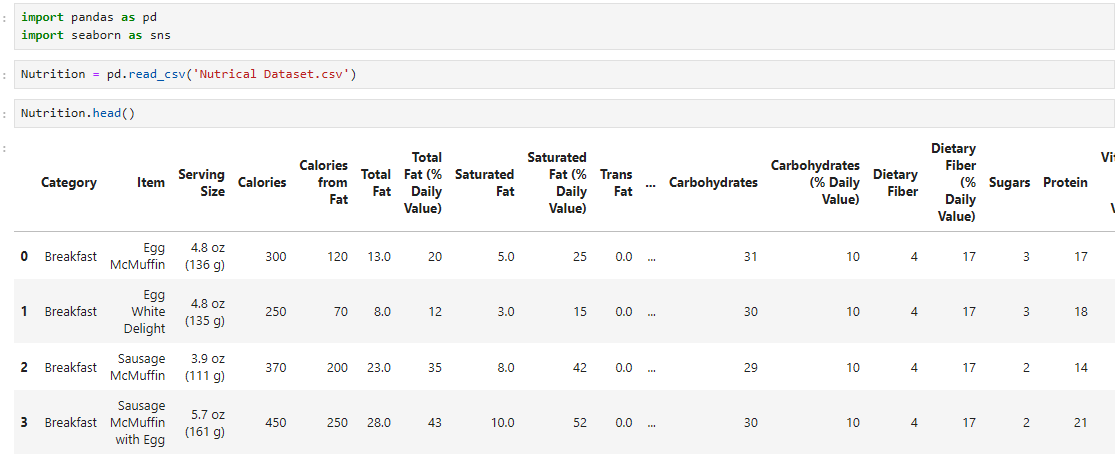
*import pandas as pd*

*import seaborn as sns*

*Nutrition = pd.read\_csv('Nutrical Dataset.csv')*

*Nutrition.shape*

*Nutrition.info()*

****

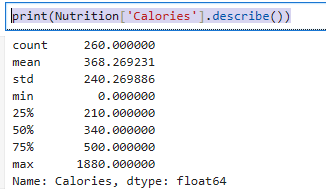
**Exploratory Data Analysis (EDA):**

1. **Analyse the distribution of calorie counts across menu items.**

* We’ve loaded the dataset, analysed the trends. Let’s understand a statistical summary for the calories in the dataset.

***Syntax:***

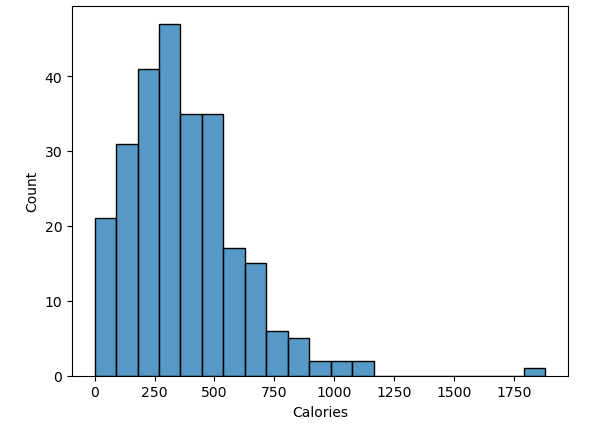
*print(Nutrition['Calories'].describe())*

**

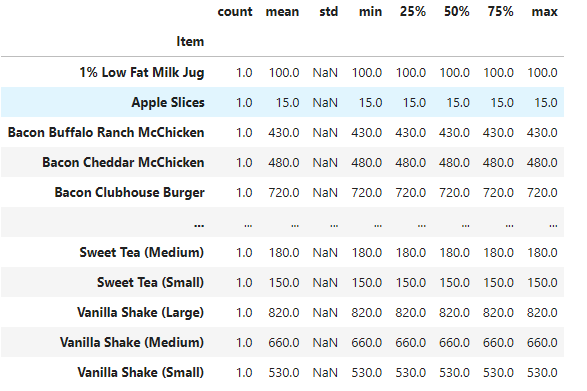
* Univariate Analysis: Analysing Calorie Distribution:
* To visualize the distribution of calories, plot a histogram.

