Obesity Trends: Development of Python-based prediction model to aid possible solutions

Project Report

Sejal Kumar
sejal.kumar@student.biss.no
Student, British International School of Stavanger (BISS) Gausel

Table of Contents

Abstract

Obesity has emerged to be one of the major health hazards that are constantly increasing in the 21st century. Globally, the number of obese people has more than doubled since 1990. Due to its potential health consequences, obesity has become an alarming health threat which may become harder to control in the future as these rates continue to rise. Therefore, this report aims to develop solutions to prevent obesity becoming a health crisis. Python programming language has been used to develop a prediction model to forecast global obesity trends using current obesity data, so that effective and actionable solutions can be proposed to control rise in obesity levels.

A comprehensive literature review was conducted to identify the factors that are influencing obesity. The interrelationship between these factors have been visualized using a fishbone diagram. Quantitative data on adult obesity levels from 1975 to 2016 was collected from 'Our World in Data' for almost all the countries in the world. To develop a prediction model, linear regression approach was adopted employing Python based machine learning techniques. The model has been developed in 3 stages to accurately predict obesity trends from 2017 to 2050 using the country-based obesity data. Initially, a simple linear regression model was developed for a specific country. In the second stage, the model was introduced with polynomial features, through which the model was able to produce curves and create more accurate predictions. In the last stage, the model was generalised to produce trends for a total of 190 countries with 98.8 % overall accuracy. This data was then used to compare predictions of different countries within a region or similar economy and identify the common factors which are influencing the rise in obesity levels. On the contrary, research was done to find out if effective solutions had been implemented in countries where the prediction model suggested peak and thereafter downfall in obesity levels. The model has thus aided in proposing solutions that can control rising obesity trends of the countries whose predictions indicated alarming obesity rates by 2050.

The predictions revealed a complex interplay of factors contributing to future obesity, including genetics, health conditions, socioeconomic status, culture, government regulations, urban designing, lifestyle, and eating habits. An analysis of the predicted trends showed significant variations in obesity rates among countries. The middle eastern countries, such as Egypt and Saudi Arabia, exhibited the highest increases, while countries like South Korea and Singapore maintained relatively low obesity rates. Other countries like USA and Canada displayed a peak in their significantly high obesity trends, leading to a decreasing trend in the near future.

The study reveals a strong correlation between a country's stage of development and its obesity trends. As nations progress through various developmental stages, their ability and willingness to implement effective obesity prevention strategies evolves. In the early stages of development, countries typically prioritize economic growth, often leading to rising obesity rates as Western dietary patterns and sedentary lifestyles are adopted. In addition, locals are not fully aware of the health consequences to obesity therefore unhealthy eating habits are adapted. Countries may also simply not have enough economic power and resources to implement effective preventive measures. However, as countries advance and accumulate wealth, they become increasingly aware of the health consequences of obesity and the need for prevention. This awareness, coupled with greater financial resources, enables the implementation of more comprehensive government regulations and policies aimed at combating obesity. Developed nations can invest in public health initiatives, urban planning that promotes physical activity, technologies to promote weight loss and education programs about healthy lifestyles. However, developing countries can be aided by World Bank, WHO or non-profit organisations, to inspire governments on policy changes and implementation of community programs. Depending on the speed and the scale of implementation of the solutions recommended, the predicted trends will adjust in future.

1. Introduction

1.1. Background & Research Objectives

Obesity, categorised as excessive fat accumulation, has emerged to be one of the most prevalent health issues in the 21st century. The World Health Organisation (WHO) reports that adult obesity rates have more than doubled since 1990 with over 890 million adults classified as obese in 2022. (World Health Organization, 2024). As a student interested in the healthcare field, I was shocked coming across the escalating obesity rates, highlighting the need for solutions.

This prompts me to raise my first research question:

1. What are the factors contributing to Obesity?

My fascination with coding and desire to learn Python have led me to explore how data science can help in better understanding of this pressing issue.

This leads me to my second research question:

2. How can a predictive model help in developing effective obesity prevention strategies?

My aim is to develop a Python based predictive model, which can forecast obesity trends that will help in formulating effective prevention strategies. By combining healthcare with my interest in coding, I intend to analyse possible solutions of obesity through data-driven evidence and innovative solutions in this project.

1.2. Problem Statement

Obesity arises from the imbalance between energy expenditure and energy intake. Globally, changes in dietary habits, such as the increased consumption of high-calorie, nutrient-poor foods, combined with declining physical activity levels, have led to significant rise in obesity. As a result, the health consequences due to this crisis are intensifying. In 2019 alone, WHO claims that obesity was indirectly responsible for approximately 5 million deaths from noncommunicable diseases (NCDs). (World Health Organization, 2020).

As shown in Figure 1 below, the risk that obesity poses to the society is relentlessly increasing. This statistic emphasizes the crucial need for further studies and effective solutions to decrease the rising trend of obesity. Beyond the immediate health consequences, obesity contributes to a variety of chronic diseases, including diabetes, cardiovascular diseases, and certain types of cancer (ACDPA, 2019), placing an enormous burden on healthcare systems globally. The economic impact is equally affected, with billions of dollars spent each year on obesity-related healthcare costs and lost productivity. Global costs of overweight and obesity are predicted to reach 3 trillion USD per year by 2030 and over 18 trillion USD by 2060, as the World Health Organisation reported in 2024. (World Health Organisation, 2024).

Obesity has proved to be a fatal health threat to the society as it was also ranked as the second leading cause of preventable death after smoking in the United States of America. (Hurt et al., 2011). With the fact that deaths due to obesity can be prevented, it is crucial to conduct further studies to enhance our understanding and develop effective preventive solutions. However, effective solutions can only be developed through a proper understanding of the factors that contribute to obesity.

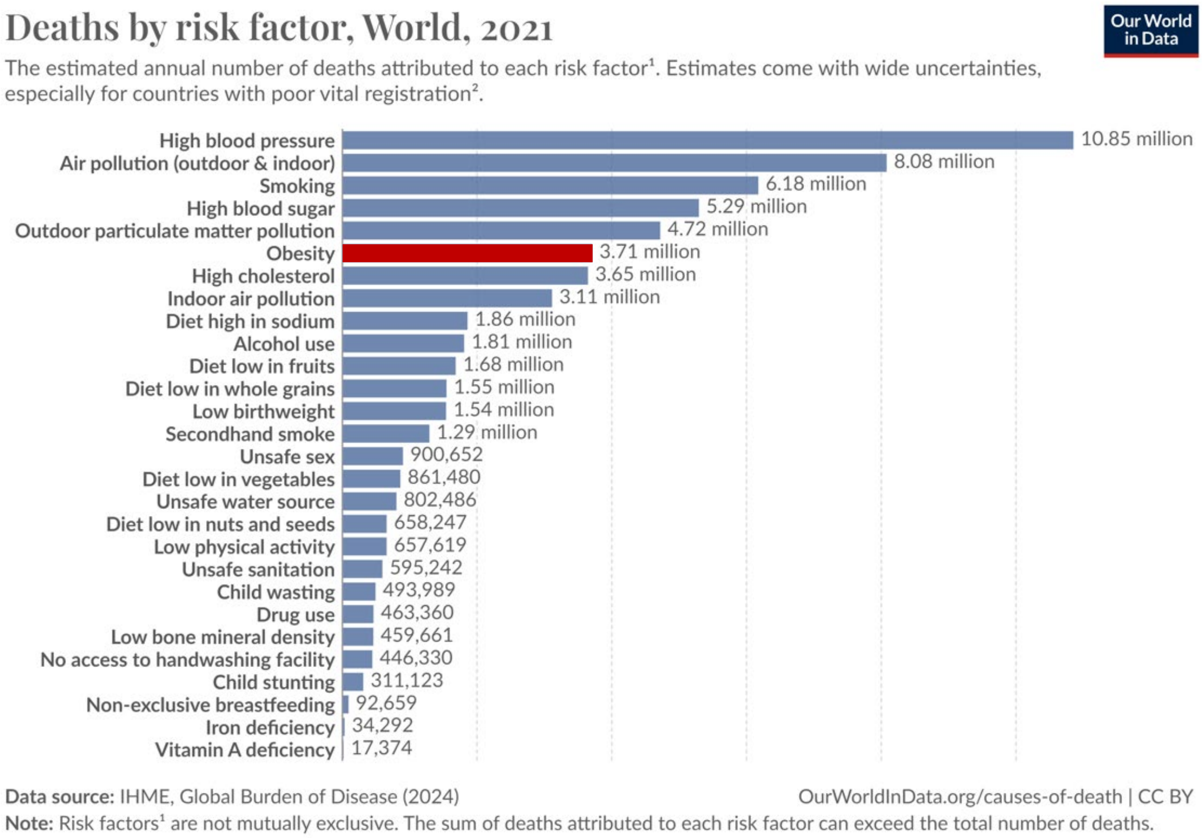

Figure 1: Ranking of deaths by risk factor (Ritchie and Roser, 2024)

2. Factors Contributing to Obesity

Obesity is not caused by a single aspect, but a range of interconnected factors, where the examination of these factors can help us understand the root causes of obesity and identify possible solutions to overcome this crisis. Although there has been a lot of work done on this topic, the identification of all the factors and presentation of the inter-relationship between these factors is missing. I have tried to identify all the major factors influencing obesity and present their relationships in a fishbone diagram (as shown below in Figure 2).

