NTNU

Kunnskap for en bedre verden

Department of Computer Science

TDT4259 - Applied Data Science

Improving Public Health in the UK Through Data Science

Authors:
Hedda Wasskog Aamo (522175)

Julie Åsbo Foss (525334)

Lukas Øystein Normann Stjernen (526027)

Mina Jørstad Uthus (506483)

Solvei Yang (527736)

Maren Aksdal Åsvang (504590)

27.11.2023

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Recent years have challenged households worldwide with falling incomes, rising living costs, widened inequality, and welfare reforms that have made it increasingly difficult to afford a healthy diet. Households living in poverty are consistently linked to poorer diet quality and dietary related health issues. They are unable to afford good quality and high nutrition foods (Thompson 2021).

Poverty can be characterised as dynamic and uncertain, and those living in poverty can experience difficulty planning for the future. Low income households therefore find it challenging to plan and perform dietary resilience. These households suffer from chronic uncertainty, such as insecure housing conditions or not knowing how much money one has disposable from week to week (Thompson 2021). This can force households into short-term and potentially harmful dietary practices (Whittle et al. 2020). Socio-economic differences related to diet can lead to health inequalities and outcomes such as obesity, type 2 diabetes and malnutrition (Thompson 2021).

A health survey conducted in England in 2021 found obesity among 26% of adults (NHS 2022). Diabetes UK links health issues to poverty by stating that people living in poverty are more than twice as likely to develop type 2 diabetes compared to those with average income (DiabetesUK 2021). Approximately one in every three children aged 1 to 15 in the UK is classified as overweight. There is strong evidence in the UK that obesity disproportionately falls on children from poorer households. The disproportion is likely explained by difference in lifestyle and dietary intake, as healthy foods tend to cost more than unhealthy foods. A study from 2018 concluded that adolescents from poorer households should be considered an important target group for interventions aimed at improving diet and health (Noonan 2018).

The IFS Deaton Review of Inequalities documented and analyzed how wealth and income inequality has evolved in the UK the past decades. The results show that the gap between the poorest 10% of the population and the richest 10% has widened significantly since the data was first collected in 1961. Especially those within the top 1% of UK earners have seen a significant increase in income in comparison to the average salary development in the country. The data also reveals that the UK is ranked among the highest in terms of income inequality compared to other developed countries, and that the UK has been experiencing a steady growth in relative poverty among working-age adults (Brewer and Wernham 2022).

To illustrate the development of income inequality in the UK, data gathered from The Office of National Statistics (ONS 2020) has been used to create Figure 1. The graph shows the wage trends of deciles 1, 5 and 10 from 2003 to 2022. Deciles are created by dividing the total income into ten equal-sized groups based on the total equivalised income. Decile 1, 5, and 10 represent the poorest, middle-class, and richest, respectively. By comparing these three income groups, a clearer picture of the income disparity in the UK is obtained. This, combined with rising living costs, underlines the importance of investigating health consequences based on income level.

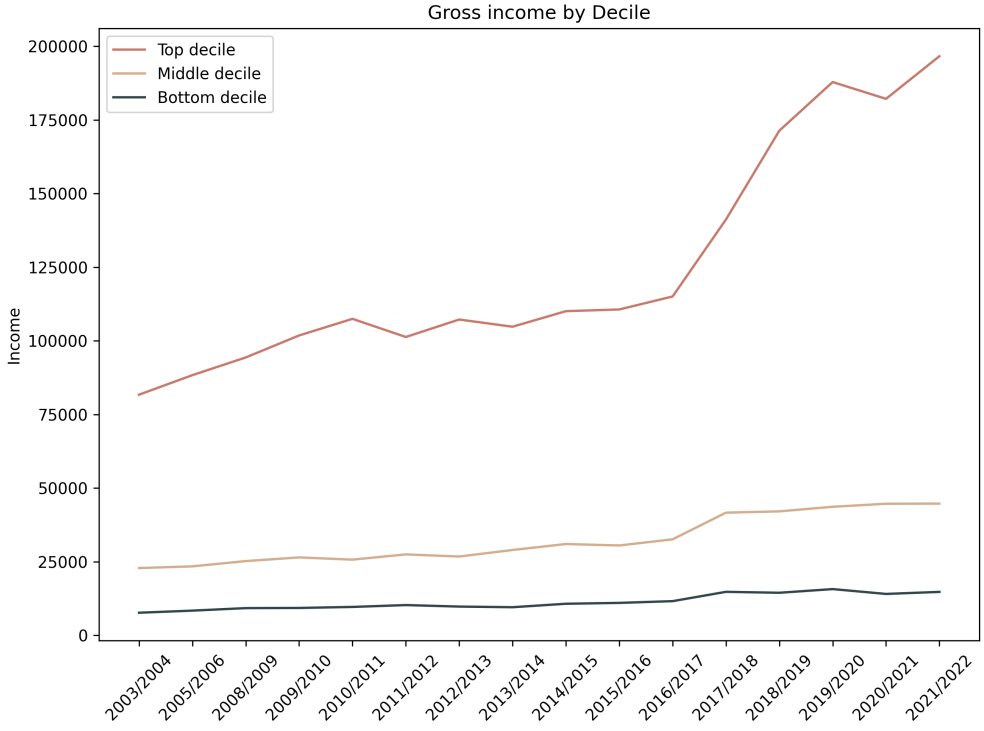


Figure 1: Gross income by decile in GBP

Hypothesis:
Due to the aforementioned developments in income inequality and the effects of poverty on diet, this paper will analyze the correlation and eventual causality of the two topics, specifically in the UK. Due to the nature of the topic and task, a hypothesis-driven approach will be used. The hypothesis is therefore as follows:

Household income has affected the lower deciles health, diet and food consumption in the UK.

1.2 Problem description

The objective of this report is to provide the UK Government with a clear understanding of the situation and problem regarding income and nutritional intake, with actionable measures to counteract negative health effects. By analyzing food consumption data provided by the UK Government, the ultimate goal is to present precise insights that can assist the UK Government in addressing inequality and the health-related problems stemming from it. These inquiries will provide valuable guidance on where and what improvements should be prioritized by the UK Government to improve the current situation.

To reach the objective, the following questions need to be answered:

1. How much of each food group does each income decile eat?
2. Which nutritional food groups are the lower deciles in lack of?
3. Which unhealthy food groups are the lower deciles in excess of?
4. Is there a significant difference in food consumption based on income level?

1.3 Team

The team consists of six students, five students studying at NTNU Trondheim and one exchange student from DTU Copenhagen. Table 1 gives an overview of the team members.

Member	Study program	Additional Responsibility
Hedda	BSc digital business development and MSc management of technology.	Language and formal requirements.
Julie	BSc digital business development and MSc management of technology.	Action planner.
Lukas	BSc data engineering and MSc computer science and engineering.	Illustrator.
Maren	BSc bio-engineering and MSc management of technology.	Data validator.
Mina	BSc digital business development and MSc management of technology.	Data collector.
Solvei	MSc industrial economics and technology management, specialization in computer science and AI.	Meeting manager.

Table 1: Team members

1.3.1 Roles and responsibilities

All team members have participated in roles of data scientists and analyzers. This decision was made due to all team members expressing their interest for taking part in the entire process. As listed in Table 1, every team member was also assigned an additional responsibility. This was done to ensure that every aspect of the project was adequately covered. The team has had frequent meetings throughout the project period, contributing to an even workflow and ensuring information flow among all members. Joint programming sessions were held during some of the meetings, to ensure that everyone could participate in and learn data collecting and processing methods despite their academic background.

2 Background

2.1 Project objective

As described in Section 1.2, many health issues in the UK stem from bad diets. The purpose of this project is therefore to help the UK Government implement measures that can help mitigate consumption of unhealthy food. Mitigation can be achieved by allocating resources where they are most needed, and preparing an action plan to target those with high consumption of unhealthy food. The objective of the report and how it will be achieved are listed below.

What does the project want to achieve?
Create a specified action plan for the UK Government to improve public health among the people in the UK through diet.

How to achieve the project objective

- By identifying the differences between the income groups, and placing them into an understandable context, the report aims to answer the posed questions from Section 1.2. These need to be answered in order to develop recommendations.
- Develop recommendations regarding food consumption to improve overall public health, based on theory, experiences from other countries, and findings from the data. Combine

these recommendations into a comprehensive action plan that will be given to the UK Government to carry out.

2.2 Scope

The report will focus on food consumption trends among income deciles 1, 5 and 10, spanning the years 2001 to 2019. Deciles 1, 5 and 10 were chosen to represent lower, middle and upper class respectively. Dividing the population into these classes enables comparative analysis of food consumption across different income levels.

The main reason for choosing the time span 2001-2019 was due to the availability and accessibility of the relevant data. Before 2001, food consumption data was recorded by the UK Government, but only published in physical form. Due to the time- and resource constraints of this project, only the data available digitally in spreadsheet form is included in the analysis. The food data for the years after 2019 have not been made available online, and was therefore not included.

Although food consumption data for the previous four years has not been included in the analysis, the team argues that the analysis still provides an overall and general understanding of food consumption trends among chosen deciles. Especially when considering factors such as the Covid-19 pandemic, Brexit, and high inflation rates in Europe, one can argue that the years after 2019 may have large deviations from general trends. As this report looks at the overall trends in food intake and income, excluding years with unusual deviations from normal conditions shouldn't significantly impact the report's results. Rather, a project specifically investigating the impacts of these world-impacting events should be done, and is discussed further in Section 5.5.

In the first phase of the analysis, only the main *Food Categories* from the food dataset - introduced in Section 3.1 - will be analyzed. These food categories are based on a categorization scheme that the UK Government has used since the 1940s for data collection on food consumption, rather than the more traditional five major food groups. The reason for only looking at the main food categories is for the project to be more resource-efficient, as well as allowing for a more general overview of the food consumption patterns in the data. Where it is appropriate, the analysis divides the food category into its constituent subgroups to see what has affected the main category most.

Information about food consumption does not take into account eating-out data, meaning that the report only considers expenditure on food brought into the home, such as groceries or take-away. Geographical data on lowest income households is not taken into consideration through the analysis, but is lightly discussed in Section 5.5.

This report will present a descriptive analysis rather than a predictive one. The nature of the overall project objective makes descriptive analysis a suitable approach, since the focus is on interpreting existing data to create concrete recommendations and an action plan. Rather than predicting future trends, the chosen approach is to compare future data with defined Key Performance Indicators (KPIs), to evaluate the impact of implemented measures and emerging trends.

2.3 Data strategy

As described in Section 2.1, the purpose of this project is to provide an action plan to the UK Government to improve public health, by analysing food consumption and developing recommendations. To reach the desired outcome, a data strategy is necessary, as it provides a systematic framework from problem identification to solution deployment. It helps make sure that no crucial steps are skipped, and fosters comprehensive understanding and analysis of the data for all parties involved (Sargo 2023).

For this project, multiple data strategies were considered. As the team members had limited knowledge on data strategies beforehand, only the most popular and well-documented strategies were considered. The evaluated strategies were BAM (*Business Analytics Model*) (Hindle and

Vidgen 2018), SEMMA (*Sample-Explore-Modify-Model-Assess*), and CRISP-DM (*Cross Industry Standard Process for Data Mining*) (Azevedo and Santos 2008).

The team landed on using the CRISP-DM framework, for a number of reasons. Firstly, CRISP-DM encourages focus on business goals, to ensure that project outputs provide tangible benefits to the organization. In other words, it helps ensure that the business goals remain at the centre of the project throughout. CRISP-DM also provides an iterative approach, including frequent opportunities to evaluate the progress of the project against its original objectives. Finally, the CRISP-DM methodology is both technology and problem-neutral. Whatever the nature of the project, CRISP-DM will still provide a framework with enough structure to be useful. This proved to be the biggest factor in deciding which design strategy to use, as BAM and SEMMA were perceived as being either too technical or not technical enough (Sridharan 2019).