***Syntax:***

*sns.histplot(Nutrition['Calories'])*

**

*Nutrition.groupby('Item')['Calories'].describe()*

**

* To understand the spread of calorie values, we go with Box plot:

***Syntax:***

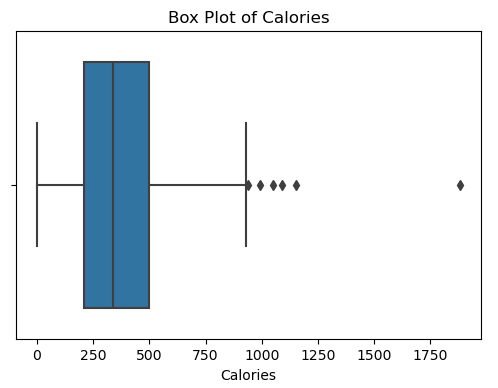
*import matplotlib.pyplot as plt*

*plt.figure(figsize=(6,4))*

*sns.boxplot(x=Nutrition['Calories'])*

*plt.title('Box Plot of Calories')*

*plt.show()*

**

* To understand the unique items in the “Items” columns, we perform the following syntax:

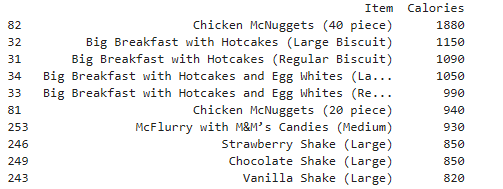
*Print(‘The unique items are: ’, Nutrition[‘Items’].unique())*

* We get almost all the items present, as the dataset consist of unique items vs calories values.
* Items, with the maximum calories:

***Syntax:***

*top\_items = Nutrition.nlargest(10, 'Calories')*

*print(top\_items[['Item', 'Calories']])*

**

* Bivariate analysis: Relationship between “Items” and “Calories”: Creating a bar plot of Items and avg. calorie count.

***Syntax:***

*plt.figure(figsize=(8,4))*

*top\_items = Nutrition.groupby('Item')['Calories'].mean().nlargest(10)*

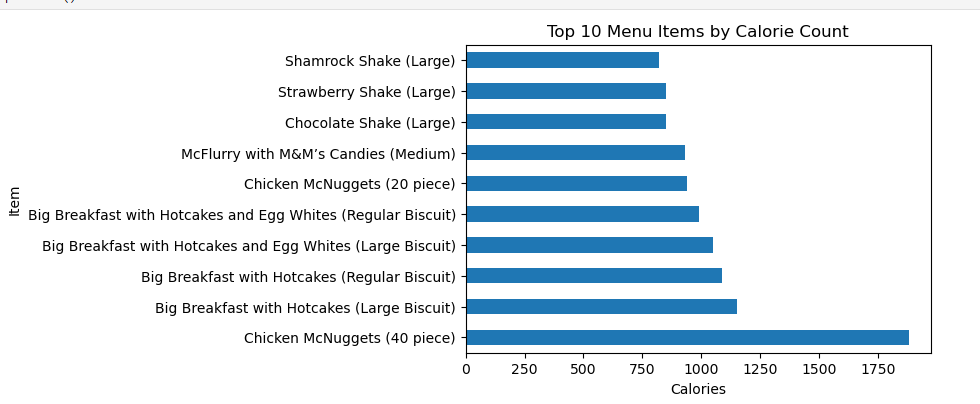
*top\_items.plot(kind='barh')*

*plt.title('Top 10 Menu Items by Calorie Count')*

*plt.xlabel('Calories')*

*plt.ylabel('Item')*

*plt.show()*



* Checking for any extremely high calories rates, so to consider Outliers. To retrieve the Z-scores we perform the following.