2.1. Genetics

Genetic Inheritance Leading to Obesity

Genetic bias to obesity refers to the genetically inherited traits that increase an individual's likelihood of becoming obese. For instance, variations in the fat mass and obesity-associated (FTO) gene have been linked to affect appetite and higher body mass index (BMI). (Ruth and Giles, 2021)**.** Some genes such as the GPR158 gene affect the metabolic rate through its variants like rs11014566. This variant is known to cause lower energy expenditure, higher BMI, and higher fat percentages leading to weight gain. (Watson, 2021).

2.2. Health

Medication

According to the National Heart, Lung, and Blood Institute (NHLBI), certain medications that aim to treat health conditions often contribute to significant weight gain because such treatments result in side effects that disrupt chemical signalling to the brain which leads to increased appetite (National Heart, Lung, and Blood Institute, 2022). Alternatively, these drugs can impact metabolic processes by having side effects leading to increased fat storage and reduced energy expenditure, (Verhaegen and Van Gaal, 2019).

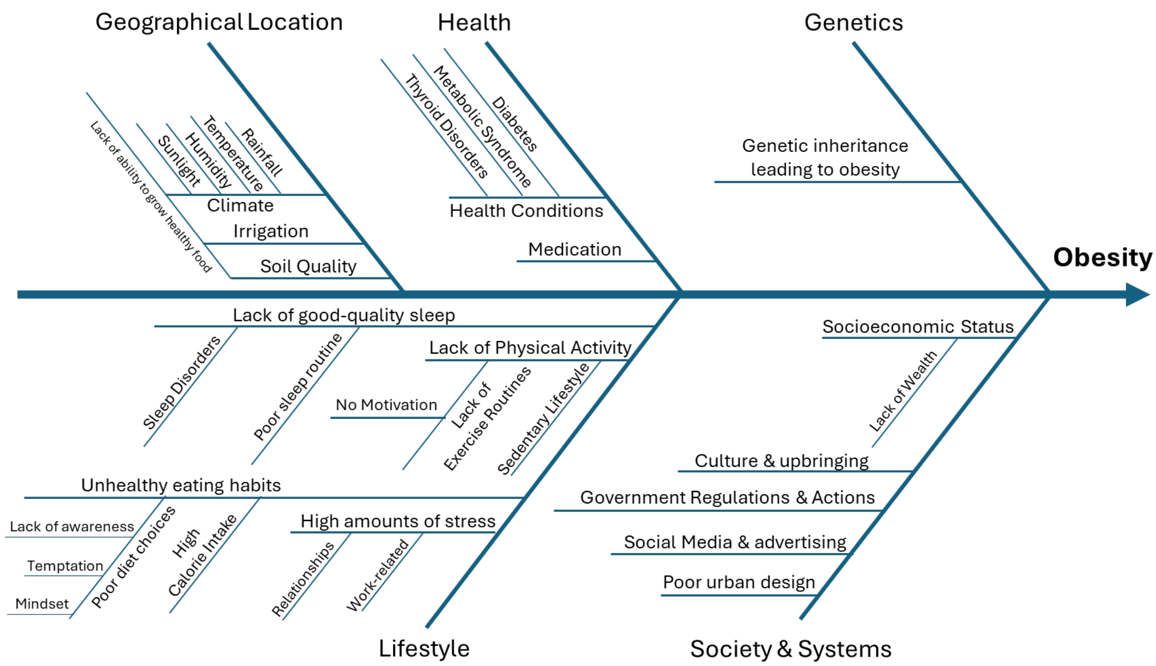

Figure 2: Fishbone diagram of factors leading to obesity.

Health Conditions

Diabetes:

Diabetes is referred to as a health condition that affects the body's usage of blood sugar (glucose) (Mayo Clinic, 2024). Chronic Diabetes condition concern type 1 and type 2 diabetes which have a deficiency of insulin and ineffective insulin respectively. Therefore, some people who are diabetic need medication with doses of insulin if they cannot manage their blood sugar with a healthy diet and exercise. Insulin resistance can lead to inefficient glucose uptake by cells, causing the body to store more fat, reduce energy expenditure and slowed metabolism (National Institute of Diabetes and Digestive and Kidney Diseases, 2018). The likelihood of developing type 2 diabetes through genetic inheritance is more compared to type 1 diabetes. However, lifestyle such as eating and exercise habits can also be contributing factors (NLM and U.S. Department of Health and Human Services, 2017).

Metabolic Syndrome:

Metabolic syndrome is a group of health problems that increase the risk of type 2 diabetes which can lead to obesity primarily through insulin resistance. (Mayo Clinic, 2021)

Thyroid disorders:

Thyroid disorders, particularly hypothyroidism, can lead to obesity due to their effects on metabolic rate. The thyroid gland produces hormones that regulate the body's metabolism. When the thyroid is underactive (hypothyroidism), it doesn't produce enough hormones, resulting in a slowed metabolic rate which intern causes fat accumulation, (Sanyal and Raychaudhuri, 2016). Genetics play a key role in thyroid disorders. Studies suggest that around 67% of thyroid hormone concentration is due to genetic inheritance, (Panicker, 2011).

2.3. Geographical Location

Geographical locations have a significant contribution to obesity, specifically due to economic factors. Biomes with favourable climatic conditions (ideal temperatures for farming, higher rainfall, humidity, and sunlight), are regions with significant agriculture which provides the availability of cheap and fresh foods within these regions. However, in colder, dryer, darker and arid biomes, it is not possible to grow such vegetation, especially in winters, forcing these regions to import food from other countries with favourable biomes, thus increasing the pricing of imported fresh food. This causes locals with low socioeconomic status, to rely on cheaper processed and frozen foods leading to unhealthy eating habits, making them prone to obesity.

Moreover, favourable biomes provide suitable climate to have an active outdoor lifestyle, which is more restrictive in colder biomes, especially in winter leading to sedentary lifestyles and increasing obesity.

2.4. Society & Systems

Socioeconomic Status

Low socioeconomic status (SES) significantly contributes to obesity by limiting access to healthy foods. Financial constraints often lead to reliance on inexpensive, calorie-dense options. These factors create an environment that promotes unhealthy eating habits and increases obesity risk.

Culture & Upbringing

Japanese children are known to be among the healthiest in the world, partly due to their traditional diet, eating habits and their attitude towards food. (Doran, 2023). Japanese culture encourages fresh and unprocessed foods in children's diet (Shubrook, 2022). Thisdorn cultural approach to food likely contributes to healthier eating habits among Japanese locals which they have adapted early from childhood. The difference in how food is perceived and presented to children in Japan compared to countries like the United States is a notable factor of obesity rates in both countries (this will be further discussed in chapter 3).

Government Regulations and Actions

Government regulations and actions play a crucial role in addressing public health issues, with smoking control being a prime example. In the UK, the government implemented taxation on cigarettes and the Health Act 2006, which banned smoking in all enclosed public spaces and workplaces. Professor Linda Bauld from the University of Bath notes that this regulation led to immediate health benefits, including a reduction in heart attack rates and improved air quality in public spaces. (Bauld, 2011)

Social Media & Advertising

Social media significantly impacts obesity rates. Dr. Jaime Sidani with several other researchers found a strong association between social media use and concerning eating habits. (Sidani et al., 2016). Social media contributes to obesity by promoting sedentary behaviour and exposing users to constant junk food advertising through the brand itself or other influencers.

Poor Urban Design

Poor urban design, particularly the lack of accessible recreational areas, such as sidewalks and parks, can significantly contribute to obesity. Through limiting physical activity opportunities, locals are unable to do such activities even with high motivation yet are discouraged in walking and outdoor exercise through poor urban planning and area layouts.

2.5. Lifestyle

Lack of physical activity

A lifestyle with sitting and laying down for long periods with very little body movement and exercise is defined as a sedentary lifestyle. A sedentary lifestyle can lead to obesity primarily through consistent higher calorie consumption compared to the calorie expenditure. Additionally, many people with lack of exercise routines result with obesity. This is often due to lack of motivation to consistently have an exercise routine, making it harder to adopt healthier habits.

High amounts of stress

Chronic stress, whether from work or relationships, can significantly contribute to obesity through stress eating. Prolonged stress increases cortisol levels, which heightens cravings for high-fat, sugary foods. (Torres and Nowson, 2007) This hormonal change makes "comfort foods" more appealing, resulting in weight gain over time. This creates a cycle where weight gain causes additional stress, leading to further overeating and escalating obesity.

Lack of good-quality sleep

Insufficient or poor-quality sleep, whether due to sleep disorders or irregular sleep routines, can significantly contribute to obesity by disrupting hormones and increasing hunger. Sleep deprivation raises levels of ghrelin, the "hunger hormone," and lowers levels of leptin, the "fullness hormone." (Knutson and Van Cauter, 2008). This imbalance makes you feel hungrier and less satisfied after eating.

Unhealthy Eating habits

Unhealthy eating habits, characterized by high-calorie intake and poor diet choices, are major contributors to obesity. The widespread availability and temptation of processed and unhealthy foods, often high in sugars and unhealthy fats, lead to excessive calorie consumption. Furthermore, a lack of awareness about the long-term health impacts of these foods is a common reason leading to poor dietary choices. However, such choices can also be due to a mindset of lack of concern towards healthy eating, particularly in environments where unhealthy options are normalized.

