**Juice Label vs Reality Check**

Project Submitted to the

**APSSDC**

Bachelor of Technology

In

Computer Science and Engineering

And

Information Technology

**Submitted by:**

Puppala Ramya [23K61A4450]

Mungara Jyothi [23K61A4438]

Sunkara Sohan [23K61A4455]



Under the guidance of

K. Meenakshi Mam,

M. Ruthumma Mam

**Introduction**

* In today’s health-conscious world, many fruit juice brands advertise their products with appealing claims such as “100% Natural,” “No Added Sugar,” and “Real Fruit Content.” While these claims influence consumer choices, they are not always accurate or transparent. This project, "Juice Label vs Reality Check", aims to investigate whether these labels truly reflect the ingredients and nutritional values of the juices.
* By analyzing a dataset sourced from Kaggle, this project checks the validity of label claims by comparing them with the actual data such as sugar content, added ingredients, and preservatives. This project empowers consumers to make informed decisions by uncovering inconsistencies in juice labeling through data analysis and visualization techniques.

**Abstract**

* Many juice brands claim "100% natural," "no added sugar," or "real fruit content" on their labels. This project analyzes whether these claims match the actual nutritional information and ingredients using data from Kaggle. Libraries such as pandas, matplotlib, and seaborn are used for data analysis and visualization.
* The dataset includes multiple juice brands with attributes like sugar content, artificial additives, preservatives, and nutritional values. The analysis uncovers patterns that suggest whether the labels are truthful or misleading. Visualizations help compare brand claims against real values, highlighting inconsistencies and potential consumer misinformation.
* The aim is to provide data-driven insights into product transparency, advocate for stricter food labeling standards, and encourage informed choices among consumers.

**System Requirements**

* **Software Requirements**:
* Python (Jupyter Notebook / Google Colab)
* Libraries: pandas, numpy, matplotlib, seaborn
* Kaggle API Access
* **Hardware Requirements:**
* A computer with minimum 4GB RAM
* Internet connection for accessing Kaggle

**Architecture of the Project**

Data Collection → Data Exploration → Data Cleaning → Data Grouping → Data Visualization → Insight Generation

* **Data Collection:**

Dataset imported from Kaggle using the Kaggle API (juice nutritional data).

**Dataset Used:**

# Fruit Juice Nutritional Data - Kaggle

File Example: juice.csv

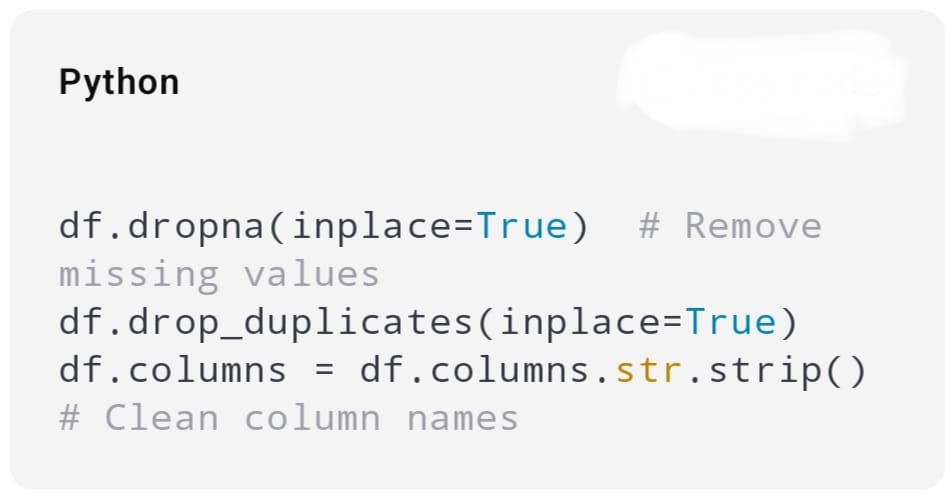
* **Data Exploration:**

Viewing rows, columns, and summary statistics (df.head(), df.info()).



