

Starbucks



Statistical Learning Project

Alberto Calabrese
Eleonora Mesaglio
Greta d'Amore Grelli

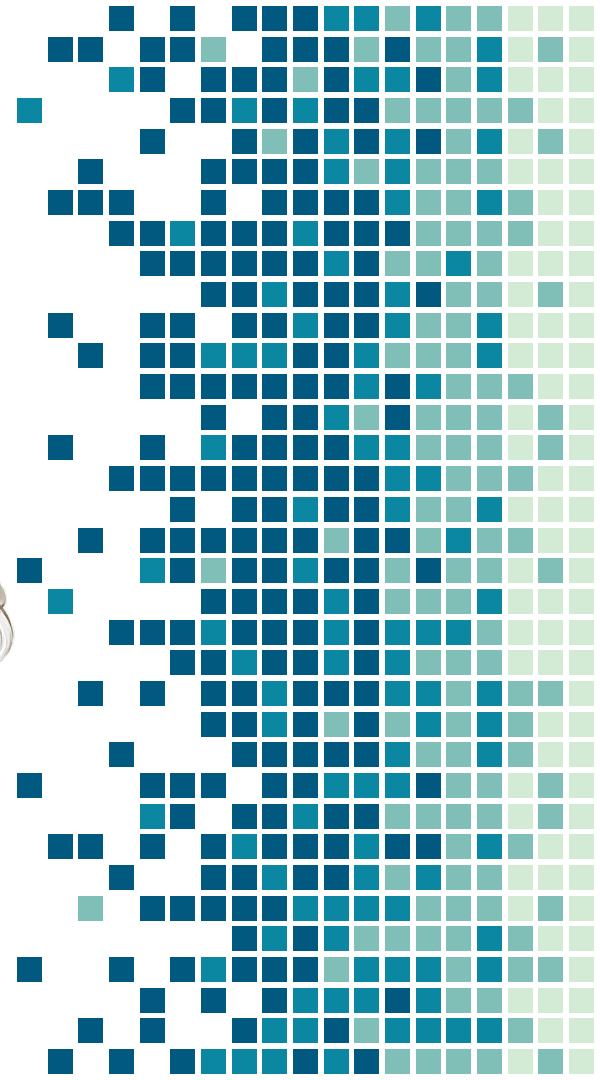


What is Starbucks ?

Global coffeehouse chain known for its specialty coffee drinks, teas, and pastries.

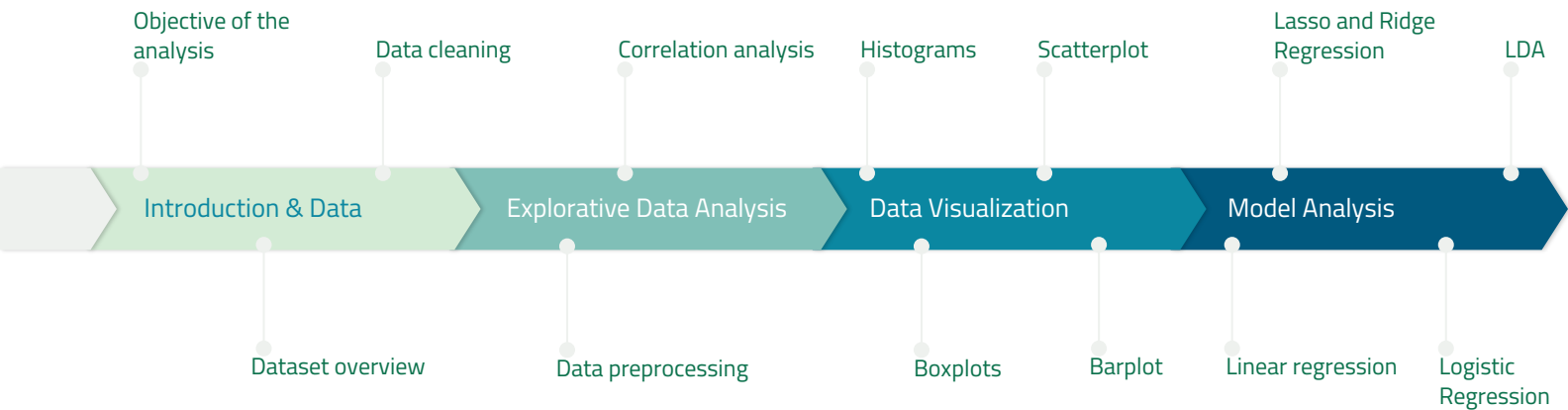
Founded in **Seattle** in 1971

Noted for its **customer-centric approach** and ethically sourced coffee beans.