3. Current Trends in Obesity

After identifying the different factors contributing to obesity, I was curious to investigate the changes in obesity levels over the years for different countries. By analysing obesity trends among different countries or regions, it might be possible to find some correlation between obesity trends and some of the identified factors contributing to obesity levels, such as differences in diet, lifestyle, socioeconomic status, regional climatic conditions and cultural practices.

3.1. Data Collection

I started researching if obesity data of different countries is freely available through any authentic source. After going through several government and scientific websites, I finally found authentic data available through *'Our World in Data'*, (Ritchie and Roser, 2024). This source is a credible organisation that uses the work of thousands of Oxford researchers to present data and showcase their understanding of the large problems the world is facing. *Our World in Data* is a project owned by Global Change Data Lab, a registered charity in England and Wales, who publishes and maintains the website and the data tools. It is a non-profit organisation with a purpose to provide authentic data accessible to all for research purposes. The adult obesity data provided by *Our World in Data*, is sourced from World Health Organization – Global Health Observatory (2024).

3.1.1. Data Description

Quantitative data related to adult obesity levels has been extracted from *Our World in Data*. The data provides annual changes in obesity percentages from 1975 to 2016. Unfortunately, I was unable to find data from 2017 onwards. However, 41 years data between 1975 to 2016 can provide very good insights for trend development and analysis. Data values are classified in Body Mass Index (BMI).

BMI is a widely used measurement for assessing an individual's body weight in relation to their height, providing a numerical value that categorizes individuals into weight groups. The following formula is used to calculate the BMI:

$$BMI = \frac{Weight\ (kg)}{Square\ of\ Height\ (m^2)}$$

The idea behind this formula is to standardize weight across different heights, providing a consistent way to categorize individuals as underweight, normal weight, overweight, or obese. With this formula the height differences can be accounted. This means that taller individuals naturally weigh more due to their larger overall body size, not necessarily because of excess body fat. BMI is primarily used to classify individuals as underweight (<18.5), normal weight (18.6-24.9), overweight (25-29.9), or obese (>30), which can help in evaluating potential health risks associated with body fat levels, (Ramsay Health Care, 2024).

Despite its widespread use, BMI has limitations, particularly because it does not differentiate between muscle and fat mass, nor does it account for variations due to age, sex, or ethnicity. This can lead to misclassification, especially in people with high muscle mass or varying body compositions, (Harvey and James, 2023). However, BMI remains a useful initial assessment tool for identifying individuals who may require further medical evaluation or lifestyle changes. Despite its limitations, understanding BMI trends across populations remains crucial, particularly when comparing obesity rates globally.

3.2. Trends Visualization

Initially, I extracted annual obesity percentage data for Norway from 1975 to 2016, to understand the overall obesity condition of my country.

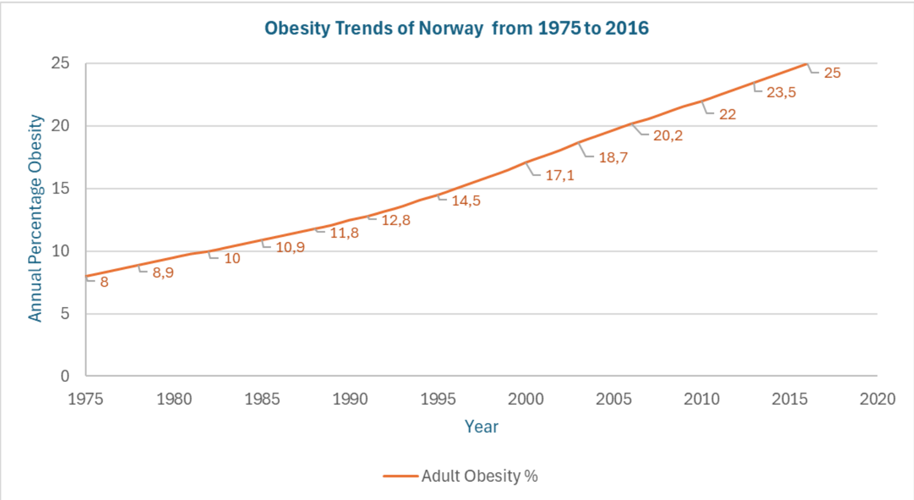

Figure 3: Graph displaying average obesity trends of Norway from 1975 to 2016.

The graph (Figure 3) shows the percentage of obesity in Norwegian adults aged 18 and above. There is a significant increase of obesity rates from 8% to 25% over a period of 41 years. From 1975 to 1993 the graph has a gradual increase. However, from 1994 to 2016, it increases more rapidly. This can be attributed to various factors. After discovery of Oil in Norway in 1969, it took a few years to make more discoveries, generate wealth and job opportunities. By the 1990s 'the oil fund' was established so that the government could invest surplus profits, (Oslo Kommune, 2023). Moreover, Norwegians got exposed to American culture and food when they came to establish the Norwegian oil industry. During that time many processed food chains also opened up in Norway, which led to a major shift to unhealthy eating habits. Sedentary behaviours in many and easy access to high-calorie foods, changed the lifestyle, explaining the significant rise in obesity levels, (Norwegian Institute of Public Health, 2024).

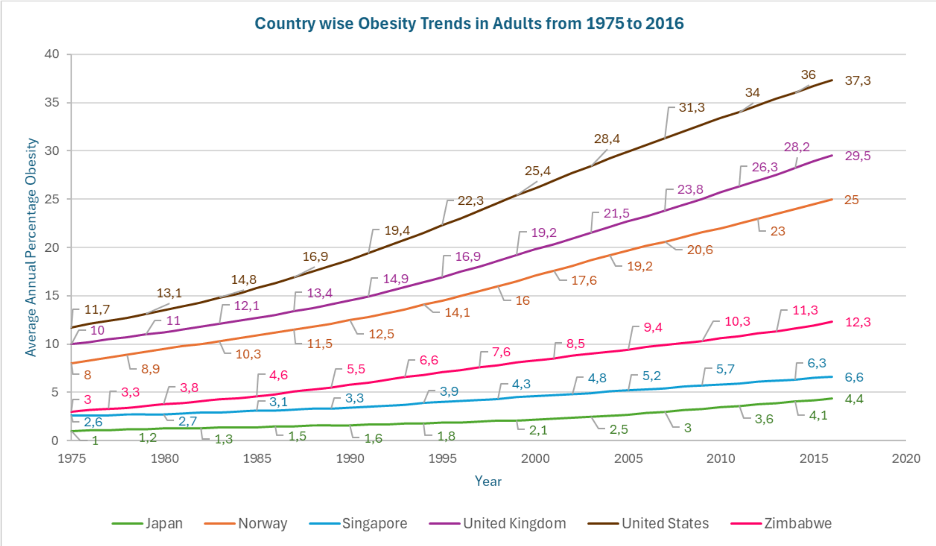

Figure 4: Graph displaying adult obesity levels of Japan, Norway, Singapore, United Kingdom, United States of America, Zimbabwe through 1975 to 2016.

Figure 4 shows the changes in obesity percentage among adults aged 18 and above, between 1975 to 2016. The countries were carefully chosen to show contrasting comparison of some of the most dominant factors effecting obesity.

The varying cultures across the displayed countries in Figure 4 may explain the variation of their obesity trends. The rise in obesity in the US can largely be attributed to the culture of fast-food and sedentary lifestyles. This has been compounded by the fact that fast-food restaurants are often more accessible and cheaper than healthier food options, creating the "obesogenic environment" involving easy access to high-calorie, low-nutrient foods with larger portion sizes, (Waidmann et. al., 2022). Similarly in the UK, the culture of a sedentary lifestyle combined with increasing numbers of office jobs contribute to the rising obesity rates. Furthermore, Norway's obesity rates can be attributed to the change in eating habits after the exposure to western cultures. In Zimbabwe, cultural perceptions that associate larger body sizes with wealth and good living play a key role in the country's obesity rates. (Moyo, 2021). In contrast, traditional Asian dietary habits, which are prevalent in Singapore, tend to be lower in calories and fat, contributing to a lower increase in obesity rates. Similarly, Japan's dietary habits also involve traditional Japanese meals. The Japanese diet typically includes a high intake of fish, soy products, and green tea, while being low in consumption of animal fat, meat and processed foods. This contributes to a lower calorie intake compared to Western diets, (Gabriel, Ninomiya and Uneyama, 2018). Secondly, the culture in Japan places a strong emphasis on food quality and portion control, which helps maintain healthy eating habits. Overall, a strong correlation between obesity rates and culture can be seen through this analysis. While countries who have continued with traditional food choices and lifestyle remain at lower obesity rates, countries that have adapted to western cultures have displayed a more rapid increase in obesity.

Urban planning is also a contrasting factor within the 6 displayed countries in Figure 4 that has a significant impact on obesity. The United States face a declining level in physical activity particularly due to its urban planning and technological advancements that discourages physical activities like walking, cycling, jogging etc. (lack of parks, walkways, sidewalks, etc.). However, when compared to countries like Singapore and Japan, a clear contrast is visible. The urban design of leisure areas in Singapore encourages walking, cycling and other forms of physical activity, (Pico, 2024), similar to Japan's urban design and city plan that encourages physical activity. In addition, the high cost of car ownership and the convenience of public transportation in Japan leads to more walking as part of daily life. Overall, this implies that urban planning has a significant impact on the lifestyle and physical activity levels of locals.

