Data Wellness Technical Report

Authors: Janeel Abrahams
Aditya Palapati
Jin Hee Lee
Shervorn Mathews

Table of Contents

Introduction

The relationship between food consumption and health is a topic of significant importance in today's society. The nutritional composition of foods plays a crucial role in determining their impact on our overall well-being. In this report, we will analyze the nutrient composition of various food groups using food datasets sourced online. By examining key factors such as energy content, protein, fat, carbohydrates, and other essential nutrients, we aim to gain insights into the relationship between food and health outcomes.

The findings from this report can contribute to our understanding of the role of food in health outcomes, inform dietary recommendations, and guide interventions aimed at addressing diet-related health issues. By identifying foods that are high in unhealthy components and exploring potential nutrient imbalances, we can develop strategies to promote healthier dietary choices and combat nutrition-related health concerns.

Food Analysis

Understanding the nutrient profiles of different food groups can help us identify dietary patterns that promote optimal health or contribute to health issues. By examining the distribution of nutrients, identifying foods that exceed recommended allowances, and exploring correlations between nutrient levels, we can uncover valuable insights into the impact of food consumption on our well-being. Additionally, we will assess the prevalence of malnutrition by examining undernourishment trends over time using a global undernutrition dataset. This analysis will shed light on the persistent challenges related to food security and their impact on population health.

To begin, undernourishment is under the umbrella of malnutrition, this is the insufficient intake of nutrients which can result in diseases including but not limited to cardiovascular diseases,obesity and type 2 diabetes. After reviewing the hunger and undernourishment dataset obtained from worldindata.org it was seen that Somalia and Angola had the most consistent trend of undernourishment regardless of their food supply. The graph below shows the growth of the population over time and the trend in food supplies for Zimbabwe,Afghanistan and the US. These countries were chosen because they had a consistent trend of undernourishment. The graph below shows the trend of undernourishment between the years of 2002 to 2017. The United States was added to show the comparison of trends in countries with different percentages. The project also does a deep dive on food consumption and health related diseases in the United States.

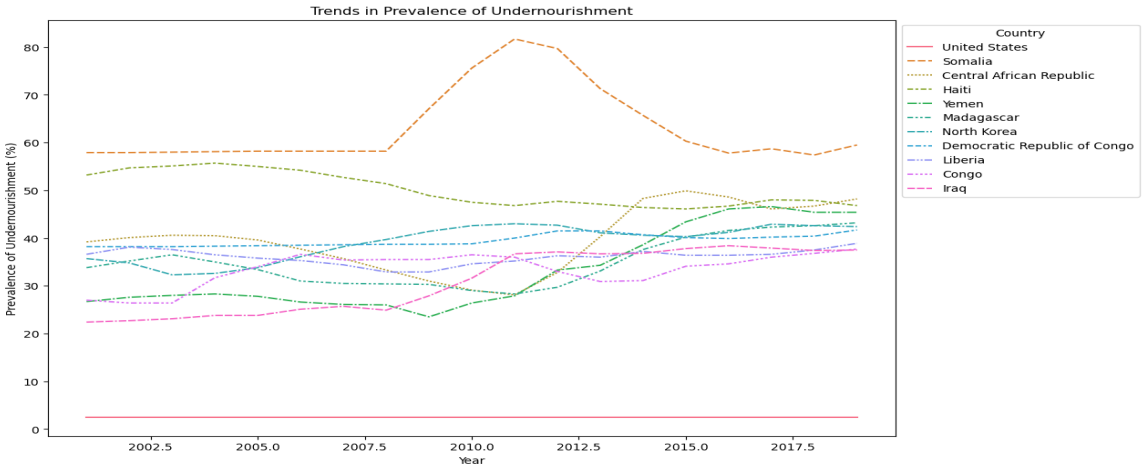

Fig 1.1- Prevalence of Undernourishment

After observing Fig1.1 above the group decided to compare the population growth over time with the food supply for the United states and African countries. The goal was to see if the trend was consistent across African countries as well as explore how it differs from the trend found for the US. The graphs below show the population growth overtime as well as average Food Supply in Calories per capita.