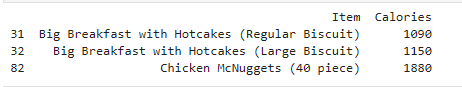
***Syntax:***

from scipy import stats

z\_scores = stats.zscore(Nutrition['Calories'])

outliers = Nutrition[abs(z\_scores) > 3] # Items with z-scores greater than 3

print(outliers[['Item', 'Calories']])

****

* **Insights:**

1. Calorie Distribution: For the given dataset, the skewness of calories is said to be 1.452 which indicates positive skewness (right skewness). Most of the calories are concentrated on the lower end. The distribution is asymmetry.
2. Skewness greater than 1 suggests a highly skewed distribution. There are few high-calorie items comparatively.

***Syntax:***

*skew\_calories = Nutrition['Calories'].skew()*

*print("Skewness of Calories:", skew\_calories)*

*O/P : Skewness of Calories: 1.4524981730922621*

1. The top menu items with the highest calories rates are Chicken McNuggests, Big breakfast with Hotcakes – regular and large biscuits.
2. The menu with the lowest calorie rate is Diet Coke (Small).

***Syntax:***

*min\_calorie\_item = Nutrition.loc[Nutrition['Calories'].idxmin(), 'Item']*

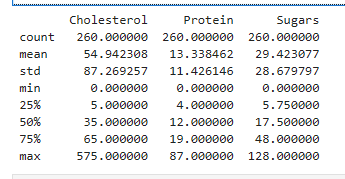
*print("Item with the lowest calorie:", min\_calorie\_item)*

**Explore the nutritional content of different items.**

* Let’s understand a statistical summary for the proteins, cholesterol and sugars in the dataset.

*Syntax:*

*print(Nutrition[['Cholesterol', 'Protein', 'Sugars']].describe())*



* Univariate Analysis: Analysing Calorie Distribution:
* To visualize the distribution of calories, plot a histogram.

*Syntax:*

*plt.figure(figsize=(4,4))*

*sns.histplot(Nutrition['Protein'])*

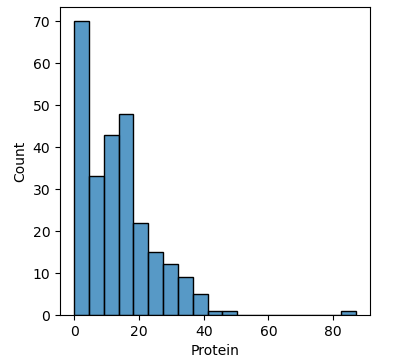
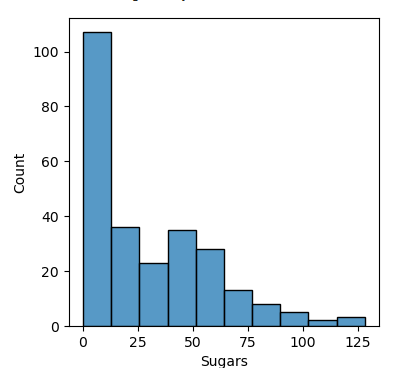
*plt.figure(figsize=(4,4))*

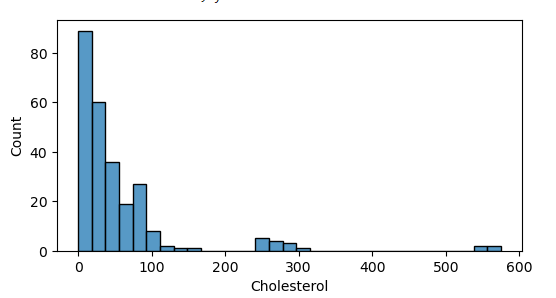
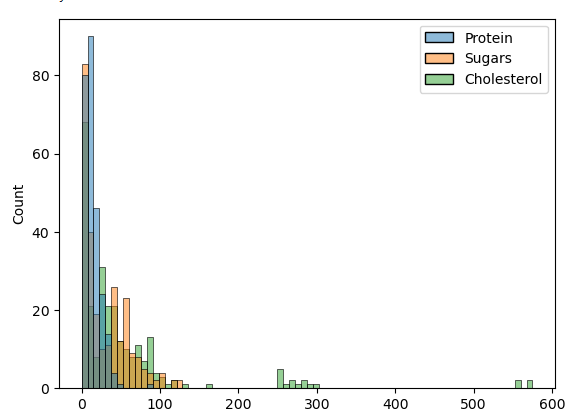
*sns.histplot(Nutrition['Sugars'])*