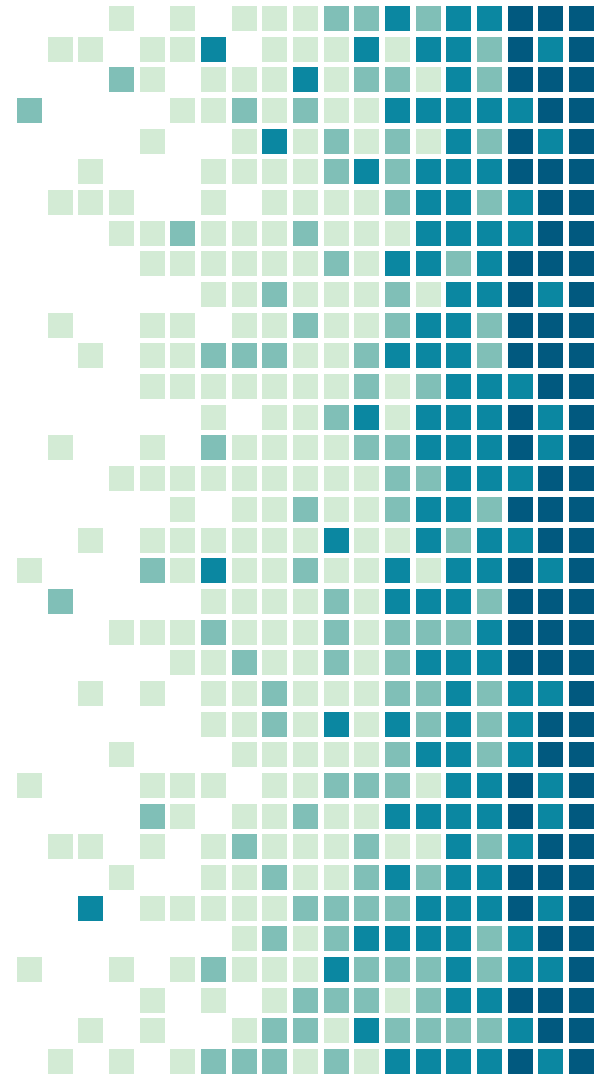
Content





1. Introduction & Data

Objective of the analysis | Data



Objective of the analysis



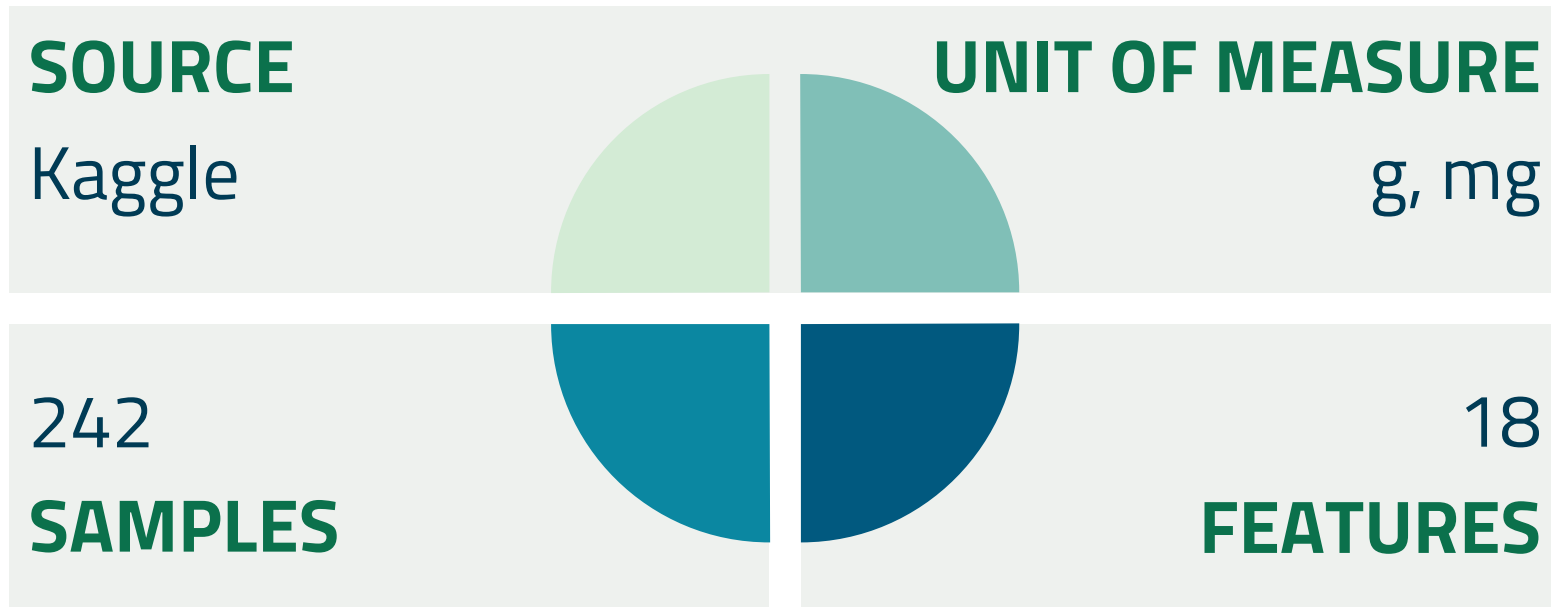
Goal:

- Gain a comprehensive understanding of the data;
- Build models for accurate predictions and classifications.