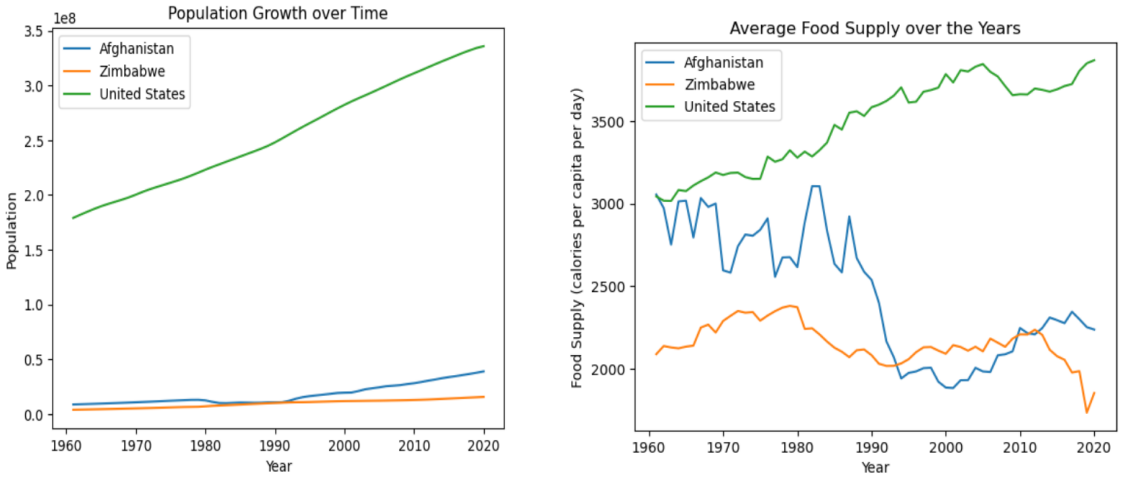

Fig 1.2- Population vs Food Supply

After observing Figure 1.2 above it can be concluded that there are similar trends of increasing population but decreasing food supplies across the African continent, some countries' food supplies fluctuate more than others but ultimately it decreases. While for the US there are also fluctuations but the opposite is true.

US Deep Dive

In an attempt to understand food consumption in the US we analyzed the food content. During our analysis we found that the calorie content of 96.2% of foods in the dataset were fatty calories with only 3.8% being non-fatty calories. We also saw a direct correlation between Calories and Fat in the dataset. This let us review the recommended daily allowances of the nutritional content as assigned by the Food and Nutrition board here in the US. Figure 3 below shows our findings.

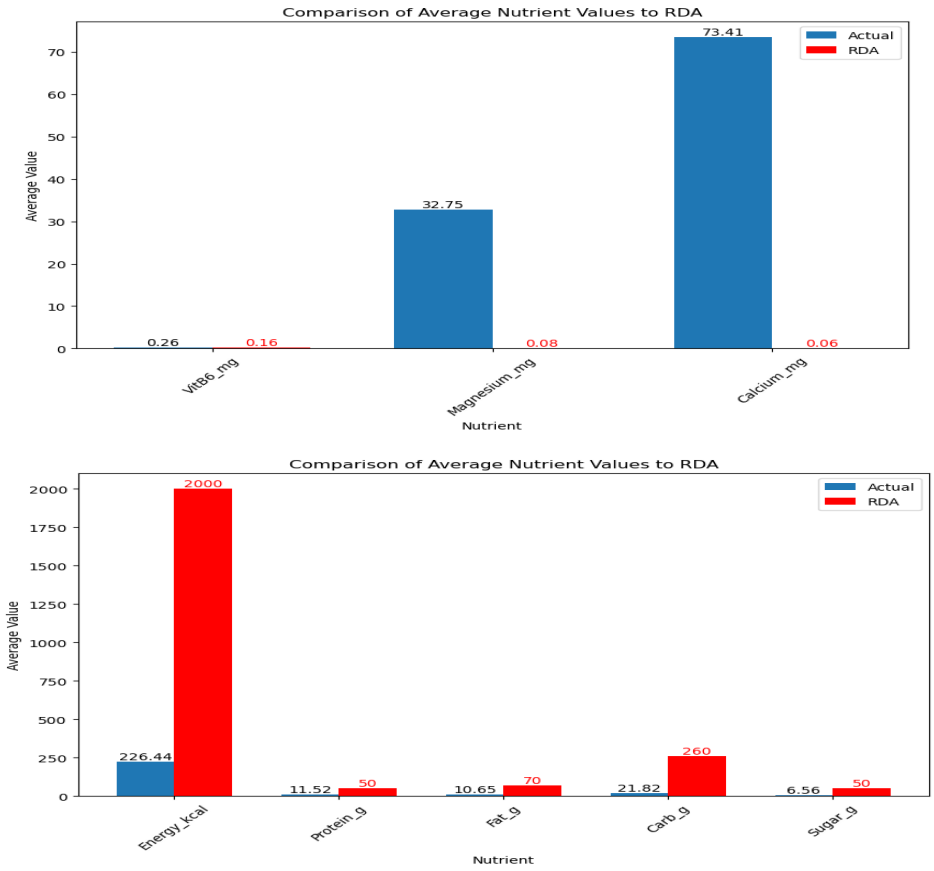

Fig 1.3 - RDA vs Actual nutritional values

In Figure 1.3 we can see that for some nutrients and vitamins like Vitamin B6, Magnesium and Calcium there is an excess of them in food when compared to the RDA value. On the other hand, for nutrients like Protein,Fats, Carbs and Sugar it is the opposite. The foods provide insufficient nutritional values, which is good if we aren't eating them in bulk. This can allow us to eat an array of different foods to get the recommended daily allowance of nutrients. This result led us to review the foods and pick out the 5 most problematic foods. These were mainly fats and oils with some dessert food items. As stated before there is a direct correlation between fats and calories, so it was no surprise to see that these foods had a lot of calories. Another non-surprising fact was that the dessert item had the most sugar having only 52.54 grams of sugar while the fatty items had between 0 - 0.6 grams of sugar.

The discovery led us to examine the crops.

