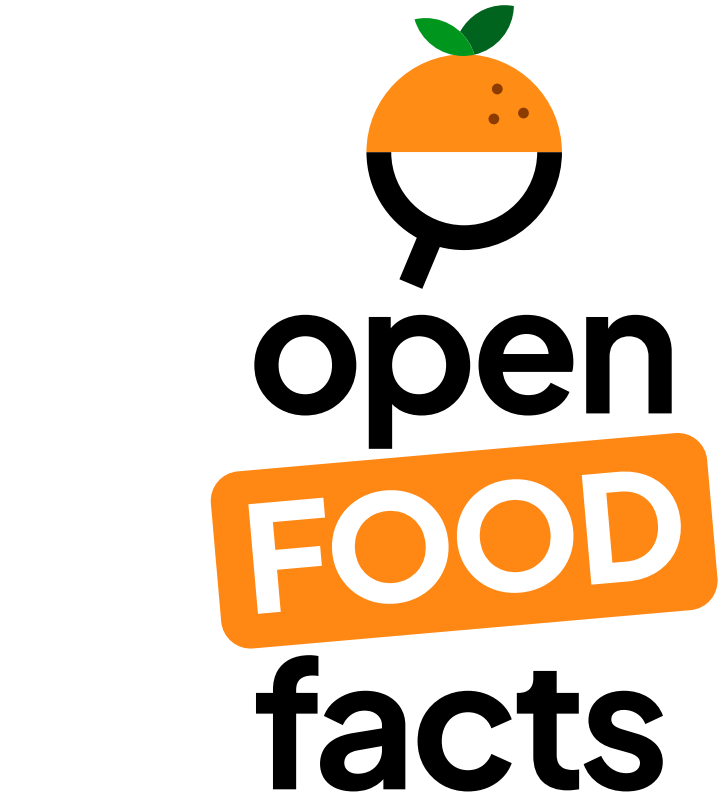


Open Food Fact Data Analysis

Discover Relation Between Food Product Nutrients with Country Diabetes Level

About The Dataset - Main

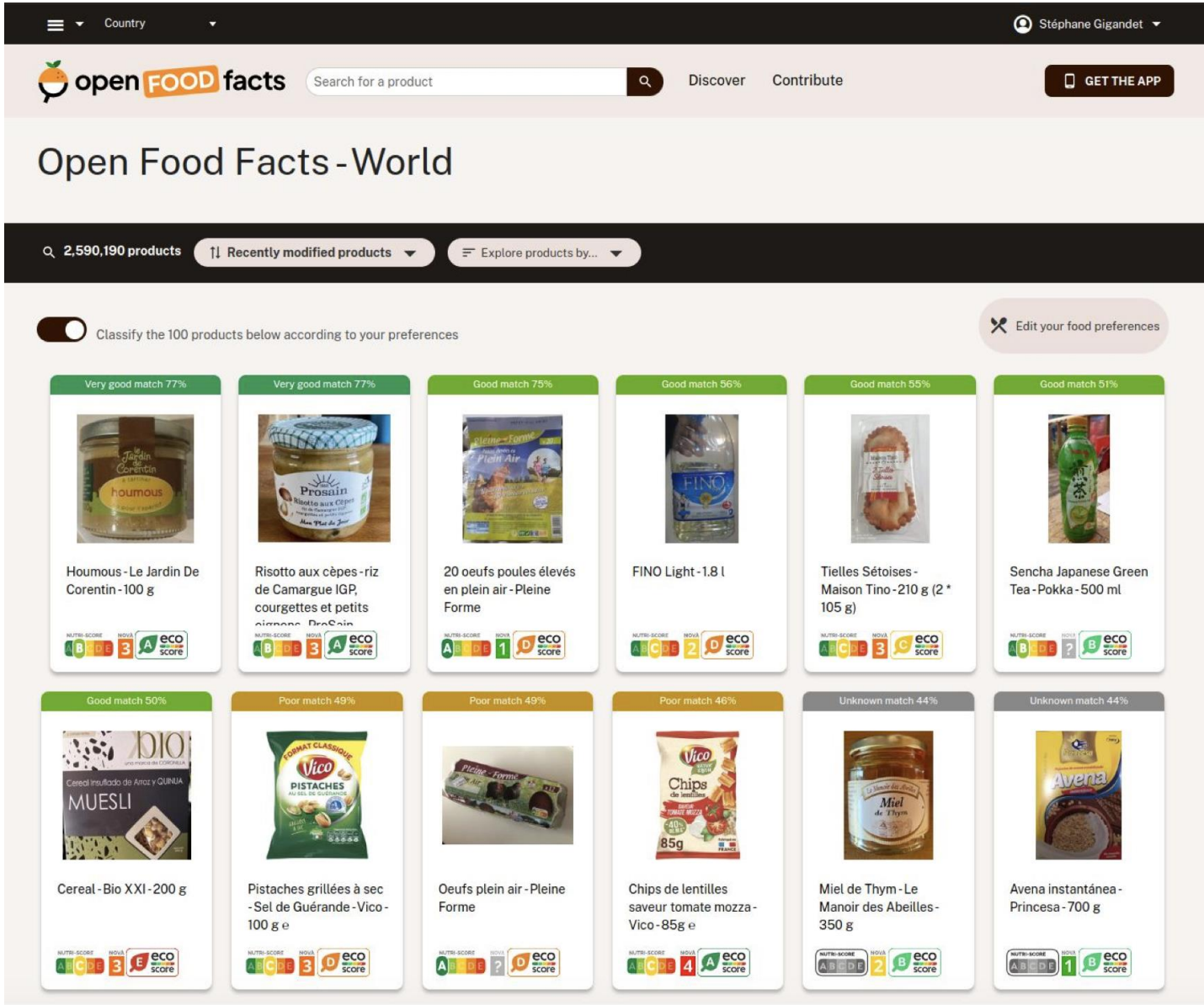
open FOOD facts

A Wikipedia for Food Products

5.9 GB .parquet file size
3.6 M row of products from
180+ countries in **open data**

Data Sources:

1. Crowdsourcing
2. Food Industry



Datasets Column:

category (e.g., snacks, beverages),
product_name,
nutriscore_grade,
sugars_100g,
additives_n,
year_entry,
country,
....

About The Dataset - Secondary

International Diabetes Federation

- Established in 1950 in Amsterdam
- Non-profit umbrella organization for 240 diabetes association globally.
- Vision to provide access to affordable, quality diabetes care and education worldwide.
- Provide **country-level diabetes rate** data

Where are the estimates for the IDF Diabetes Atlas sourced from?

The data in the *IDF Diabetes Atlas 10th edition* come from a variety of sources such as peer-reviewed scientific papers, and national and regional health surveys. Official reports by international organisations, such as the World Health Organization (WHO), were also assessed for their quality that was defined in consensus with an international expert panel. Data sources that passed strict selection criteria were included in the data analysis.

Global

Diabetes data report 2000 — 2045

At a glance	2000	2011	2021	2030	2045
Diabetes estimates (20-79 y)					
People with diabetes, in 1,000s	151,000.0	366,000.0	536,600.0	642,800.0	783,700.0
Age-adjusted comparative prevalence of diabetes, %	4.6	8.5	9.8	10.8	11.2
People with undiagnosed diabetes, in 1,000s	-	183,000.0	-	-	-
Proportion of people with undiagnosed diabetes, %	-	50.0	44.7	-	-

537 million adults are living with diabetes

3 in 4 adults with diabetes live in low- and middle-income countries

6.7 million deaths due to diabetes in 2021

Key global findings 2021

The **IDF Diabetes Atlas 10th edition** reports a continued global increase in diabetes prevalence, confirming diabetes as a significant global challenge to the health and well-being of individuals, families and societies.

