Fast Food Product's Nutrition Value

How different nutritional value elements interact

Ionescu Ariana-Georgiana

The problem

Dataset

The dataset contains products from 6 major fast food chains and nutritional values for each.

Analysis

What insights can we obtain about each restaurant chain and its products?

Predictions

Would it be possible, with the clean data, to make predictions about which fast food chain products are from?

Taking a closer look at the initial dataset

Company	Category	Product	Per Serve Size	Numerical columns
The name of one out of the 6 companies	The product category	The product name	The size of one serving	Nutrition Value features such as: protein, fiber, total fat.

Number of products from each Fast Food restaurant
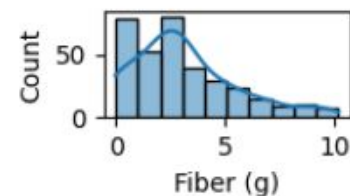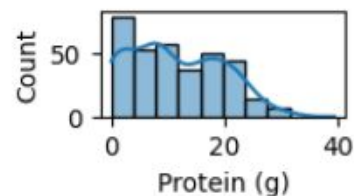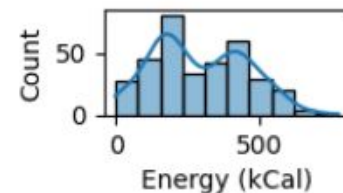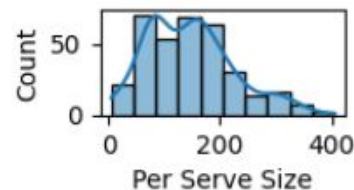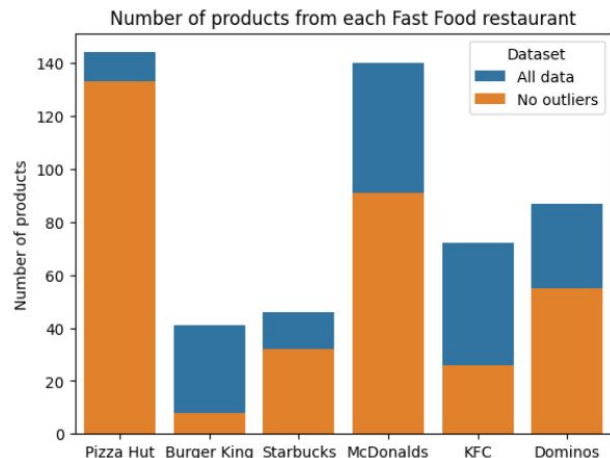
Data preprocessing