Crops Analysis

Human health and food has been a hot topic for years and will be for years to come. Nutrition only covers one aspect of it, but to get a better understanding of it we must look towards where consumers source their food. Without agriculture there is little food or nutrition, but, at the same time, availability of food from agriculture does not always translate to good nutrition. To bridge the disconnect between agriculture and nutrition we must work together across sectors to create a food system that is sensitive to nutritional outcomes.

The nation's food supply relies upon sufficient sources of healthy plants. A range of variables affect plant health, including the surrounding environment and the extent to which they are protected from pests and disease.Healthy plants are vital to crop production and to the quality and cost of the nation's supply of food, fuel, and fiber. Using data from the Food and Agriculture Organization of the United Nations, analysis of trends within the US from 1961 to the 2000s we can see that the area harvested over the years has been relatively stable.

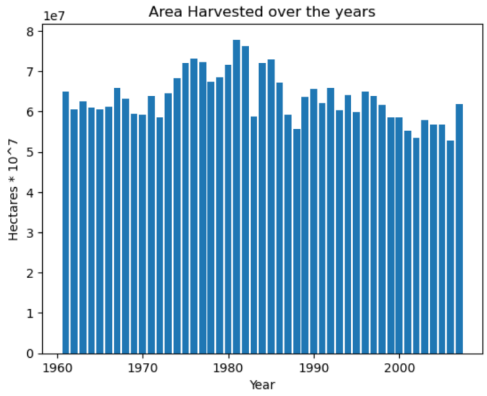

Fig 2.1- Area Harvested across the US (1961-2007)

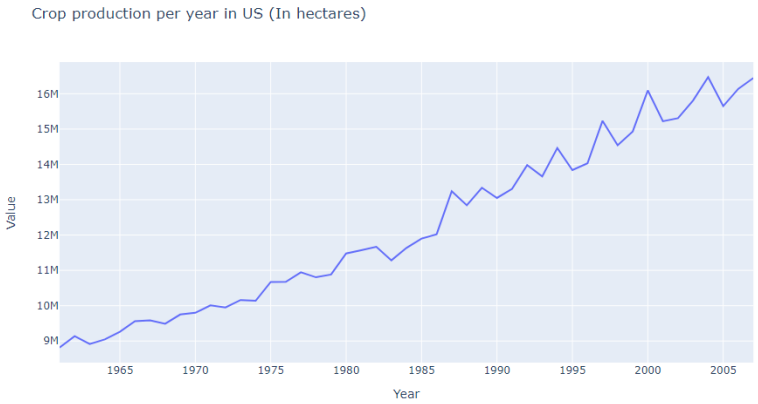

Fig 2.2- Crop production per year in the US (1961-2007

At the same time, the crop production has been steadily increasing in an upwards trend and continues to do so. Crop production is very important but the yields give us a better understanding of just how much of these crops become available to consumers. Through both charts we can see while some crop yields have stayed stagnant over the years of analysis, others have followed upward trends as well with sugarcane and tomatoes providing the greatest yield.

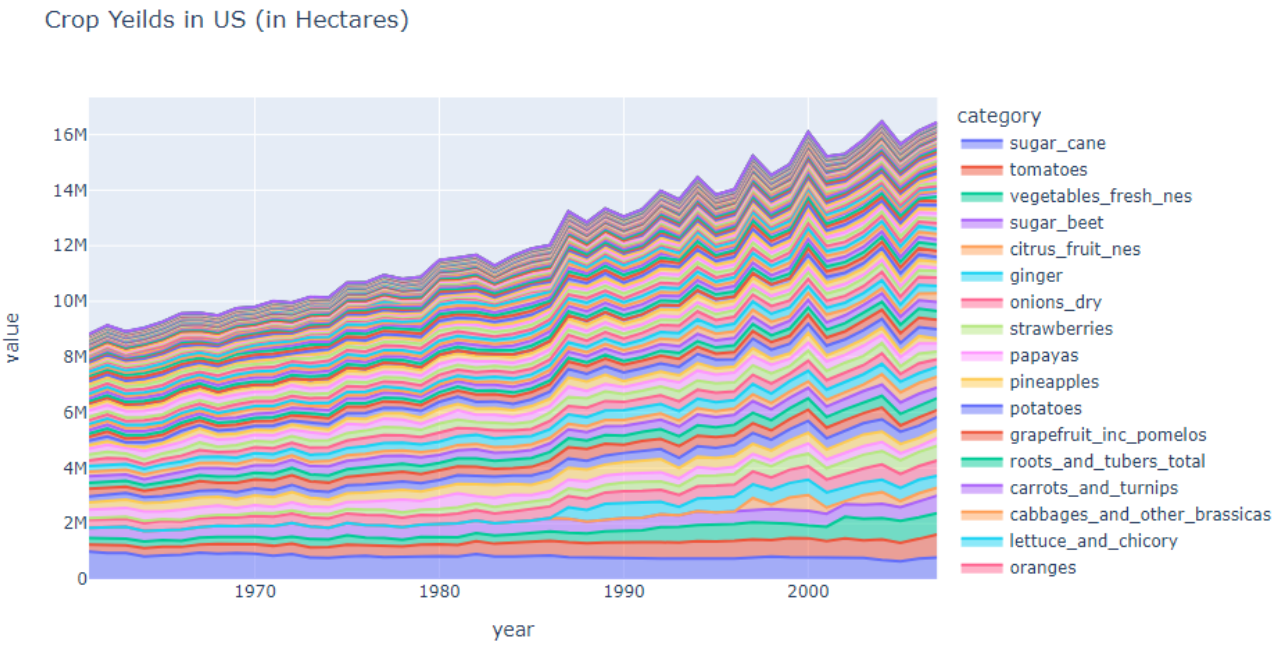

Fig 2.3- Crop Yields in the US by crop (1961-2007)