Download the IDF Diabetes Atlas 10th Edition and other resources.

View all the latest national and regional data in our data portal

Diabetes around the world in 2021:

- **537 million** adults (20-79 years) are living with diabetes - 1 in 10. This number is predicted to rise to **643 million** by 2030 and **783 million** by 2045.
- **Over 3 in 4** adults with diabetes live in low- and middle-income countries.
- Diabetes is responsible for **6.7 million** deaths in 2021 - 1 every 5 seconds.
- Diabetes caused at least **USD 966 billion** dollars in health expenditure – a 316% increase over the last 15 years.
- **541** million adults have Impaired Glucose Tolerance (IGT), which places them at high risk of type 2 diabetes..

Research Question

Sugar Content

1. What is the annual trend in **average sugar content** in sweet food products across all countries?
2. Which product categories have the **highest average sugar content** across all countries in Europe?

Diabetes Rate

3. How is **sugar content correlated with country diabetes levels** globally, and how does this relationship vary in Europe?

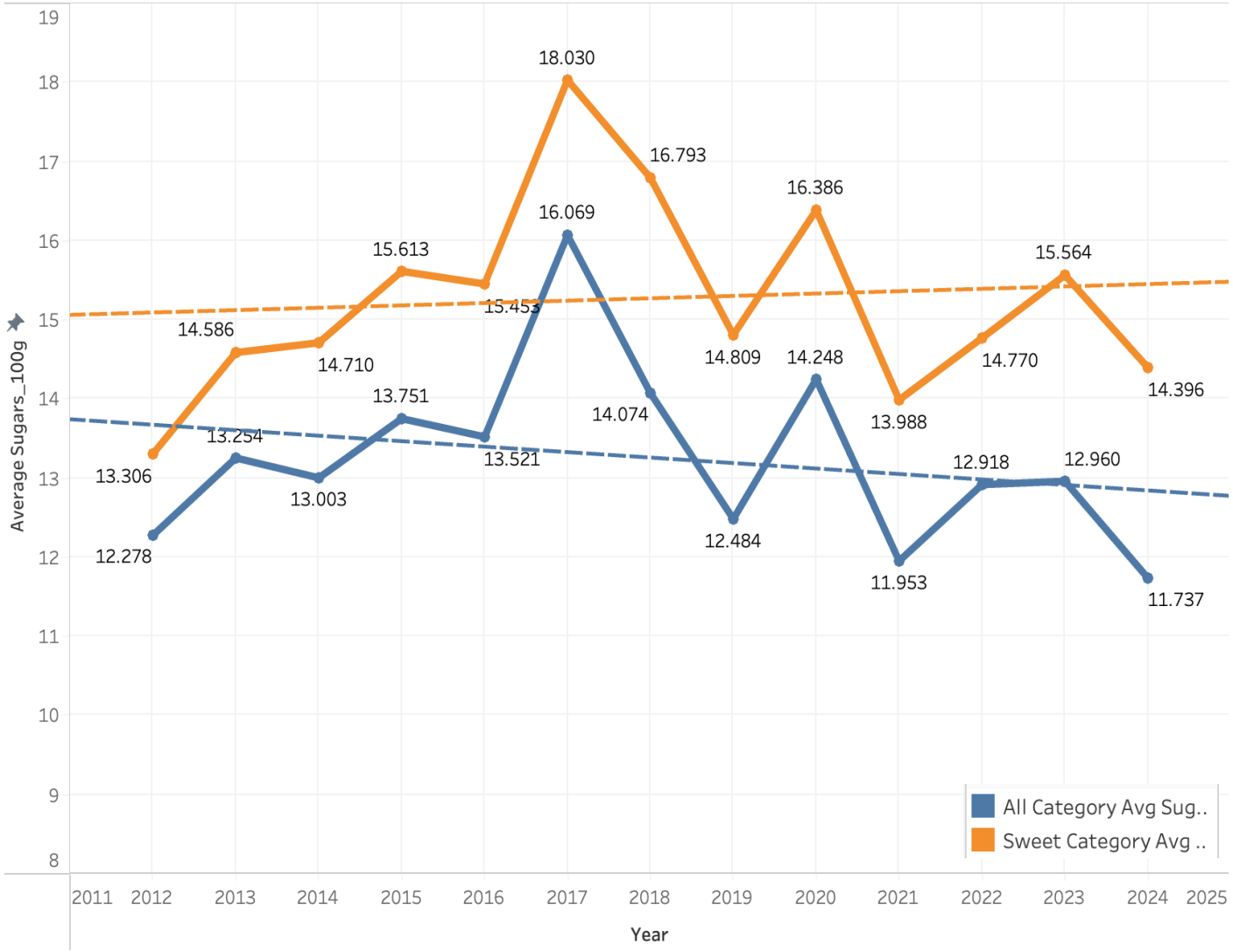
Additives Content

4. What is the annual trend in the **average count of additives** in food products?
5. Which product categories had the **highest average additive content**?

Relationship & Prediction

6. What is the relationship between sugar content, additives, and Nutri-Score grades in Europe?
7. Can a predictive model accurately estimate unknown Nutri-Score grades and what are the most significant predictors for Nutri-Score estimation?