Changing column types

The Per Serve Size column has grams and milliliters, we convert it to just numerical data.
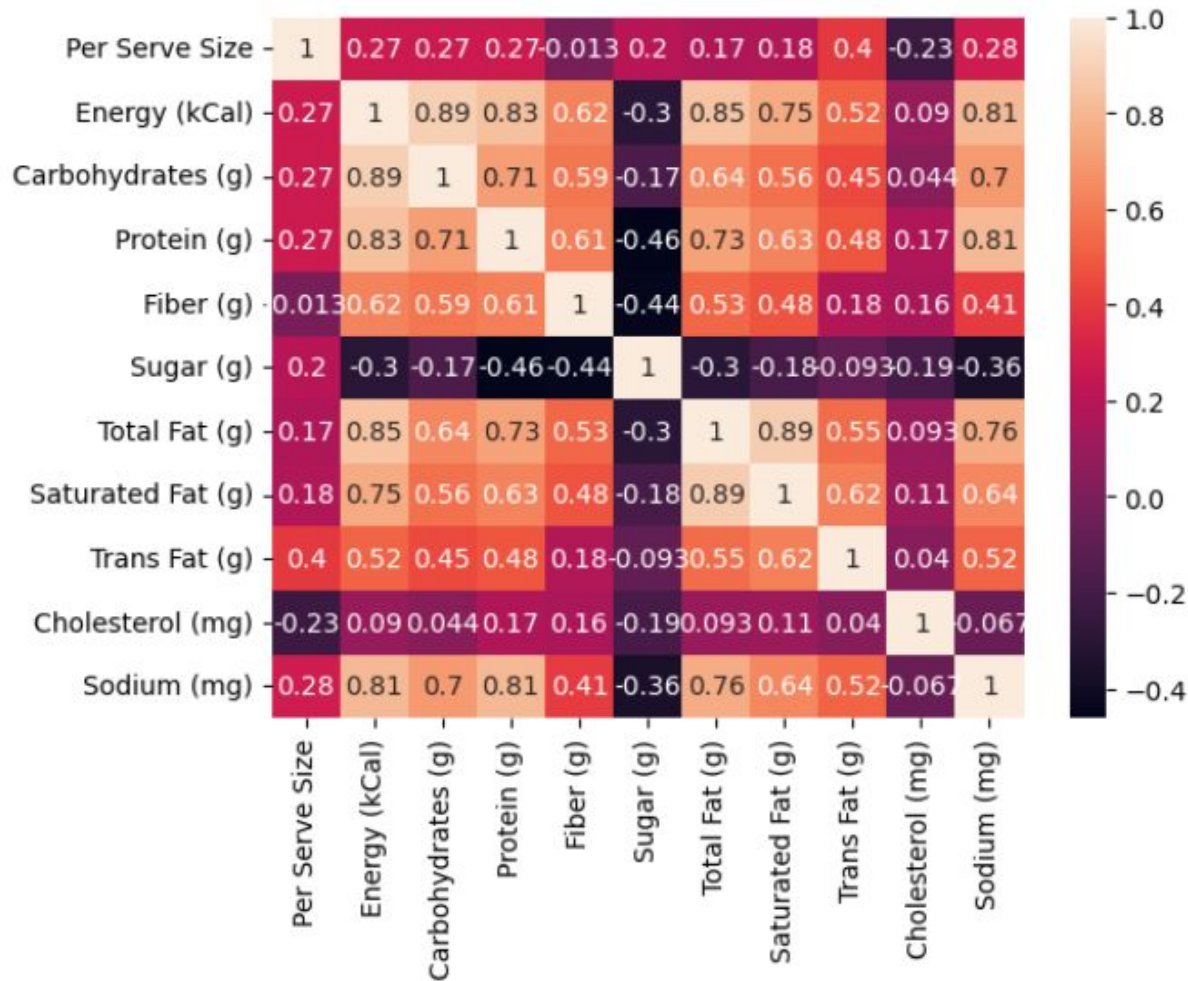
Replacing null values

Multiple columns have null values that we replace with the value 0, to not remove rows.

Removing outliers

During a simple distribution visualization we can see that there are many outliers for each column. We need a database without them for classifying.

Correlation between different features

After cleaning the data and removing the outliers it would be be interesting to look at the correlation between the features to see in greater detail how they interact with each other.


|	Per Serve Size	Energy (kCal)	Carbohydrates (g)	Protein (g)	Fiber (g)	Sugar (g)	Total Fat (g)	Saturated Fat (g)	Trans Fat (g)	Cholesterol (mg)	Sodium (mg)
Per Serve Size	1	0.27	0.27	0.27	-0.013	0.2	0.17	0.18	0.4	-0.23	0.28
Energy (kCal)	0.27	1	0.89	0.83	0.62	-0.3	0.85	0.75	0.52	0.09	0.81
Carbohydrates (g)	0.27	0.89	1	0.71	0.59	-0.17	0.64	0.56	0.45	0.044	0.7
Protein (g)	0.27	0.83	0.71	1	0.61	-0.46	0.73	0.63	0.48	0.17	0.81
Fiber (g)	-0.013	0.62	0.59	0.61	1	-0.44	0.53	0.48	0.18	0.16	0.41
Sugar (g)	0.2	-0.3	-0.17	-0.46	-0.44	1	-0.3	-0.18	-0.0093	-0.19	-0.36
Total Fat (g)	0.17	0.85	0.64	0.73	0.53	-0.3	1	0.89	0.55	0.093	0.76
Saturated Fat (g)	0.18	0.75	0.56	0.63	0.48	-0.18	0.89	1	0.62	0.11	0.64
Trans Fat (g)	0.4	0.52	0.45	0.48	0.18	-0.0093	0.55	0.62	1	0.04	0.52
Cholesterol (mg)	-0.23	0.09	0.0044	0.17	0.16	-0.19	0.093	0.11	0.04	1	0.067
Sodium (mg)	0.28	0.81	0.7	0.81	0.41	-0.36	0.76	0.64	0.52	-0.067	1

