**Project Title:Snack Analysis**

**Nutritional Snack Analysis Report**

**Tools Used**:PySpark,Pandas,Excel,Tabluea

**Data Sources:**

Starbucks-menu-nutrition-drinks.xlsx (Starbucks)

Starbucks-menu-nutrition-food.xlsx (Starbucks)

starbucks\_drinkMenu\_expanded.xlsx (Starbucks)

menu.xlsx (McDonald's)

**1.Introduction**

This project aims is to evaluate and compare snack items from **Starbucks** and **McDonald's** to determine which snacks is healthier based on their nutritional value.This analysis uses PySpark to handle and process large datasets also It followes World Health Organization (WHO) guidelines for evaluating healthier food choice.

**2.Data Extraction**

* Below Data is Extracted from Git repository:

Starbucks Drink Nutrition Data(*starbucks-menu-nutrition-drinks.xlsx*)

Starbucks FoodNutritional Data(*starbucks-menu-nutrition-food.xlsx*)

Starbucks Expanded Drink Menu(*starbucks\_drinkMenu\_expanded.xlsx*)

McDonald’s Menu (*menu.xlsx*)

All datasets were read using *pandas.read\_excel()* and then converted to Spark DataFrame.

**3.Data Cleaning and Transformation**

**1.Type Conversion**:

Some columns had mixed data types (e.g., Total Fat (g) and Caffeine (mg)) in (*starbucks\_drinkMenu\_expanded.xlsx)* excel file.

They were converted to numeric using: pd.to\_numeric()

**2.Column Selection & Renaming:**

From each dataset, only relevant nutritional columns were selected:like Calories, Fat (g), Carbs (g), Fiber (g), Protein, Sodium (mg),Sugar.  
Food Name was kept same across all datasets using aliasing.

**3.Cleaned and Merged Datasets:**

Added only those column which are common in all files so that analysis result will be fair.Cleaned some columns as spark need proper names of column to merge data and some column has list of products as unnamed column so aliase them as Food Name.

s1:cleaned Starbucks Drinks

s2:cleaned Starbucks Food

s3:cleaned and renamed Starbucks Extended DrinksMenu files column

Now,Each dataset has Food Name, Calories, Fat (g), Carbs (g), Fiber (g), Protein (g) columns with standardized names.

Finally Merged all dataset using:  
 tdf = s1.unionByName(s2, allowMissingColumns=True)

cdf=tdf.unionByName(s3,allowMissingColumns=True)

**4.Added Source Column**:

Added Source Column to find out the particular product belongs to which shop.

**5. Integrated Starbucks and McDonald’s Items** **and Added Category Column:**

Integrated Starbucks and McDonald’s Items and Categorized Source items as Drink or Food by adding Category column.

**6.**Finally **Calculated the average of Nutriants Values** to compare which food chain offers Healtheir snacks  
  
Using:

from pyspark.sql.functions import avg

healthy\_nutritional\_summary = combined\_df.groupBy("Source", "Category").agg(

avg("Calories").alias("Average Calories"),

avg("Fat (g)").alias("Average Fat (g)"),

avg("Carbs (g)").alias("Average Carbs (g)"),

avg("Fiber (g)").alias("Average Fiber (g)"),

avg("Protein (g)").alias("Average Protein (g)"),

avg("Sodium (mg)").alias("Average Sodium (mg)"),

avg("Sugars (g)").alias("Average Sugars (g)")

)

healthy\_nutritional\_summary.show()