Annual Trend of Sugar Content in all Region

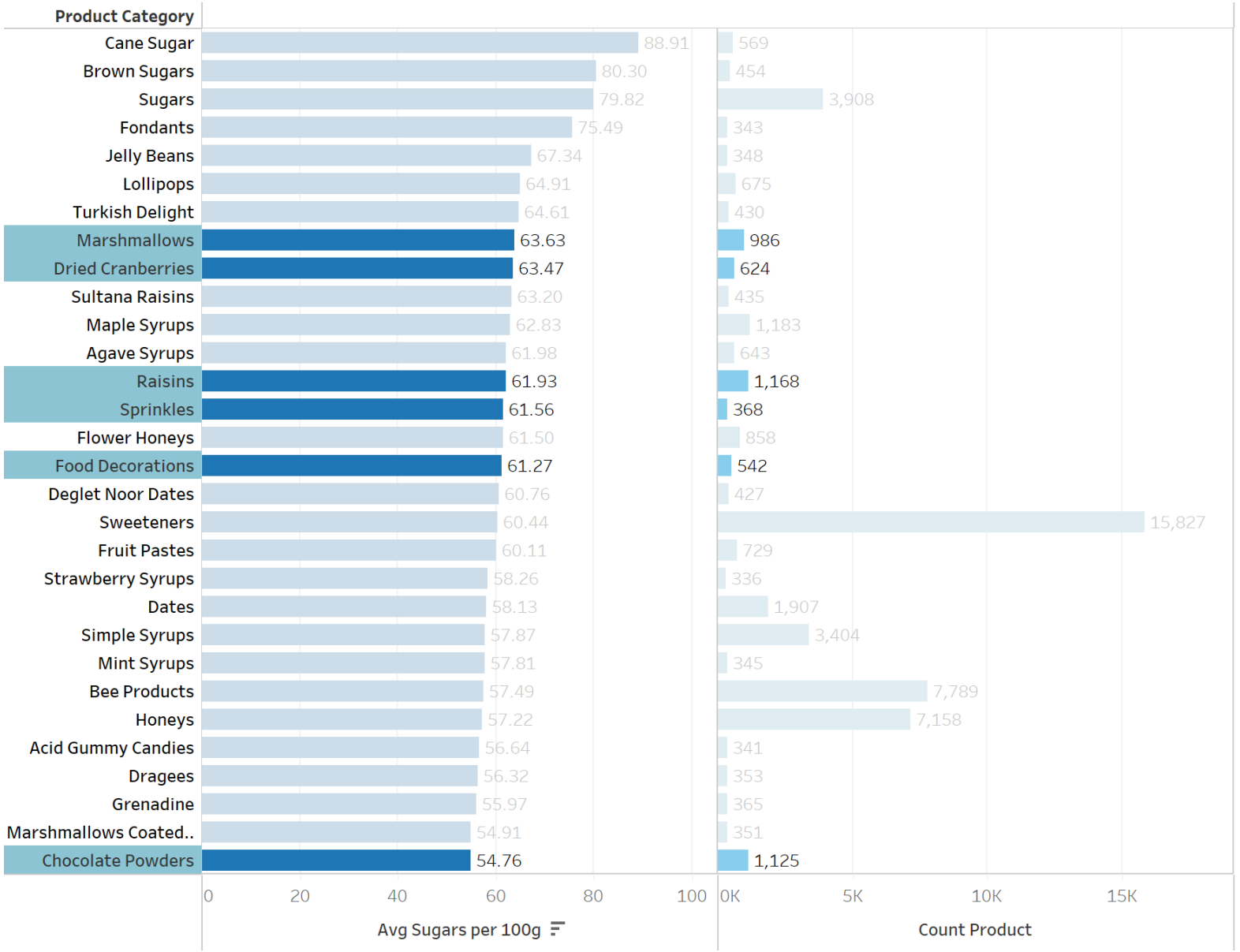


- The sweet product categories (average sugar content >5g)* with a **slightly increasing sugar content trend**, while all product categories display a **decreasing trend**
- Both of them showing **insignificant trends**
- The average sugar content in all product categories is approximately **13.2g/100g**, while in sweet product categories, it is higher at **15.2g/100g**
- This represents a **13.18% difference** in sugar content for sweet products compared to the all product category

*Sweet product definition based on EU Nutrition & Health Claims Regulation legislation (EC) 1924/2006.

Top-30 Product Category* with Highest Sugar Content

UNIVERSITY OF TWENTE.



- In general there is no surprised here, top product category with highest sugar content is dominated by "sugar", "sweetener", "candy", and "honey" related products.
- Some interesting product category that included in this list including:
  - Marshmallows
  - Dried Cranberries
  - Raisins
  - Sprinkles
  - Food Decorations
  - Fruit Pastes

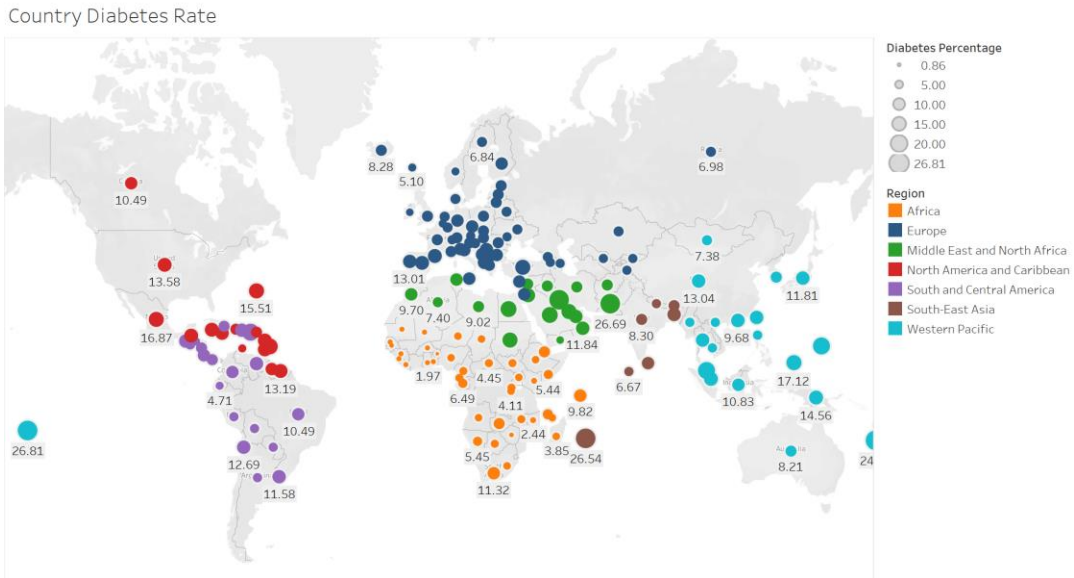
*Already filtered to top-500 count product category to remove insignificant product category

Sugar Content Correlation with Country Diabetes Rate

Highest Correlated Product Category in Europe

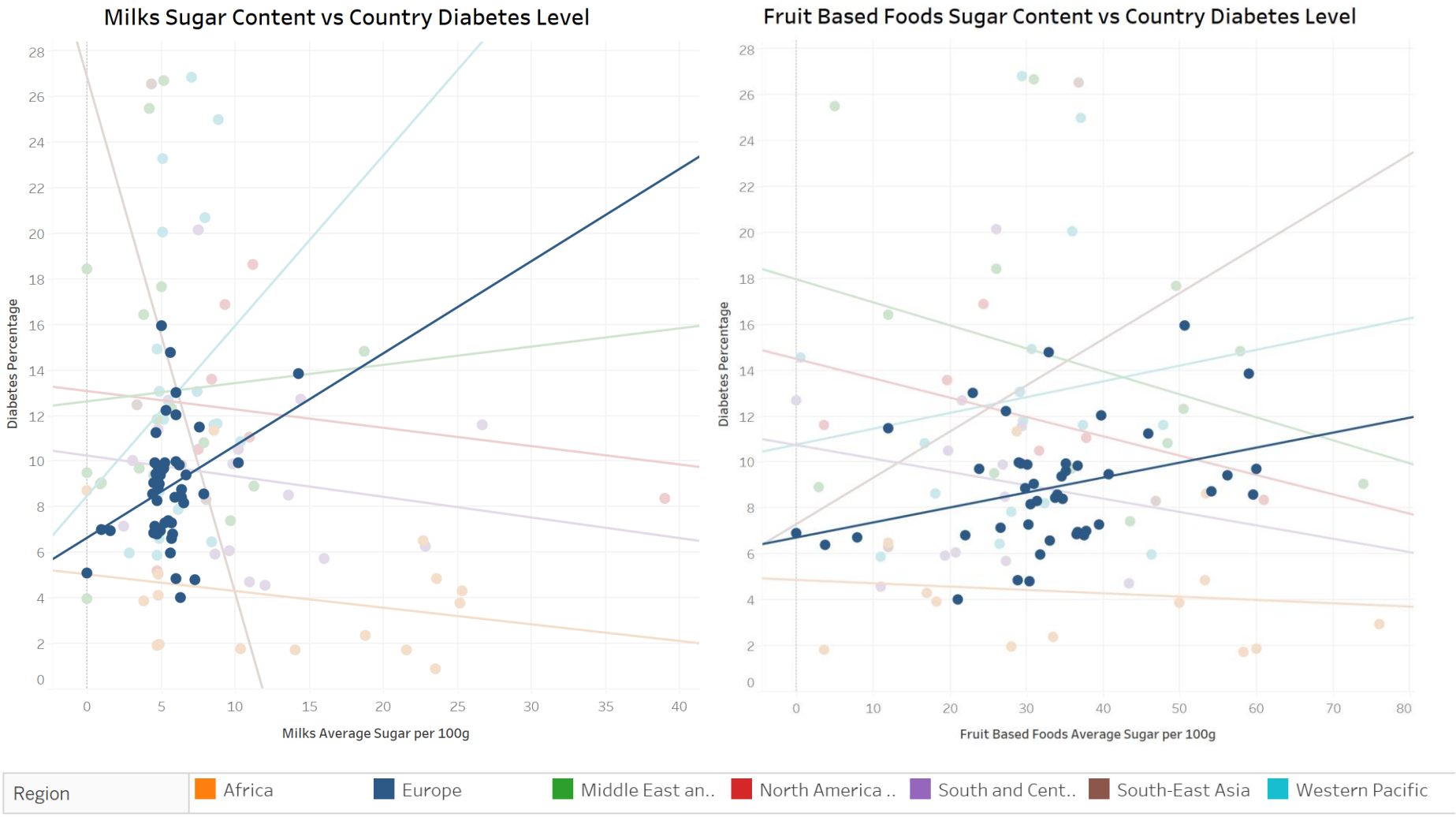
Product Category	Correlation	
	Europe	All Country
Milks	0.4181	0.0127
Fruits Based Foods	0.4099	0.1033
Dried Products To Be Rehydrated	0.3622	-0.0414
Fruits And Vegetables Based Foods	0.3586	0.0339
Sweet Pastries And Pies	0.3522	0.0994
Viennoiseries	0.3507	0.0992
Jams	0.3498	0.1774
Legume Butters	0.3481	0.1970
Peanut Butters	0.3476	0.1969
Fruit And Vegetable Preserves	0.3454	0.1842
Cereal Bars	0.3296	0.1645
Dried Products	0.2997	0.0098
Bars	0.2835	0.2881
Milk Substitutes	0.2827	0.1374
Fruits	0.2766	0.1665
Milk Chocolates With Hazelnuts	0.2757	0.0611
Orange Juices	0.2722	0.0835
Puffed Cereals	0.2672	0.0106
Cereal Flakes	0.2632	0.2052
Energy Drinks	0.2618	0.0762