Through soil data analysis within the US we can ascertain whether certain fruits, vegetables, or grains would thrive in certain conditions. By taking the data we are able to create a model considering different growing aspects, such as soil pH levels, nitrogen, phosphorous, and potassium concentrations as well as temperature, rainfall, and humidity. A logistic regression, decision tree, and random forest model has been created followed by cross validation techniques to ensure the models are low in bias and variance. To visually show this, a confusion matrix was created on our random forest model to show just how accurate it is. Through the use of this modeling we can assist farmers to determine the best crops to be grown while considering their nutritional content as well.

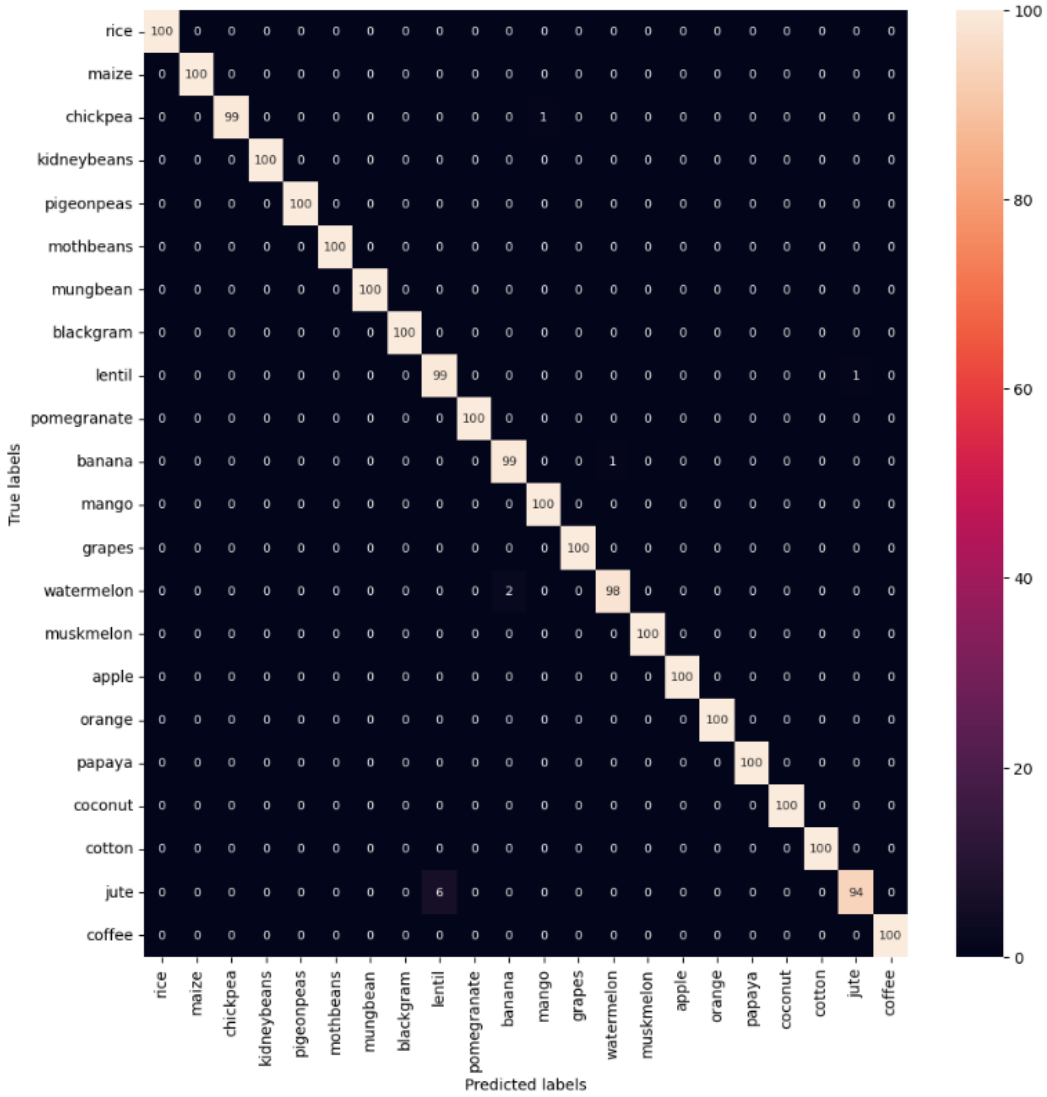

Fig 2.4- Confusion Matrix on Random Forest Model

Crops and nutrition is just one factor to human health, but now we will look at a variety of food related health problems in the US.