Dataset



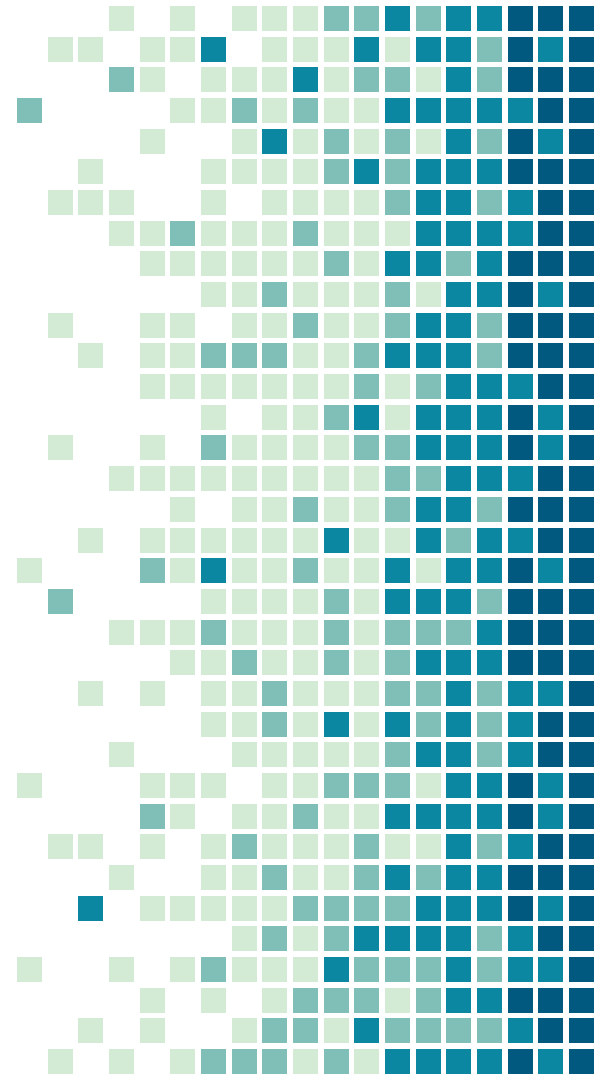
Introduction &
Data





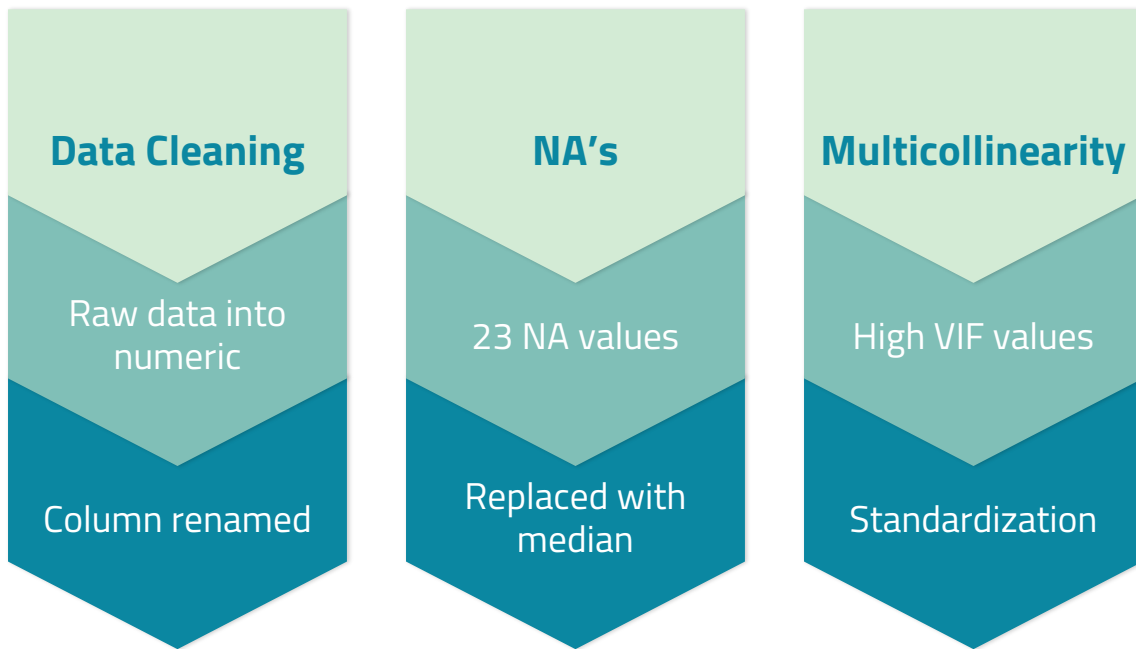
2. Explorative Data Analysis

Data preprocessing | Correlation Analysis





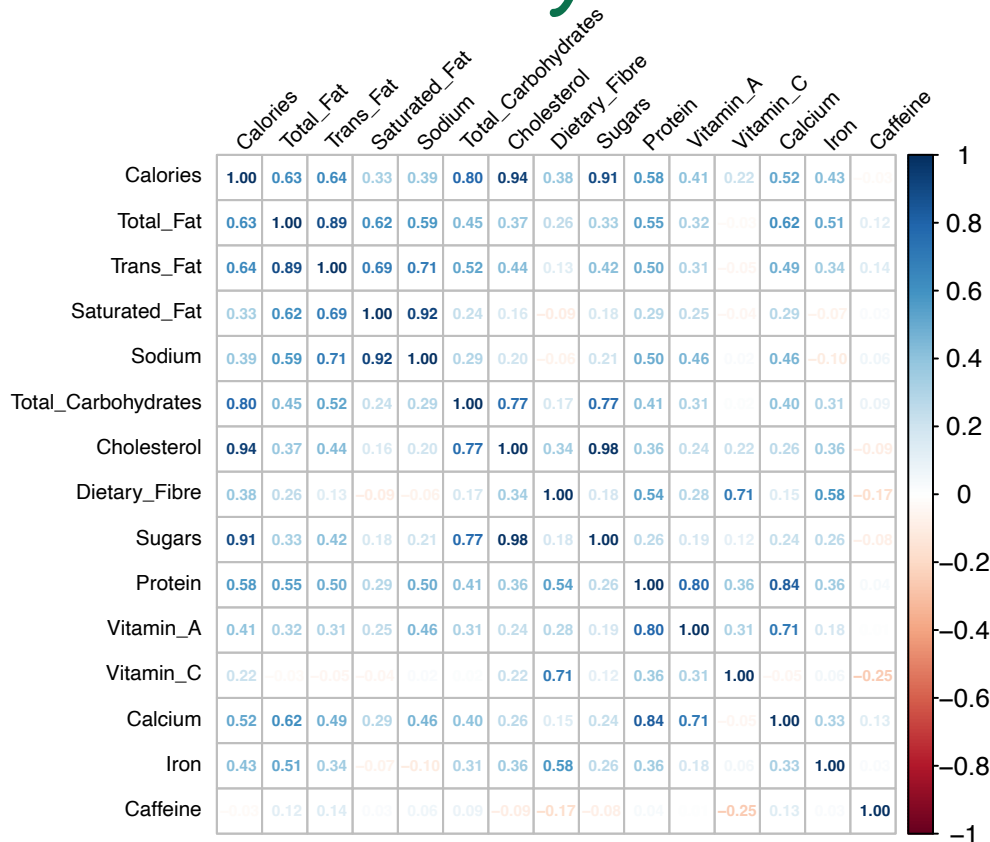
Problems with the data & data preprocessing