- For **all region** the highest correlation is for “Bars” category with only 0.288, which is **relatively low correlated.**
- **Other region** have higher correlation, but the reliability is not accountable due to **too many missing data**.
- So we **focus in Europe** as this region have the most complete data compare to other region
- Top-2 highest correlated product category in Europe has a moderate correlation, which is **“Milks” & “Fruit Based Foods”.**



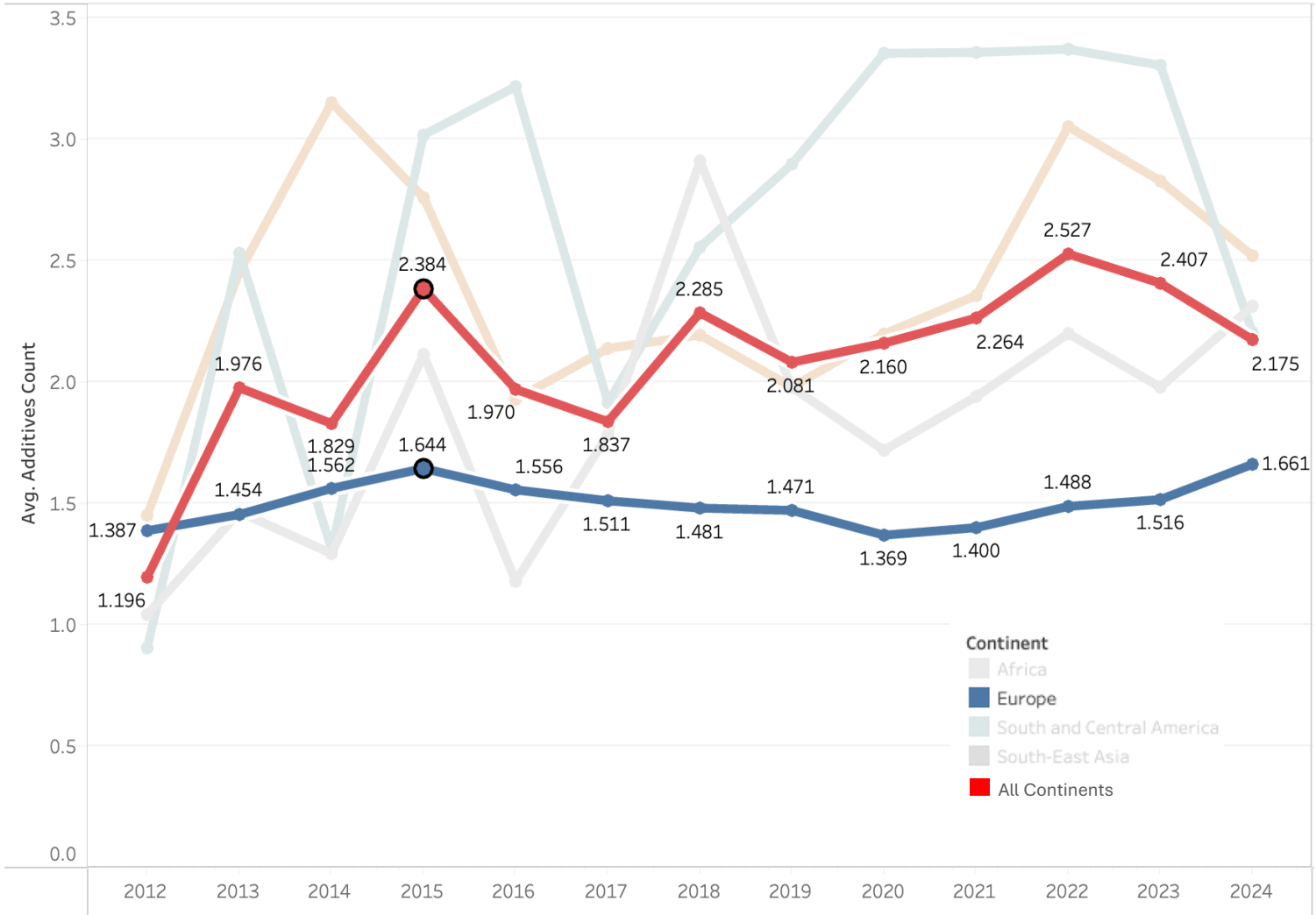
Sugar Content Correlation with Country Diabetes Rate

UNIVERSITY OF TWENTE.

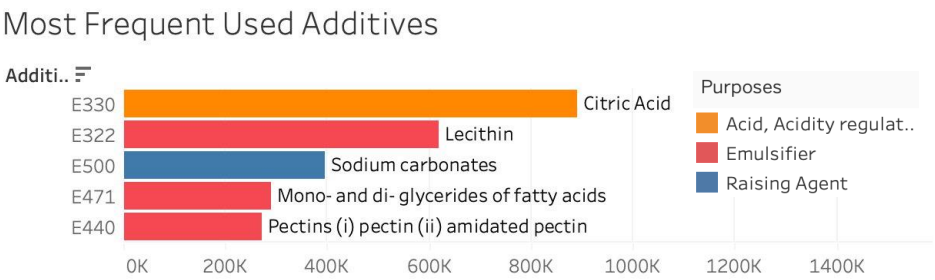


- This graph shown **scatter plot of 2 highest correlated product category in Europe**, Milks & Fruit Based Foods.
- **X-axis is average sugar** content per 100g
- **Y-axis is country diabetes level**
- As we can see from the trend line that **in Europe**, these 2 product sugar content is **positively correlated** with country diabetes level.
- Other region that have similar trend (with higher diabetes level) is Western Pacific.

