MT-XYZ – Probability & Statistics Assignment

Statistical Analysis of McDonald’s Menu

# Group Members:

## Zain Ammad – CS-19079

## Taimoor Jawaid – CS-19083

## Rohan – CS-19091

## Sohaib Ahmed Abbasi – CS-19096

# Initial data exploration / understanding

Getting required libraries for data analysis

import pandas as pd

import matplotlib.pyplot as plt

Reading the data set and show initial information related to it

dataset = pd.read\_csv('McDonalds Menu.csv')  # read the data set

dataset.info()  # show the information of the data set

output of last line:

A picture containing text, plaque

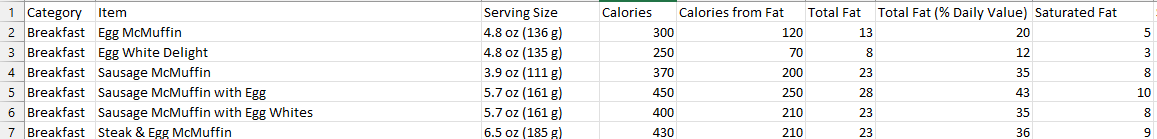
Description automatically generated

This shows all the columns of the data set, how many values are non-null (not empty) in each column (Non-Null Count), and data type of each column (Dtype) (Dtype object means text/string, int64 means integer number and float64 real number.)

From this we get the following information:

* There are 260 rows and 24 columns in the dataset.
* No column has null value (i.e no cell is empty).
* For each menu item, the dataset has its category, name of the item, size of one serving, followed by a bunch of nutritional information like calories, total fat, protein, etc.

Here is a sample of the dataset (first few rows and some columns shown):



Now let’s answer some questions related to this dataset and find out some interesting insights.

# Q1: What is the average number of calories in each category of the menu (and how do those averages compare with overall average calories of the entire menu)?

Calorie of a food item is a measure of amount of energy one gains by having that food. Let’s find out the mean number of calories in each category of McDonalds and how do those individuals averages compare to overall mean calories of the entire menu.

First lets find out the mean calories overall.

# find mean calories overall

menu\_mean = dataset['Calories'].mean()

print(f"Mean calories: {menu\_mean}")

Output:



So the average amount of calories per serving in a McDonalds meal is around 368.27 calories. That isn’t a lot. How about the most number of calories in an item? And what item is it? Let’s find that out.

most\_calorie\_item = dataset.nlargest(1, 'Calories')

print(f"Most caloric item: {most\_calorie\_item['Item'].values[0]} with {most\_calorie\_item['Calories'].values[0]} calories")

Output:



Wow there’s an item with 1880 calories per serving! That’s a lot for one person to eat in a meal! But it makes sense because that item is 40 pieces of Chicken nuggets, which is surely meant for multiple people to eat, not one.

Now let’s find out average Calories of each category

# make a new data frame with categories column and average calories for each category

categories = dataset.groupby('Category')['Calories'].mean().round(2)

categories.reset\_index(inplace=False)

Output:

A picture containing text, screen, scoreboard, screenshot

Description automatically generated

Let’s sort these values and plot them for better understanding

# sort categories by calories

categories = categories.sort\_values(ascending=False)

# plot categories as bar chart

categories.plot(kind='bar')

for index, data in enumerate(categories):

    plt.text(x=index-0.25 , y =data+2 , s=f"{round(data, 2)}" , fontdict=dict(fontsize=8))

# plot menu mean line on top of categories and add label to it

plt.axhline(y=menu\_mean, color='r', linestyle='--')

plt.text(x=5, y=menu\_mean+10, s=f"Menu average: {round(menu\_mean, 2)}", fontdict=dict(fontsize=8))

Output:

Chart, bar chart

Description automatically generated

This shows that Chicken & Fish category on average has the most calories per item and beverages is the category with the least average per item calories. Also a surprising fact is discovered that Smoothies & Shakes categories is 2nd in the list of most average calories.

# Q2: What are the least and most calorie dense (eatable) food items in each category.

Calorie density refers to the amount of calories per unit food. This is a very important fact to know specially if one is looking to loose fat as fat loss occurs due to a calorie deficit (more calories lost/used up compared to consumed). A low calorie dense food can be had in higher amount for equal or less calories than a high calorie dense food which can lead to the person eating such a food staying full and satisfied for longer and help them stay in a calorie deficit.

So let’s find out the top 3 most calorie dense food items in each category which one should try and avoid and top 3 least calorie dense food items in each category that one should look to have if having a meal at McDonald’s.