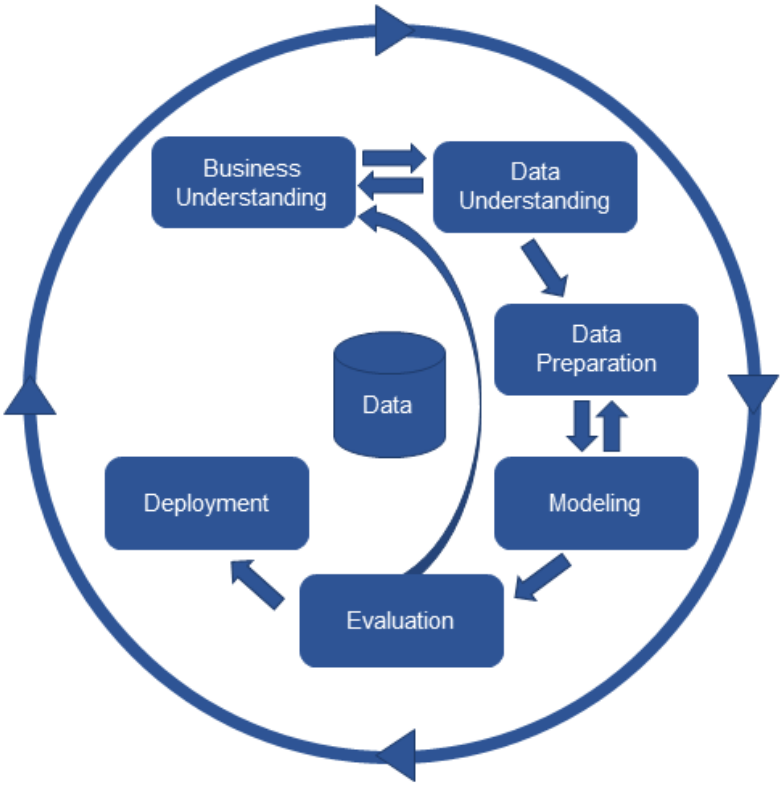


Figure 2: CRISP-DM framework
(Taylor 2018)

Figure 2 depicts the different phases in the CRISP-DM model. It emphasizes adaptive transitions between project phases, as seen from the arrows pointing back-and-forth between different phases. This means using the knowledge gained from later phases to go back and reiterate the decisions made in previous phases. The following sections describe what was done in each phase for this project.

2.3.1 Phase 1: Business Understanding

The purpose of the Business Understanding phase is to understand the project objectives and requirements from a business perspective. This knowledge is then formulated as a data mining problem and a preliminary plan is developed.

In this project, the UK Government was considered the business organization and main stakeholder, with income and health as the main points of focus. It was therefore crucial in the Business Understanding phase to understand the income demographics in the UK, and what food habits

they have. This phase was therefore used to collect more in-depth information on the correlation between health issues and poverty.

2.3.2 Phase 2: Data Understanding

To be able to reach the defined objective, spending time understanding and becoming familiar with the data to be able to best utilize it was prioritized. In this phase, it was discovered that analyzing all available data provided in the dataset would be too time-consuming compared to the benefits it could provide. It was therefore decided in this phase that only the main food categories would be analyzed, as discussed in Section 2.2. If any of the food groups were found to exhibit noticeable trends, a verdict would be made on whether the food group should be inspected further with more detailed analysis.

2.3.3 Phase 3: Data Preparation

The preparation part of the framework involves preparing and constructing a final dataset by transforming and cleaning the initial raw data (Azevedo and Santos 2008). The preparation step is important to ensure that the data going into the modeling or visualizing stage, can give an accurate representation of the data. By having a thorough preparation stage, the quality and integrity of outcomes and findings increase. There were many inconsistencies found in the dataset, that may have caused issues if not detected in this phase. How the datasets were cleaned and prepared is described further in Section 3.1.

2.3.4 Phase 4: Modeling

After preparing datasets for analysis, it is necessary to select appropriate modeling techniques based on the defined objectives (Hotz 2023). CRISP-DM guidelines suggest iterations of the modeling stage to improve the model throughout the CRISP-DM lifecycle.

This stage was initiated with a planning phase where the team agreed on which techniques to use, what information from the dataset to display, which kinds of charts or diagrams to use in visualization, and what would be the goal of each model. Section 2.1 is a result of this phase. As recommended by CRISP-DM guidelines, the modeling stage was iterative to ensure the models aligned with the overall objective of the report. The iterations consisted mostly of choosing which food groups affect the health of lower deciles the most through developing graphs for each group.

In this phase the team settled on displaying the data mainly using line charts, and to choose specific food categories to model based on the main sources of the UK population's dietary problems.

2.3.5 Phase 5: Evaluation

The evaluation stage involves choosing which chart meets the requirements and decide what to do with the results (Wirth and Hipp n.d.). It is important to evaluate all steps executed to build the chart to ensure it meets all defined objectives.

Charts for each of the main food categories were developed to examine which ones deviated the most from health recommendations given by the UK Government. This method provided visualisations of the different deciles intake, which ensured that the categories with the largest differences was examined and included. The food categories selected for further analysis were those with the most significant deviations, and recommendations for measures to be implemented by the UK Government were then developed based on these findings.

2.3.6 Phase 6: Deployment

The deployment stage involves translating data-driven insights into practical implementation. The knowledge gained from the different types of models must be organized and presented in a manner that is understandable for stakeholders and other important actors (Wirth and Hipp n.d.).

Applying design thinking principles, as outlined in Section 2.3.7, the project concentrated on prioritizing the user perspective in presenting data findings and outlining deployment steps. As mentioned, the objective is to help the UK Government by offering concrete recommendations for the future. The first step was therefore to convert results from the analysis into actionable recommendations. Based on this, a step-by-step action plan for the UK Government, including time frame and descriptions, were provided, as seen in section Section 5.1.

2.3.7 Design Thinking

CRISP-DM was used in combination with *design thinking*, a problem-solving approach that generates solutions based on empathic and user-driven methodologies. Empathising with the users to better understand their needs is a crucial part of design thinking (Pham et al. 2022). To ensure that the findings and recommendations are understood by the main stakeholder, the UK Government, the team used these user-driven principles to keep stakeholders' views in mind throughout the CRISP-DM process.

The human-centered approach of design thinking is normally obtained through collaborative work where stakeholders are involved. Since the project's main stakeholder was the UK Government, it was not possible to directly involve them in the problem-solving process. An effort was made to obtain a similar human-centered approach through extensive research and understanding of the UK in relation to poverty and health, the two main aspects of the hypothesis. Although the UK Government could not be directly involved in the problem solving, direct communication was made between the project team and a representative from the UK Government Department of Environment, Food, and Rural Affairs to ensure thorough understanding on how data was collected and systematized. By combining the CRISP-DM framework with aspects from design thinking, the team was able to work systematically towards providing recommendations targeted specifically to the UK Government.

3 Method

The main dataset used for this project was the "Purchased quantities of household food & drink by income decile" dataset. This dataset was used as the basis for the analysis, as it included all necessary information regarding the food consumption for the different deciles. In addition, the dataset "Effects of taxes and benefits on household income" was used to supplement with information regarding the income development in the UK. This provided context and background for the project, as described in Section 1.1. The following sections will provide more information on the contents of the datasets, and how they were processed.

3.1 Datasets

3.1.1 Family Food

Family Food is an annual publication by the Department for Environment, Food & Rural Affairs in the UK (DEFRA 2023). The data in this publication is collected from a sample of households in the UK, and data is gathered by having participants self-survey their food consumption using provided diaries, supported by till receipts of all purchases. Food consumption data is collected over a two week period, where all known quantities are recorded in the diaries. If there are any gaps in the entries, estimates are made. The estimates are based on data collected in the Family

Food Module of the Living Costs and Food Survey, as well as on adjusted data collected in the National Food Survey (DEFRA 2023).

The dataset used for this project ("Purchased quantities of household food & drink by income decile") is one of the datasets included in the Family Food publication. It provides data on the food consumption of each decile in the period 2001-2019, categorized by food categories. Only this specific dataset has been processed and cleansed, as explained in the following section.

Processing the Family Food dataset
To process the used dataset, the data was first downloaded from the official UK Government source as an excel file. Due to inconsistencies in the data, some manual labor was necessary to prepare the data before processing it in python:

1. Only the relevant spreadsheets (decile 1, 5 and 10) were chosen and made into their own files.
2. Parts of the spreadsheet which could clutter the analysis were removed, such as the headline and information at the top of the spreadsheets, that did not offer anything to the analysis.
3. As there were some inconsistencies in how the years were written, the members responsible for data processing wrote all years in a consistent manner (e.g 2001 instead of 2001-02).
4. To ensure that the dataset only consisted of needed information, unwanted columns were removed.
5. Some food categories and food groups were wrongly placed in the spreadsheet, so these were cleaned up to be consistent. For instance, within the food category "milk and milk products excluding cheese", "cheese" and "carcase meat" were listed as food groups, when actually being their own food categories.

After the data was cleansed, only attributes needed for the analysis remained. This left the dataset with these features for each decile:

- Year (2001-2019)
- Unit (in ml or g)
- Food category (e.g milk, meat, fish...) or Food group (e.g. fresh fruit, white fish, ...)

3.1.2 Income

The supporting income dataset ("Effects of taxes and benefits on household income") is published by the Office for National Statistics. This data is published yearly as an excel workbook, and contains different types of data on household income, split into different sheets. These sheets contain data for financial years, by quintile and decile groups, country and region and tenure type, as well as incomes, taxes and benefits, and household characteristics of retired and non-retired households in the UK (ONS 2023a). The data in the dataset relevant for this project was the sheet "Average household incomes, taxes and benefits of all individuals by equivalised household original income decile group".

Processing the income dataset
The income dataset was collected by downloading all available excel data, and processing it in python using the Pandas library. Due to some inconsistencies in the data, some manual preprocessing had to be performed in order to make the Python code run seamlessly. Some of these inconsistencies were:

- The sheet where the relevant income data is located varied for each year. Because of this, the excel workbook for each year had to be manually investigated in order to know which sheet to refer to in the Python code.

- The indices for the data was inconsistent between the years, which lead to more manual investigation to find the correct index offsets.

When the preprocessing stage was finished, the data for each year was collected in a Pandas dataframe and plotted using the library Matplotlib. The resulting plot showed the development of final and gross income for the chosen decile groups. The code for processing the income dataset can be found in the appendix.

3.2 Tools

The team used Python as programming language to work with the datasets, as it offered many possibilities due to its broad field of use. To benefit from its wide range of applications, the team used two Python libraries, *pandas* and *matplotlib*. Pandas is a Python library that offers the user a high-level building block, working well for doing practical and real world data analysis (pandas n.d.). Pandas was used to analyze the data, while *pyplot* from the matplotlib library was used to visualize the findings. Pandas was used to create dataframes from the files containing the datasets. The built-in functionality for reading csv and excel files was the primary utilization of Pandas in this context, but it was also used to create dataframes containing the result data to make visualization easier.

For visualization of the data, different types of charts, such as line-, bar-, and pie-charts were tried out and discussed. After some trial and error, the team discovered that the line charts provided visualizations that were the most intuitive for readers. Line charts are effective in displaying trends and variations over time. A limited number of lines in each diagram ensures the reader gets a comprehensive and accurate understanding of the presented plots. In addition, some other chart types have been used, selected based on the specific point being conveyed. To provide accessible visualizations, the project team chose colors proven to be easy to separate for people with colorblindness (Kilin 2022). The colors chosen for the visualisations are therefore #31454A (decile 1), #D3AD8D (decile 5), and #C6796C (decile 10).

Github was used to share data and code. Google Drive was used as the main tool to keep meeting documents and resources available to all team members. Both these tools ensured that all members were up to date on the current state of the project.

4 Analysis, Evaluation and Interpretation

4.1 Analysis and Evaluation

The main food categories fruit and vegetables, fish, and sugar and preserves are the focus of the analysis. These categories were chosen as they contribute to a healthy diet as recommended by the UK Government (NHS 2023c), and because these charts showed noticeable differences between the deciles and deviated from health recommendations. The following section will provide an analysis and interpretation of these food categories. Each of the categories will first be analyzed based on the empirical data gathered from the Family Foods dataset, then interpreted in light of how the found food consumption habits affect the health of the UK population. The findings from the presented charts are substantiated by incorporating research on relevant fields. As mentioned in Section 2.3.5, the team chose not to include all charts that were made. The categories that are not included in the analysis were omitted. These omitted charts can be found in Github, linked to in the appendix.