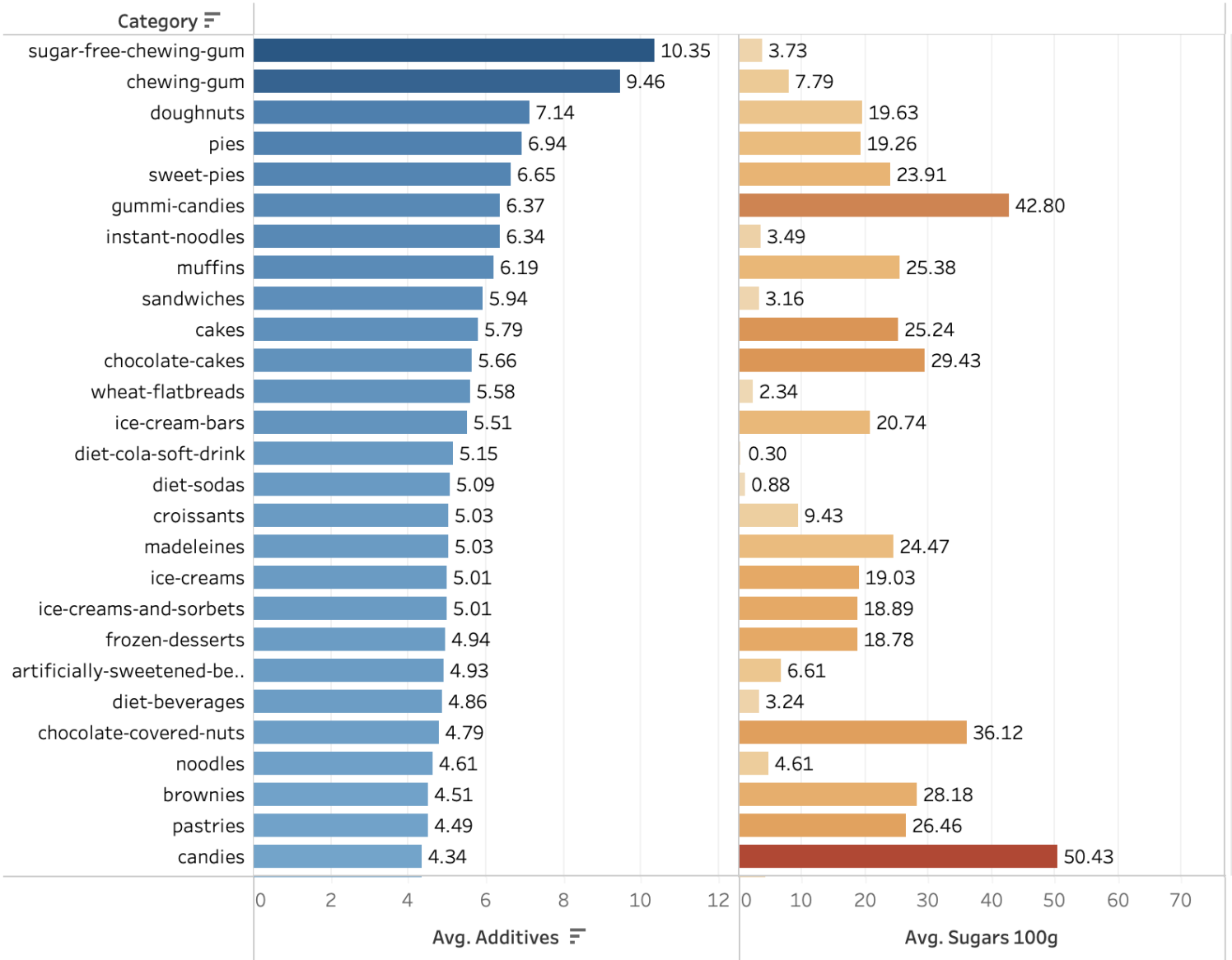
Annual Trend of Additives Content by Region



- Global average in additives content shows relatively high fluctuations, ranging between **n=1.2** and **n=2.5** over the years.
- **Europe** demonstrate steady trend in average additives content between **n=1.3 and n=1.6** additives from 2012-2024
- Europe consistently ranks as one of the **lowest regions** in amount of additives content
- Most high used additives is **Citric Acid** (E330), and several types of **Emulsifier** (E322, E471, E440)