Government initiatives additionally have a significant role in controlling obesity rates. Comprehensive government initiatives in Singapore, such as the National Healthy Lifestyle Program, promote healthy eating and physical activity through public education and supportive environments, (Goh, Gan and Pang, 2012). Furthermore, government policies in Japan such as mandatory health checks for employees, help monitor and manage weight gain, (McLaren, 2022).

Lastly, the geographical location, climate and ease of producing local food also plays an important role in rising obesity trends. Tropical countries like Singapore have favourable climate year-round for active outdoor lifestyle, whereas Northern parts of US, UK, Norway and other northern European countries have limitations to mobility in Winter. Local crop and vegetable cultivation is also limited during winter months, forcing to import from other countries, thereby impacting the pricing and freshness of food. On the other hand, cheap, high-calorie food chains are readily available, escalating obesity levels.

3.3. Correlation Analysis

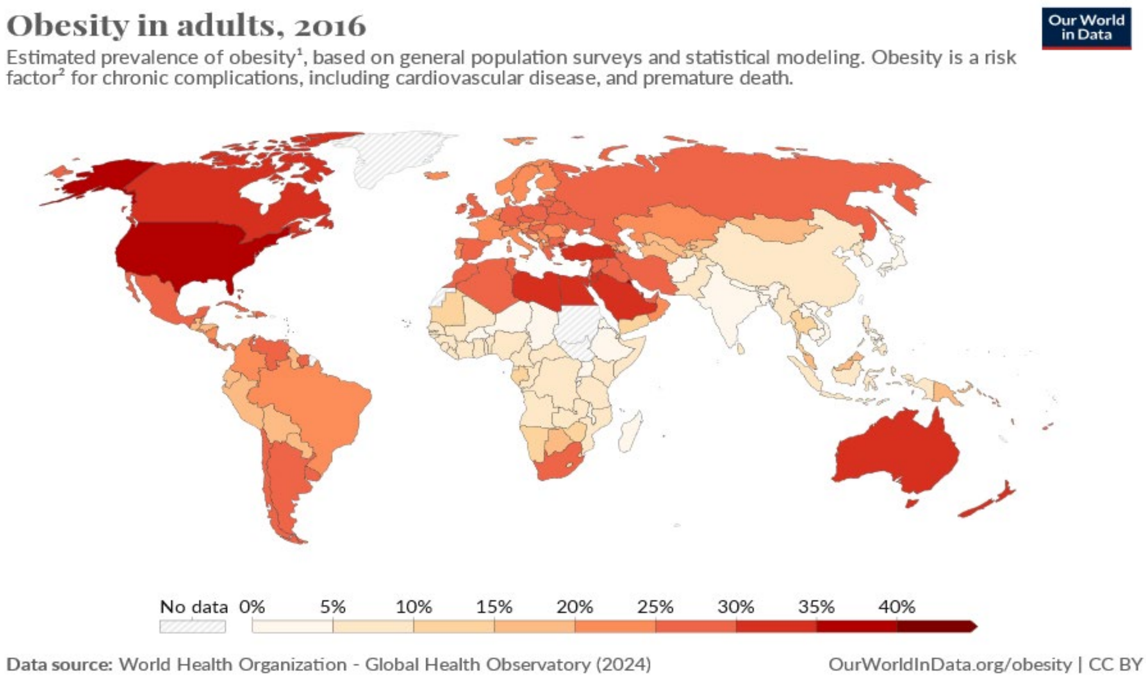

Figure 5: Colour-coded map displaying higher adult obesity percentages among darker regions and lower adult obesity percentage in lighter regions in 2016, (Ritchie and Roser, 2024).

The growing problem of obesity around the world is a complex issue influenced by many interconnected factors. One major contributor is urbanization, which changes how people live. In urban areas, people often lead more sedentary lifestyles, meaning they are less physically active through technological advancements and office-based jobs. This trend is especially noticeable in high-income countries like the United States, Canada, New Zealand and Australia (as shown in Figure 5), where reliance on cars and less outdoor activity have become common. Another factor playing an important role is easy access to high-calorie, low-nutrient foods further promoted by excessive advertising through platforms like social media. Many people in these countries are not properly informed about the ill-effects of fast-food consumption, whereas for some it is simply difficult to resist temptation.

The Pacific Island nations, including Nauru, Samoa, Tonga, Marshall Islands and others, have some of the highest obesity rates globally, often exceeding 40%. Several factors contribute to this situation. Firstly, genetic predispositions play a role, as Pacific Islanders tend to have larger body frames and a genetic inclination towards weight gain, according to journalist Meera Senthilingam. This can also be attributed to a certain extent to the nuclear testing at Bikini Atoll. Due to the large amounts of radiation, nearby islands such as Marshall Islands faced severe effects. Radioactive iodine, a biproduct of the nuclear testing, caused thyroid related issues within locals, causing obesity. Furthermore, the agricultural land was contaminated with radioactivity causing inability to consume locally grown food. Therefore, western nations started providing high calorie canned food to the effected islands which accelerated obesity, (Wang et al., 2019). Cultural perceptions that view larger body sizes as a sign of beauty, has also exacerbated the obesity crisis in these regions, (Senthilingam, 2015). In Nauru, rapid industrialisation with phosphate mining resulted in reduction of cultivable land. Therefore, locals were dependant on imported high calorie canned food, (McLennan and Ulijaszek, 2014). Lack of awareness towards the health consequences of obesity related to high calorie canned food consumption may also be attributed to such high levels in obesity.

In contrast, many countries in South Asia, Southeast Asia and East Asia such as, Nepal, India, Singapore, China, Japan, South Korea, etc., have lower obesity rates, despite varying levels of economic development. For example, Japan, a highly developed nation, maintains low obesity rates due to the factors discussed in section 3.2. Such cultural lifestyles may have an influence within several countries in these regions contributing to a lower obesity rate. Furthermore, the cultural emphasis on maintaining a healthy weight and social disgrace associated with obesity also discourage excessive weight gain in these regions.

Additionally, in less developed countries in Africa (for example, Uganda, Madagascar, Ethiopia, Niger, etc.,) and some of the Southeast Asian countries (for example, Vietnam, Cambodia, Bangladesh, Bhutan, Laos, etc.,) have limited access to processed foods in addition to a lifestyle that often involves more physical labour such as agriculture. This largely contributes to lower obesity rates.

South American countries, such as Argentina, Brazil, Bolivia, Colombia, Chile, etc., are facing ‘double edged sword’ according to (Popkin and Reardon, 2018). This means that while rapid urbanization is occurring throughout this region, flooding of processed food chains is oppositely arising explaining the relatively high obesity rates. 2016 figures show the obesity rates between 20-30%, which is rapidly increasing.

Middle eastern countries, such as UAE, Egypt, Saudi Arabia, Lebanon, Jordan, etc., with an average of around 25% to 30% obesity rate, can be attributed to the animal based high calorie diet within their cultures, sedentary lifestyle and economic changes, (Okati-Aliabad, et. al., 2022). In addition, adverse weather conditions discourage outdoor active lifestyle further promoting the sedentary lifestyle leading to such obesity rates.

Through this thorough analysis, it can be stated that complex interrelated factors are responsible for the increasing obesity rates. Out of the main factors identified in Figure 2, it can be agreed that geographical location and genetics are the factors which cannot be effectively controlled compared to other factors like lifestyle, society & systems and health (to some extent).

Furthermore, it is important to predict the current obesity trends in the future, assuming no change in the current factors affecting obesity levels. By doing this, we can get an overview of the countries which will be affected the most. The dominating factors of those countries will hold the key to effective solution development.

4. Python based prediction model development

Python is one of the most popular, fastest growing and user-friendly programming languages. It is flexible and can run on any modern computer. Python is a general-purpose programming language which can be used in almost every field, especially for advanced data analysis and model development. The python libraries contain a collection of codes that can be used in simple and complex ways to predict data, (Vilmate LLC, 2019). One of these libraries involve linear regression, which is the most common library for prediction model development. Apart from this several other libraries are used in the model such as libraries for data analysis, data visualisation, mathematical functions, etc. The unique combination of these libraries assists in overall model development.

Linear regression is used to predict the values of a dependant variable based on its correlation to the independent variable where its goal is to draw a straight line that best fits the relationship between the independent variable (years) and the dependant variable (actual obesity percentage), (Yale University, 2019). Excel offers the same algorithm however, python's efficiency in handling larger data sets is incomparable. In this model development, data from the year 1975 to 2016 of several different countries is used, of which the trendlines need to be predicted for future. This process requires automation in the model so that it can automatically predict the trendline further with good accuracy, for any given data points forming a trend. This is not possible to achieve using excel, which requires manual intervention at every stage in addition to its limitations with accuracy. Through the developed python-based prediction model, it is possible to get accurate predictions. Section 4.2 will provide more detailed information on the model development and the challenges faced.

4.1. Data Preprocessing

To prepare data for predicting obesity rates using Python, a few data preprocessing steps were necessary. First, Python's machine learning algorithms typically worked best with numerical data, therefore country names needed to be converted into a numerical format. This was done by assigning each country a different number, creating a representation of the countries that the algorithm could understand.

In addition, the dataset contained information that was not relevant or could potentially skew the results. This included regional data (which might have represented averages across multiple countries) and data of countries that no longer existed, like former Sudan. These entries needed to be identified and removed to ensure the dataset only contained correct, country-specific information.

4.2. Model Development

The model development approach taken in this project is to build a basic model first and then start adding complexities to it so that its accuracy increases, and it becomes closer to reality. Finally, model generalization is performed to expand the scope of the model so that it can predict any obesity trend of any country with good accuracy.