Nutri-Score

We introduce a new dimension to help us analyze the data further

The Nutri-Score is a color-coded label (A to E) that rates the nutritional quality of food, with A (green) being the healthiest and E (red) the least healthy.

It takes into account different nutrition value content: protein, fiber, saturated fats, sugar, salt, calorie density and if the product is solid or a beverage.

Distribution of Nutri-Scores

Nutritional Value Labels for each Company
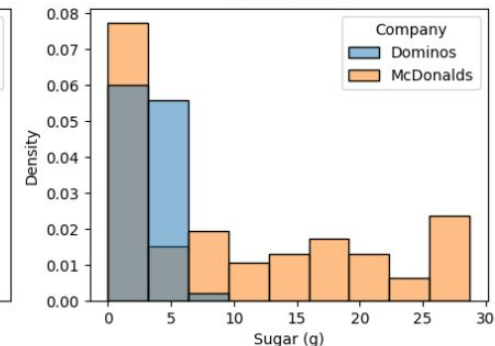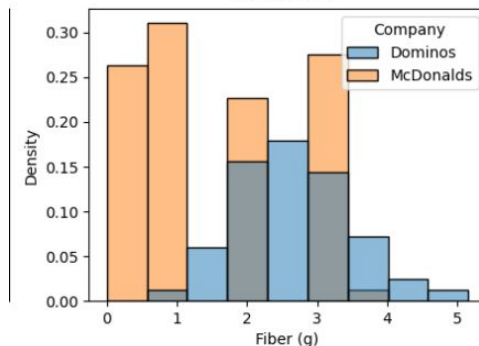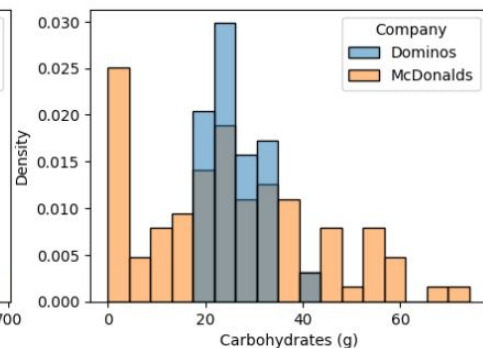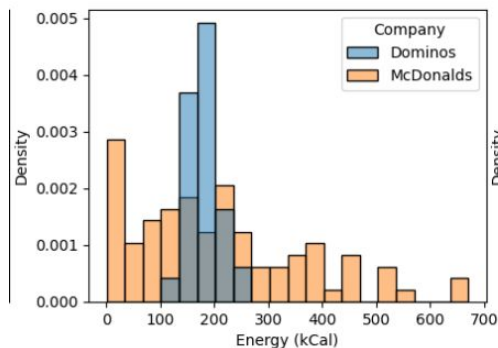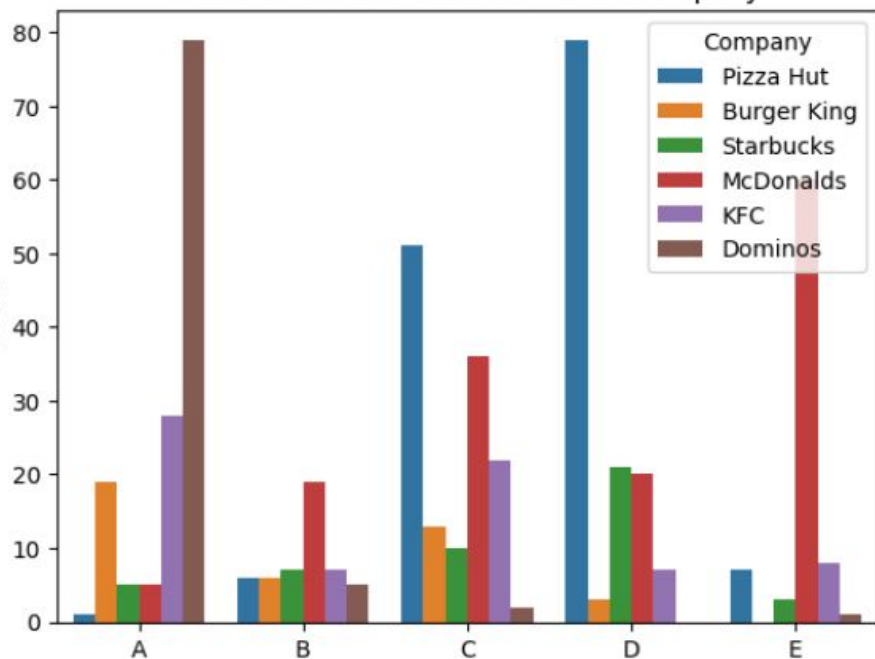
Making company predictions based on the nutrition value features