*plt.figure(figsize=(6,3))*

*sns.histplot(Nutrition['Cholesterol'])*

*sns.histplot(Nutrition[['Protein', 'Sugars', 'Cholesterol']])*

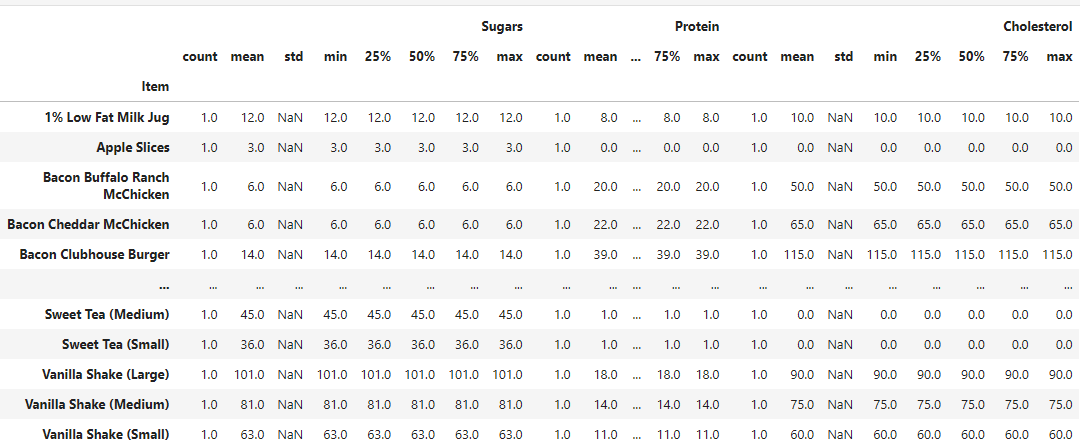
* *

* *

* Let’s describe Cholesterol, Proteins and sugar in the menu.

*Syntax*

*Nutrition.groupby('Item')[['Sugars', 'Protein', 'Cholesterol']].describe()*

**

* To understand the spread of calorie values, we go with Box plot:

***Syntax:***

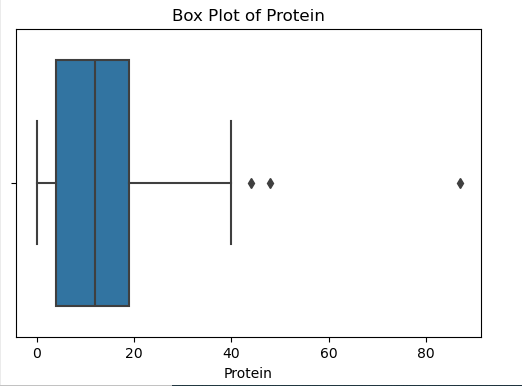
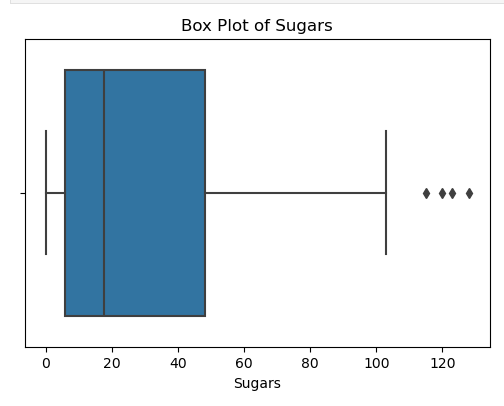
*import matplotlib.pyplot as plt*