Diseases Analysis

The research pertains to the relationship between food consumption and health issues.

- Major health problems in the USA: examining cases across different regions and climate zones, grouped by states. (see figure 3.1)
- Potential connection between food choices(consumption) and major health problems in the USA
- The rise in chronic diseases:increase in cases of obesity, diabetes, heart diseases, high blood pressure, and certain types of cancer (see figure 3.2)

- These conditions are closely linked to poor dietary choices, such as excessive consumption of sugary beverages, foods with high calories(fat has most calories (9 kcal per grams))
- Top 5 problematic foods are processed oils and sugar such as butter, cream, high fructose corn syrup
- Among other health problems, obesity shows significant increase among states. (see figure 3.2)
- Increasing Obesity Rates: Data reveals that obesity increased by 269% from 2004 to 2017 (see figure 3.3)

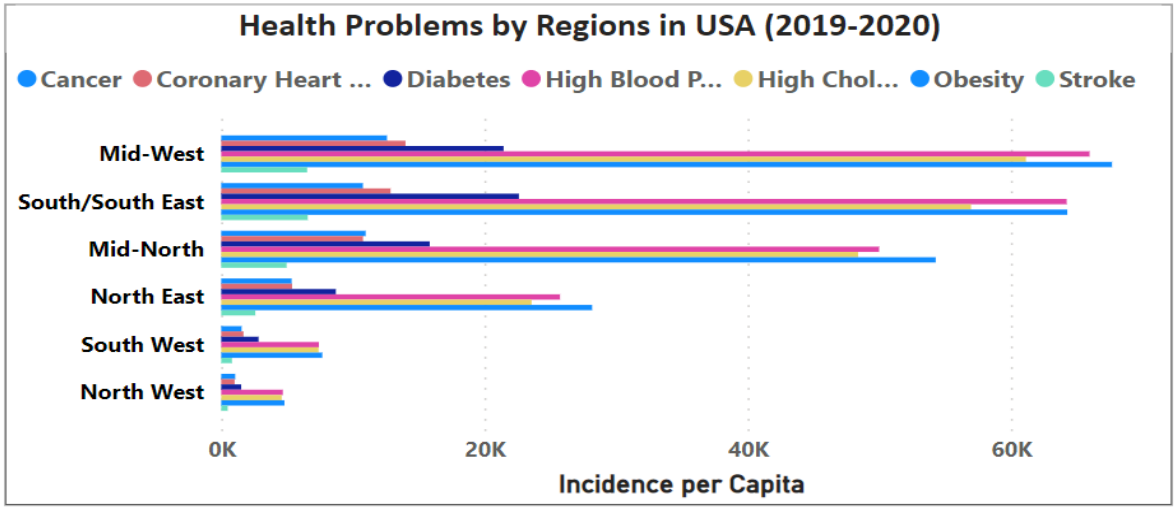

Fig 3.1: Health problems by Regions in the USA

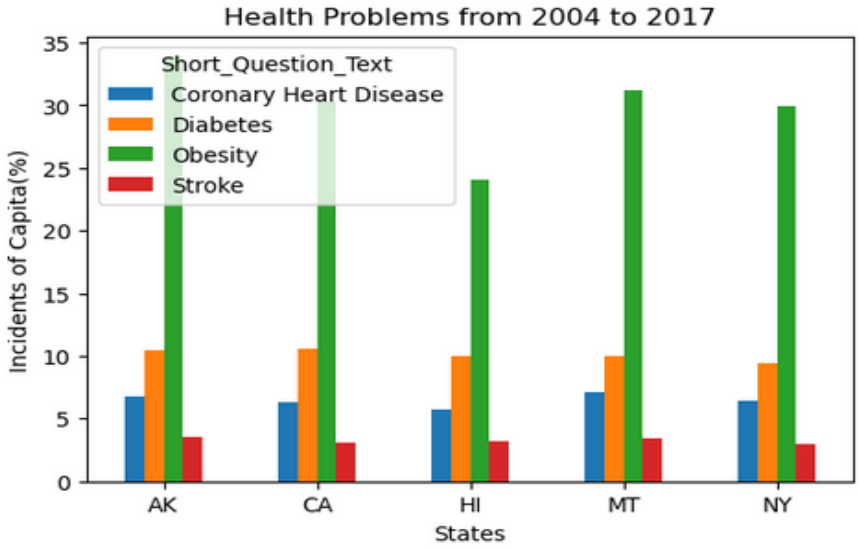

Fig 3.2: Obesity vs other health problems

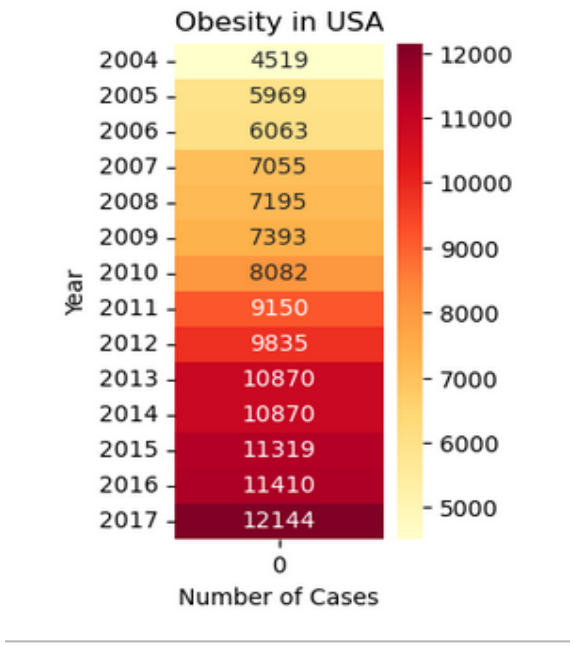

Fig 3.3: Obesity from 2004 to 2017

Obesity Analysis

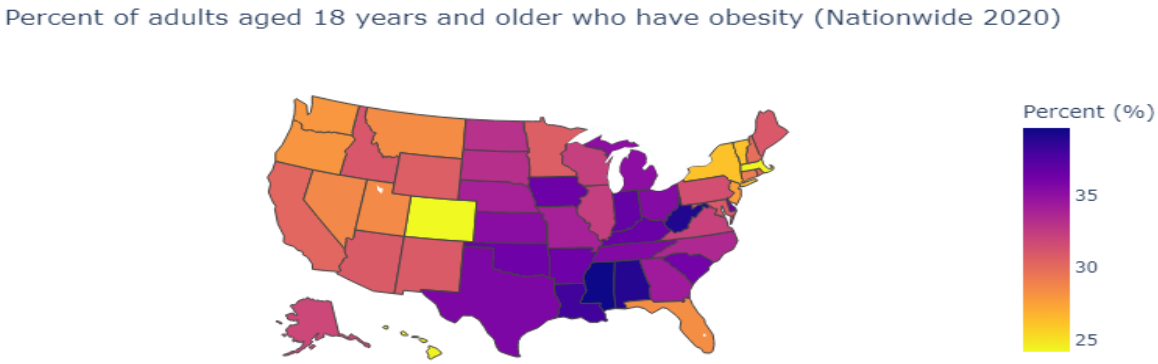

Fig 4.1:

The presented figure portrays the prevalence of obesity across US states, delineating a discernible pattern wherein the southeastern region exhibits a notably higher incidence of obesity. Conversely, the western states demonstrate comparatively lower rates of obesity. Intriguingly, the data highlights outliers in the form of Colorado and Massachusetts, where the prevalence of obesity is strikingly minimal, while Alabama and Mississippi emerge as states grappling with the highest obesity rates. This stark contrast invites a compelling investigation into the comparative statistics of these regions, promising insightful findings.

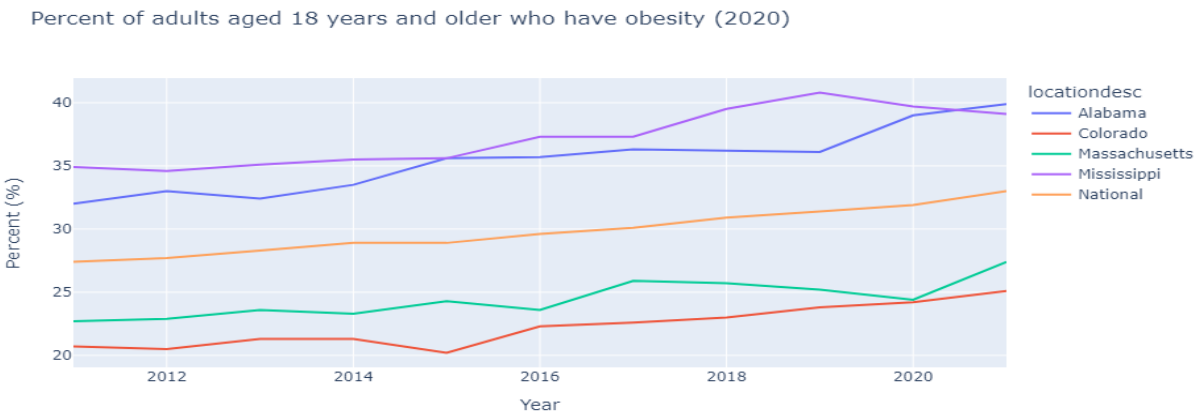

Fig 4.2:

The provided figure presents a comparative analysis of obesity percentages in Alabama, Colorado, Massachusetts, Mississippi, and the national average, spanning the years 2011 to 2021. Notably, a consistent 15% disparity persists between states with higher obesity rates (Alabama and Mississippi) and those with lower rates (Colorado and Massachusetts) throughout the examined period. Furthermore, both the higher and lower states exhibit a proximity of approximately 7% deviation from the national average. Remarkably, an average incremental rise of 5% is discernible in the obesity rates over the course of the ten-year duration.