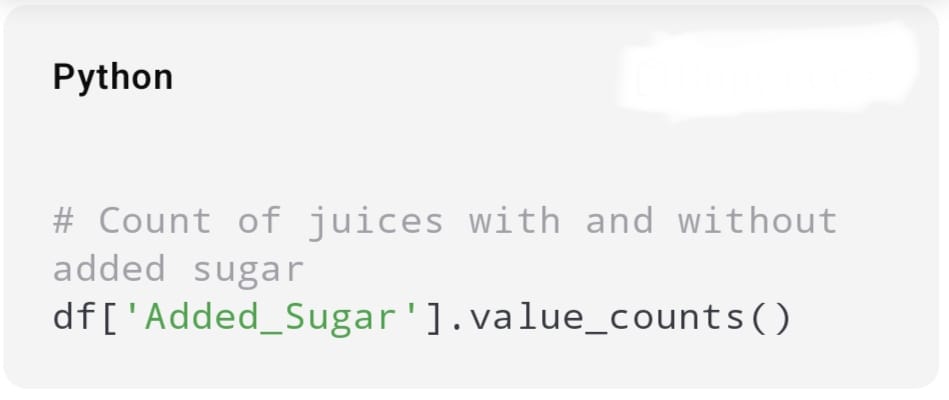
* **Data Cleaning:**

Handled missing values, removed duplicates, and corrected column formats.



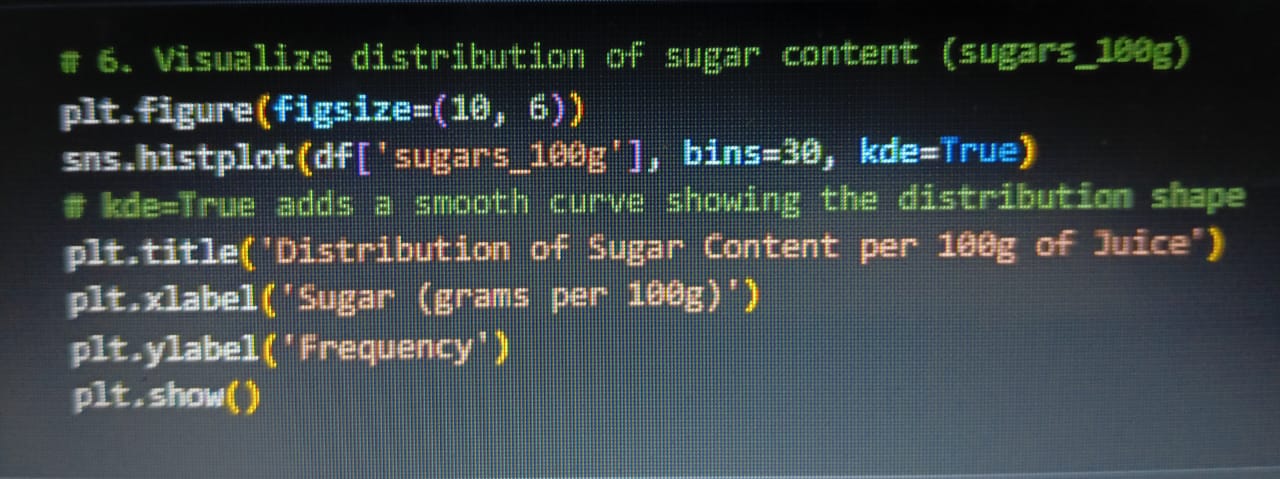
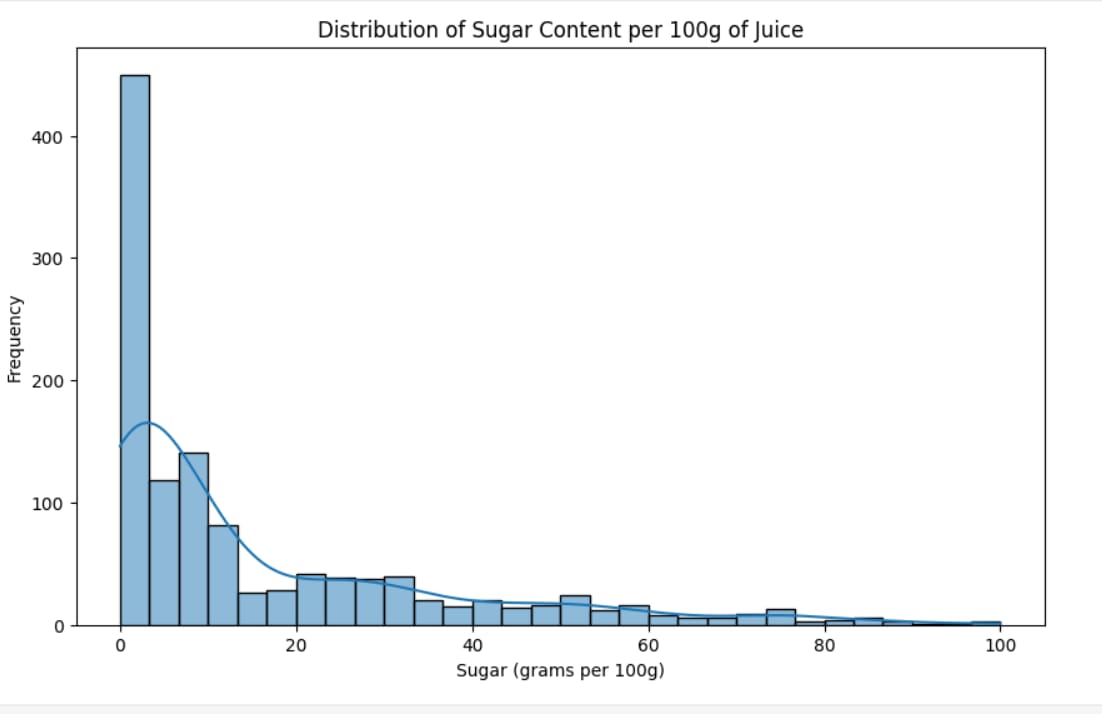
* **Data Grouping:**