*plt.figure(figsize=(6,4))*

*sns.boxplot(x=Nutrition['x'])*

*plt.title('Box Plot of Calories')*

*plt.show()*

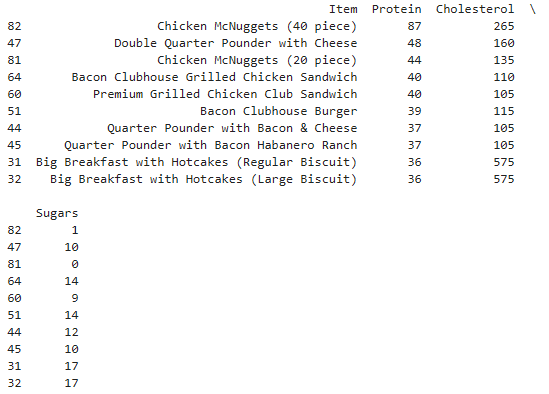
* *

* Let’s understand the top items with high Protein, Sugars and cholesterol levels:

*Syntax:*

*top\_items = Nutrition.nlargest(10, ['Protein', 'Sugars', 'Cholesterol'])*

*print(top\_items[['Item', 'Protein', 'Sugars', 'Cholesterol']])*

**

* Bivariate analysis: To understand the relationship between ‘Items’ and “Cholesterol, Proteins and Sugars”

*Syntax:*

*plt.figure(figsize=(8,6))*

*top\_items = Nutrition.groupby('Item')[['Protein', 'Sugars', 'Cholesterol']].mean().nlargest(10, 'Sugars')*

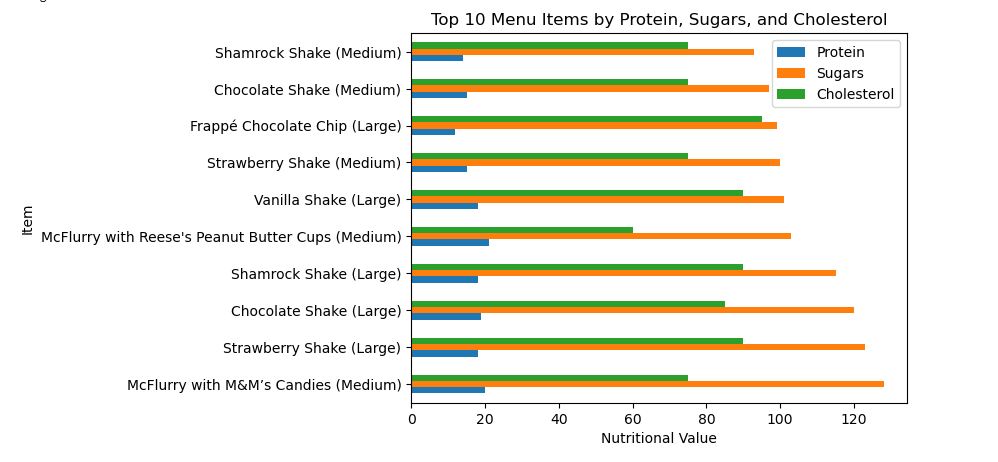
*top\_items.plot(kind='barh', stacked=False)*

*plt.title('Top 10 Menu Items by Protein, Sugars, and Cholesterol')*

*plt.xlabel('Nutritional Value')*

*plt.ylabel('Item')*

*plt.show()*

**

* Checking for any extremely high Sugar, Protein and cholesterol rates, so to consider Outliers. To retrieve the Z-scores we perform the following.
* *Items having z-score > 3:*