4.2 Interpretation

4.2.1 Fruit and vegetables

Nutrient-rich and healthy foods are reported to be three times costlier than their less nutritious, calorie-dense counterparts. Consequently, families facing financial challenges often rely on cheaper, less nutritious, and higher-calorie food options due to the structure of the UK's food system (Scott et al. 2018). This is particularly true for fruit and vegetables, which form the cornerstone of a healthy diet. Diets which are rich in fruits and vegetables (henceforth referred to as FV), are associated with reducing risks of different types of chronic diseases, type 2 diabetes and cancer (Wang et al. 2016), and are therefore important for healthy living.

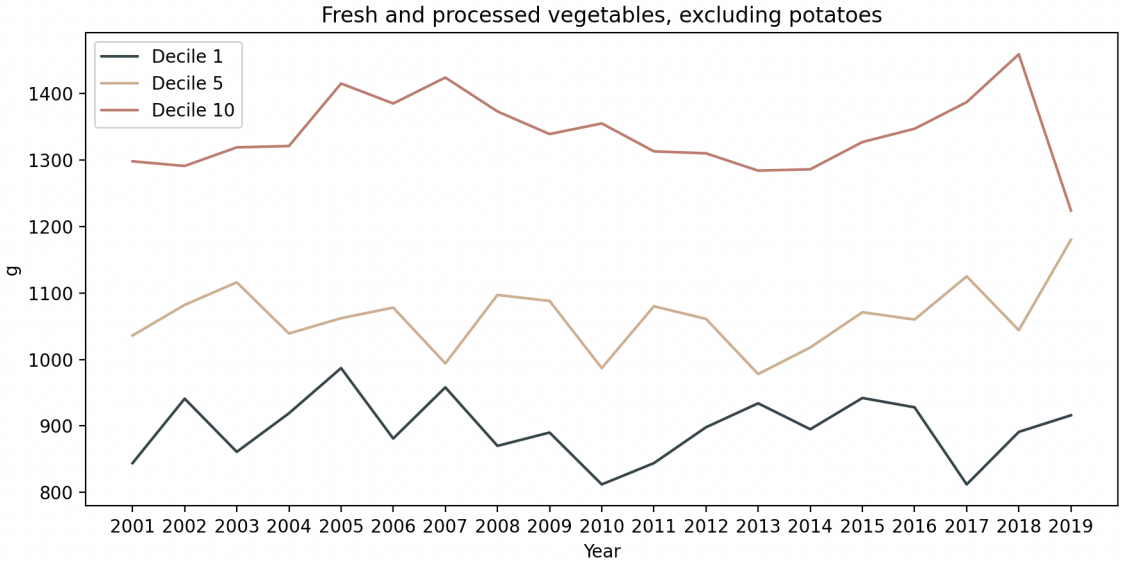


Figure 3: Fresh and processed vegetables, excluding potatoes

Figure 3 depicts the FV intake of decile 1, 5, and 10, not including potatoes. This chart shows a clear difference in the amount of FV eaten based on income level - the top 10% of the UK population eat approximately double the amount of FV in comparison to the bottom 10%. The World Health Organization (WHO) recommends that the daily required intake of FV is 400g a day (Rickman et al. 2007). Table 2 shows the calculated daily FV intake of the different deciles, calculated from the FV average from 2001-2019. As can be seen from the table, both decile 1 and 5 clearly lie below the recommended intake of 400g, while decile 10 is just above the recommended amount. This is a significant observation, as vegetables contain several vitamins and minerals, many of which are antioxidants that help reduce the risk of multiple diseases. Insufficient amount of FV in the diets of decile 1 and 5 poses therefore a significant health risk to a large portion of the UK population.

Decile	Average Weekly FV (g)	Daily FV Intake (g)
1	1773	253
5	2150	307
10	2857	408

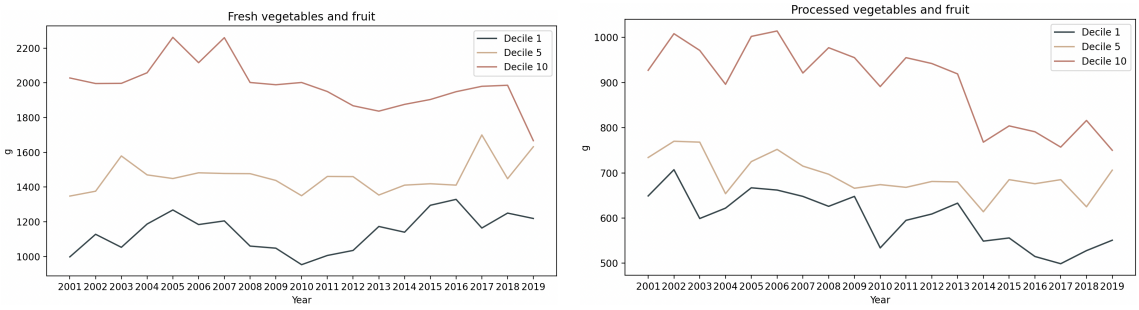
Table 2: Average weekly and daily intake of FV

Some European countries also suggest eating up to 650g of FV a day to reap the full nutritional benefits. This suggests that all income groups could benefit from eating more fruits and vegetables (EuropeanCommision 2023-08-09), not only the lower deciles.

Fresh vs Processed FV

An intriguing angle to explore within the FV category involves examining the contrasts between the consumption of fresh and processed FV. Studies have shown that processed FV often has more concentrated fiber due to the water being removed during the processing (Slavin and Lloyd 2012), but lower amounts of vitamins and minerals (Rickman et al. 2007) than fresh FV.

As can be seen from Figure 4, the tenth decile eats more of both processed and fresh FV. However, the difference between decile 1 and decile 10 is less stark in processed FV (an approximate difference of 250g) in comparison with fresh FV (an approximate difference of 800g). This further supports the argument that lower income households struggle with eating enough FV. As lower income households eat proportionally more processed FV which also have less of the health risk-reducing nutrients, they are at a greater risk of health issues such as cardiovascular diseases and cancer. This can be correlated with data from The Health Foundation for UK which shows that people from lower income households tend to be more prone to illness (Tinson 2020).



(a) Fresh FV

(b) Processed FV

Figure 4: Graphs of the subcategories for fresh and processed FV

Potatoes

Potatoes have been a significant and reliable part of the Western diet for many hundred years (Robertson et al. 2018). Today, potatoes continue to play a crucial role in daily nutrition and are recommended as part of the UK diet (NHS n.d).

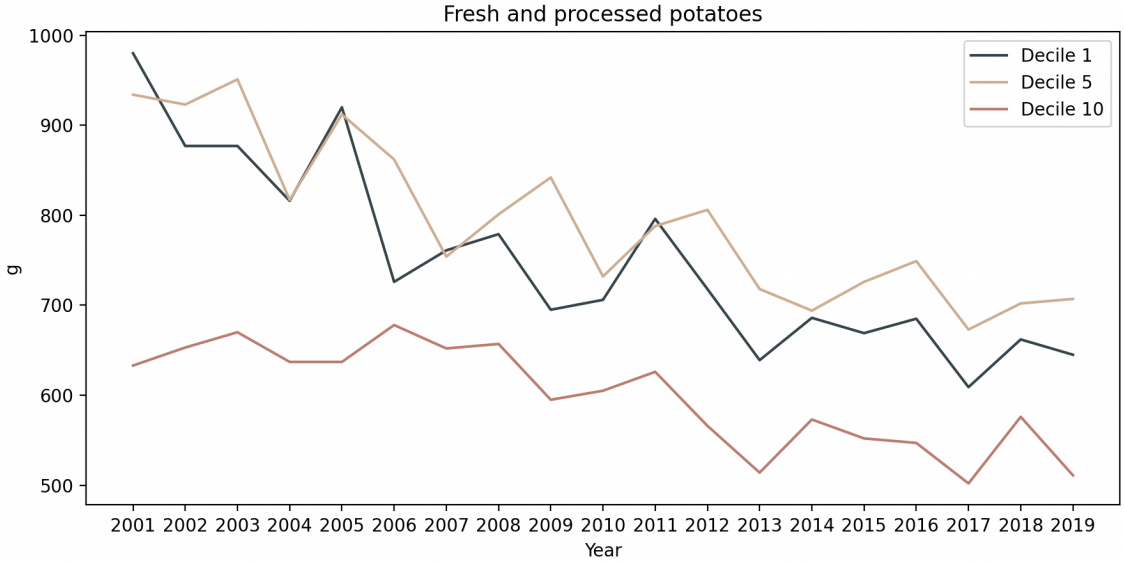


Figure 5: Fresh and processed potatoes

Figure 5 shows consumption of fresh and processed potatoes in deciles 1, 5 and 10. There is a visible distinction between the three income levels consumption, where the top 10% of the UK population eat significantly less potatoes than the bottom and middle 10%. All deciles have had

a general decrease in their consumption from 2001 until 2019. Researchers point to associations with weight gain, fear of chronic diseases, change in preference, or introduction of substitutes as possible explanations for this downwards trend (Robertson et al. 2018).

The difference in potato intake can be explained in many different ways, one of which is income. Potatoes are known for being a cheap and filling part of peoples diets (Mann 2011), which may be the reason as to why the middle and lower deciles may turn to the potato rather than other food groups that (are less filling but) cost more. This is consistent with literature, that indicates that lower income households and those with lower education often choose the potato over other starchy substitutes like rice (Touvier et al. 2010). The upper decile, however, does not need to search for the cheapest alternative, and the potato may therefore be a less attractive choice.

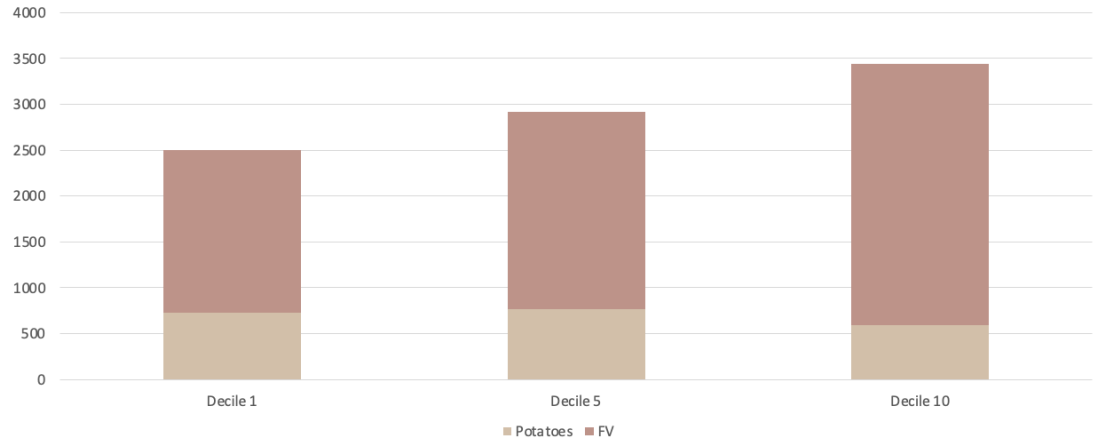


Figure 6: Comparison between FV and potato intake between deciles

Although the middle- and lower deciles of the population consume more potatoes, it needs to be addressed that potatoes does not replace the five-a-day FV that are recommended by the National Health Service (NHS). This is due to the fact that potatoes are nutritionally very different from traditional vegetables, and can therefore not be considered a substitute. Rather, potatoes are more similar to refined carbohydrates (Onque 2023). Figure 6 visualizes the proportion of FV and potatoes consumed by each decile. This figure shows that in comparison to decile 10, decile 1 and 5 consume significantly less FV, but also more potatoes - meaning that these deciles seemingly use the potato as a substitute for FV in their diet. This further emphasizes and supports the argument that lower deciles are in need of more FV in their daily nutritional intake.