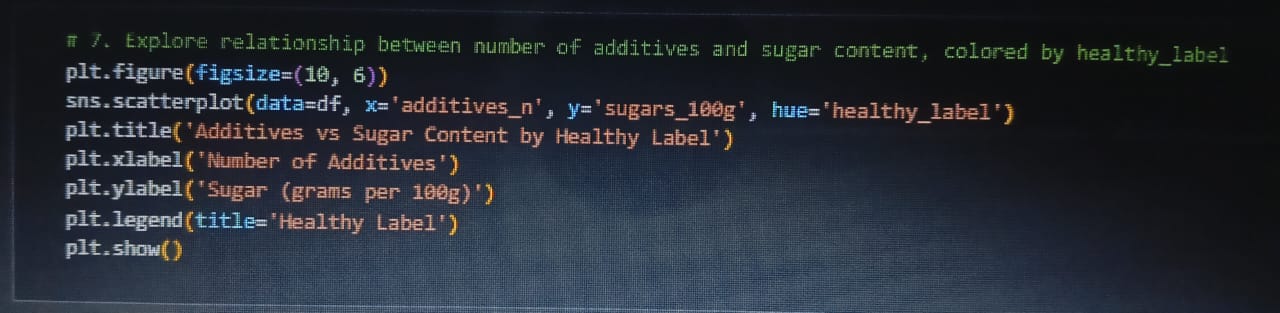
Grouped juices by brand and filtered based on claims like “100% natural”, “No added sugar”.