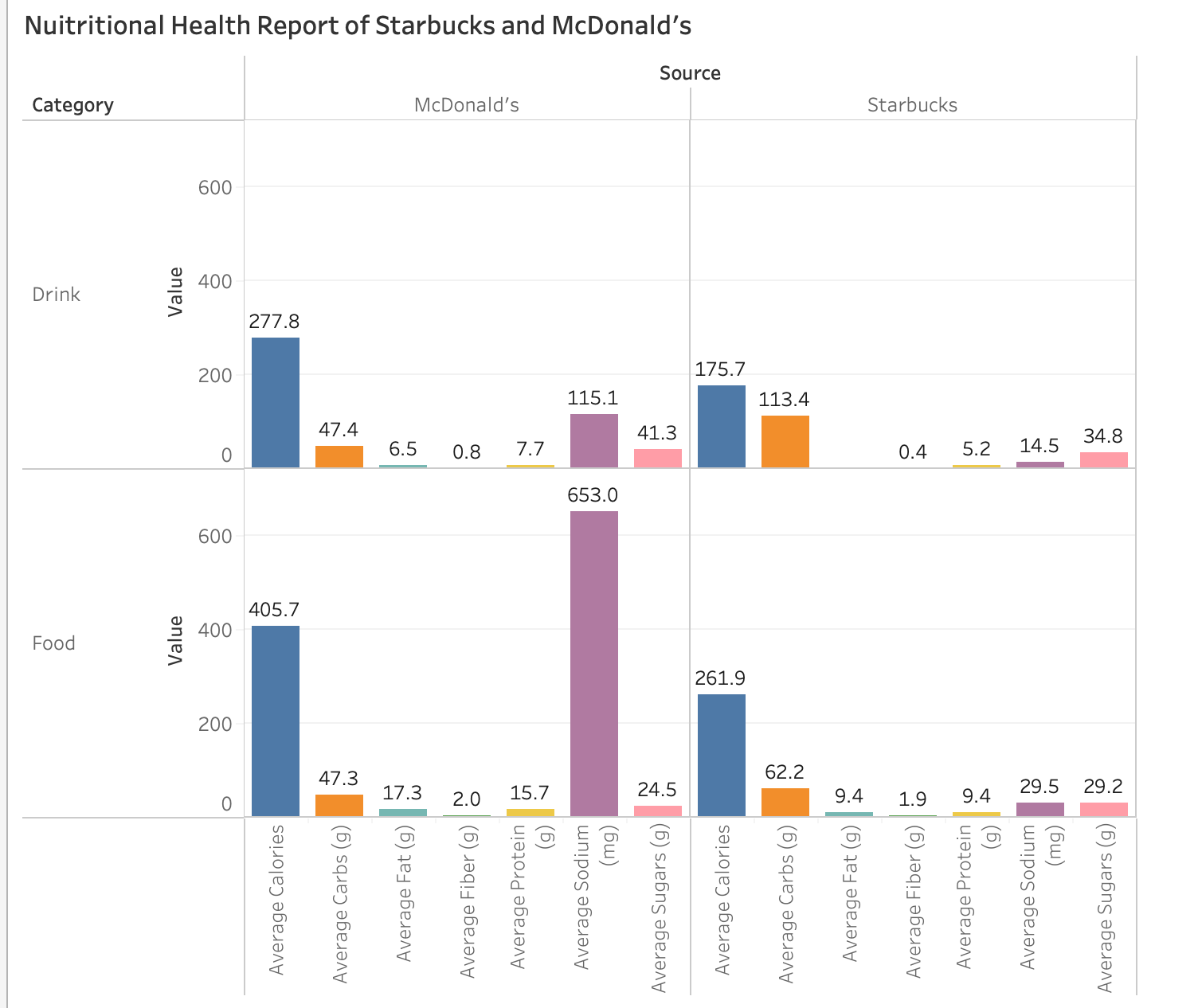
**Final Output :**

**Average Nutritional Values by Source and Category**

| **Source** | **Category** | **Avg. Calories** | **Avg. Fat (g)** | **Avg. Carbs (g)** | **Avg. Fiber (g)** | **Avg. Protein (g)** | **Avg. Sodium (mg)** | **Avg. Sugars (g)** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Starbucks | Drink | 175.65 | NaN | 113.43 | 0.42 | 5.23 | 14.52 | 34.79 |
| Starbucks | Food | 261.92 | 9.39 | 62.21 | 1.93 | 9.43 | 29.48 | 29.19 |
| McDonald's | Food | 405.65 | 17.32 | 47.31 | 1.96 | 15.67 | 652.96 | 24.53 |
| McDonald's | Drink | 277.76 | 6.52 | 47.43 | 0.83 | 7.70 | 115.13 | 41.26 |

7.This Result is saved in healthy\_nutritional\_summary.csv file.

**8.** **Visual representaiotn of Starbucks and McDonald’s Nutrientional Values:**

****

**Visual Analysis Summary:**

**Alignment with WHO Guidelines**

WHO Guidelines for Healthy Snacks:

* Low in fat, sugars, and sodium
* Moderate in calories
* Rich in fiber and protein

**Drinks:**

| Brand | Calories | Carbs (g) | Fat (g) | Fiber (g) | Protein (g) | Sodium (mg) | Sugars (g) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| McDonald’s | 277.8 | 47.4 | 6.5 | 0.8 | 7.7 | 115.1 | 41.3 |
| Starbucks | 175.7 | 113.4 | 0.4 | 5.2 | 14.5 | 34.8 | 29.2 |

#### 

**Food:**

| Brand | Calories | Carbs (g) | Fat (g) | Fiber (g) | Protein (g) | Sodium (mg) | Sugars (g) |
| --- | --- | --- | --- | --- | --- | --- | --- |
| McDonald’s | 405.7 | 47.3 | 17.3 | 2.0 | 15.7 | 653.0 | 24.5 |
| Starbucks | 261.9 | 62.2 | 9.4 | 1.9 | 9.4 | 29.5 | 29.2 |

**Best performer overall: Starbucks Food**

* Sodium (29.5 mg) is extremely low compared to McDonald’s value(653 mg)
* Fat (9.4g) is low compared to McDonald’s value(17.3 g)
* However, Sugar (29.2g) is slightly above the mcDonald’s(24.5g).
* Fiber (1.9g) is still below WHO’s healthy minimum of 3g.

**Least healthy option: McDonald’s Food**

* Very high fat (17.3g) and sodium (653mg).
* Fiber (2.0g) and sugar (24.5g) both fall short of WHO’s health goals.

**Drink Comparison:**

* Starbucks drinks have much lower sodium and fat, but extremely high carbs (113.4g) and sugars (29.2g) due to sweetened beverages.
* McDonald’s drinks have slightly lower sugar (41.3g) than Starbucks but still exceed WHO’s 5g sugar limit by a large margin.

**Final Healthiness Ranking (Best to Worst):**

1. Starbucks Food – lowest sodium and fat, moderate overall.
2. McDonald’s Drink – moderate fat and sodium, but sugar still high.
3. Starbucks Drink – very low sodium and fat, but sugar and carbs are a concern.
4. McDonald’s Food – fails almost all health thresholds.

**Conclusion: Starbucks is the Healthier Choice**

* **Drinks**: Starbucks drinks are significantly healthier across nearly all categories: they contain fewer calories, less fat, lower sodium, and more fiber and protein than McDonald’s drinks.
* **Food**: Starbucks food is lower in calories, fat, and sodium—making it a better option for those managing heart health and weight. However, McDonald’s food items offer more protein and slightly more fiber
* **Starbucks** drinks and food generally align better with WHO recommendations, especially due to their **lower sodium and fat content**, and **higher fiber in drinks**.
* **McDonald’s** has higher **sodium, fat**, and **calories**, particularly in food items, which may contribute to health risks if consumed frequently.

While both brands serve items that may not qualify as “ideal” healthy snacks by WHO standards, **Starbucks offers relatively more balanced and lighter options**, making it the **healthier overall choice** in this comparison.