4.2.2 Fish

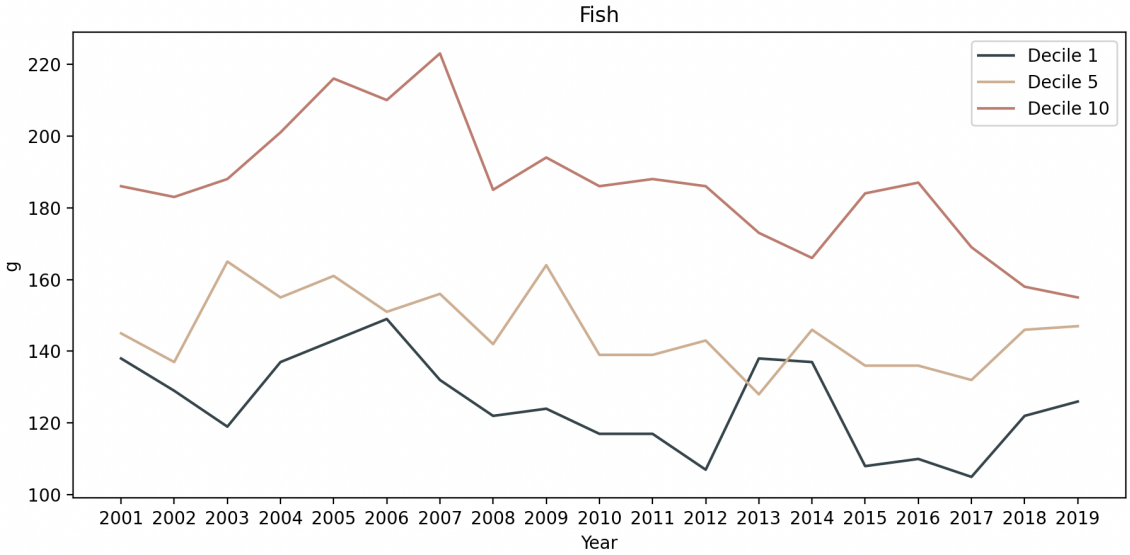


Figure 7: Fish

The fish category includes all fish and shellfish, whether they are fresh, frozen, or tinned. Figure 7 shows that, despite a declining trend, decile 10 has been consuming the most fish of the three income groups. In spite of this, the upper decile still ate less fish than recommended, even at the peak in 2007. The NHS guidelines for food recommend the population to aim for 2 portions of 140 grams of fish every week, where one should be oily fish (NHS 2023c). The reason the NHS specifically recommends one portion of oily fish, is because of its relatively high content of healthy omega-3 fatty acids (NHS 2023a). This recommendation raises an interesting question of how the consumption of oily fish vs non-oily fish distributes among the income groups. One way of investigating this is comparing the charts for the subcategories salmon (oily) and white fish (non-oily).

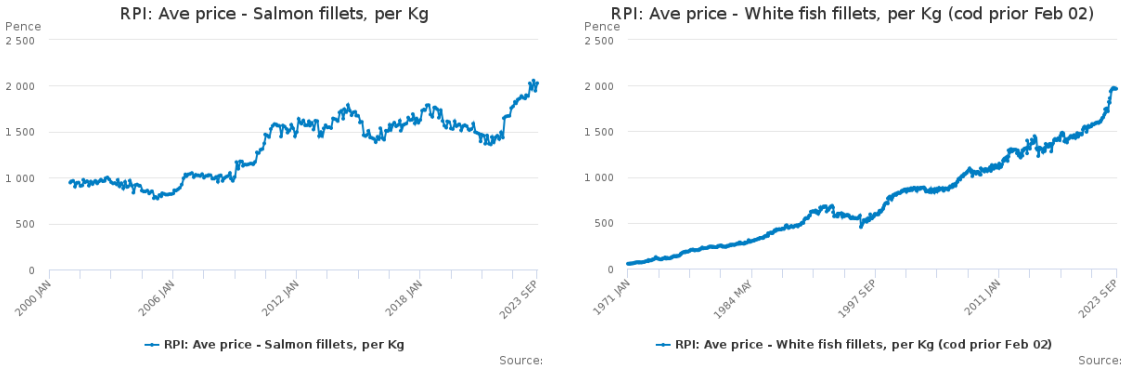
(a) Salmon

(b) White fish

Figure 8: Graphs of the subcategories for salmon and white fish

Figure 8 shows charts for the consumption of salmon and white fish. The chart for white fish contains a lot of noise, but shows a declining trend in all three income groups. The chart for salmon shows a different picture; on average the upper decile eat more than twice as much salmon than the middle and lower decile. This comparison shows that the middle and lower deciles eat

less oily fish than the richest decile, and thus do not gain as much health benefits from the fish they eat compared to the top earners. This, in addition to the fact that none of the three income groups meet the recommended weekly amount of fish, shows that public health could be improved if all income groups increased their consumption of fish.



(a) Salmon fillets, adapted from (ONS 2023-10-18a) (b) White fish, adapted from (ONS 2023-10-18b)

Figure 9: Average price for salmon and white fish fillets per kg

The difference in consumption of white fish and salmon is highly likely to be related to the price of each of them. Figure 9 shows that salmon has been more expensive than white fish for the vast majority of the years measured. This is consistent with the pattern in Figure 8, where the richest decile eat the most salmon.

4.2.3 Sugar and preserves

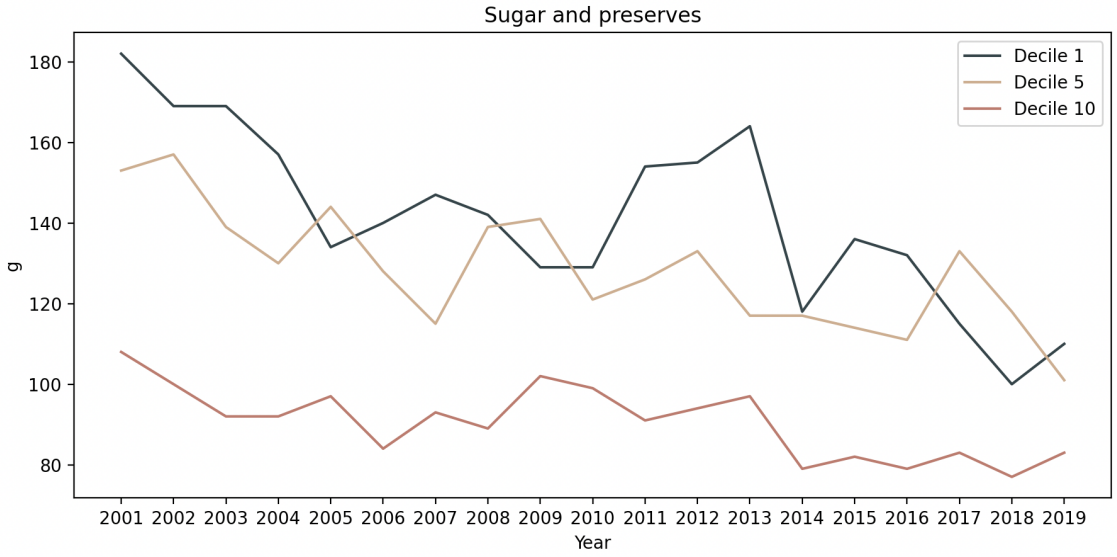


Figure 10: Sugar and preserves

The food category "Sugar and preserves" includes sugar, jams and fruit curds, marmalade, syrup and honey. As seen in Figure 10, decile 10 consumes the least within this category, while deciles 5 and 1 have a higher consumption and overlap over the years. Decile 1 seems to consume the most, but decile 5 have in some time periods exceeded decile 1.

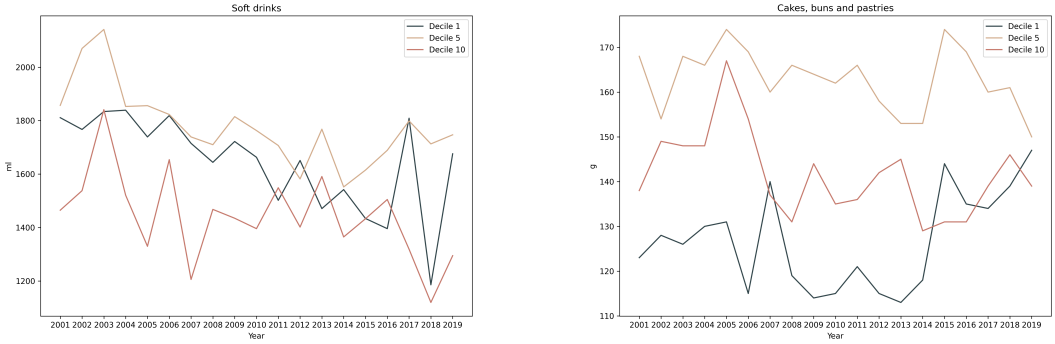
Sugar and preserves consumption has decreased among all deciles, although the poorest decile has seen the biggest reduction (down 39,56%). This could be a result of the UK Government's

measures to improve public health, for example by releasing reports like the *Sugar reduction* report in 2015 (Tedstone et al. 2015), and through the Eatwell guide that shows how to achieve a healthy diet (NHS 2023c). In later years, the sugar tax introduced in 2018 has also seemingly had an effect on sugar intake (Treasury 2018).

The NHS recommend no more than 30 grams of free sugars a day, and define free sugar as sugars added to drinks or food, or sugars in honey, syrups and nectars (NHS 2023b). Figure 10 shows that decile 1 in 2019 had a weekly consumption of 94g free sugars (sugar, syrup and honey), meaning that the daily intake of free sugar was approximately 13.4g. Seen isolated, this is well under NHS's daily intake recommendations. However, taking other sources of free sugar into account, this becomes a problem. When adding free sugar intake from soft drinks and cakes, buns and pastries, the daily intake in 2019 for decile 1 increases to 34.2 grams. This is calculated based based on Pepsi having approximately 11 grams of sugar per 100 ml (Pepsi n.d.), and an average of 36 grams of sugar in cakes and pastries (Clarke 2018). Using the same calculation method, decile 5 and 10 were found to be within the recommended amount of maximum sugar intake, as shown in Table 3, although decile 5 is close. The calculated consumption amount does not consider other sources of free sugar such as confectionery, juice and cereals, meaning that daily intake of free sugar is likely even higher than the calculated amount, and that also decile 5 possibly consumes more than what is recommended. However, only analysing sugar, soft drinks, and cakes, buns and pastries reveals a daily intake difference of over 10 grams between decile 1 and 10.

Decile	Daily Sugar intake (g)
1	34,2
5	29,5
10	23,5

Table 3: Daily sugar intake (2019) estimated from food groups Sugar and preserves, Soft drinks and Cakes, buns and pastries.



(a) Soft drinks

(b) Cakes, buns and pastries

Figure 11: Supplementing graphs for soft drinks and cakes, buns and pastries

The findings that lower income households consume too much sugar aligns with research that suggest a correlation between income levels and sugar intake. Research find that lower socio-economic groups are more exposed to advertisements for sugar and sugary drinks than others (UK-Parliament 2022). Such exposure and temptation are factors that could potentially contribute to Public Health England's (PHE) findings from 2015, that indicate that children growing up in underprivileged areas are twice as likely to be overweight than those from privileged areas. Sugar intake is also found to be higher amongst adults and teenagers in low income groups (Tedstone et al. 2015). Marketing bias towards unhealthy foods in deprived areas should therefore be a matter of concern for the UK Government.

Although decile 5 was found to be just within the recommended amount of maximum sugar intake, they still have a considerably high intake of soft drinks, cakes, buns and pastries as seen in Figure 11

and in Table 3. Actions should therefore be targeted at both deciles 1 and 5. It is however worth mentioning that decile 5's intake of soft drinks mainly consists of low calorie soft drinks, while decile 1's intake is mainly not low-calorie. Most of the recommendations for reduced sugar intake will therefore focus on decile 1, though other deciles can also benefit from them.

As mentioned previously, high sugar intake can lead to weight gain that increase the chances of getting heart diseases, type 2 diabetes, strokes and some cancers (Tedstone et al. 2015). Estimations show that obesity costs the NHS in England over £ 6bn every year, and by 2050 this is expected to reach £ 10bn. Diabetes costs the NHS an additional £ 10bn each year (Scott et al. 2018). By implementing actions to reduce sugar intake, the UK Government could cut such costs related to dietary illnesses, allowing for resources to be spent elsewhere. This underlines the importance of implementing further measures to decrease the sugar intake in the UK population.

4.3 Key results

To summarize the interpretations of this analysis, the main findings are presented for each food category. This provides the base for developing recommendations for the government to improve overall health in the UK.

Fruits and vegetables

- Decile 1 and 5 do not eat enough FV according to the WHO's standards, though all deciles can benefit from increased FV consumption.
- Decile 1 and 5 eat proportionally more processed FV, which are less nutritious than fresh FV.
- Decile 1 and 5 eat more potatoes, though this cannot be used to replace other traditional FV.