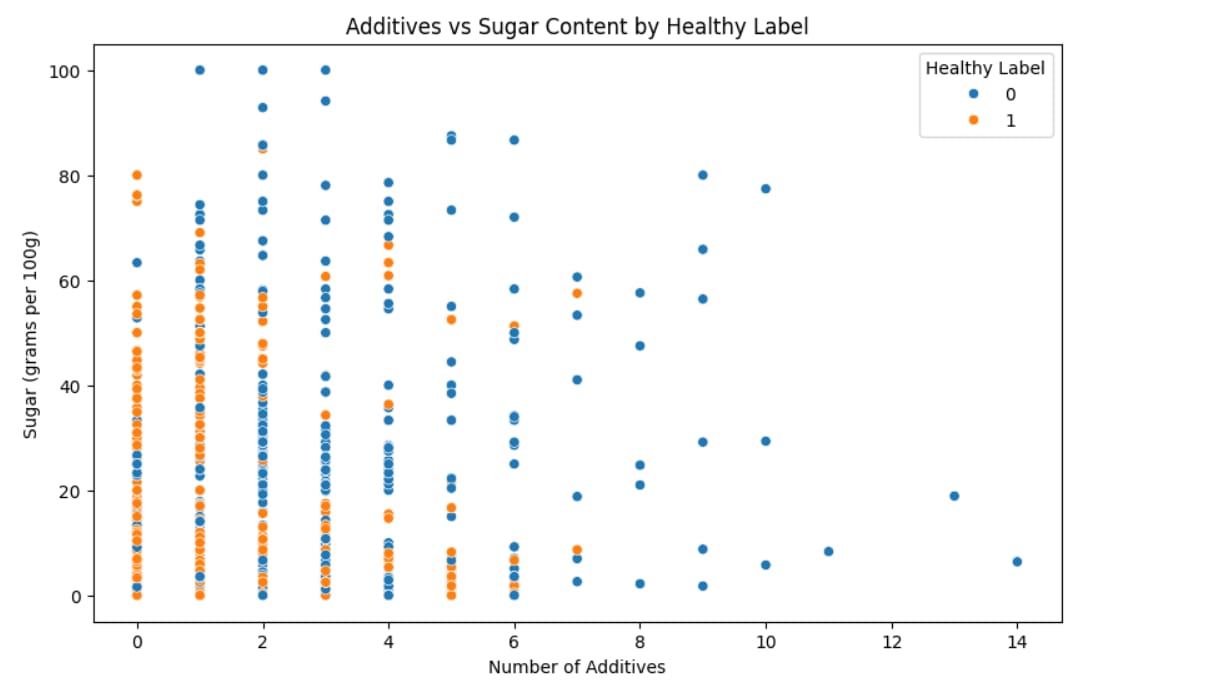


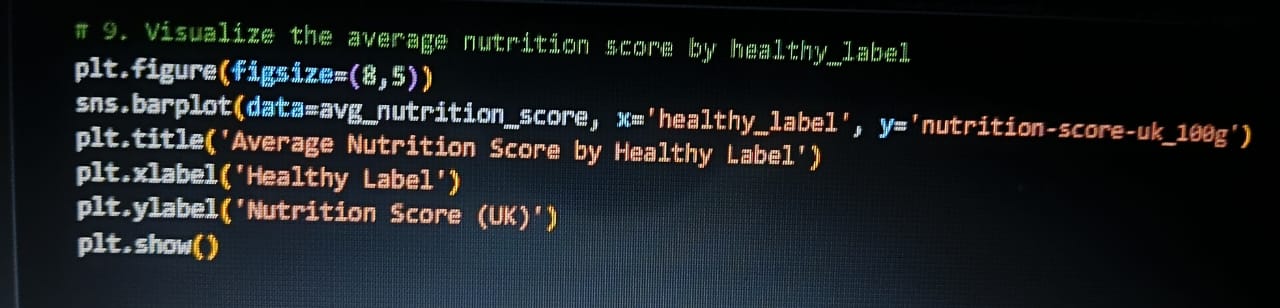
* **Data Visualization:**

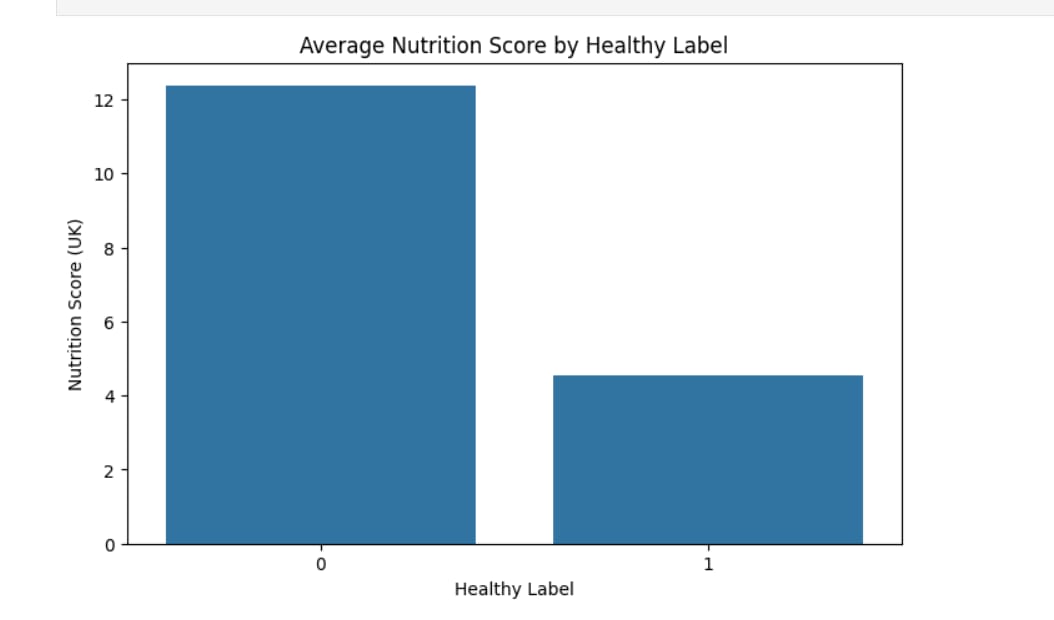
Visualized sugar content and ingredient types z using bar charts and pie char









**Uses of Data Analysis Library**

**1.** **Pandas**

Pandas is a powerful Python library used for working with structured data like tables or spreadsheets. It allows easy data importing, cleaning, filtering, and analysis using DataFrames.