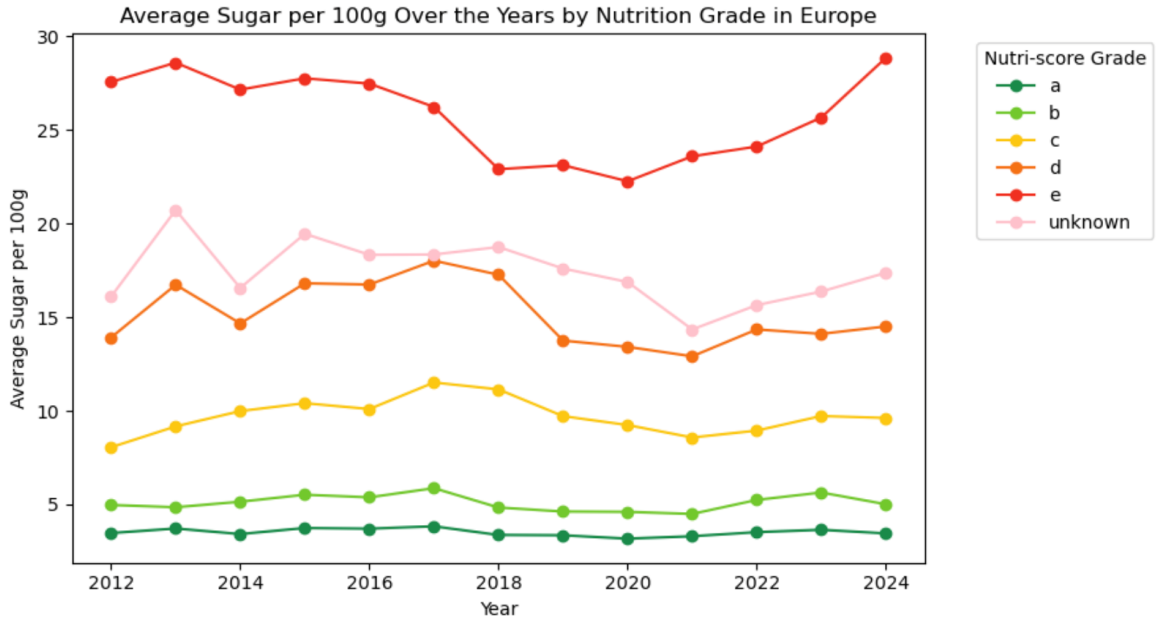


Product Category with High Additives



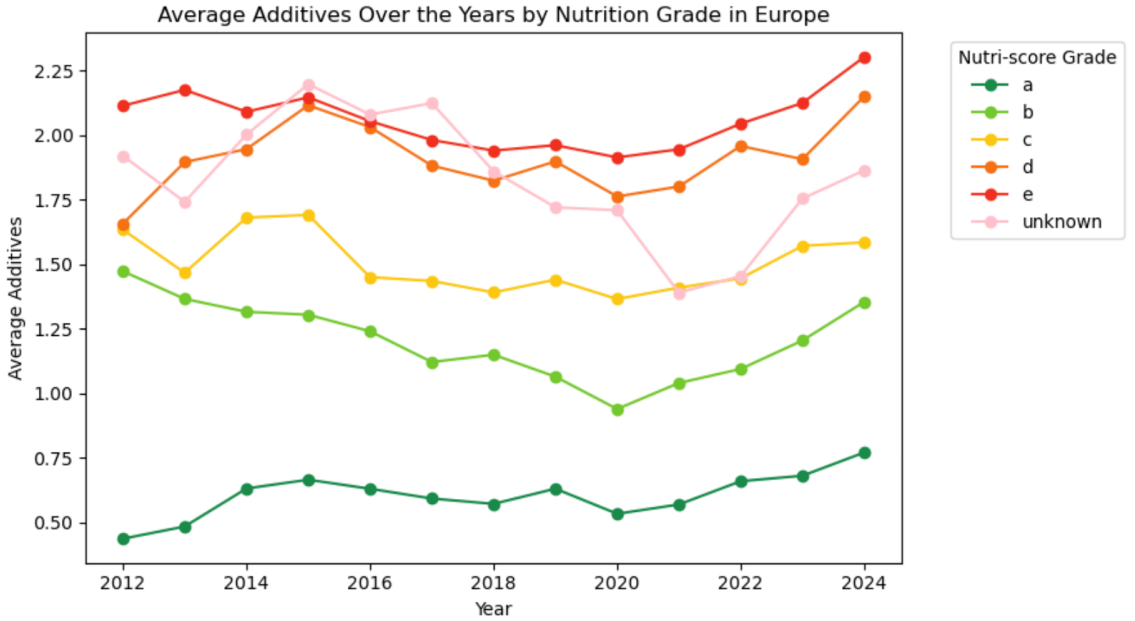
- High additive product show in categories like **confectionery, beverages, and instant food**
- Surprisingly, the results demonstrate that **high additive content does not always match high sugar content**
- Categories with 'Sugar-Free', has highest additives compared to normal product that reflects certain products often **rely on chemical enhancements for flavor and texture**

Nutriscore Grade vs Avg Sugar vs Additives

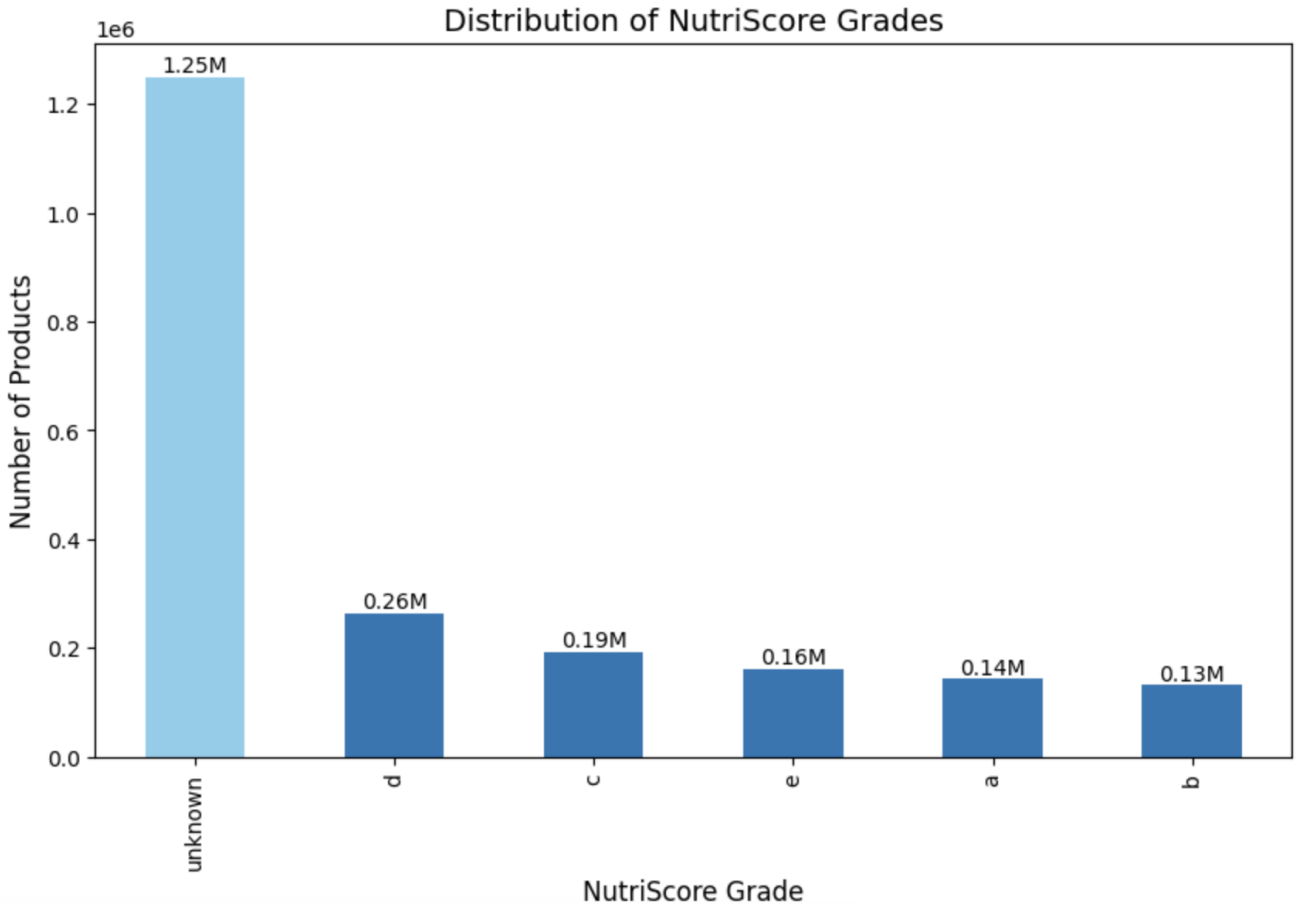


- The foods that have average sugar per 100g **low** are categorized as foods with a nutri-score grade **A**.
- The **higher** the average sugar level in food, the more the food is categorized as a less good product to the worst from **grade B to E**.

- The amount of additive content in food also affects the nutri-score grade of the food.
- The average amount of additive of products with nutri-score **grade A** is relatively **low**.
- Nutri-score **grades B, C, D**, and **E** of the products have an average amount of additive of around **1.5 to 2.1.**



Nutrition Score Grade Prediction



NutriScore is a grading system that ranks products from A (healthiest) to E (least healthy) based on nutritional quality. It helps consumers make informed food choices.