Correlation Analysis



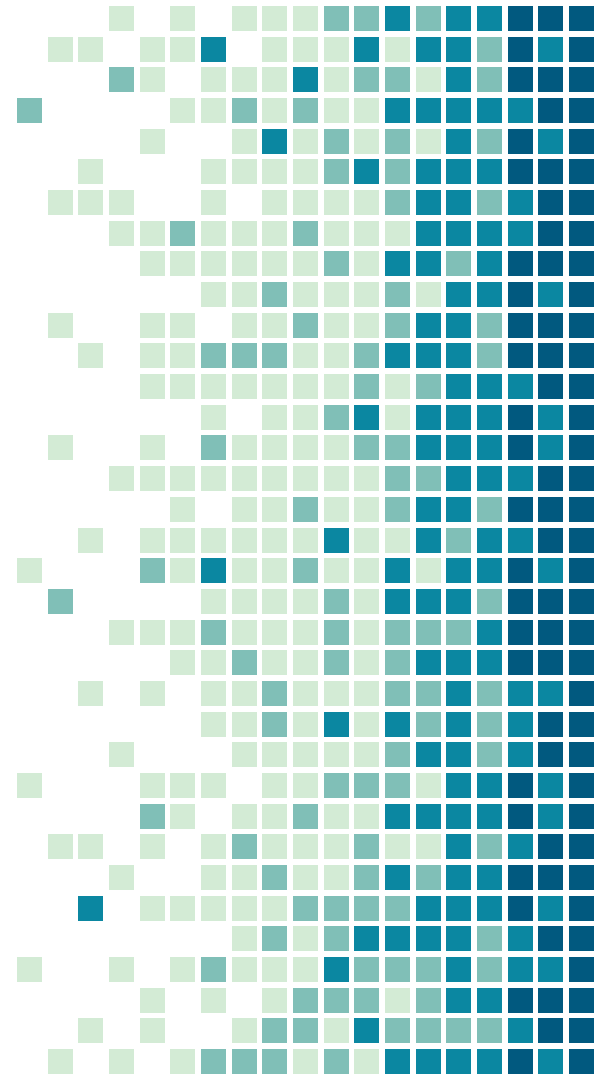
Explorative
Data Analysis



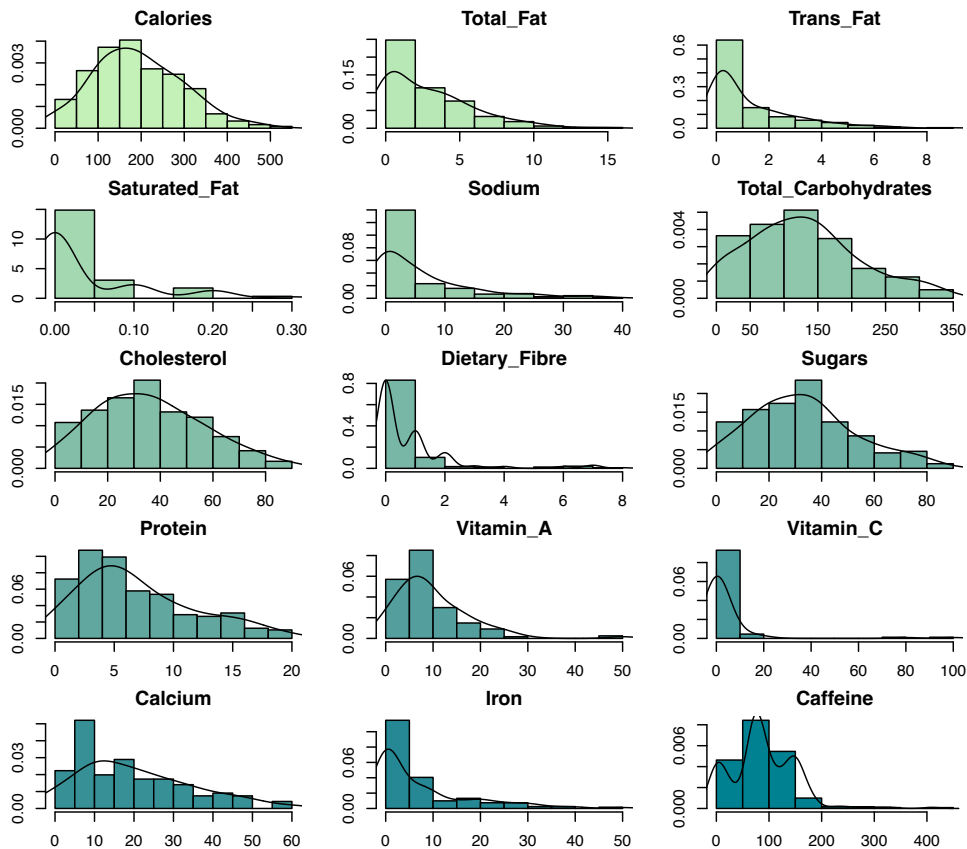


3. Data Visualization

Histograms | Boxplot | Scatterplot



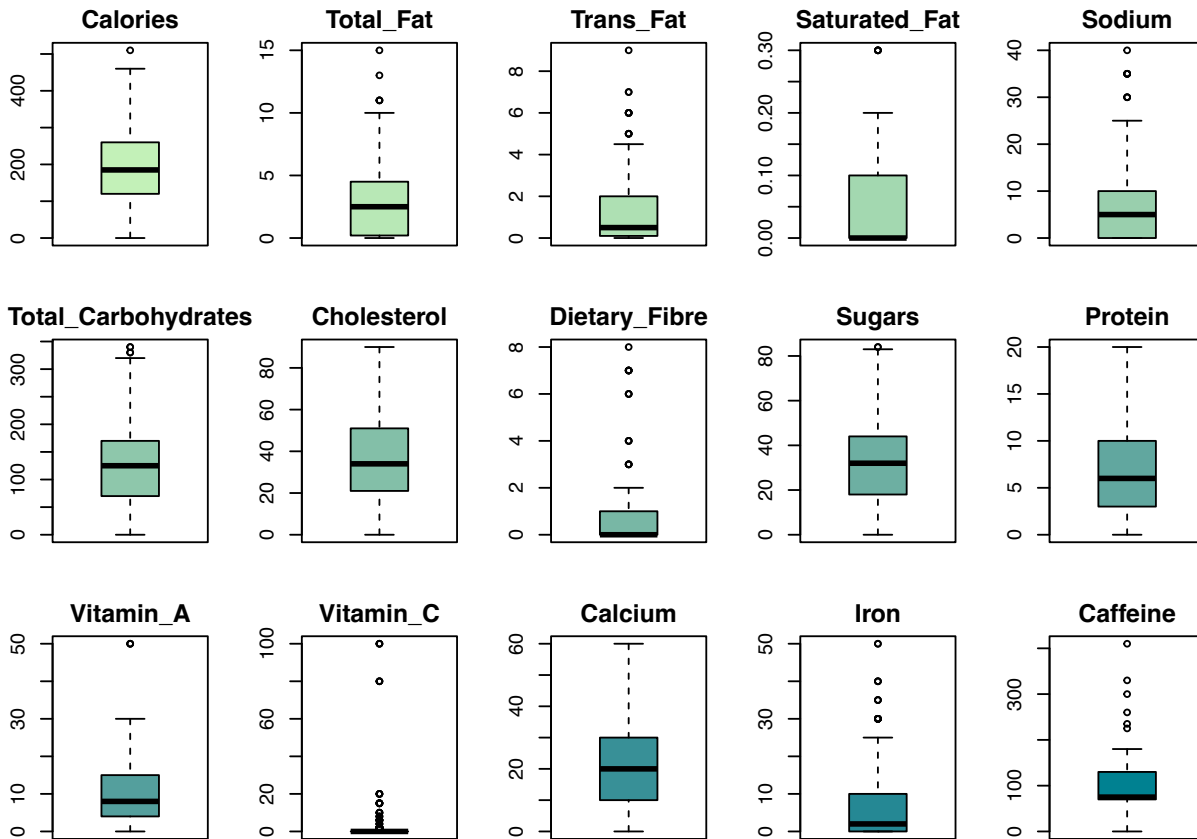
Histograms



Boxplot



Data
Visualization

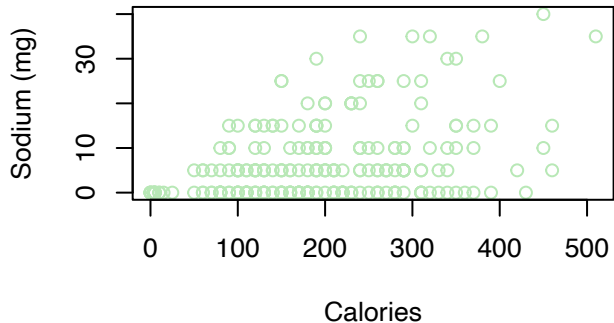


Scatterplot

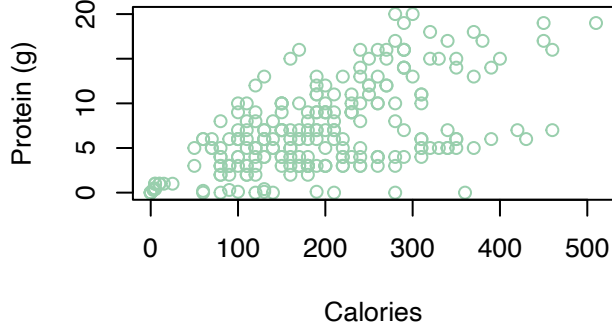


Data
Visualization

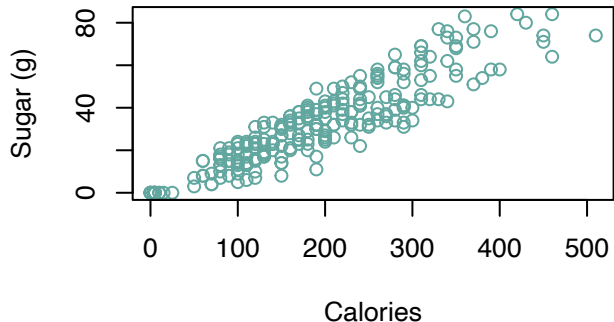