Fish

- Decile 10 eats the most fish, but none of the income groups meet the recommended amount.
- Oily fish is recommended to make up for half of weekly fish intake. Decile 10 eats considerably more oily fish (salmon) than the rest. This is most likely due to price. The result of this is that rich people get more health benefits from the fish they eat than the rest.

Sugar and preserves

- Decrease in sugar consumption among all deciles, though the lowest decile has seen the largest decrease, a result of different implemented measures.
- The free sugar intake of decile 1 is too high.
- The free sugar intake of each of the decile groups is likely higher than the calculated amount, due to the exclusion of other sources of sugar.

5 Recommendations and Deployment

Recommendations and deployment serve as cornerstones for the reports overall project objective. This section will present a set of recommendations developed through thorough data analysis and research. They embody a forward-thinking approach and are designed for tangible improvements in public health. Deployment will be presented through an action plan that the UK Government can directly set in motion. The action plan is presented together with a supplementing time plan and measuring metrics to monitor and evaluate effects of actions. To conclude the section, limitations of method and datasets will be discussed.

5.1 Recommendations

This section presents the recommendations developed based on the main findings presented in Section 4.3. The recommendations will therefore cover the three food categories: FV, fish, and sugar.

To begin, a comprehensive information campaign is proposed. This campaign is meant to educate the UK population on the benefits of increasing their intake of FV and fish, as well as reducing their intake of sugar. Because the campaign is intended to increase consumption of FV, fish, and sugar substitutes, the campaign must be accompanied by measures that will make these food categories more available to all UK citizens. Therefore, other recommendations will focus on how to increase accessibility of FV, fish, and sugar substitutes, and how to reduce sugar consumption.

5.1.1 Information campaign for decreased sugar and increased FV and fish

Providing information on what a healthy diet is, and how to achieve it, is seen as the first step to convince UK citizens to adopt more healthy eating habits. In this project, three key takeaways have been identified to better public health: reducing sugar intake, increasing FV intake, and increasing fish intake.

Target group
The first recommendation for the UK Government will therefore be to launch an information campaign aimed at individuals from lower and middle deciles. Section 4.2.1 revealed that both lower and middle deciles have potential to better their diets. They both lie bellow recommended FV intake, and decile 1 has a too high intake of sugar. The campaign will use targeted messaging to effectively address challenges faced by these socioeconomic groups.

Channels
To resonate and reach a broad audience within the target group, the campaign will be multi-faceted by combining traditional and digital communication channels. Using several channels can increase exposure and result in greater campaign effects (Snyder 2007).

A large portion of those living in poverty are elderly (Age-UK 2023), meaning many of those in decile 1 are most easily targeted through traditional channels. These channels include print ads, pamphlets and televised commercials. To reach younger individuals and encourage them to build healthy eating habits, another channel will be education lectures in schools. Print ads will be physically displayed in supermarkets in geographical areas with primarily lower and middle deciles. The goal is to reach consumers as they are about to shop for groceries, to hopefully affect their shopping decisions. Pamphlets with the same content as the print ads should be handed out or made available at public services, for instance doctor offices, dentists and teaching institutions. This will connect the campaign to trustworthy and reliable sources. Each communication channel recommended should include a reference to a link provided by the UK Government, offering additional information. The campaign should also be promoted as part of educational programs in primary and lower secondary schools all over the UK. Although lower and middle deciles are the main target groups, it is beneficial to educate all UK citizens from an early age on benefits of a healthy diet.

The digital channels will consist of the social media platforms Facebook and Instagram, as these are the most used platforms in the UK (besides chatting services like WhatsApp and Messenger) (Dixon 2023). Ads distributed through digital communication channels should resemble the pamphlets and print ads for an overall cohesive campaign message. Social media offers key features like broad reach and potential for engagement, and are rich grounds for promoting behavioural changes (McGloin and Eslami 2015). Since the campaign aims at reaching a broad audience and promote dietary changes, this is considered a valuable channel. It will also help the UK Government to engage with the target group which could result in discovering unknown needs and challenges.

Message and content
The overall message of the campaign will be the same for all channels. A suggested message to focus on is how nutritious food can create "*a healthier, happier you*". Positive motivations are

often used in persuasive strategies, and refers to a perceived reward (*Communication in the Real World: An Introduction to Communication Studies* 2016). By choosing to use positive motivation in the information campaign, the goal is to convey a message that choosing a healthier diet leads to positive results.

Research suggest that consumers favour campaigns that focus on practical and useful information with simple and activating messages, rather than strictly informative campaigns (C. et al. 2022). The campaign should therefore be conveyed in a simple manner, and provide useful and practical information. Print ads, pamphlets, televised commercials and social media ads will display the slogan "a healthier, happier you", together with either easy-to-follow recipes or tips for cheap and healthy groceries. Tips should highlight sugar substitutes like nuts and fruits. Both in digital, televised and print ads, the visual content will include eye-catching graphics and pictures displaying nutritious meals and sugar substitutes. Since findings and research suggest correlation between lower income and unhealthy diets, this message will relate to the target group by providing information on budget friendly, nutritious food options. This message can easily be conveyed and tailored to the different channels; print ads, pamphlets and social media platforms. For educational material the message will be the same, and the focus will still be to motivate and encourage healthier diets.

5.1.2 Other measures for decreasing sugar intake

Although sugar consumption is decreasing among UK citizens, sugar intake among those from low income areas is still at an unhealthy level (as discussed in Section 4.2.3). PHEs *Sugar Reduction* report from 2015 presented several key actions targeted at reducing sugar consumption in the UK. These include regulations on marketing of high sugar foods, reducing the amount of sugar in everyday food and drinks, and to ensure education on diet and health to those who influence food choices (catering, fitness, leisure, local authorities) (Tedstone et al. 2015). The actions taken as a result of this report has from 2015 to 2020 resulted in a reduction in both the sugar content in supermarket products, as well as sugar consumption from soft drinks. Nevertheless, the UK has yet to achieve its goal of reducing the consumption of sugar for adults to less than 30g a day - comprising only 5% of their total energy intake (NHS 2023b).

Following the reduction in sugar intake as a result of PHE's measures, many of the same recommendations will be made in this report. This is done in hopes that continuing these measures will, in turn, continue to reduce sugar consumption. In addition, this report will also provide recommendations on promoting sugar substitutes, e.g. fruit, to make the transition process to a less sugar-heavy diet easier.

Regulations on marketing
There is a clear promotional bias towards unhealthy foods in the UK, compared to other foods, as research find that lower socio-economic groups are more exposed than others to promotional activity that encourages sales on sugary foods and drinks (UK-Parliament 2022). Brands often use famous characters to target children, and supermarkets often offer price discounts on fatty and sugary items. Therefore, the first recommended action for reducing sugar consumption will be stricter regulations on marketing of high sugar food and drinks. Approximately 6% of purchased sugar could be prevented if promotion of products with high sugar content was banned (Tedstone et al. 2015).

To foster a healthier society and effectively regulate marketing of unhealthy foods targeted at children and individuals with low income, several regulations should be implemented. Firstly, there should be limitations on advertising to children by restricting the use of child-friendly characters, animations, and celebrities in association with unhealthy foods. Secondly, restrictions should be imposed on when unhealthy food advertisements can be broadcasted on television to minimize exposure during peak children's viewing times.

To reduce exposure towards lower deciles, supermarkets in areas with a large percentage of low income households should be given restrictions on which type of foods can be discounted and promoted, based on certain health criteria that should be developed together with the NHS.

Placement of products
Another recommendation related to reducing exposure of sugary foods and drinks is to reconsider where these food items are placed in store. PHE suggest that regulating the placement of products in supermarkets as a future action, in which unhealthy food items must be removed from till points and end-aisles (Tedstone et al. 2015). The report recommends such regulations to be enacted by the government. However, a less imposing recommendation would be to strongly incentivise supermarkets to contribute to the improvement of public health through their choice of product placements. This by contributing where they can; through placement of product. The goal is the same as for the previous recommendation; to reduce temptation and exposure.

Reduce sugar content in products
Reducing the sugar content in popular products has contributed to the reduction of sugar intake among the UK population (Tedstone et al. 2015). Therefore, this recommendation is furthered by the project team. This can either be done through formal regulations, or through an encouragement campaign that motivate companies to reduce use of sugar by showing business benefits of offering healthier options. These benefits could include positive brand association, or increased trust between brand and consumer because of healthy morals. Of course, the greatest incentive for businesses is ensuring that having a reduced sugar content in products is the more economically sound choice, which is best achieved through taxes and other formal regulations.

Promoting sugar substitutes
As part of the effort to decrease sugar consumption in the UK, there is a push to encourage the use of sugar substitutes such as fruits and nuts. This initiative is partially addressed through the information campaign, which emphasizes why these alternatives are superior sources of energy. To enhance motivation further, there should be a general increase in the accessibility of fruits and nuts for the population. The proposed methods for achieving this can be implemented in a manner akin to the measures outlined for increasing FV and fish consumption, as detailed in the subsequent sections.

5.1.3 Other measures for increasing FV and fish consumption

As seen in Section 4.2.1 and Section 4.2.2, there are large differences in the consumption of both FV and fish between the deciles. The analysis shows that the lower and middle deciles are in lack of these food categories in their diet. Using this information as a foundation, this section presents the recommendations formulated by the project team to increase the consumption of FV and fish in the UK population.

Coupons
As an immediate measure, the UK Government should start distributing coupons to low-income households. These coupons will be directly sent via post to decile 1, as this is where the change is most needed. Coupons can also be disseminated through online newspapers, social media, and grocery store apps (like Sainsbury's Groceries app and The Co-op App) for wider accessibility. These coupons, intended for use in various grocery stores across the UK, aim to encourage increased consumption of FV and fish among all income groups. Coupons offer discounts on FV and fish specifically for lower-income groups, making these essential food items more affordable and enticing. Given the struggle of lower-income households to incorporate FV and fish into their diets, these coupons aim to alleviate financial barriers, ensuring wider access to these groceries across society. This initiative stems from the recognition that 52% of UK households with children are unable to afford a socially acceptable diet meeting both nutritional and social needs (Scott et al. 2018), as outlined by the Minimum Income Standard (R. et al. 2019). The effectiveness of this action is expected to increase with larger discount margins, encouraging greater coupon utilization.

Free FV and school lunches
Offering free FV at schools ensures that children from disadvantaged backgrounds incorporate these vital elements into their daily diets, as well as helping them ingest the recommended five-a-day. Introducing FV early in life establishes enduring healthy eating habits, underscoring the commitment of the UK Government and schools to cultivating an environment that prioritizes health and well-being, thereby setting a positive example for both children and their families. In

addition, the UK Government should provide free school lunches for public primary- and lower secondary schools. School lunches play a pivotal role in delivering well-balanced, nutritious meals that encompass a variety of food options, including essential nutrients from fish and FV. By offering such meals universally, regardless of financial backgrounds, schools guarantee equal access to essential nutrition for all children throughout the school day.

Healthy lunch meals for businesses
The project team proposes that, similar to school lunches, the UK Government should offer incentives for businesses to implement healthier lunches at the workplace. Within the public sector, this initiative can be executed directly by ensuring that employees have access to nutritious lunches. In the private sector, the government could facilitate subsidies for health-conscious catering companies or specifically subsidize the provision of healthy alternatives offered by catering services. This approach would ensure that purchasing healthy lunches for employees becomes financially advantageous.

Subsidies on FV and fish
Another recommended action is to subsidize FV and fish. Subsidies involve the government covering part of the expenses, enabling producers to offer products at a reduced price, especially benefiting lower-income households (Clements and Parry 2018). Research shows that subsidies on FV could increase consumption by 15%, and would reduce the gap between FV consumed and the recommended intake by 1/3 (Pancrazi et al. 2022). This underscores the substantial impact of food pricing on consumer behaviour and, consequently, public health outcomes. As research demonstrates the beneficial outcome of subsidies on FV, this leads the project team to propose a parallel increase in fish consumption as a result of similar subsidies on fish.

5.1.4 Action plan

Aligned with the project objective, all recommendations are summarized in an action plan. This plan can be used as a strategic roadmap for the UK Government in future work towards a better public health. The plan presents key actions with a short description, along with how and where it should be implemented. To visualise how the actions can be overlapping and that the approach is cohesive and synchronized, a time plan is also provided.