4.2.1. Basic Linear Regression Model Development

In order to initiate the model development process and experiment with the data and codes, I originally started developing a prediction model designed to predict the obesity rates of a single country. I had chosen this country to be Norway and entered the following codes:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
```

I have imported the libraries: pandas, matplotlib, train test split and linear regression. These are general libraries that help in model development. Libraries are chosen based on what is required to achieve through the model.

```
# Load the original dataset for changes in Obesity% for Norway
file_path = 'obesity-in-norway.csv'  # Update with your file path
data = pd.read_csv(file_path)

# Extract the features and target variable
X = data[['Year']]
y = data['Actual Obesity %']
```

The code above uses pandas, an imported library that allows python to read the named csv file, which contains data related to obesity. This csv file is labelled as 'data'. While reading the data in python, specification of the features (input) and target (output) is required. The feature and target in my provided data set is the 'year' and the 'actual obesity percentage' respectively. While plotting and training of the data, the 'x' values can be considered as 'year', and the 'y' values can be considered as 'obesity percentage'.

```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

With the above code, a new set of codes (algorithm) 'train_test_split' is used from the library 'sklearn'. This algorithm is used for splitting the data into training, testing and validation data sets in a machine learning prediction model. The logic behind this method is that when the model is first developed, several errors can be prevalent in the model causing its accuracy to be very poor. Therefore, instead of entering all the data into the prediction model, 80% is used as the 'training data' which is entered into the model, and 20% is used as the 'testing data' which evaluates the accuracy of the training data. The value of the random state (42) is chosen arbitrarily. After the evaluation, certain hyperparameters can be added for the purpose of tuning the model for better efficiency. Once the model is run again, the validation data set is used to compare the accuracy of the predicted values.

```
# Create the Linear Regression model
model = LinearRegression()

# Train the model
model.fit(X_train, y_train)

# Generate predictions for all years from the minimum year in the dataset to 2050
years_to_predict = pd.DataFrame({'Year': range(data['Year'].min(), 2051)})
```

These codes are responsible for 3 aspects of the model. The first code creates the model through linear regression from 'sklearn'. The second code trains the model using the x and y values. The third code generates predictions of the obesity percentages based on its correlation to the years within the testing data.

```python
# Plot the results
plt.figure(figsize=(10, 6))

# Actual data points
plt.scatter(comparison_df['Year'], comparison_df['Actual Obesity %'], color='blue', label='Actual Obesity %', s=10)

# Line connecting predicted points
plt.plot(comparison_df['Year'], comparison_df['Predicted_Obesity'], color='red', linewidth=1, label='Predicted Trend')

plt.xlabel('Year')
plt.ylabel('Obesity %')
plt.title('Actual vs Predicted Obesity % up to year 2050')
plt.legend()
plt.grid(True)
plt.show()
```

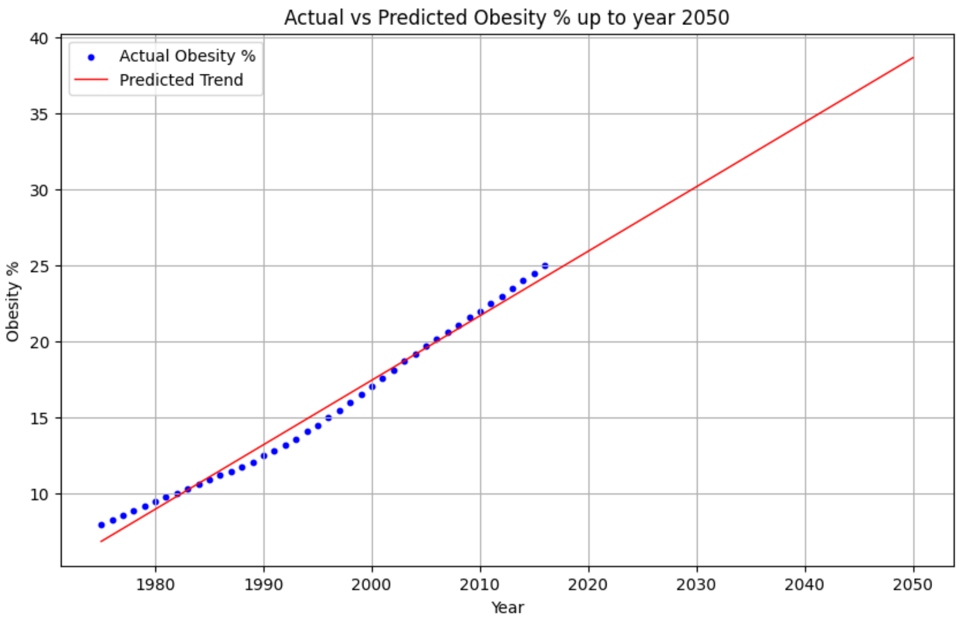

Figure 6: Norwegian predicted obesity trends from 1975 to 2050 through a linear relationship.

Next, I have used the codes above to plot the results using the 'matplotlib' library. This has produced a graph which shows the correlation between the x and y values (Figure 6). The red trendline is the based on the predicted values from a linear regression model. Prediction of obesity percentage is done up to the year 2050.

```python
# Predictions for all the years
predictions = model.predict(years_to_predict)

# Combine actual and predicted values into a DataFrame
# Filter the actual values from the original dataset
actual_values = data[['Year', 'Actual Obesity %']]

# Adjust the display settings to show all rows
pd.set_option('display.max_rows', None)

# Merge the actual and predicted values
comparison_df = pd.merge(
    actual_values,
    years_to_predict.assign(Predicted_Obesity=predictions),
    how='right',
    on='Year'
)

# Display both actual and predicted obesity %
print(comparison_df)
```

```
Year Actual Obesity % Predicted_Obesity
1975              8.0          6.868215
1976              8.3           7.29209
1977              8.6          7.715966
1978              8.9          8.139842
1979              9.2          8.563717
 ...              ...               ...
2001             17.6         17.888981
2002             18.1         18.312856
2003             18.7         18.736732
2004             19.2         19.160608
2005             19.7         19.584483
2006             20.2         20.008359
 ...              ...               ...
2010             22.0         21.703861
2011             22.5         22.127737
2012             23.0         22.551613
2013             23.5         22.975488
2014             24.0         23.399364
2015             24.5          23.82324
2016             25.0         24.247115
 ...              ...               ...
2045              NaN         36.539508
2046              NaN         36.963384
2047              NaN         37.387259
2048              NaN         37.811135
2049              NaN         38.235011
2050              NaN         38.658886
```

By running the above codes, the actual and predicted obesity percentages get displayed. The displayed results are also shown above.

The results given from the model are not accurate enough as it is a linear regression model with a linear predicted trend. For the results to be more accurate, the model must be able to adapt to the plotted curves.

4.2.2. Improved Regression Model Development

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error
```

In the improved model, three new libraries, 'PolynomialFeatures' from 'sklearn.preprocessing', 'numpy' and 'mean _absolute_error' from 'sklearn.metrics' are imported along with the previously imported libraries. I imported the library called 'PolynomialFeatures' to transform the original data by raising them to powers in stated degrees. For example, a feature 'x', of degree 5 would create 'x^5'. This allows the model to capture non-linear relationships that a simple linear model might miss. In this case, as I am predicting the obesity percentages over time, using polynomial features; it lets the model consider not only the year but also the year squared or cubed, helping to identify trends that may change direction or intensity over time. The 'numpy' library is imported to efficiently execute advanced mathematical operations related to large data sets. In the present case, 'numpy' library is used by the model to understand different polynomial equations related to different curves in the trend.

The 'mean _absolute_error' library helps to calculate the accuracy of all the predictions together. Mean absolute percentage error (MAPE) is calculated as the average of absolute percentage errors and represented through the following formula:

$$MAPE = \frac{\sum \frac{|A-P|}{A} \times 100}{n}$$

where, A is the actual data; P is the predicted data; n is the number of observations/iterations and the vertical bar symbolises absolute values (Institute of Business Forecasting & Planning, 2024).

```
# Load the dataset
data = pd.read_csv('obesity-in-norway.csv')

# Define a function to train the model and make predictions
def train_and_predict(data):
    years = np.arange(1975, 2051)  # Prediction range

    # Extract the features (X) and the target (y)
    X = data[['Year']]
    y = data['Actual Obesity %']

    # Split the data into training and testing sets
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

    # Create polynomial features up to degree 4
    poly = PolynomialFeatures(degree=4)
    X_train_poly = poly.fit_transform(X_train)

    # Train the linear regression model on the polynomial features
    model = LinearRegression()
    model.fit(X_train_poly, y_train)

    # Prepare the prediction data
    X_pred = pd.DataFrame({'Year': years})
    X_pred_poly = poly.transform(X_pred)
    y_pred = model.predict(X_pred_poly)

    return pd.DataFrame({'Year': years, 'Predicted_Obesity %': y_pred})

# Get predictions
predictions_df = train_and_predict(data)
```

Through the above given codes, the model is assigned a function to train and make predictions through a specified year range of 1975 to 2050. Consecutively, the features and targets (x values and y values) are extracted from the loaded file. Thereafter, the polynomial features are added and set to a degree (power) of 4. The model is then created through linear regression, specifying the values (x and y) that must be used to train the model to make predictions. Finally, the last code asks for the predictions from the training data to interpret its accuracy.