Relation between Calories and Sodium



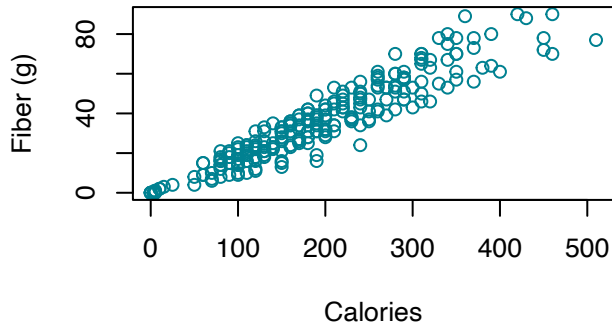
Relation between Calories and Protein



Relation between Calories and Sugars



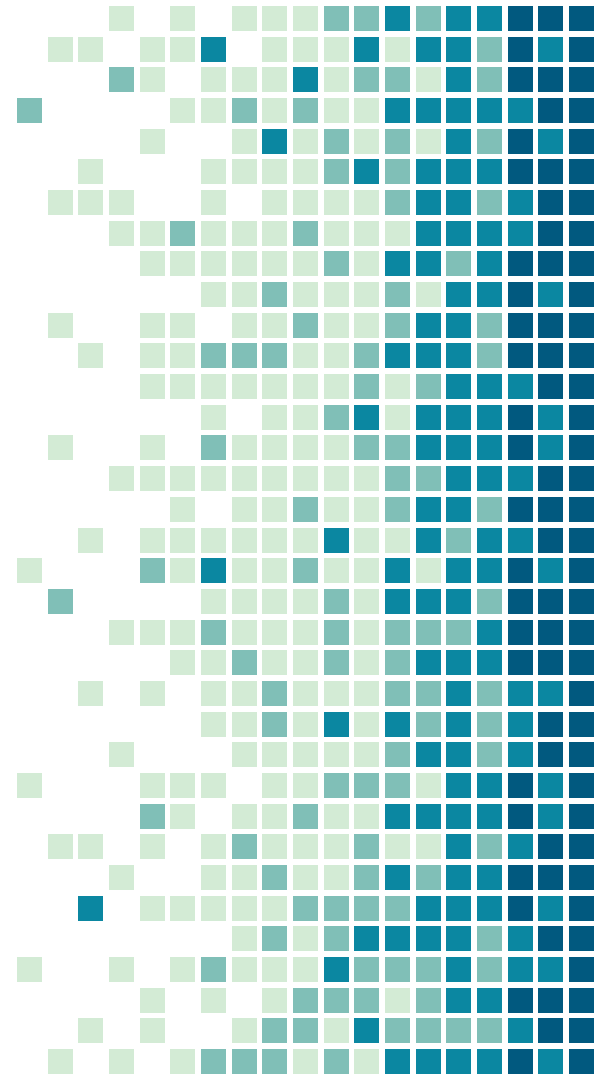
Relation between Calories and Fiber





4. Model Analysis

Regression Analysis | Classification Analysis

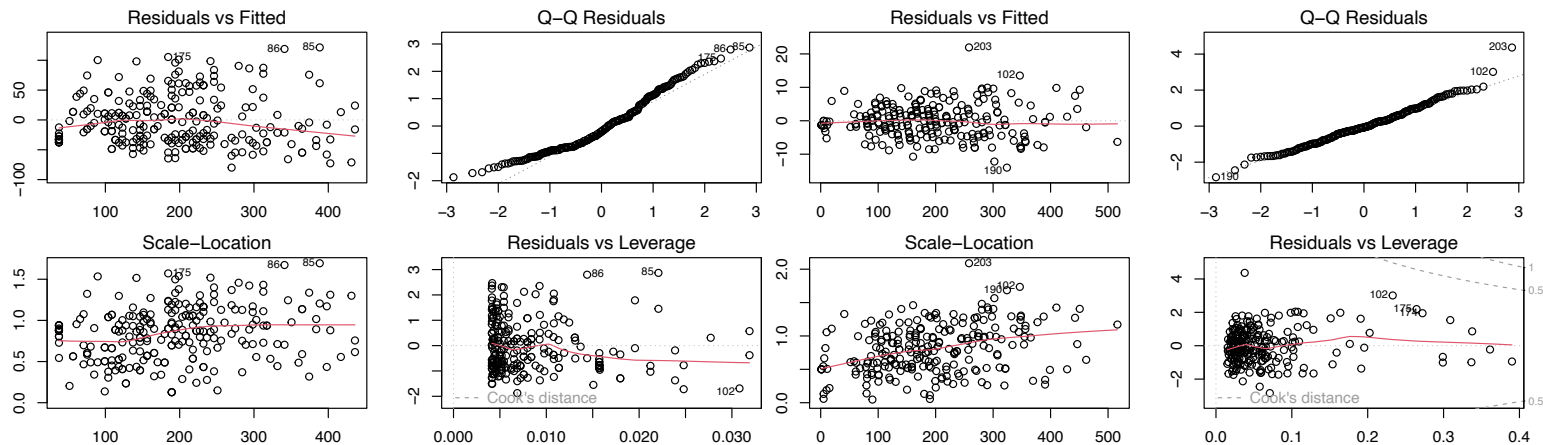


Linear Regression

Simple and Multiple



Regression
Analysis



	AIC	BIC	R^2	Adjusted R^2