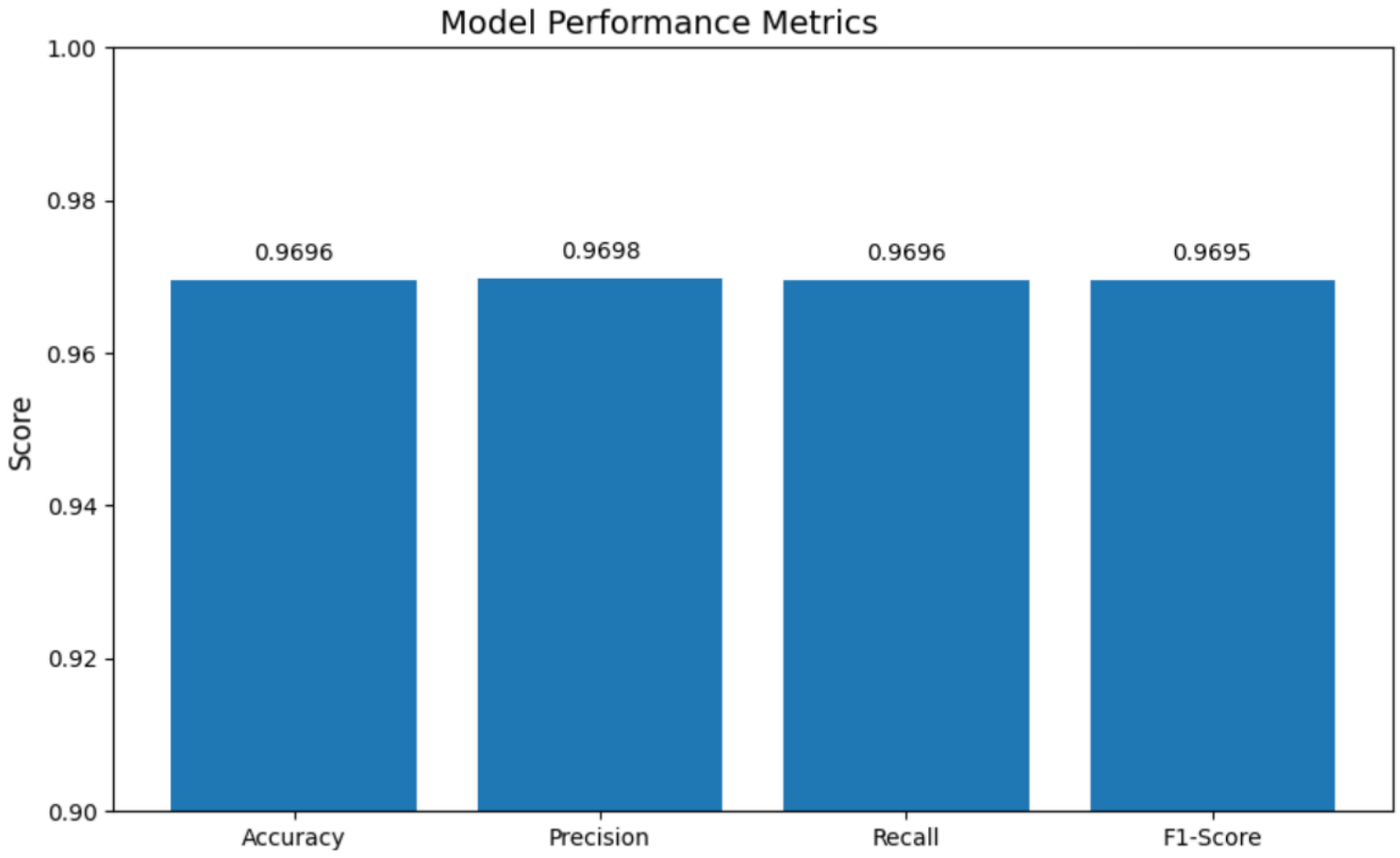
Total products analyzed is about 2 Millions, where **58.26% had unknown nutrition grade.**

The presence of unknown NutriScore grades creates gaps in the dataset. Predicting these grades helps transform an incomplete dataset into a more comprehensive and usable resource for analysis.

Nutrition Score Grade Prediction

UNIVERSITY OF TWENTE.

Classification Performance



Model Used:

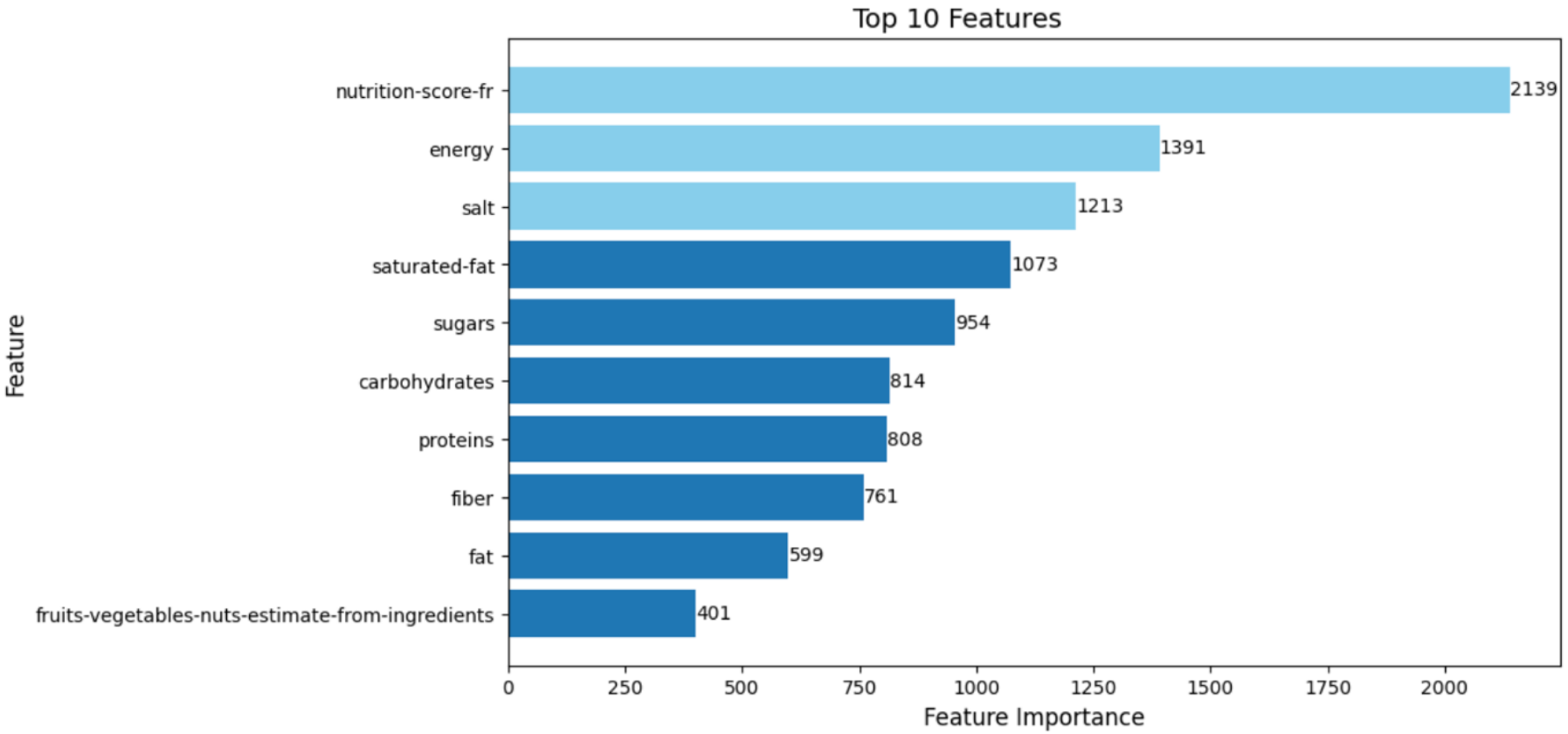
- The analysis was conducted using the **Gradient Boosting** model, known for its efficiency and strong performance on large datasets, also it can handled imbalanced data automatically.
- The model was trained and evaluated to predict missing NutriScore grades.