*Syntax:*

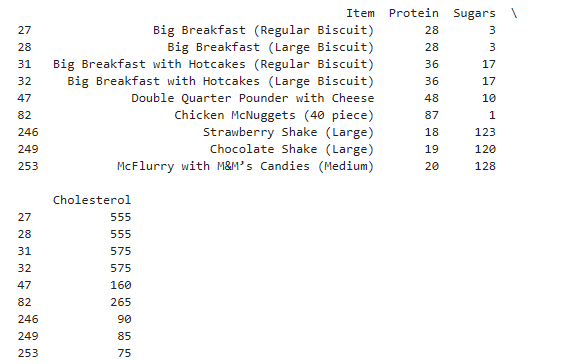
*from scipy import stats*

*columns = ['Protein', 'Sugars', 'Cholesterol']*

*z\_scores = Nutrition[columns].apply(stats.zscore)*

*outliers = Nutrition[(z\_scores.abs() > 3).any(axis=1)]*

*print(outliers[['Item', 'Protein', 'Sugars', 'Cholesterol']])*

**

* **Insights:**

1. Nutrients Distribution: For the given dataset, the skewness of Protein is 1.57, Sugars is 1.03 and for Cholesterol is 3.79.
2. The skewness follows the pattern: Sugars< Protein < Cholesterol, being highly skewed. The distribution is asymmetry.
3. There are few items with high cholesterol, making the dataset to be highly- right skewed.

***Syntax:***

*columns = ['Protein', 'Sugars', 'Cholesterol']*

*skew\_Nutrition = Nutrition[columns].skew()*

*print("The skewness for :", skew\_Nutrition)*

*O/P : The skewness for : Protein 1.579924*

*Sugars 1.031940*

*Cholesterol 3.798907*

1. The menu with the better healthy nutrients having less sugar, cholesterol and proteins are:

***Syntax:***

*columns = ['Cholesterol', 'Protein', 'Sugars']*

*min\_calorie\_item = Nutrition.loc[Nutrition[columns].idxmin(), 'Item']*

*print("Item with the better health ratio:", min\_calorie\_item)*

*O/P: Item with the better health ratio:*

*38 Hash Brown*

*101 Apple Slices*

*38 Hash Brown*

**Nutrition-Based Insights:**

1. **Identify menu items with the highest and lowest calorie counts.**

* The top menu items with the highest calories rates are Chicken McNuggests, Big breakfast with Hotcakes – regular and large biscuits.
* The menu with the lowest calorie rate is Diet Coke (Small).

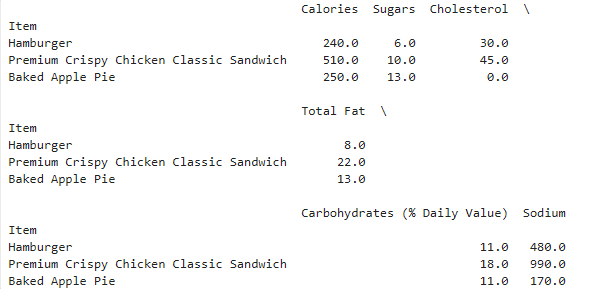
1. **Determine the average nutritional content of popular menu categories.**

***Syntax:***

*average\_nutrition = Nutrition.groupby('Item')[['Calories', 'Sugars', 'Cholesterol', 'Total Fat', 'Carbohydrates (% Daily Value)', 'Sodium']].mean()*

*popular\_items\_avg = average\_nutrition.loc[['Hamburger', 'Premium Crispy Chicken Classic Sandwich', 'Baked Apple Pie']]*

*print(popular\_items\_avg)*

**

**DOCUMENTATION:**

* From the above dataset, we found Chicken Nuggets possess high calorie counts and diet coke holds the least amount. One who watches their diet should also observe the nutrient contents. Thanks to EDA.
* Calories and Cholesterol levels are highly right-skewed, resulting that certain items possess hold higher level of such nutrients – which aren’t good for health concerns.
* One to enjoy such a meal, they can hold a better frequency on consuming meals from McDonald’s between intervals, possibly monthly once.
* If customers are following this trend, it would be a huge decline on the revenue of the restaurant. Hence, it’d be high time for the restaurant to concentrate on the nutrient values and provide better meal to their customers.
* The restaurant should primarily focus on calorie and cholesterol contents in their dishes, resulting in a healthy meal and increasing better revenues.