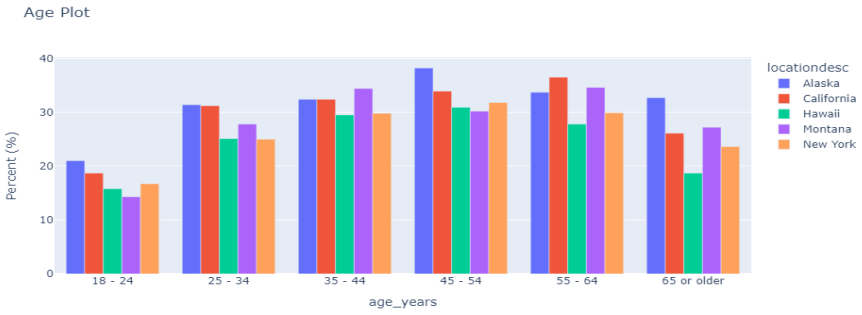

Fig 4.3:

The accompanying figure delves deeper into the analysis of obesity by examining different age groups within the previously mentioned states. A distinct trend emerges, revealing that younger age groups exhibit lower obesity percentages compared to their older counterparts. As age progresses, the prevalence of obesity steadily increases until reaching a peak within the (45 - 54) age group. However, intriguingly, the rates then begin to decline among older age groups.

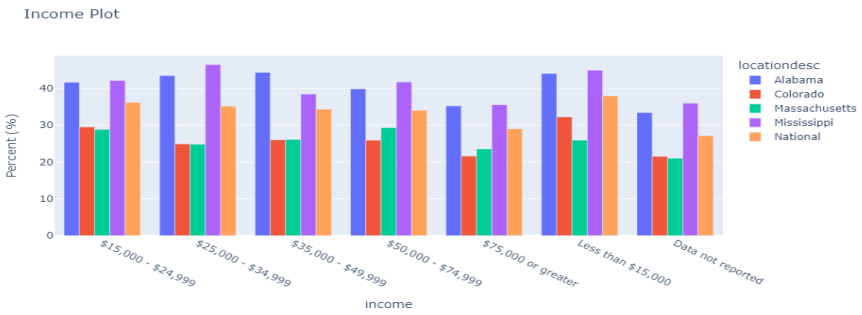

Fig 4.4:

The subsequent figure investigates the relationship between income levels and obesity, indicating no discernible impact of income on obesity prevalence. Despite variations in income across the states examined, there is no clear correlation between higher income levels and lower obesity rates, or vice versa.

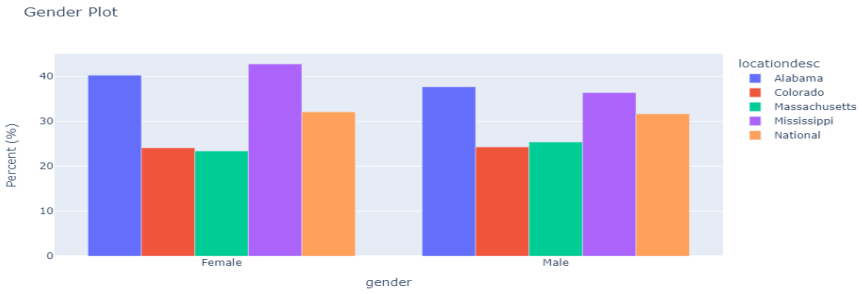

Fig 4.5:

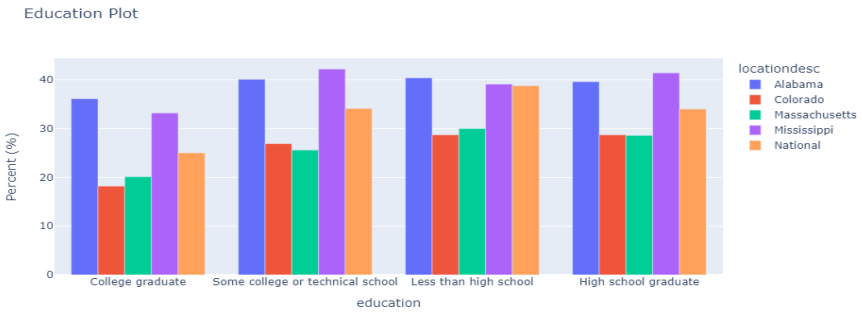

Fig 4.6:

The following figure delves into the relationship between education levels and obesity percentages, revealing a noteworthy pattern. It indicates that individuals with higher levels of education tend to exhibit lower obesity rates. This finding suggests a potential association between education and healthier lifestyle choices, including dietary habits and physical activity, which may contribute to lower obesity prevalence among those with higher education levels. Further exploration of the underlying factors driving this relationship, such as knowledge,

socioeconomic status, access to resources, and health literacy, can provide valuable insights for designing effective interventions and educational programs to combat obesity.