Simple linear regression	2509	2519	0.827	0.826

	AIC	BIC	R^2	Adjusted R^2
Multiple linear regression	1494	1550	0.997	0.997

Linear Regression

Backward selection | Multicollinearity



Regression
Analysis

	AIC	BIC	R^2	Adjusted R^2
Backward model	1492	1544	0.997	0.997

High VIF Values

LOG Standardization

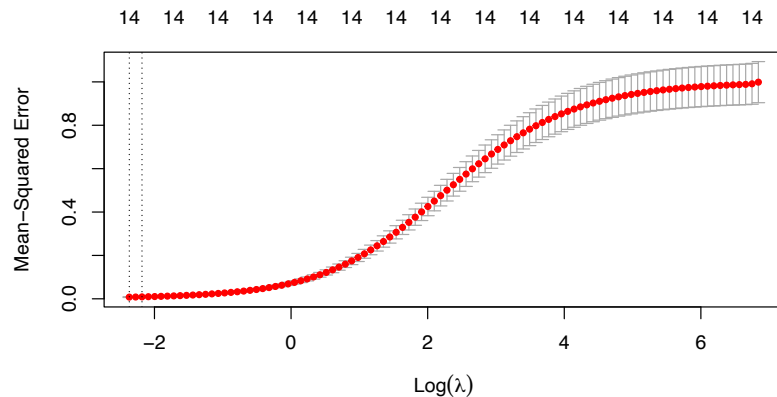
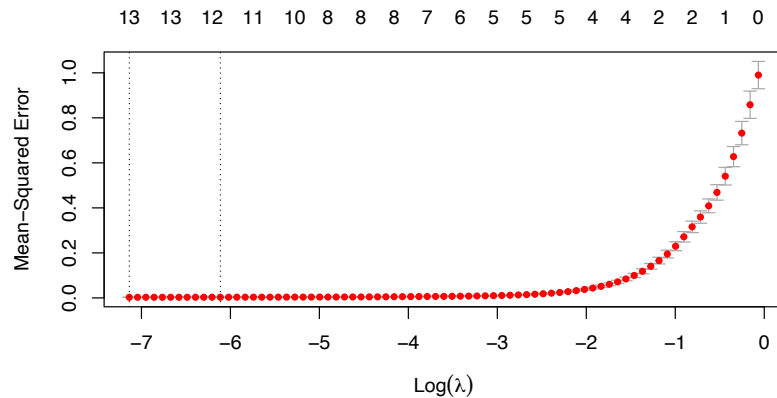
	AIC	BIC	R^2	Adjusted R^2
Log standardized Backward model	460	488	0.63	0.62