First we will separate out only eatable food items from the menu (i.e items whose categories aren’t 'Beverages', 'Coffee & Tea', or 'Smoothies & Shakes'.

eatables = dataset[~dataset['Category'].isin(['Beverages', 'Coffee & Tea', 'Smoothies & Shakes'])]

Next, we need to do some data cleaning. As it can be seen from the sample of the dataset shown before, the Serving Size column has values in ounces and grams both. Let’s clean it up by only having values of grams in the serving size column.

eatables['Serving Size'] = eatables['Serving Size'].str.split('(', expand=True)[1].str.split('g', expand=True)[0].astype(float)

eatables['Serving Size']

Output: (this is how the Serving Size column is now)

A picture containing text, meter, device

Description automatically generated

Now let’s find out calorie density of each item (we will add a new column for it). Calorie density is the amount of caloires per unit food as discussed above. So we can obtain calorie density for each food by dividing amount of calories per serving by the size of one serving, as done in the following code:

# add a new column for calorie density of each non-drink item

eatables['Calorie Density'] = eatables['Calories'] / eatables['Serving Size']

# show calorie density column

eatables['Calorie Density']

Output:

Text

Description automatically generated

Now let’s find out top 3 most calorie dense food in each category.

# show Item column for top 3 items by calorie density for each Category

most\_calorie\_dense = eatables.sort\_values(by=["Category", 'Calorie Density'], ascending=[True, False]).groupby('Category')[["Item", "Calorie Density", "Category"]].head(3)

most\_calorie\_dense

Output:

Graphical user interface, text

Description automatically generated

These show 3 items in each category that one should try and avoid if they want to consume low calorie dense foods

Next let’s find out top 3 least calorie dense food in each category

# show Item column for bottom 3 items by calorie density for each Category

least\_calorie\_dense = eatables.sort\_values(by=["Category", 'Calorie Density'], ascending=[True, True]).groupby('Category')[["Item", "Calorie Density", "Category"]].head(3)

least\_calorie\_dense

Output:

Graphical user interface, text, application

Description automatically generated

These are the foods that one should be looking to have at McDonald’s if they want to consume low calorie dense foods.

Here is a comparison of the calorie density of the items shown in the two tables above shown using a scatter plot.

plt.scatter(most\_calorie\_dense['Item'], most\_calorie\_dense['Calorie Density'], color='red')

plt.scatter(least\_calorie\_dense['Item'], least\_calorie\_dense['Calorie Density'], color='blue')

plt.xticks([])

plt.xlabel('Items')

plt.ylabel('Calorie Density')

plt.title('Scatter Plot of Items and Calorie Density')

plt.legend(['Most Calorie Dense', 'Least Calorie Dense'])

plt.show()

Output:

# Chart, scatter chart Description automatically generated

# Q3: Are grilled chicken options better than crispy chicken ones?

One may think that grilled chicken options would be a healthier choice than crispy chicken ones, but is that actually true? Let’s find out by comparing various nutritional facts for grilled and crispy chicken menu options.

First let’s separate out grilled and crispy chicken options

# seperate out grilled chicken and crispy chicken items

grilled\_chicken\_items = dataset[dataset['Item'].str.contains('Grilled')]

crispy\_chicken\_items = dataset[dataset['Item'].str.contains('Crispy')]

Let’s see a sample of both of these sub-datasets

grilled\_chicken\_items.sample(5)["Item"]

Output:

Text

Description automatically generated

crispy\_chicken\_items.sample(5)["Item"]

Output:

Text

Description automatically generated

Let us focus our comparison of the two types of chicken items on the following 2 categories of nutritional values:

1. Good nutritional values (that we want more of): Dietary Fiber and Protein
2. Bad nutritional values (that we want less of): Total Fat, Calories, Saturated Fat, Cholesterol, and Sugars

First lets see which chicken type has less of the bad nutritional values.

plt.boxplot([grilled\_chicken\_items[["Total Fat", "Calories", "Saturated Fat", "Cholesterol", "Sugars"]].sum(axis=1),

            crispy\_chicken\_items[["Total Fat", "Calories", "Saturated Fat", "Cholesterol", "Sugars"]].sum(axis=1)])

plt.xticks([1, 2], ['Grilled Chicken', 'Crispy Chicken'])

plt.title(f'Sum of Total Fat, Calories, Saturated Fat, Cholestrol, and Sugars of grilled and crispy chicken items')

plt.show()

Output:

Chart, box and whisker chart

Description automatically generated

This shows that most grilled chicken items have less of the bad nutrition values so most of them are better, although some of them have more of the bad nutritional values than some crispy chicken items (as shown by the maximum value of grilled chicken being higher than minimum of crispy chicken). On average though grilled chicken items are better.

Next let’s see how the two types compare with the amount of good nutritional values

plt.boxplot([grilled\_chicken\_items[["Dietary Fiber", "Protein"]].sum(axis=1),

            crispy\_chicken\_items[["Dietary Fiber", "Protein"]].sum(axis=1)])

plt.xticks([1, 2], ['Grilled Chicken', 'Crispy Chicken'])

plt.title(f'Sum of Dietary Fiber and Protein  of grilled and crispy chicken items')

plt.show()

Output:

Chart, box and whisker chart

Description automatically generated

Grilled chicken wins here again as most items in that category have more of the good nutritional values (although not all) and average is also better.

This analysis shows that indeed, for most cases grilled chicken items are better than crispy chicken ones but one must still choose carefully as some grilled chicken items are worse than some crispy chicken items.