Cross-Validation:

- Evaluate model consistency with 5-fold cross-validation
- Result: **Mean accuracy of 97.94% with minimal variation across folds.**
- The high and consistent performance makes the model suitable for predicting NutriScore grades accurately

Nutrition Score Grade Prediction

Explainability



Understanding feature importance helps explain the model's predictions by identifying which factors contribute most to the NutriScore grades

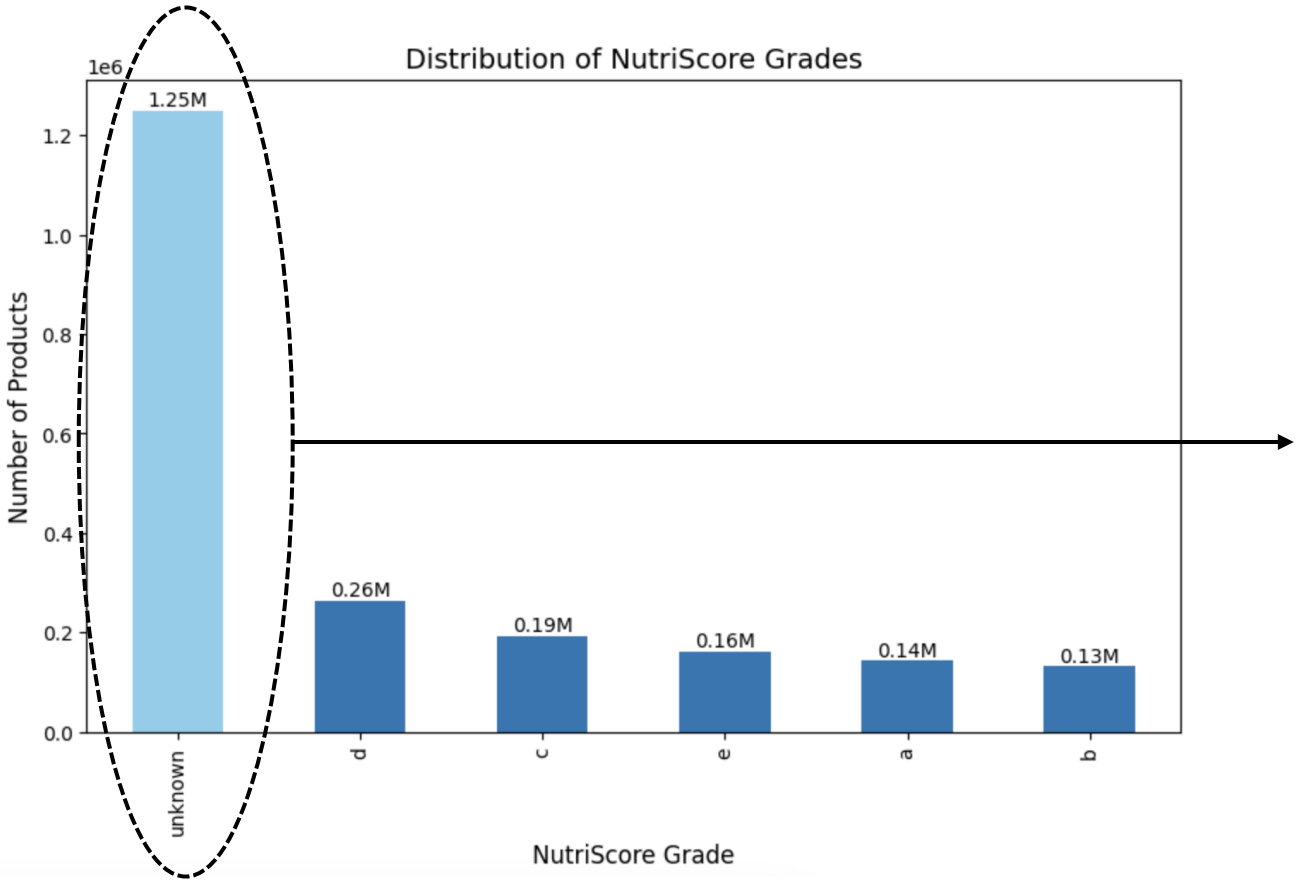
The feature **nutrition-score-fr** is the most important, with significantly higher importance than others

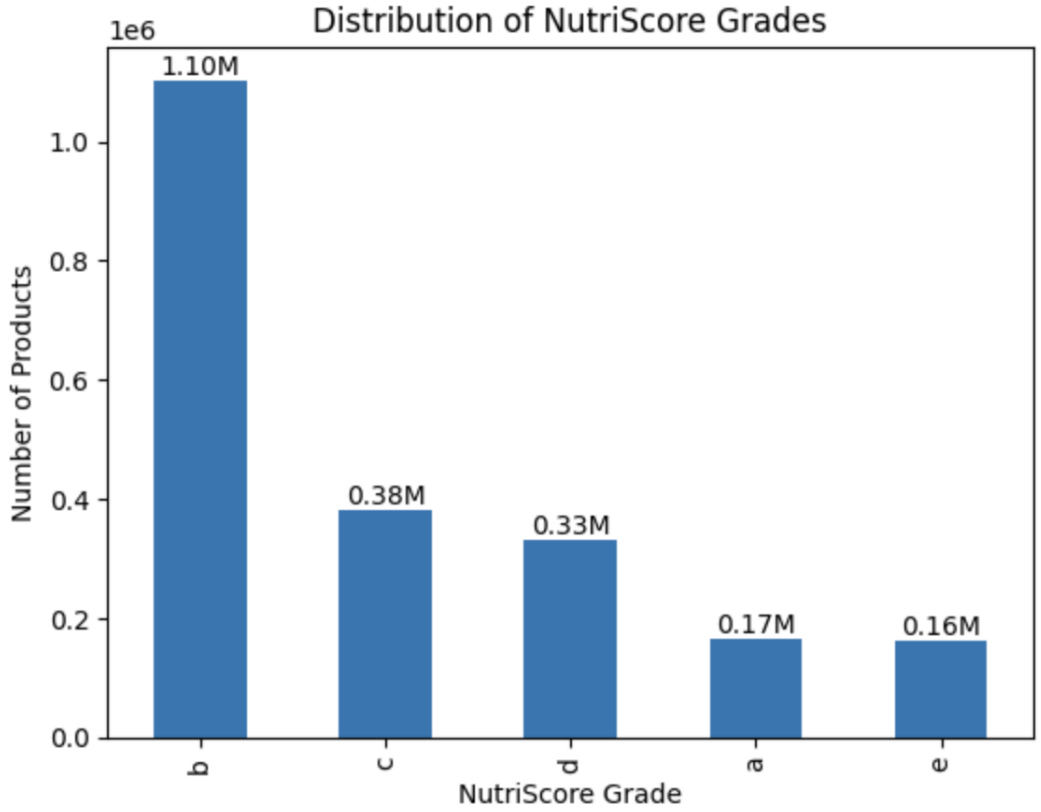
Energy and Salt are the next most impactful features, aligning with the NutriScore calculation methodology

Nutrition Score Grade Prediction

UNIVERSITY OF TWENTE.

Prediction in unknown nutriscore





Initial count of 'unknown' nutriscore_grade: **1,25 Million products**

Most of unknown nutriscore grade are categorized as “B” and most of them are ”sandwich product, meat, fruit juice”

Conclusion

UNIVERSITY OF TWENTE.

Sugar Content & Diabetes Rate

1. The trend of sugar levels is **slightly increasing on sweet product category**, but it's statistically insignificant
2. Globally, there is no high correlation between product sugar content and diabetes rate. While in Europe the highest correlated product categories such as '**Milks' and 'Fruit-Based Foods'** shows moderate correlation.

Additives Content

3. Europe shows a **stable trend in additives content** over the years, maintaining an average of 1.3 to 1.6 additives per product. This consistency aligns with Europe's stricter food regulations and public health policies
4. Category product with highest additives content is **confectionery, beverages, and instant food**

Relationship & Prediction

5. Key predictors such as **sugar content, salt and saturated fat,** were identified **as critical factors in determining NutriScore grades**
6. Despite 58.26% of products in the dataset lacked NutriScore grades, by implementing **Gradient Boosting** model, we accurately predicted missing grades with **97.94%** accuracy
7. The model's prediction showed **Grade B** as the **most common category** for missing entries, revealed moderate nutritional quality in ungraded products

Thank you

QnA

UNIVERSITY OF TWENTE.