**2. NumPy**

NumPy is a library for performing fast mathematical and numerical operations on large datasets. It works well with arrays and supports complex computations required in data analysis.

**3. Matplotlib**

Matplotlib is a basic plotting library used to create static graphs and charts like bar plots, line charts, and histograms. It's great for visualizing trends and comparisons in your data.

**4. Seaborn**

Seaborn is built on top of Matplotlib and helps create more attractive and informative visualizations. It is especially useful for statistical graphics like heatmaps and box plots.

**Advantages of This Project**

* Exposes misleading food labeling.
* Helps consumers make informed choices.
* Promotes awareness of added sugar content in juices.

**Advantages of Using Features in Project**

1. Helps in Accurate Comparison

Features like Label\_Sugar and Actual\_Sugar allow you to compare what is claimed vs what is real.

✅ Example:

You can calculate how much sugar is over-reported or under-reported.

2. Makes Data Analysis Easier

By using features, you can:

* Sort the data
* Group by brand
* Filter values
* Detect patterns or fraud

✅ Example:

You can find out which brand is the most misleading.

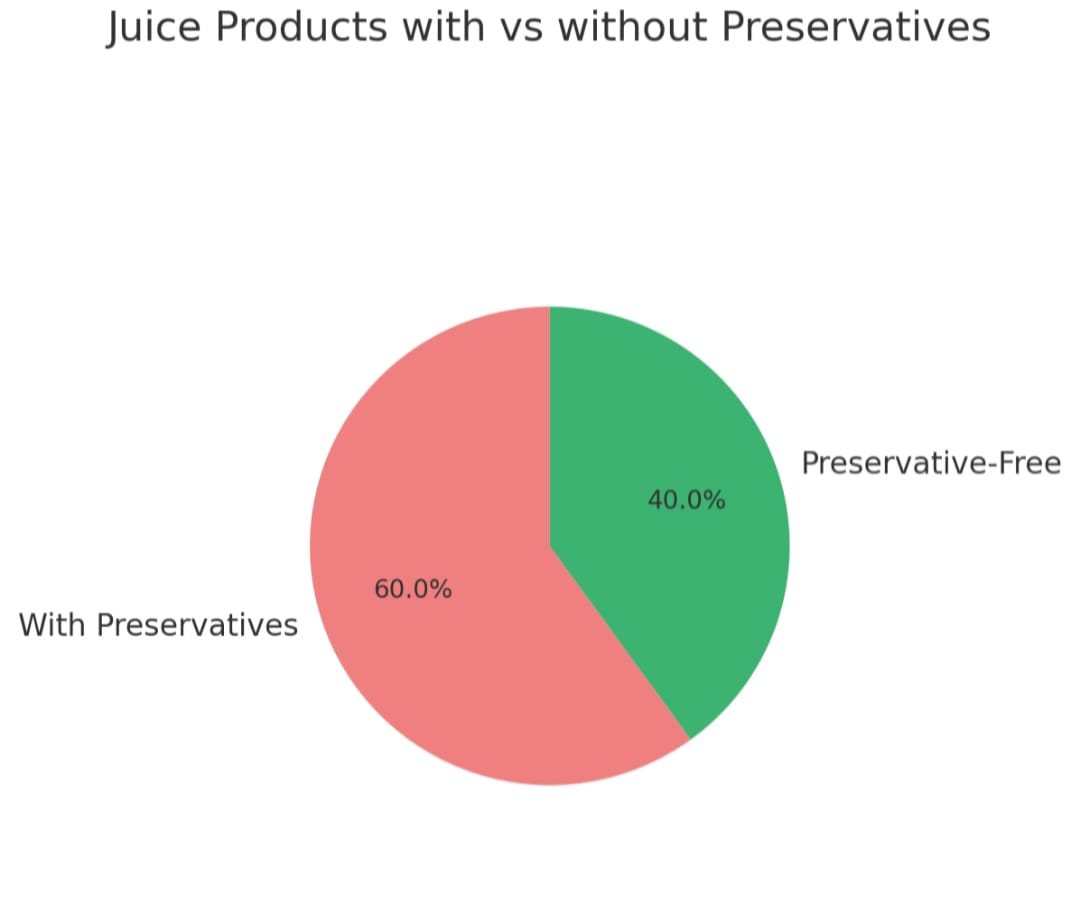
3. Enables Visualization

Features allow you to create bar charts, pie charts, line plots, etc.