```
# Generate predictions for all years from the minimum year in the dataset to 2050
years_to_predict = pd.DataFrame({'Year': range(data['Year'].min(), 2051)})

# Combine actual and predicted values into a DataFrame
# Filter the actual values from the original dataset
actual_values = data[['Year', 'Actual Obesity %']]

# Adjust the display settings to show all rows
pd.set_option('display.max_rows', None)

# Merge the actual and predicted values
comparison_df = pd.merge(
    years_to_predict,
    pd.merge(actual_values, predictions_df, on='Year', how='right'),
    on='Year',
    how='left'
)

# Display both actual and predicted obesity %
print(comparison_df)
```

```
Year Actual Obesity % Predicted_Obesity %
1975             8.0            8.315128
1976             8.3            8.474669
1977             8.6            8.654785
1978             8.9            8.854859
1979             9.2            9.074271
 ...             ...                 ...
2001            17.6           17.528715
2002            18.1           18.019821
2003            18.7           18.515118
2004            19.2           19.013961
2005            19.7           19.515707
 ...             ...                 ...
2045             NaN           34.450183
2046             NaN            34.50428
2047             NaN           34.533345
2048             NaN           34.536694
2049             NaN           34.513639
2050             NaN           34.463493
```

With the above codes, the predictions until 2050 are generated and displayed alongside the actual obesity percentages so that it is easier to compare them. The displayed results are also shown above.

```
# Extract the validation dataset for years 2006 to 2016
validation_df = comparison_df[(comparison_df['Year'] >= 2006) & (comparison_df['Year'] <= 2016)]

# Print the validation dataset
print("\nValidation Dataset (2006-2016):")
print(validation_df)

# Calculate accuracy metrics
actual = comparison_df['Actual Obesity %'].dropna()
predicted = comparison_df['Predicted_Obesity %'].dropna()

# Ensure actual and predicted arrays are aligned
mask = ~pd.isna(actual) & ~pd.isna(predicted)
actual = actual[mask]
predicted = predicted[mask]

# Calculate the metrics
mae = mean_absolute_error(actual, predicted)

# Calculate Absolute Percentage Error (APE) and Mean Absolute Percentage Error (MAPE)
ape = np.abs((actual - predicted) / actual) * 100
mape = np.mean(ape)

# Print the metrics
print(f"\nMean Absolute Percentage Error (MAPE): {mape:.2f}%")
```

```
Validation Dataset (2006-2016):
    Year  Actual Obesity %  Predicted_Obesity %
31  2006              20.2            20.019711
32  2007              20.6            20.525326
33  2008              21.1            21.031905
34  2009              21.6            21.538801
35  2010              22.0            22.045363
36  2011              22.5            22.550942
37  2012              23.0            23.054887
38  2013              23.5            23.556546
39  2014              24.0            24.055265
40  2015              24.5            24.550392
41  2016              25.0            25.041270

Mean Absolute Percentage Error (MAPE): 0.93%
```

By using the codes above, the model identifies the validation data set from 2006 to 2016. The validation set consists of existing data that the model is not provided, so that it is used to compare the predicted values of the model to determine the model's accuracy. The validation dataset is displayed above. Thereafter, the mean absolute percentage error is calculated using the actual obesity % data, predicted obesity % data and the number of years this data was compared in the model. The MAPE varies with the stated polynomial feature degree. A degree of 4 displays minimal error of 0.93%. This means the model is having an accuracy of 99.07%.

```
# Plot the results
plt.figure(figsize=(10, 6))

# Actual data points
plt.scatter(comparison_df['Year'], comparison_df['Actual Obesity %'], color='blue', label='Actual Obesity %', s=10)

# Line connecting predicted points
plt.plot(comparison_df['Year'], comparison_df['Predicted_Obesity %'], color='red', linewidth=1, label='Predicted Trend')

plt.xlabel('Year')
plt.ylabel('Obesity %')
plt.title('Actual vs Predicted Obesity % up to year 2050')
plt.legend()
plt.grid(True)
plt.show()
```

Thereafter, the codes plot the results using the actual data and predicted data onto a graph (Figure 7).

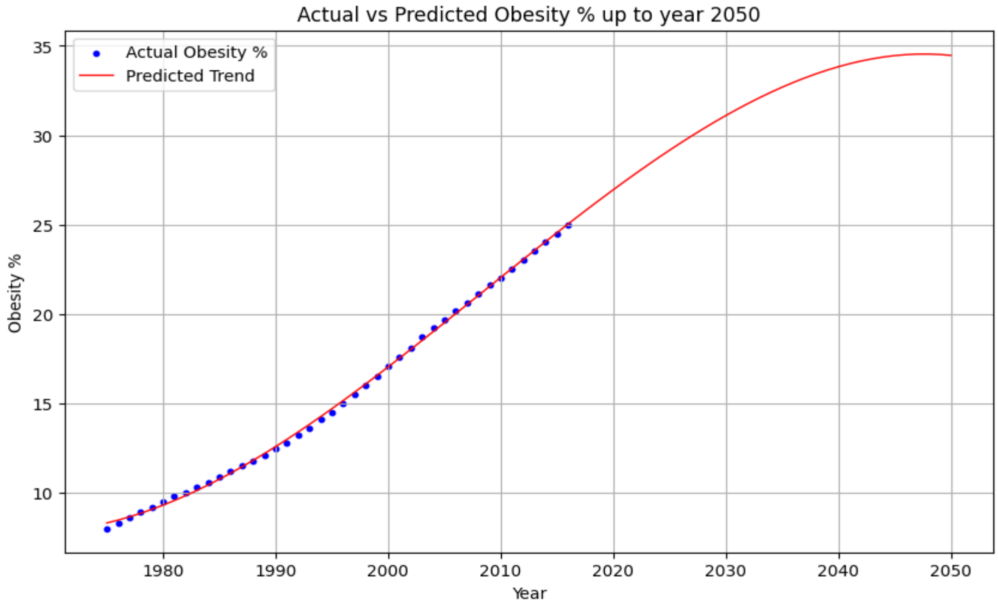

Figure 7: Norwegian predicted obesity trends from 1975 to 2050 through a linear relationship with polynomial features.

Although the model which has been developed for Norwegian obesity data is of high accuracy, it is not generalized in a way such that any other country data can be fed into the model to get accurate future obesity predictions. Therefore, to generalize the model, new set of codes are required to include multiple country data.

4.2.3. Model Generalisation

For model generalisation, I have used the same libraries as previously used in section 4.2.2.

```
# Load the dataset
data = pd.read_csv('Obesity_country_year_data.csv')

# Load the country mapping dataset
country_mapping_df = pd.read_csv('country_mapping.csv')

# Merge the country mapping with the main dataset
data = pd.merge(data, country_mapping_df, left_on='Country Number', right_on='Country Number', how='left')
```

The above code loads 2 datasets. The first dataset provides the obesity data for each country labelled as numbers. The second dataset provides the assigned number to each of these countries. The 2 files are then merged so that the codes can identify the countries labelled as numbers. In total, adult obesity data of 190 countries are into this Python model.

```
# Function to train the model and make predictions for each country
def train_and_predict_by_country(data):
    countries = data['Country Name'].unique()
    years = np.arange(1975, 2051)

    predictions = pd.DataFrame()
    actual_values = pd.DataFrame()

    for country in countries:
        country_data = data[data['Country Name'] == country]

        if country_data.empty:
            print(f"No data for country: {country}")
            continue

        X = country_data[['Year']]
        y = country_data['Actual_Obesity %']
```

Through this code, I train the model to make predictions for the loaded dataset from the year 1975 to 2050. I have kept the data from 2006 to 2016 to be used as the validation dataset to compare the predicted data to the actual data. The code then provides the function to split the data within the data sets of each country. In the next code I specify the x (independent variable) and y (dependent variable) values in the dataset.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

The code above provides the function to split the data into training and testing data. The data is split randomly at a rate of 42 where 80% of the data is training data and 20% is testing data. The value of the random state (42) is chosen arbitrarily.

```
    poly = PolynomialFeatures(degree=4)
    X_train_poly = poly.fit_transform(X_train)
    X_test_poly = poly.transform(X_test)

    model = LinearRegression()
    model.fit(X_train_poly, y_train)

    X_pred = pd.DataFrame({'Year': years})
    X_pred_poly = poly.transform(X_pred)
    y_pred = model.predict(X_pred_poly)

    country_predictions = pd.DataFrame({
        'Country Name': country,
        'Year': years,
        'Predicted_Obesity %': y_pred
    })
    predictions = pd.concat([predictions, country_predictions], axis=0)

    country_actual = country_data[['Country Name', 'Year', 'Actual_Obesity %']]
    actual_values = pd.concat([actual_values, country_actual], axis=0)

return predictions, actual_values
```

Through the code above, I apply polynomial features, which has been prepared and set to a degree of 4. In the following code, the actual model is created using linear regression functionality, where the defined x and y values are used to predict future values. The final codes program the model to generate predictions where the actual and predicted values are defined.

