

# Food Consumption Characteristics of the European Union

## Interactive Visualization Report - Data Visualization 2021 - Group Project 2

Nova IMS – Master of Data Science and Advanced Analytics

**Group G – Authors:** Maximilian Maukner (m20200645@novaims@unl.pt); Ehsan Meisami Fard (m20201050@novaims@unl.pt); Franz Michael Frank (m20200618@novaims@unl.pt); Steffen Hillmann (m20200589@novaims.unl.pt)

### Introduction

There has been a number of forces that has been driving changes in global food consumption. In the past decades, consumers have shifted to more lavish forms of nutrients. This is mostly due to the increasing food consumption power among most customers in the world. Despite the relationship of the buyers purchasing power and certain food consumption patterns, the interrelationships between the climate and geographical location of a country and its food consumption characteristics has been a thought-provoking subject to explore. More importantly, the consumption patterns are highly correlated with the overall health issues, such as obesity, diabetes and cardiovascular diseases which ultimately play a major part in determining life expectancy.

The present project intends to take a closer look at the similarities and dissimilarities in regard to food consumption and health issues among countries within the European Union which could function as a cornerstone to further studies.

### Dataset Description

The main dataset used originates from the Food and Agriculture Organization of the United Nations [1] which entails the data concerning the food supply in kilograms per capita per year for 179 countries. Yet this project aspires to explore countries within the European Union for the sake of a narrower scope of research as well as preventing incomplete information within the dataset. Furthermore, information regarding life expectancy, health expenditure, GDP, diabetes diseases and obesity per country have been obtained from Our World in Data [2], Opendatasoft [3] as well as Eurostat [4].

### Visualization and Interaction Choices

The present dashboard intends to maximize the number of available interactions and options for the user. In total there are six visualizations with all of them allowing interaction with the user.

#### **Food Consumption per country**

In the first visualization, the eating and drinking habits (kilograms per capita per year) are displayed through a geo map while illustrating the countries with the highest and lowest consumption of the chosen food category which is dependent on the subject that the user eagers to analyze. Besides, showing the mean and standard deviation offers the user more intuition to better understand the average consumption of the chosen nutrition category.

#### **General Health Information per Country**

Next, the bar chart displays the difference between the countries in term of health issues as well life expectancy and health expenditure of the respective country. The exhibition of the differences regarding the variables mentioned provokes further questions that can be a subject of discussion and research. For instance, the high variation in cardiovascular death rates despite identical results in terms of obesity and health expenditure where obesity hints at the food consumption and

exercise habits while health expenditure exposes the priority of health for the government as well as the economic aspects of the respective country.

### **Health Issues and Food Categories**

Subsequent to the map, a heatmap has been developed to investigate the correlations between the five health issue variables, which are obesity, diabetes prevalence, cardiovascular death rate, life expectancy and health expenditure, and the eighteen food variables.

The results prompt a set of new research questions that entail the interrelationship of each food category with the health variables listed.

### **Correlation Between Food Consumption and Health**

Similar to the plot mentioned previously, this visualization investigates the correlation between health issues and food variables, however, the scatter plot includes a third dimension which is the GDP per capita of the country illustrated as the size of the dots within the scatter plot. In addition to the option of using health and food variables, users are able to choose between boxplots or violin plot. While a box plot only shows summary statistics such as median and interquartile ranges, the violin plot shows the full distribution of the data.

### **K-Means Clustering**

Lastly, a clustering has been implemented which consists of two visualizations. On the left, users can select the clusters to inspect while on the right the resulting clusters can be compared with respect to a chosen variable. The clusters intend to minimize the amount of information by categorizing similar data items together which facilitates interpretations, while at the same time finding similarities between countries.

### **Technical Aspects**

The interactive dashboard was implemented using the open-source Python library *Dash*. In this process, a Python script was written within the programming environment *Visual Studio Code*, in which the various *plotly* figures were assembled in respective *HyperText Markup Language* (HTML) containers to create the desired dashboard. For the appropriate layout of the dashboard, a *Cascading Style Sheets* (CSS) file was integrated into the Python code. Subsequently, all the scripts collectively, including the individual data files in *csv* format, were uploaded to a *GitHub* repository. For the final release of the interactive, web-based dashboard, the cloud platform *Heroku* was used. Therefore, the entire *GitHub* repository was connected to *Heroku*, which enabled the ultimate deployment of the dashboard application.

### **Discussion**

The project intends to give the user the flexibility to explore food consumption patterns as well as health diseases in countries in the European Union. Additionally, it has the potential to function as the cornerstone for future work regarding the interrelationship between health and food consumption. It's crucial to note that any correlation does not necessarily mean causation. The domain of study is complicated as it includes aspects such as history, ethnology, sociology, agriculture, ecology, political economy and cultural. Therefore, additional research is highly advisable prior to making any conclusion or assumption.

Lastly, due to the size of the sample at hand which includes 27 countries, it is essential to take the weight and the influential impact of outliers into consideration while interpreting the results.

## **Dashboard**

The interactive dashboard can be accessed through the following link:

<https://health-eu.herokuapp.com>

## **References**

[1] Food and Agriculture Organization of the United Nations:

<http://www.fao.org/faostat/en/#data/FBS>

[2] Our World in Data:

<https://covid.ourworldindata.org/data/owid-covid-data.csv?v=2021-03-11>

[3] Opendatasoft:

<https://data.opendatasoft.com/explore/dataset/european-union-countries@public/export/>

[4] Eurostat:

[https://ec.europa.eu/eurostat/databrowser/view/HLTH\\_SHA11\\_HF\\_custom\\_227597/bookmark/table?lang=en&bookmarkId=1530a1e6-767e-4661-9e15-0ed2f7fae0d5](https://ec.europa.eu/eurostat/databrowser/view/HLTH_SHA11_HF_custom_227597/bookmark/table?lang=en&bookmarkId=1530a1e6-767e-4661-9e15-0ed2f7fae0d5)

## **GitHub**

The developed code as well as the used data can be accessed through the following link to the GitHub repository:

[https://github.com/FranzMichaelFrank/health\\_eu](https://github.com/FranzMichaelFrank/health_eu)