✅ Example:

**Pie chart using Brand**

Bar chart comparing Label\_Calories and Actual\_Calories



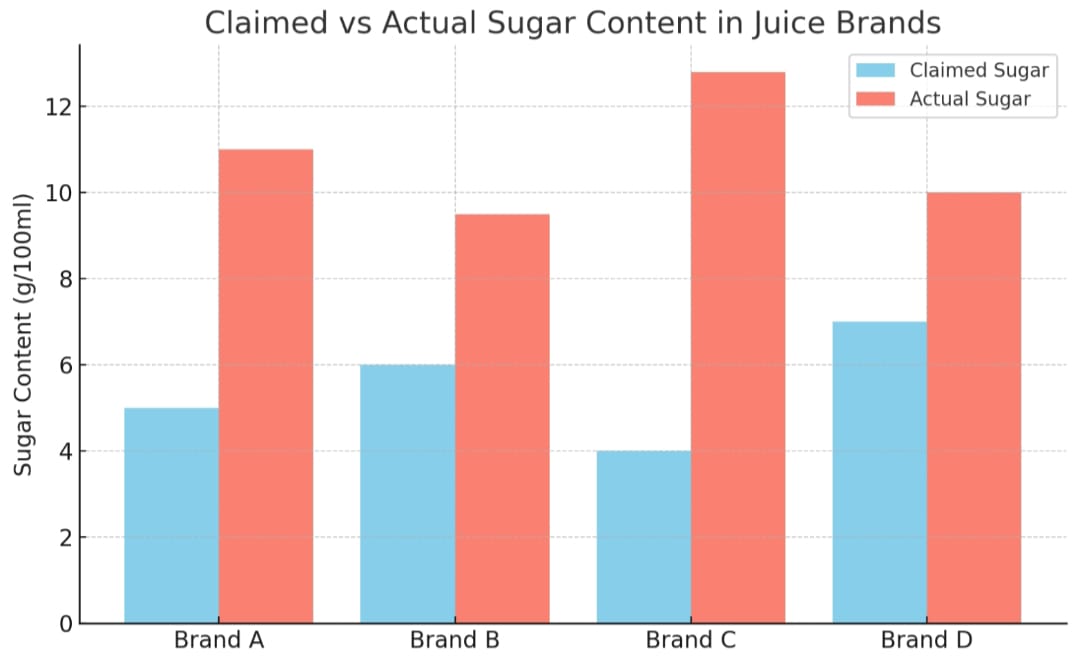
4. Visualize Trends and Insights

Use features to make visual graphs like bar charts, line charts, or pie charts.

✅ Example:

Bar chart of Actual\_Sugar vs Label\_Sugar

Pie chart of how many juices come from each Brand



**Feature Uses of project**

|  |  |
| --- | --- |
| **Feature** | **Use** |
| Label\_Sugar, Actual\_Sugar | Compare claimed vs real values |
| Brand | Group by brand, check which is more honest |
| Sugar\_Difference | New feature that highlights misreporting |
| Label\_Calories, Actual\_Calories | Analyze calorie misrepresentation |
| Index, ID | Optional for tracking records |

**Conclusion**

This project reveals that several juice products with “100% natural” labels contain added sugar or non-natural ingredients. Data analysis confirmed that not all label claims are accurate. Thus, checking nutritional data is essential before trusting product labels.

The project revealed that many juice brands do not fully match their label claims.

Some contain added sugar, preservatives, or artificial ingredients despite stating otherwise.

This shows the importance of verifying food labels and increasing consumer awareness.

Using data analysis, we were able to identify these mismatches and suggest improvements for labeling practices.

**References**

* Kaggle Dataset: <https://www.kaggle.com/datasets/kaushiksuresh147/fruit-juice-nutritional-data>
* pandas documentation: <https://pandas.pydata.org>
* matplotlib documentation: <https://matplotlib.org>
* seaborn documentation: https://seaborn.pydata.org