```
# Get predictions for all countries
predictions_df, actual_values_df = train_and_predict_by_country(data)

# Merge actual and predicted values for comparison
comparison_df = pd.merge(actual_values_df, predictions_df, on=['Country Name', 'Year'], how='right')

# Display both actual and predicted obesity %
print(comparison_df)
```

```
Country Name  Year  Actual_Obesity %  Predicted_Obesity %
 Afghanistan  1975               0.4             0.395973
 Afghanistan  1976               0.4             0.429309
 Afghanistan  1977               0.5             0.463037
 Afghanistan  1978               0.5             0.497387
 Afghanistan  1979               0.5             0.532588
         ...   ...               ...                  ...
    Zimbabwe  2046               NaN            14.958741
    Zimbabwe  2047               NaN            14.881086
    Zimbabwe  2048               NaN            14.788727
    Zimbabwe  2049               NaN            14.681342
    Zimbabwe  2050               NaN            14.558608
```

The above code is written to display the predictions through the trained data. It is not possible to show all the data therefore truncated data is shown in this report.

```
# Extract the validation dataset for years 2006 to 2016 for all countries
validation_df = comparison_df[(comparison_df['Year'] >= 2006) & (comparison_df['Year'] <= 2016)]

# Print the validation dataset
print("\nValidation Dataset (2006-2016):")
print(validation_df)
```

```
Validation Dataset (2006-2016):
      Country Name  Year  Actual_Obesity %  Predicted_Obesity %
31     Afghanistan  2006               2.6             2.640403
32     Afghanistan  2007               2.8             2.792276
33     Afghanistan  2008               2.9             2.951648
34     Afghanistan  2009               3.1             3.118751
35     Afghanistan  2010               3.3             3.293818
...            ...   ...               ...                  ...
14401     Zimbabwe  2012              11.1            11.156198
14402     Zimbabwe  2013              11.3            11.386337
14403     Zimbabwe  2014              11.6            11.612715
14404     Zimbabwe  2015              11.9            11.835001
14405     Zimbabwe  2016              12.3            12.052866
```

The final code displays the predicted data from the developed generalised model for each country during the validation years which provides a comparison of the actual data and predicted data.

```python
# Calculate accuracy metrics for each country
def calculate_accuracy_by_country(comparison_df):
    countries = comparison_df['Country Name'].unique()
    results = []

    for country in countries:
        country_data = comparison_df[comparison_df['Country Name'] == country]
        actual = country_data['Actual_Obesity %'].dropna()
        predicted = country_data['Predicted_Obesity %'].dropna()

        if actual.empty or predicted.empty:
            print(f"No valid data for accuracy calculation for country: {country}")
            continue

        mask = ~pd.isna(actual) & ~pd.isna(predicted)
        actual = actual[mask]
        predicted = predicted[mask]

        mae = mean_absolute_error(actual, predicted)
        ape = np.abs((actual - predicted) / actual) * 100
        mape = np.mean(ape)
        Accuracy = 100 - mape

        results.append({'Country Name': country, 'MAPE': mape, 'Accuracy': Accuracy})

    return pd.DataFrame(results)

# Calculate and display accuracy by country
accuracy_df = calculate_accuracy_by_country(comparison_df)
print("\nAccuracy by Country:")
print(accuracy_df)

# Compute overall accuracy metrics
def calculate_overall_accuracy(comparison_df):
    actual = comparison_df['Actual_Obesity %'].dropna()
    predicted = comparison_df['Predicted_Obesity %'].dropna()

    if actual.empty or predicted.empty:
        print("No valid data for overall accuracy calculation")
        return None

    mask = ~pd.isna(actual) & ~pd.isna(predicted)
    actual = actual[mask]
    predicted = predicted[mask]

    overall_mae = mean_absolute_error(actual, predicted)
    ape = np.abs((actual - predicted) / actual) * 100
    overall_mape = np.mean(ape)

    return overall_mape

overall_mape = calculate_overall_accuracy(comparison_df)
if overall_mape is not None:
    Overall_Accuracy = 100 - overall_mape
    print(f"Overall MAPE: {overall_mape}%")
    print(f"Overall Accuracy: {Overall_Accuracy}%")
```

```
Accuracy by Country:
      Country Name      MAPE   Accuracy
0      Afghanistan  2.118082  97.881918
1          Albania  0.933990  99.066010
2          Algeria  0.655865  99.344135
3          Andorra  0.233776  99.766224
4           Angola  1.245693  98.754307
..             ...       ...        ...
185      Venezuela  0.363689  99.636311
186        Vietnam  7.598263  92.401737
187          Yemen  1.341813  98.658187
188         Zambia  1.932617  98.067383
189       Zimbabwe  1.004558  98.995442
```

```
Overall MAPE: 1.1652082515582194%
Overall Accuracy: 98.83479174844177%
```

The above set of codes ensure that the accuracies of the predicted values for each country can be calculated from their similarities to the validation data sets through mean percentage absolute error (MPAE), where the average of MPAE gives the overall accuracy of the model.

```python
# Plot the actual vs predicted obesity % for all countries
plt.figure(figsize=(15, 10))
for country in predictions_df['Country Name'].unique():
    country_actual = comparison_df[comparison_df['Country Name'] == country]
    plt.plot(country_actual['Year'], country_actual['Actual_Obesity %'], '.', label=f'{country} Actual')
    plt.plot(country_actual['Year'], country_actual['Predicted_Obesity %'], '-', label=f'{country} Predicted')