Do these 6 fast food restaurants have different enough nutrition values so that it would be possible to predict which company a product is from based on the features?

Using a Random Forest Classification algorithm, only on the numerical data, we are able to predict the company with 96% accuracy.

Further Analysis

From the plot about Nutri-Score we were able to deduce that McDonald's does not have the healthiest options available. Using data about eating habits and obesity percentage across the US and the number of McDonald's locations in each state could give us further insight into obesity causes.

We can see some relevance, but not very high.

Results of the Health Questionary

Number of McDonalds locations across the US
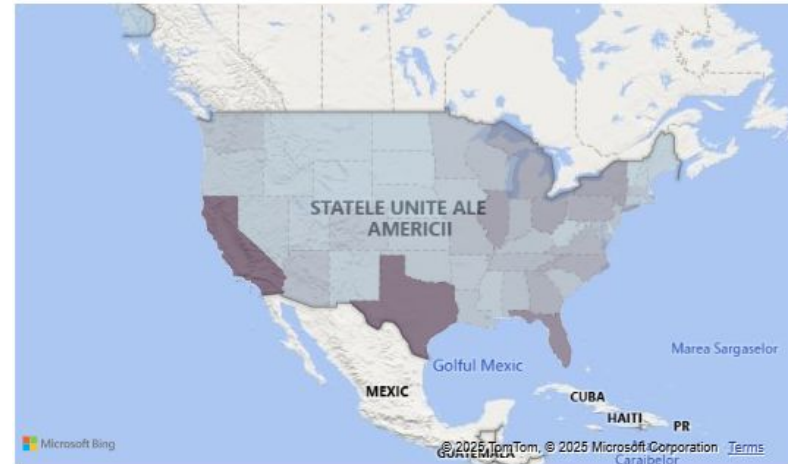
Thank you!

Dataset sources:

Nutritional values:

https://www.kaggle.com/datasets/rakkesharv/fast-food-joint-nutrition-values-dataset

Health Survey:

https://data.cdc.gov/Nutrition-Physical-Activity-and-Obesity/Nutrition-Physical-Activity-and-Obesity-Behavioral/hn4x-zwk7/about_data

McDonald's across the US:

https://www.kaggle.com/datasets/mdmdata/mcdonalds-locations-united-states