In order to limit the funding needed to begin taking actions, the action plan proposes to not launch all activities at once. Section 4.2.1 found that the UK Government have implemented several actions to reduce sugar intake, and the first proposed actions are therefore not focused on sugar reduction. The goal is to begin a changing process where efforts are lower today. All actions are intended to last until the end of a 10 year period to give consumers time to develop new healthy habits, and to be able to measure dietary changes through defined KPIs, further explained in Section 6.1.

The first proposed activity is the information campaign, which will function as a key motivator for all deciles. The campaign is suggested to start as soon as possible, although it is recognized that it will require preparation time. Since the campaign consists of different measures (e.g. print ads, posters, school programs), each measure will be ready to launch at different times, and the goal is to rotate the different measures throughout a 10 year period. The suggested school lecture should however be a permanent measure.

After campaign launch, it is recommended to focus on children to establish healthy dietary habits from an early age. Free FV in schools and healthier school meals is therefore the next recommended action. The following action after that is to offer subsidies to suppliers and coupons to consumers to make FV and fish more affordable to lower and middle deciles.

Measures regarding increasing FV and fish consumption are orange/yellow, while measures targeting sugar reduction are red. Since the information campaign aims to achieve both of these effects, it is designated with its own distinct color.

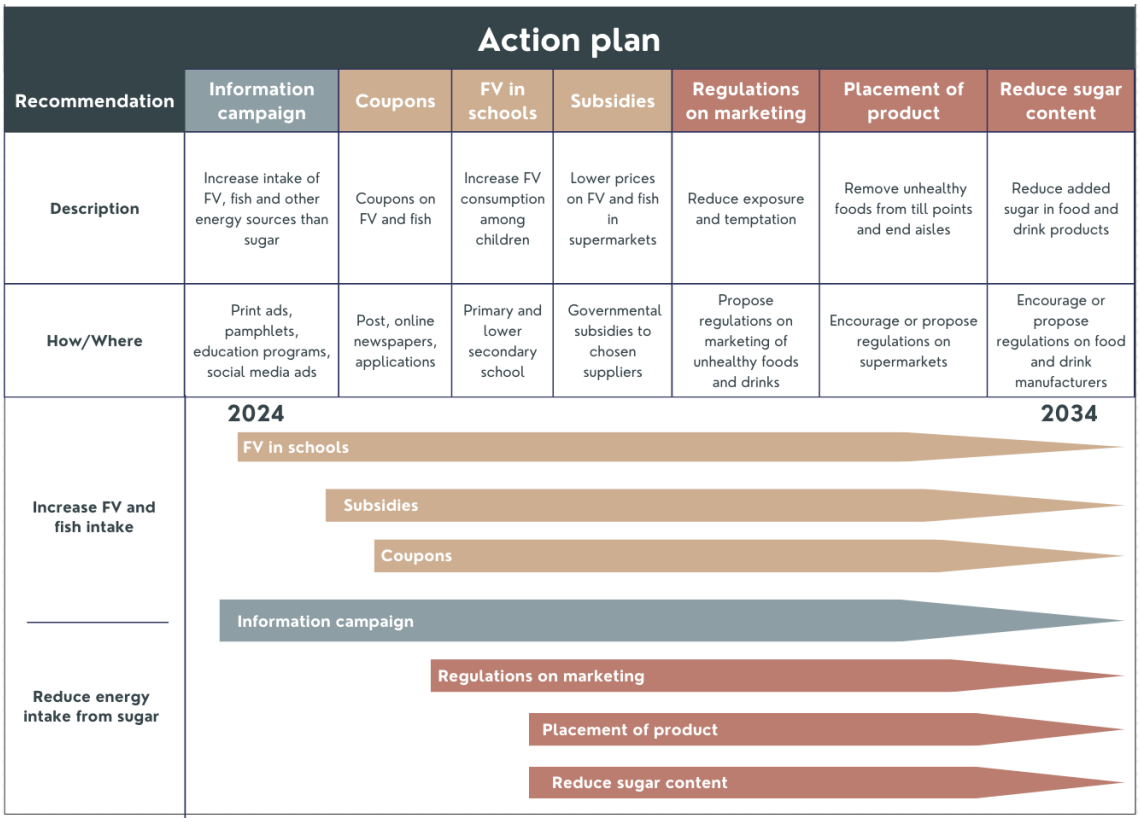


Figure 12: Action plan

Stakeholders

For the proposed action plan, the primary stakeholder is the UK Government, with support and expertise from the NHS and PHE. The UK Government is mainly responsible for implementing and overseeing the outlined measures. The NHS and PHE should function as a human intervention that can adjust actions if needed, and support in measuring the effects of the actions. The impact of each action extends beyond the primary stakeholder, and will affect or be affected by other actors. Supermarkets, food manufacturers and schools emerge as additional stakeholders. In actions regarding sugary food supply chains, or coupons and subsidies, food manufacturers and supermarkets are directly affected. In discussing school meals, primary and lower secondary schools are involved. Collaborative efforts are therefore beneficial, and necessary, in launching the action plan, as the proposed actions have an interconnectedness between several actors.

Measuring metrics

To achieve effective and impactful public health initiatives, measuring metrics play a crucial role in assessing the impact and success of the proposed action plan. The metrics offer a quantitative lens on the effectiveness of suggested actions. They are developed to address and evaluate each recommended action, to enable the stakeholder to change, reconsider or adjust measures after launch.

- Measure public awareness on healthy foods as a result of the information campaign through surveys and awareness tracking.
- Track redemption rate of coupons through integrating the coupon system with supermarket's POS systems for automatic tracking of redemption at the point of sale.
- Measure increase in expenditure on FV and fish as a result of coupons and subsidies lowering prices by analyzing sales data.
- Measure compliance rates among advertisers and food manufacturers with the imposed regulations and encouragements thorough audits and inspections.

- Measure decrease in expenditure on unhealthy sugary foods and drinks through insight in sales data.
- Measure quantifiable reduction in sugar content in unhealthy food and drinks.
- Measure and monitor meal quality based on nutrition value in school meals in primary and lower secondary school.
- Randomized school visits to monitor their focus on promoting FV and fish, and healthy foods in general.

Additional success criteria are presented together with KPIs in Section 6.1.

5.2 Limitations of Method

CRISP-DM Framework
The application of a CRISP-DM framework, along with the utilization of Pandas and Matplotlib in Python, provided valuable insight into food consumption datasets and enabled the creation of line charts for visual representation. However, these methods also had certain limitations that might have affected the robustness of the findings.

As explained in Section 2.3, the team decided on using a CRISP-DM framework in problem solving. While the initial phase of the framework aims to provide the project team with a business understanding, one limitation of this method is its lack of requirement for a detailed understanding of the business, potentially resulting in outcomes that do not align with business needs. To address this, the framework was complemented with a design thinking approach.

The CRISP-DM framework, being task-focused, lacks specific guidelines for collaboration and communication (Saltz 2021). This is viewed as a limitation, as it posed challenges in coordinating efforts among team members with diverse expertise. Each meeting was therefore started off with a review of the work done so far, facilitating more constructive feedback and enhancing collaboration within the team, thereby improving the overall workflow.

Data processing method
Some limitations were also identified in the data processing stage. It became evident that the dataset were initially designed and structured for human comprehensibility rather than script utilization. This required thorough cleaning and preparation of the datasets to ensure efficient processing, as detailed in Section 3.1.

Visualizing food categories in line charts proved efficient in displaying broad differences and making the data comparable. However, choosing a linear graph representation also introduced limitations when it came to detecting or visualising subtle nuances and intricate details within the data. Initiatives were undertaken to prevent the loss of crucial details in the visualization. For instance, a bar chart was incorporated in Section 4.2.1 to elucidate specific data details by displaying subgroups within the category. Additional efforts involved the project team delving into the data for all food categories to comprehensively grasp factors influencing increases, decreases, and variations between deciles.

5.3 Evaluation of pipeline

The following section will evaluate the reliability of the pipeline, and what it means for the findings presented in this report. The pipeline is defined as the process that starts when the yearly data for the project is published, and ends when an action plan to improve public health is sent to the UK Government. The steps that have been taken to ensure the reliability of the pipeline are the following:

- Process the datasets and check for missing values to ensure validity of the data.

- Aggregate the data in a way that makes it easy to visualize, this reduces the risk of errors in the plots.
- Peer-review the plots produced by the Python code. Having several analysts look at the same data individually improves the overall quality of the findings.
- Carefully monitor the effects of the action plan throughout its execution to ensure its efficacy

It is worth noting that the data this pipeline is built upon is not without its flaws, which subsequently has an effect on the reliability of the pipeline. The flaws include small sample size and low response rates. These are further discussed in Section 5.4.

5.4 Limitations of Data and how to Improve it

When working with the project, numerous weaknesses and limitations of the dataset were identified. There is no doubt that Family Food offers valuable data for analyzing the dietary intakes of the UK population; however, it is essential to acknowledge the existing limitations within the datasets. This section looks into the specifics of these limitations and outlines potential avenues for improvement

Small sample size and low response rate
The Family Food datasets comprise a relatively small sample size in their research, especially when contrasted with the total population of the UK. The initial sample size consists of 11,484 households, with an average response rate of approximately 50%, resulting in approximately 5,000 participating households. Considering that there are 28.2 million households in the UK (ONS 2023b), the 5,000 households involved in the survey represent a mere 0.02% of the total. This minimal representation can introduce significant errors and amplify deviations in the findings.

The three most commonly cited reasons for households refusing to take part in the survey were "can't be bothered" (18%), and "genuinely too busy" (15%) (Bulman et al. 2017). One potential strategy to enhance the number of participating households by increasing the response rate is therefore the implementation of gift cards. Utilizing gift cards is a common method to boost survey responses, as it provides participants with a form of compensation, making their involvement more worthwhile and increasing overall survey engagement.

Food group categorization
The categorization of food groups in the datasets lacks intuitiveness. A representative from Family Food, contacted during the project, informed the team that the method of dividing food into groups has remained unchanged since the 1940s. This raises concerns, particularly because the diet of the UK population has evolved over the past 80 years since the initial survey. Consequently, some classifications are outdated, and newer food items may not be adequately represented in the datasets. A recommended approach for food categorization is therefore to reassess the method of grouping and dividing food, aligning the divisions more closely with contemporary academic literature.

Inconsistent labelling of years
Another aspect uncovered during the dataset analysis is inconsistency in the labeling of years. Numerous instances of inconsistency were observed, with variations such as 2001-02, 2005, and 2001718. This inconsistency poses a challenge for data analysts, as it necessitates the re-labeling of years in the dataset. While not a major issue, it detracts from the overall impression of the dataset and its design. Fortunately, this is easily remedied by adopting a consistent format for labeling years and ensuring adherence to it moving forward.

High standard error
During the analysis of the datasets, it became evident that a considerable number of graphs showed a notable increase or decrease in the years 2017/2018. In an email exchange, Family Food clarified that this phenomenon could be attributed to a greater standard error in these years. Upon examining the standard errors of the datasets, it was noted that many of them were remarkably high, with some exceeding 200. This raises concerns about the reliability of the survey data. The elevated standard errors could stem from various factors, and it is recommended that Family Food

conducts an examination to identify the reasons behind these significantly high values. Addressing these factors will be crucial for enhancing the reliability of survey data in future analyses.

Exclusion of eating-out purchases
An aspect worth mentioning in context of limitations, is the food intake from eating-out purchases. When defining the scope of this project, the team made a deliberate decision to exclude data on eating-out purchases, deeming it beyond the project's focus and scope. Consequently, the analysis and consideration of deciles' food intake from eating out were omitted, potentially leading to some relevant data being overlooked. The rationale behind concentrating on household purchases and exploring the food choices made and consumed at home was more sensible within the project's context. This decision aligns with the understanding that a significant portion of people's dietary intake occurs at home. Furthermore, focusing on household purchases was more relevant in the context of recommendations, as many of the proposed actions outlined in Section 5.1 are grounded in household purchases and cooking habits.

Survey limitations
The reliability of the survey is also weakened when selected households are tasked with meticulously recording all food brought into their homes in a physical diary. In this survey approach, designated households are required to maintain a two-week diary of their food intake for everyone over 16 years, and children between 7 and 15 years are asked to keep a simplified daily expenditure diary. This methodology introduces potential for human errors, raising concerns about the reliability of the data.