Lasso and Ridge Regression



Regression
Analysis

	R^2	MSE
Lasso Regression	0.9975	0.0024
Ridge Regression	0.9941	0.0066



Cross Validation

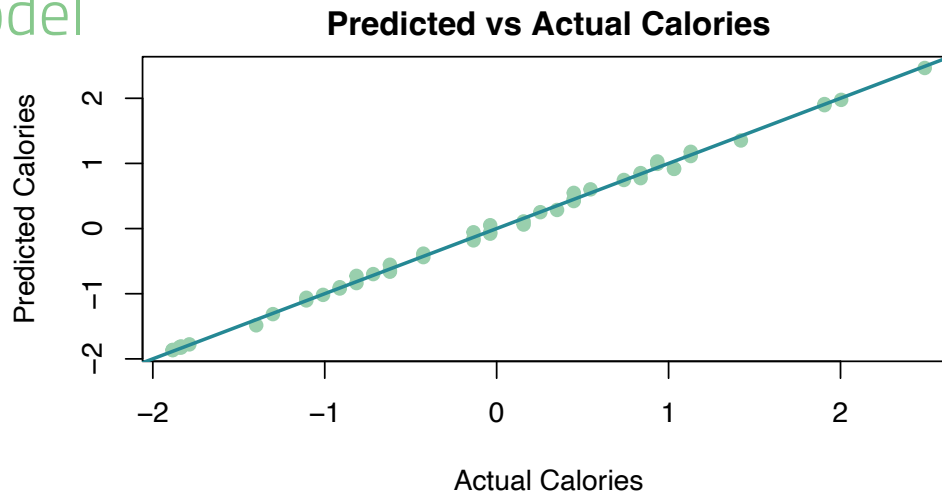
Lasso Regression Model



Regression
Analysis

Training Set
80%

Test Set
20%



	Accuracy	MSE	R^2
Lasso regression model	0.997	0.0026	0.997

Logistic Regression



Classification
Analysis

New Categorical Variable: Calories content

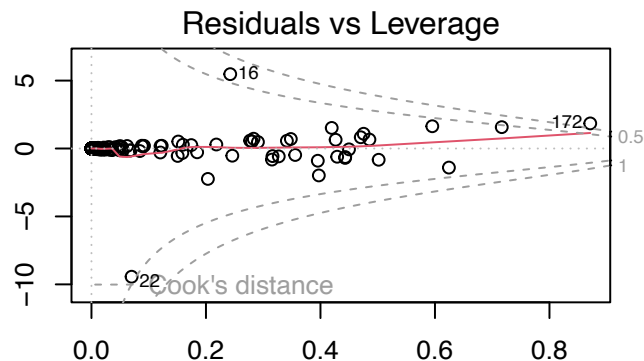
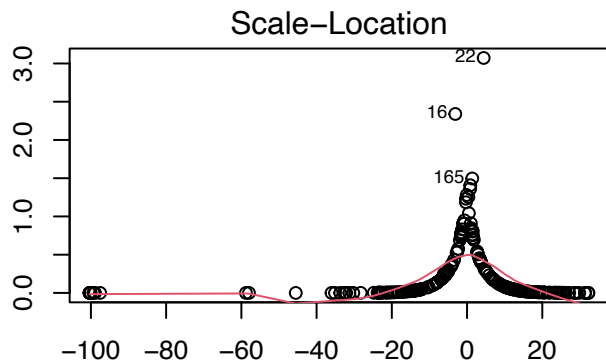
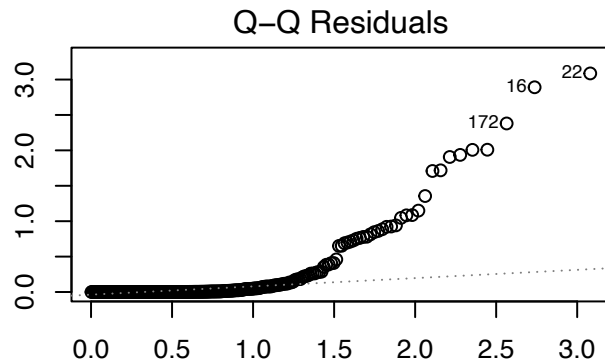
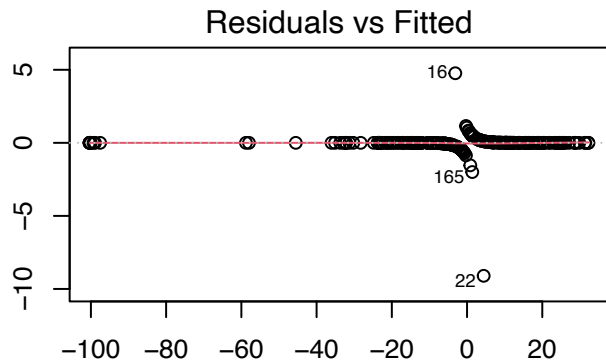
Greater than median:
High quantity of
calories