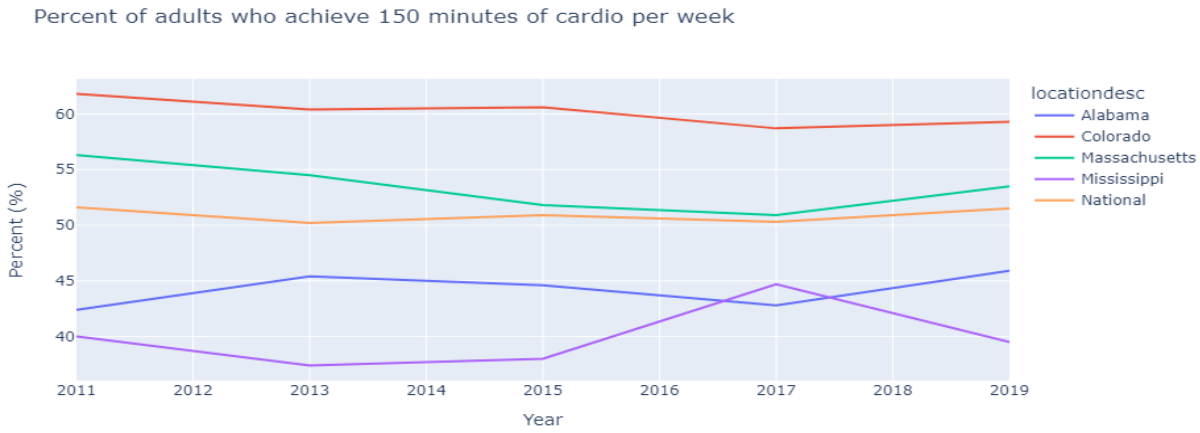

Fig 4.7:

The next figure explores the correlation between activity levels and obesity percentages among states, revealing a compelling trend. It demonstrates that states with higher activity levels tend to exhibit lower obesity rates. This finding suggests a strong association between physical activity and reduced obesity prevalence. It implies that individuals residing in states with a culture of regular physical activity, accessible recreational facilities, and supportive infrastructure for active lifestyles are more likely to maintain healthier body weights. Understanding the factors contributing to higher activity levels in these states, such as community programs, urban planning, and promotion of physical education, can offer important perspectives for developing effective strategies to address and prevent obesity at a population level.

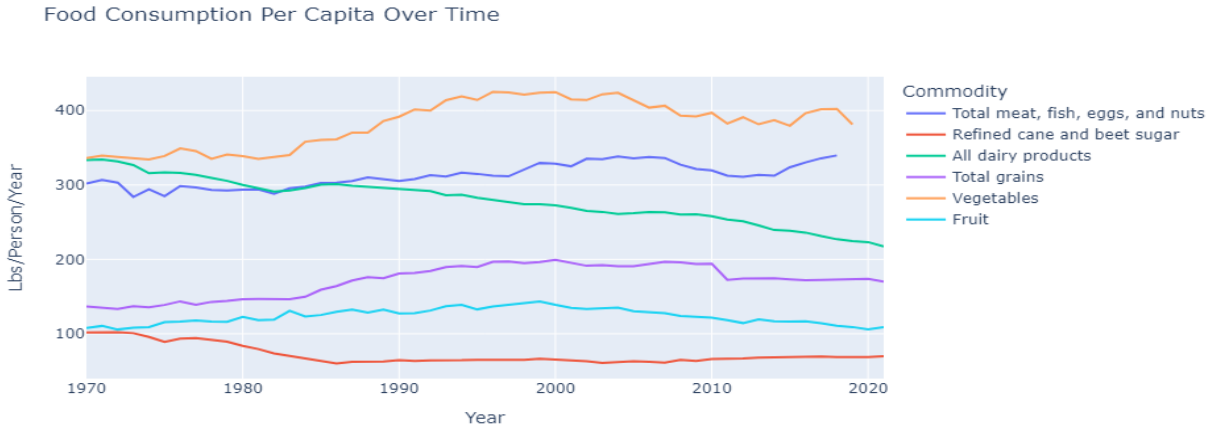

Fig 4.8:

A notable trend observed over time is the increase in the amount of food consumed per person per year, accounting for population growth. This significant rise in food consumption

serves as a potential contributing factor to the increase in obesity rates. The concurrent changes in dietary habits and portion sizes may play a role in the escalating prevalence of obesity over time. Understanding the relationship between food consumption patterns, population dynamics, and obesity trends can provide valuable insights into the complex factors driving this public health issue. Further investigation into the specific dietary factors, cultural influences, and societal norms surrounding food consumption can aid in formulating effective strategies to address and mitigate the obesity epidemic.

Conclusion

In conclusion, it is well-established that food items with high fat content generally have high calorie content, increasing the likelihood of obesity. Utilizing modeling techniques can help identify nutritious foods that can yield significant quantities for distribution to consumers. While vegetable products offer the highest nutritional value, simply consuming them in large quantities without a balanced diet and regular exercise does not effectively lower the risk of obesity-related diseases. Several factors, including location, activity levels, and sugar consumption, influence obesity rates.

Sources

https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

https://www.kaggle.com/datasets/unitednations/global-food-agriculture-statistics?select=fao_data_crops_data.csv

https://datatopics.worldbank.org/health/health

https://chronicdata.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-County-Data-20/swc5-untb

https://www.kaggle.com/datasets/GoogleNewsLab/health-searches-us-county

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

https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/