The manual recording of household intake may lead to omissions, providing an inaccurate representation of food intake. Additionally, individuals may not be entirely truthful about their food consumption due to feelings of shame or guilt. Another limitation involves estimating food amounts, as it is reasonable to assume that most participating households do not weigh all their food. These human limitations pose challenges in translating the datasets into an accurate reflection of reality. Family Food acknowledges these limitations as disclaimers in the datasets, noting that underreporting is expected, with reasons including self-consciousness among participants and simple forgetfulness in documenting food intake.

To enhance data collection for both participating households and analysts at Family Food, a recommendation is to adopt a digital survey format. This transition would provide a more user-friendly format, reduce human errors, and allow for the exploration of features that could enhance the accuracy of food documentation.

5.5 Recommendations for future analysis

To ensure the practicality of this data science project, it was essential to narrow down its scope. Consequently, this section offers recommendations for future analyses on this topic. These suggestions aim to provide additional insights into the evolving food habits of the UK population, identify influencing factors, and explore strategies for enhancing population health.

- **Including eating-out purchases.** The dataset used in the analysis only included food brought into the households. This means that food cooked at home, as well as takeout was included, but restaurant visits were not. This may have impact on the results of this discussion, as studies show that the UK population eat at restaurants approximately 1.5 times a week (OpenTable n.d), which may have significant impact on the data collected on their food intake. The UK Government already has datasets documenting eating-out purchases, so choosing to take this into account should be relatively realistic to implement.
- **Increasing timeline of analyzed data.** Through direct correspondence with one representative from the UK Government's Department of Environment, Food, and Rural Affairs, the team learnt that data on food intake has been collected since the 1940s. An addition to the current analysis would therefore be to extend the analyzed data back until the 1940s to gain a big-picture overview of food intake trends in the UK. Analyzing this together with the wage development in the UK could also provide intriguing insights. This will however

require large amounts of manual labor, as the records from before 2001 only exist in physical form, and must therefore be digitized/made into spreadsheet form before it can be processed and analyzed.

- **Take world events into account.** Many of the graphs exhibited in this analysis have had some amount of spikes and fluctuations. For the sake of the feasibility of this analysis, these fluctuations were attributed to high standard error in the data. Many of the fluctuations can presumably be explained due to this, though it could be an interesting point of analysis to look at the most prominent spikes and changes in light of large world events (such as Brexit and Covid-19) that may have affected people's eating habits.
- **Including data from recent years.** The analysis has only included the years between 2001-2019, due to the data from these years have been most accessible. As mentioned in Section 2.2, the team has argued that the lack of data for the most recent years has not been a hinder from reaching the objectives. Nonetheless, to ensure relevancy and accuracy of the project, recent data should also be included, and is therefore also recommended in the event of a follow-up project being initiated.
- **More thorough analysis of specific nutrient groups.** The analysis has for the most part focused on looking at overarching food groups consumed by the UK population. While some deeper analysis into which specific nutrients are present is done in the specific food categories the team chose to analyze (fish, FV, and sugar), this has only been done to see what effects a lack of/surplus of such food groups has on health. Therefore, future projects could focus more on looking into other food groups and nutrients specifically, which will allow for more targeted, and hopefully more effective, measures to be put in place by the UK Government.
- **Combine with geographical data on lowest income households.** The UK Government already has comprehensive data on geographical income distribution across the UK. By combining this data with the analysis done in this paper, the UK Government can obtain insights into which measures are most important for which regions. As such, the measures will become more targeted, as well as more resource-efficient, thus bettering overall public health in the UK.

6 Monitoring and Maintenance

Measuring the effects of the proposed action plan is not something that can be done immediately. Rather, the positive health outcomes resulting from the suggested recommendations will likely start to manifest only after a minimum of one year, if not longer. Consequently, it is crucial to establish a documentation strategy for maintaining the project code over time. Given the possibility of alterations to the system or implementation by the UK Government, there is no assurance that subsequent data will align consistently with earlier records. This variability could manifest in changes to food groups or a complete transformation in the format of the data.

To sustain the code for processing income data, it is important to locate the final income for the decile groups. Assuming that future data continues to be presented in Excel workbooks with consistent sheet names and data indices as they have done so far, integrating new data into the repository is relatively straightforward. It merely entails adding the corresponding Excel file to the designated folder and appending the filename to the code's array associated with the decile data sheets for all households. However, in the event that future income data adopts a different format, the code becomes deprecated and requires rewriting. A parallel scenario applies to the processing of food consumption data. If the data format undergoes significant changes in the years ahead, the codes viability may be compromised, contingent on the extent of the alterations.

The libraries, namely pandas and matplotlib, upon which data processing relies, are susceptible to potential changes or deprecation in the future. In the event of library modifications, the recommended approach for code maintenance involves consulting the updated documentation and adjusting the code accordingly. If the libraries face deprecation, identifying suitable alternatives

becomes imperative. On a positive note, there is a possibility that the existing code and libraries can continue to function in the future if they are compatible with the versions currently utilized by the team. Regularly monitoring library updates and staying informed about their development can aid in proactively addressing any changes that may impact the codebase.

The action plan is based on current data and available information based on many decades of research from credible sources. Despite this, it is important acknowledge the possibility of newer research emerging and changes in dietary recommendations, which will affect the specifics of the measures suggested in the plan. In the event of groundbreaking research, adjustments needs to be applied so that the plan is aligned with newer findings and following changed dietary recommendations.

6.1 Key Performance Indicators

Key performance indicators (KPIs) can help measure the effectiveness of the proposed recommendations and ensuring that the goals are achieved. The project team recommends evaluating the KPIs after a 10-year period, as this timeframe will allow for a thorough impact analysis of dietary recommendations on public health. The KPIs are summarized and divided into the food categories FV, fish and sugar. The percentages in the KPIs are based on the increase needed to meet the WHO and NHS recommended amounts.

FV

- All deciles should eat the recommend intake of 400 g of FV a day.
- Increase the sales of FV by 58% for decile 1.
- Increase the sales of FV by 20% for decile 5.
- Increase the sales of FV by 16% for decile 10.

Fish

- All deciles should eat the recommended 280 g of fish divided into two weekly portions (whereas one portion should be oily fish).
- Increase the sales of fish by 125% for decile 1.
- Increase the sales of fish by 90% for decile 5.
- Increase the sales of fish by 80% for decile 10.

Sugar and preserves

- Maximum of 30 grams of free sugars a day.
- Achieving less than 5% energy intake from sugar.
- Increase the price on full sugar soft drinks by 10%.
- No exposure of high sugar foods and drinks in advertisement.

Overall health

- The prevalence of childhood overweight among those aged 1-15 is reduced from one third to one fourth.
- Decrease the obesity rate among adults from 26% to 20%.

6.2 What could go wrong?

In terms of evaluating what could go wrong when executing a recommendation model and delivering an action plan to the UK Government, there are some factors and potential outcomes to consider.

Funding
As this recommendation model is designed for implementation by the UK Government, substantial funding is necessary, given its application to a significant part of the population. Therefore, the potential undesirable outcome of not securing the needed funds to execute the action plan must be acknowledged. The lack of funding could arise either from the UK Government not prioritizing the action plan's recommendations or from economic fluctuations hindering the allocation of time and resources to these suggestions.

In the event of insufficient funding, it is essential to narrow down the requested funds to an amount acceptable to the government. The team would then need to assess which measures can be omitted. Priority should be given to targeting children, as early education significantly enhances the prospects of fostering healthier eating habits. Consequently, among the proposed measures, healthy lunch initiatives for businesses would be the first to consider for reduction. Additionally, certain measures, such as product placement strategies and sugar content reduction, may be cut. This is due to the government's ability to incentivize supermarkets regarding product placement is limited, and existing sugar taxes in the UK have already shown an impact on sugar intake.

Action plan does not meet intended goals
If the UK Government does oblige to ensure sufficient funding, there is also a possibility that the action plan does not meet the intended goals. If so, factors that may have affected what went wrong needs to be investigated, and the recommendations needs to further adjust to future data collection as Family Food releases new data each year.

There is also a chance that the recommendations put forth in this project may not be adopted, and various factors could contribute to this outcome. Some of these factors may be of a legal nature, while others could involve the UK Government not regarding the recommendations seriously or opting not to prioritize them.

Limitations of Data
The source of the recommendations, i.e. the data, may also be a source of error. In Section 5.4 some concerns and criticism of the dataset were mentioned, and the reliability of the data was questioned. Despite this, it is important to stress that the source of the data is a reliable and predicable source. Given that the datasets are managed by the UK Government, there is a reasonable expectation that the content adheres to established standards

Barriers to Change
Shaping a healthier lifestyle in the UK might prove challenging considering consumer attitudes as barriers to change. One barrier might be a resistance to dietary change, due to consumers being accustomed to certain eating habits that are deeply ingrained in their routines and preferences. Another connected barrier might be the aversion to healthier options, like fish, often influenced by taste preferences. It is also worth mentioning that not everyone has time or enjoys cooking home made meals that allow for healthier options (PHE 2017). The information campaign seeks to tackle this challenge by providing easy, nutritious, tasty and tempting recipes.

The KPIs created in context of the action plan may be considered a bit ambitious. All KPIs may not be met to the extent presented in section Section 6.1 or completed as wished. This is acknowledged, but the team concluded that it is better to set multiple and higher goals to strive for improvement, and use these as an effective yardstick when evaluating the effects of the measures set in place.

6.3 Alarms

Human Intervention
To ensure valuable recommendations, human expertise was incorporated in the process of generat-

ing recommendations. Recommendations were developed based not only on findings in this report, but also on reports provided by the NHS and PHE. The project team also had correspondence with a representative from the UK Government's Department of Environment, Food, and Rural Affairs. By engaging with experts and reading reports provided by the NHS and PHE, findings were complemented with expert insight. This made it possible to enrich recommendations and aim for robust and reliable decision support.

Alarms for deviations from the general recommendation
This report has presented the recommended intake of the food categories FV, fish, and sugar. When collecting data in the future, alarms should therefore be implemented to compare the reported intake with recommended intake. Alarms can then react when intake levels deviates a certain degree from the given recommendations, or if unhealthy food categories expected to decrease have an increase instead.

To implement this alarm, the percentage difference between food consumption and recommended consumption should be calculated in the data gathering process. The percentage is then reviewed to see if deviation exceeds predefined threshold percentage. The alarm is then triggered if the consumption is far off from recommended levels.

6.4 Lessons learnt/Feedback

As mentioned in Section 5.5, there are many ways that future studies on this subject can be done. To help streamline future project processes, this section provides some lessons learnt and recommendations for future proceedings on similar projects.

The team quickly learnt the importance of establishing agreed-upon overall goals early in a process like this, as the final report is dependent on the clear establishment of objectives. Consensus-building among team members in defining overall objective can prevent unnecessary rewriting and adjustments towards the project's conclusion. Having a well-defined objective early on can serve as foundation for cohesive collaboration and can help ensure that subsections are aligned with a unified vision. This can also help minimizing divergent paths among members and increase efficiency.

Because of the comprehensiveness of a topic like public health, any project related to it should have a clearly defined scope from the start. A significant amount of time was used in this project for data gathering to make an informed decision on what scope would be best. A suggestion for future projects would therefore be to focus only on one of the recommendations from Section 5.5 to keep the scope narrow.

7 Conclusion

7.1 Conclusion on problem description

This section will answer the questions posed in Section 1.2. Table 4 shows a simple picture of the food intake trends among the income groups analyzed in this report.

How much of each food group does each income decile eat?

Decile	FV	Fish	Sugar
1	253	18	34,2
5	307	21	29,5
10	408	22	23,5

Table 4: Daily food category intake of decile 1, 5, 10 in grams based on calculated numbers retrieved from 2019.

Which nutritional food groups are the lower deciles in lack of?

The result of the analysis concludes that decile 1 and 5 are in lack of food groups FV and fish.

Which unhealthy food groups are the lower deciles in excess of?

The result of the analysis concludes that decile 1 are in excess of free sugar.

Is there a significant difference in food consumption based on income level?

The result of the analysis concludes that there is a significant difference in food consumption based on income level. This is further explained in Section 7.2.

7.2 Conclusion on hypothesis

Hypothesis:

Household income has affected the lower deciles health, diet and nutritional intake in the UK.