Lower than median:
Low quantity of
calories

Logistic Regression



Classification
Analysis



Logistic Regression



Classification
Analysis

	Accuracy	Precision	Recall	F1 Score
Cross Validation	0.91	0.92	0.92	0.92

	AIC	BIC	R^2	Residual Deviance	Null Deviance
Multiple linear regression	69.42	121.75	0.88	39.42	335.48

Linear Discriminant Analysis

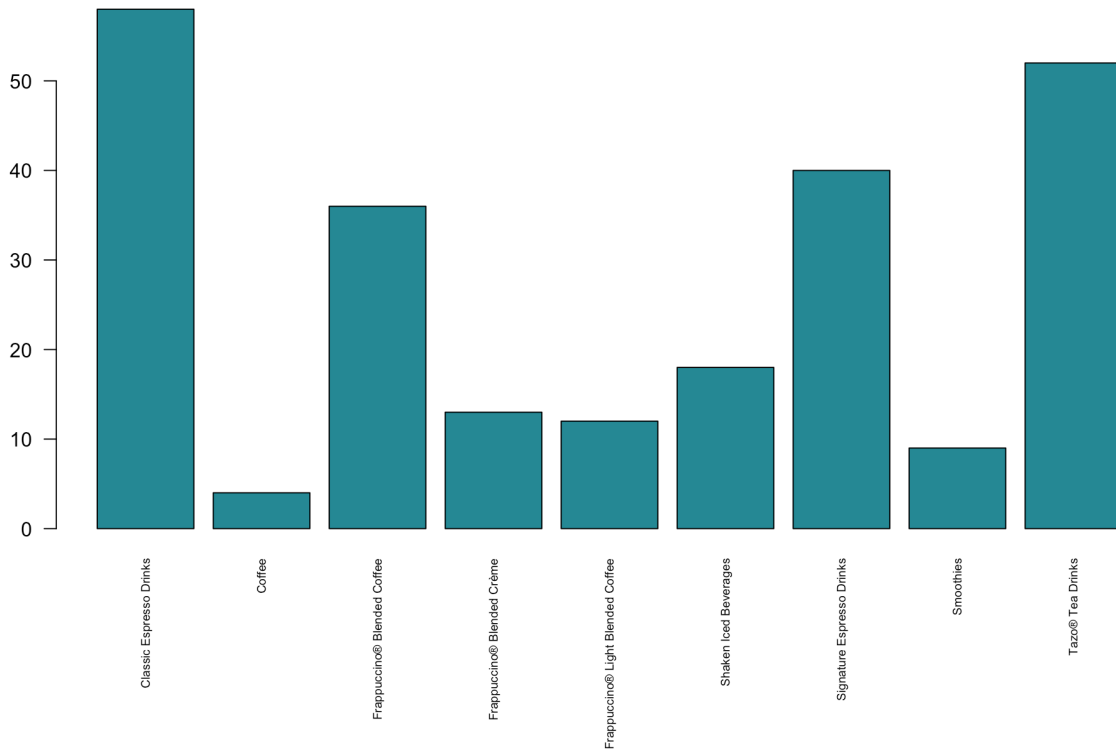


Classification
Analysis

Classification of
beverage
categories

Accuracy
80%

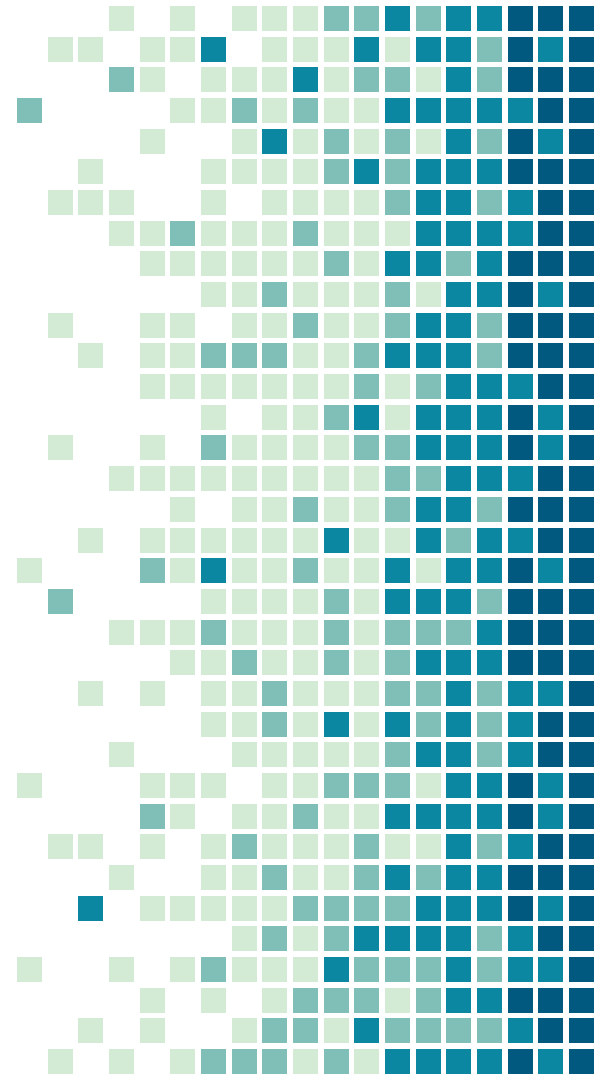
Distribution of Beverage Categories





5. Conclusions

Conclusions | Potential implementations



Conclusions



Conclusions

Lasso Regression Model

Accuracy 99.7%

Useful to predict the amount of calories in a beverage

Logistic Regression Model

Accuracy 91%

Useful to classify a beverage as calorie-dense or calorie-light

Potential Implementations



Conclusions

- Propose our model to companies developing new beverages, allowing them to **predict calorie content** based on other variables.
- Especially useful in the US, where **obesity** is a major issue, this tool can make a significant difference!

THANKS!

Any questions?

Alberto Calabrese

Eleonora Mesaglio

Greta d'Amore Grelli