This report's findings, as seen in Section 4, indicate that the lower deciles in the UK consume too much free sugar, and not enough FV and fish. Combining this with research findings further signify a correlation between income and diet by highlighting socio-economic factors that influence food choices and nutritional access. Research suggests that high sugar intake among lower deciles can be a result of marketing bias towards unhealthy foods in deprived areas, and a result of higher prices on FV and fish. The analysis and interpretation concludes that household income indeed does affect the lower deciles diet and food intake.

Bibliography

Age-UK (2023). *Poverty in later life*. URL: https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/money-matters/poverty-in-later-life-briefing-june-2023.pdf (visited on 21st Nov. 2023).

Azevedo, Ana and Manuel Santos (2008). 'KDD, semma and CRISP-DM: A parallel overview'. In: pp. 182–185.

Brewer, Mike and Tom Wernham (9th Nov. 2022). *Income and wealth inequality explained in 5 charts*. URL: https://ifs.org.uk/articles/income-and-wealth-inequality-explained-5-charts (visited on 4th Oct. 2023).

Bulman, Joanna, Robynne Davies and Oscar Carrel (Feb. 2017). *Living Costs and Food Survey Technical Report for survey year: April 2015 to March 2016*. URL: https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey/livingcostsfoodtechnicalreport2015.pdf (visited on 14th Nov. 2023).

C., Capitão et al. (2022). 'Developing healthy eating promotion mass media campaigns: A qualitative study.' In: *Frontiers in public health* 10 (931116). DOI: https://doi.org/10.3389/fpubh.2022.931116.

Clarke, David (26th July 2018). *Cakes and biscuits*. URL: https://www.actiononsugar.org/surveys/2018/cakes-and-biscuits/ (visited on 14th Nov. 2023).

Clements, Benedict J. and Ian Parry (2018). *What are subsidies?* URL: https://www.imf.org/en/Publications/fandd/issues/2018/09/what-are-subsidies-basics (visited on 20th Nov. 2023).

Communication in the Real World: An Introduction to Communication Studies (2016). University of Minnesota Libraries Publishing. Chap. 11.4.

DEFRA (14th Sept. 2023). *Family food statistics*. URL: https://www.gov.uk/government/collections/family-food-statistics (visited on 3rd Oct. 2023).

DiabetesUK (1st Oct. 2021). *Building back better with building blocks for good health*. URL: https://www.diabetes.org.uk/about_us/news/building-blocks-good-health (visited on 3rd Oct. 2023).

Dixon, Stacy Jo (2023). *Most popular social media platforms in the United Kingdom (UK) as of the third quarter 2022, by usage reach*. URL: https://www.statista.com/statistics/284506/united-kingdom-social-network-penetration/ (visited on 18th Nov. 2023).

EuropeanCommision (2023-08-09). *Health Promotion and Disease Prevention Knowledge Gateway*. URL: https://knowledge4policy.ec.europa.eu/health-promotion-knowledge-gateway/food-based-dietary-guidelines-europe-table-3_en (visited on 5th Nov. 2023).

Hindle, Giles A. and Richard Vidgen (2018). 'Developing a business analytics methodology: A case study in the foodbank sector'. In: *European Journal of Operational Research* 268 (3), pp. 36–851. DOI: https://doi.org/10.1016/j.ejor.2017.06.031.

Hotz, Nick (19th Jan. 2023). *What is CRISP DM?* URL: https://www.datascience-pm.com/crisp-dm-2/ (visited on 11th Oct. 2023).

Kilin, Ivan (2022). *The best charts for color blind viewers*. URL: https://www.datylon.com/blog/data-visualization-for-colorblind-readers?_gl=1*okd4ym*_up*MQ..*_ga*MjA4NjE4MzIxOS4xNzAwNDgxMTI4*_ga_4RCNNGGZW2*MTcwMDQ4MTEyNy4xLjAuMTcwMDQ4MTEyNy4wLjAuMA..#color-blind-palette (visited on 20th Nov. 2023).

Mann, Charles C. (2011). *How the Potato Changed the World*. URL: https://www.smithsonianmag.com/history/how-the-potato-changed-the-world-108470605/ (visited on 12th Oct. 2023).

McGloin, A. and S. Eslami (2015). 'Digital and social media opportunities for dietary behaviour change.' In: *Proceedings of the Nutrition Society* 74 (2), pp. 139–148. DOI: https://doi.org/10.1017/S0029665114001505.

NHS (5th Dec. 2022). *Health Survey for England, 2021 part 1*. URL: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/2021 (visited on 3rd Oct. 2023).

— (n.d). *5 A Day: what counts?* URL: https://www.nhs.uk/Live-well/eat-well/5-a-day/5-a-day-what-counts/ (visited on 1st Nov. 2023).

— (2023a). *Fish and shellfish*. URL: https://www.nhs.uk/live-well/eat-well/food-types/fish-and-shellfish-nutrition/ (visited on 4th Nov. 2023).

— (2023b). *Sugar: the facts*. URL: https://www.nhs.uk/live-well/eat-well/food-types/how-does-sugar-in-our-diet-affect-our-health/ (visited on 14th Nov. 2023).

NHS (2023c). *The Eatwell Guide*. URL: https://www.nhs.uk/Live-well/eat-well/food-guidelines-and-food-labels/the-eatwell-guide/ (visited on 30th Oct. 2023).

Noonan, Robert J. (2018). 'Poverty, Weight Status, and Dietary Intake among UK Adolescents'. In: *International journal of environmental research and public health* (6). DOI: https://doi.org/10.3390/ijerph15061224.

Onque, Renée (2023). *Is a potato a vegetable? There's no 'straightforward answer,' says nutritionist*. URL: https://www.cnbc.com/2023/04/07/is-a-potato-a-vegetable-heres-what-nutritionists-have-to-say.html (visited on 14th Nov. 2023).

ONS (24th Jan. 2020). *Average household income per decile by region*. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/adhocs/11191averagehouseholdincomeperdecilebyregion (visited on 16th Oct. 2023).

— (18th July 2023a). *Effects of taxes and benefits on household income*. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/datasets/theeffectsoftaxesandbenefitsonhouseholdincomefinancialyearending2014 (visited on 14th Nov. 2023).

— (2023b). *Families and households in the UK: 2022*. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2022 (visited on 18th Nov. 2023).

— (2023-10-18a). *RPI: Ave price - Salmon fillets, per Kg*. URL: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/zptx/mm23 (visited on 4th Nov. 2023).

— (2023-10-18b). *RPI: Ave price - White fish fillets, per Kg (cod prior Feb 02)*. URL: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/czol/mm23 (visited on 4th Nov. 2023).

OpenTable (n.d). *What's the average spend per head in restaurants in the UK?* URL: https://restaurant.opentable.co.uk/resources/how-much-do-brits-spend-dining-out/ (visited on 5th Nov. 2023).

Pancrazi, Roberto, Thijs van Rens and Marija Vukotic (2022). 'How distorted food prices discourage a healthy diet'. In: *Science Advances* 8 (13). DOI: https://doi.org/10.1126/sciadv.abi8807.

pandas (n.d.). *About pandas*. URL: https://pandas.pydata.org/about/ (visited on 12th Oct. 2023).

UK-Parliament (2022). 'Diet-related health inequalities'. In: URL: https://researchbriefings.files.parliament.uk/documents/POST-PN-0686/POST-PN-0686.pdf.

Pepsi (n.d.). *Pepsi Real Sugar*. URL: https://www.pepsi.com/#!products/pepsi-real-sugar (visited on 21st Nov. 2023).

Pham, Cristina Tu Anh, Stefano Magistretti and Claudio Dell'Era (2022). 'The role of design thinking in Big Data innovations'. In: *Innovation: Organization Management* (2), pp. 290–314. DOI: https://doi.org/10.1080/14479338.2021.1894942.

PHE (2017). *Health matters: obesity and the food environment*. URL: https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2 (visited on 20th Nov. 2023).

R., O'Connell et al. (2019). 'Which Types of Family are at Risk of Food Poverty in the UK? A Relative Deprivation Approach'. In: *Social Policy and Society* 18 (1), pp. 1–18. DOI: https://doi.org/10.1017/S1474746418000015.

Rickman, Joy C, Diane M Barrett and Christine M Bruhn (2007). 'Nutritional comparison of fresh, frozen and canned fruits and vegetables. Part 1. Vitamins C and B and phenolic compounds'. In: *Journal of the Science of Food and Agriculture* 87 (6), pp. 930–944. DOI: https://doi.org/10.1002/jsfa.2825.

Robertson, Tracey M. et al. (2018). 'Starchy Carbohydrates in a Healthy Diet: The Role of the Humble Potato'. In: *Nutrients* 10, p. 1764. DOI: 10.3390/nu10111764.

Saltz, J. S. (2021). 'CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps'. In: pp. 2337–2344. DOI: https://doi.org/10.1109/BigData52589.2021.9671634.

Sargo, Mike (2023). *What is a Data Strategy Why is it Important?* URL: https://www.dataideology.com/importance-of-data-strategy/ (visited on 11th Oct. 2023).

Scott, Courtney, Jennifer Sutherland and Anna Taylor (2018). *Affordability of the UK's Eatwell Guide*. URL: https://foodfoundation.org.uk/sites/default/files/2021-10/Affordability-of-the-Eatwell-Guide_Final_Web-Version.pdf (visited on 30th Oct. 2023).

Slavin, J. L. and B. Lloyd (2012). 'Health benefits of fruits and vegetables'. In: *Advances in nutrition (Bethesda, Md.)* 3 (4), pp. 506–516. DOI: https://doi.org/10.3945/an.112.002154.

Snyder, Leslie B. (2007). 'Health Communication Campaigns and Their Impact on Behavior'. In: *Journal of Nutrition Education and Behavior* 39 (2), pp. 32–40. DOI: https://doi.org/10.1016/j.jneb.2006.09.004.

Sridharan, Mithun (25th Sept. 2019). *CRISP-DM – A Framework For Data Mining And Analysis*. URL: https://thinkinsights.net/data/crisp-dm/ (visited on 11th Oct. 2023).

Taylor, James (2018). *CRISP-DM Decision Management*. URL: https://decisionmanagementsolutions.com/wp-content/uploads/2018/07/CRISP-DM-and-Decision-Management-061718.pdf (visited on 9th Nov. 2023).

Tedstone, Dr Alison et al. (2015). *Sugar Reduction - The evidence for action*. URL: https://assets.publishing.service.gov.uk/media/5a7f928c40f0b623026904b7/Sugar_reduction_The_evidence_for_action.pdf.

Thompson, Claire (2021). 'Dietary health in the context of poverty and uncertainty around the social determinants of health'. In: *Proceedings of the Nutrition Society* 81 (2), pp. 134–140. DOI: https://doi.org/10.1017/S0029665121003657.

Tinson, Adam (2020). *Living in poverty was bad for your health long before COVID-19*. URL: https://www.health.org.uk/publications/long-reads/living-in-poverty-was-bad-for-your-health-long-before-COVID-19 (visited on 14th Nov. 2023).

Touvier, M. et al. (2010). 'Variations in compliance with starchy food recommendations and consumption of types of starchy foods according to sociodemographic and socioeconomic characteristics'. In: *British Journal of Nutrition* 103 (10), pp. 1485–1492. DOI: https://doi.org/10.1017/S0007114509993345.

Treasury, HM (5th May 2018). *Soft Drinks Industry Levy comes into effect*. URL: https://www.gov.uk/government/news/soft-drinks-industry-levy-comes-into-effect (visited on 14th Nov. 2023).

Wang, Ping-Yu et al. (2016). 'Higher intake of fruits, vegetables or their fiber reduces the risk of type 2 diabetes: A meta-analysis'. In: *Journal of Diabetes Investigation* 7, pp. 56–69. DOI: https://doi.org/10.1111%2Fjdi.12376.

Whittle, Henry J. et al. (2020). 'Precarity and health: Theorizing the intersection of multiple material-need insecurities, stigma, and illness among women in the United States'. In: *Social Science Medicine*. DOI: https://doi.org/10.1016/j.socscimed.2019.112683.

Wirth, Rüdiger and Jochen Hipp (n.d.). 'CRISP-DM: Towards a standard process model for data mining'. In: *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (). DOI: https://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf.

Appendix

For code used in this project, as well plots for every food group, see:

github.com/stjernen/TDT259_UK_health