plt.xlabel('Year')
plt.ylabel('Obesity %')
plt.title('Actual vs Predicted Obesity % by Country (1975-2050)')
plt.grid(True)
plt.show()
```

Through the codes entered above, the plotting of the actual obesity % vs the predicted obesity % up to 2050 is performed for 190 countries as shown in Figure 8.

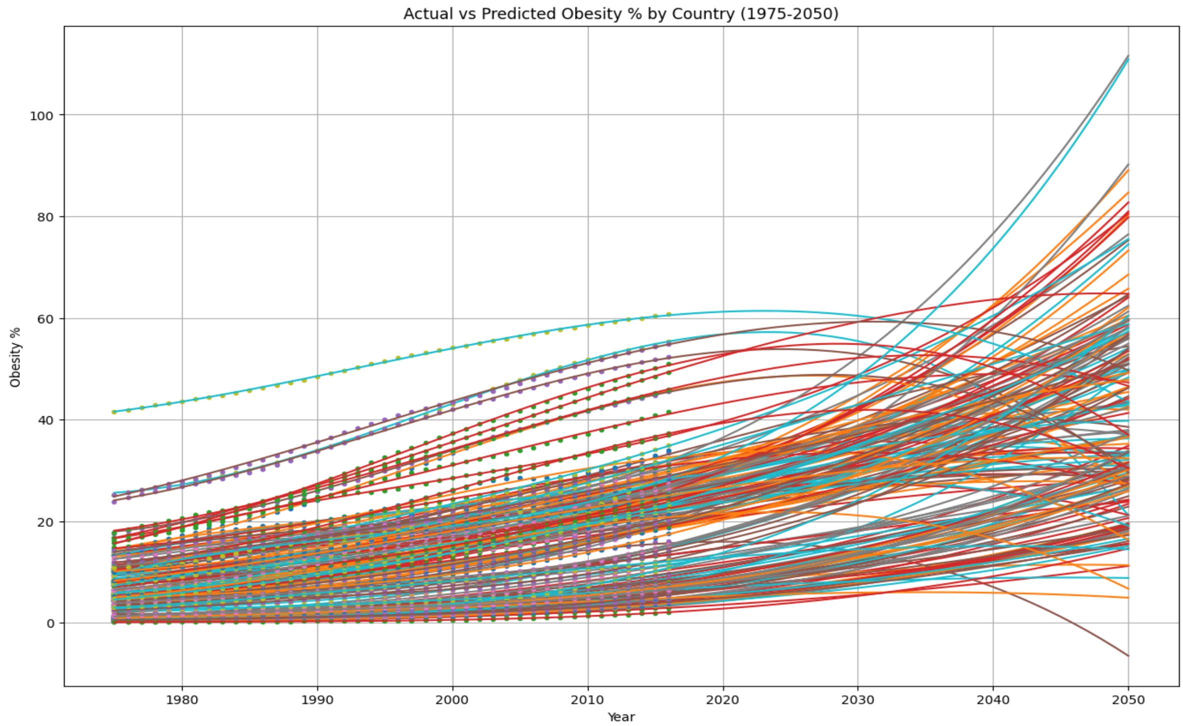

Figure 8: Predicted obesity trends of 190 countries from 1975 to 2050 developed by analysing the linear relationship with added polynomial features.

5. Results and Discussion

From the predicted results, a distinct regional correlation is evident. Certain trends are repeatedly seen within similar regions indicating that the obesity prevalence is a regional analysis. Therefore, this section will focus on justifying the different trends between different regions and draw connections to the factors that are responsible for the resulted trends.

5.1. Analysis of Predicted Results

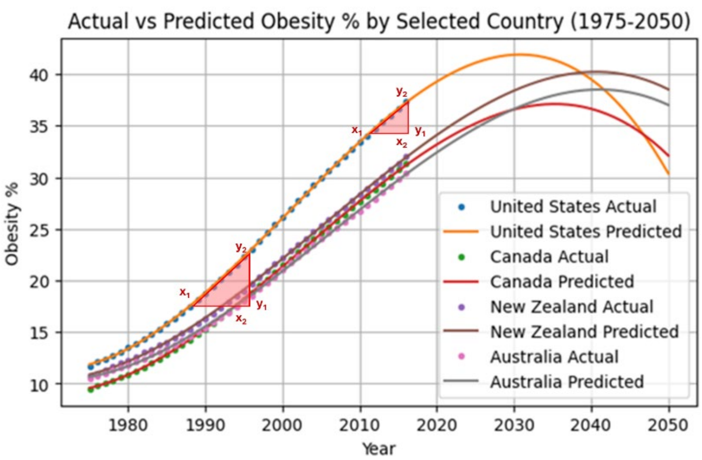

Figure 9: Actual & predicted obesity trends of USA, Canada, New Zealand and Australia from 1975 to 2050.

Developed countries like the United States, Cananda, Australia and New Zealand show a peak in their predicted obesity rates around 2030-2040. This regional connection may suggest a big factor of being economically developed. As a country becomes developed, more and more people start becoming aware of the health threats related to obesity. Literacy rate is high, resulting in more people having the power to think rationally about their health. Governments, get more spending power and implement obesity preventive measures which will slow down the rate of increase in obesity. In the US, government initiatives have been taken between 2005 and 2016 that have ultimately slowed their increasing obesity rates. In 2010, President Barrack Obama signed the Healthy Hunger – Free Kids Act, which resulted in 30 million children to healthier diets. The effect of such initiatives is clearly visible in Figure 9. If a slope between 1988-1996 and another slope between 2011-2016 is calculated, the value of slope ($\frac{y_2-y_1}{x_2-x_1}$) reduces as the trend progresses over time, explaining how some of the government initiatives taken after 2005 have affected the obesity trend of its country. These kinds of policies bring change in long term, but the government should be prepared for long term investment to see health benefits in its citizens. (Center for Science in the Public Interest, 2020)

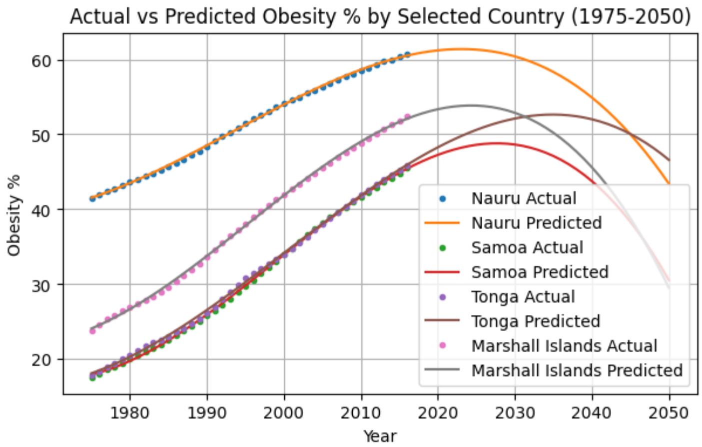

Figure 10: Actual & predicted obesity trends of Nauru, Samoa, Tonga and Marshall Islands from 1975 to 2050.

Figure 10 shows the predicted obesity trends of Pacific Island countries like Nauru, Samoa, Tonga and Marshall Islands. From this trend we can see that all the countries of this region peak near the years 2020-2040, indicating common factors that are reducing obesity levels within these Islands. In Marshall Islands the effects of radiation have slowly started to decrease as the agriculture and traditional lifestyle is gradually reviving, explaining the peak in Figure 10. Furthermore, high taxations on high calorie products in Tonga and Nauru were introduced between 2005-2016 in addition to subsidy on bottled water, (Foster et al., 2018). Various NGOs ran health awareness programs for locals during the same period to control the high obesity levels. As a result of these measures taken during 2005-2016, we ultimately see a downfall in obesity trends from 2020-2040 onwards.

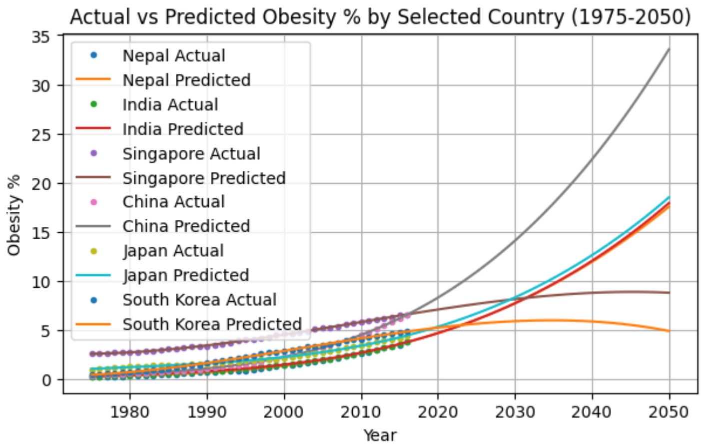

Figure 11: Actual & predicted obesity trends of Nepal, India, Singapore, China, Japan, and South Korea from 1975 to 2050.

Figure 11 displays predicted obesity data of South and East-Asian countries like Nepal, India, Singapore, China, Japan, and South Korea. This graph presents a distinct divide between the trends of Singapore and South Korea, compared to Nepal, India, China, and Japan. Such contrasting trends may be attributed to several factors including economic development and population. Countries such as India and China have massive populations making it much harder to control the obesity prevalence within the country. It will continue to increase until its economic development is sufficient to initiate regulations and preventative measure for the entire population. Countries such as Nepal have simply not reached that level of economic development which is needed to decrease such obesity trends. Singapore and South Korea are economically developed countries. An example of Singapore's development is presented through its thought-out urban planning which prevents sedentary lifestyles and maintains active lifestyles throughout the country. In addition, the work environment in Japan, China and India is very competitive, demanding and stressful. This causes lack of physical activity and constant consumption of high calorie foods like pizza, burgers, carbonated sugary drinks etc. Stress eating is also a notable factor.

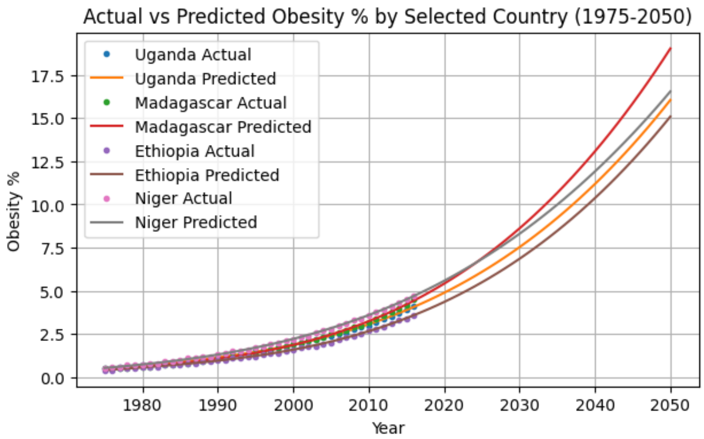

Figure 12: Actual & predicted obesity trends of Uganda, Madagascar, Ethiopia and Niger from 1975 to 2050.

Countries in the African region, such as Uganda, Madagascar, Ethiopia and Niger show predicted trends rising exponentially, reaching obesity percentages ranging from 15% to 20% in 2050, as displayed in Figure 12. These trends may be attributed to the lack of awareness to obesity and related health consequences within these regions. As a result, many tend to simply consume the food, which is easily available, unaware of its impacts on their health. In addition, urbanisation and adaption to western culture may occur in these countries in the coming years, which may be an additional cause to the predicted increase in obesity levels. World Health Organization (WHO) is trying to support African nations to develop laws and regulations to impose high taxes based on food sugar and fat content (WHO, 2022). However, until these laws are implemented, it is difficult to see the peak in obesity levels.

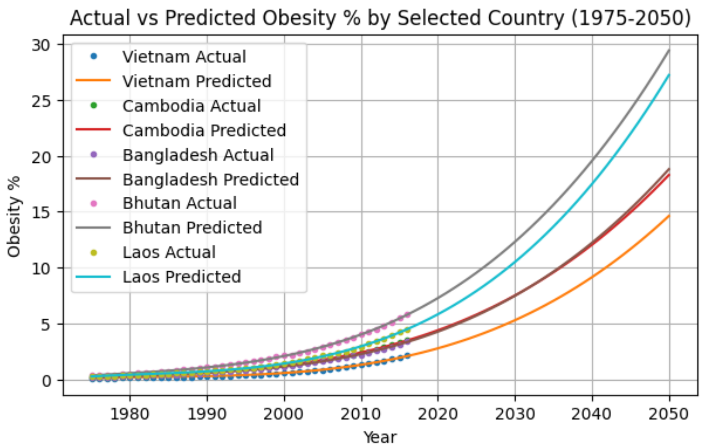

Figure 13: Actual & predicted obesity trends of Vietnam, Cambodia, Bangladesh, Bhutan, and Laos from 1975 to 2050.

Similar to the trends in Figure 12, the predicted obesity trends of Vietnam, Cambodia, Bangladesh, Bhutan, and Laos, displayed in Figure 13, show similar increasing trends rising exponentially from 1975 to 2050. It is expected that as the economies of these countries start expanding, they might potentially transition from traditional Asian diets and lifestyle to adopting Western lifestyle. Currently, these countries have not reached the stage of development to implement such preventive measures through governments with lasting impact explaining why the trends do not show a peak up to 2050.

The predicted obesity trends for South American countries like Argentina, Brazil, Bolivia, Colombia, and Chile show a consistent linear increase over time, without a clear peak or plateau in sight. This continuous upward trajectory can be attributed to several factors related to rapid urbanization and economic development in the region. As these countries are undergoing urbanization, there has been an adoption of more Western-style diets and sedentary lifestyles, contributing to rising obesity rates, (Ferrari et al., 2022). Additionally, the economic status of these countries plays a crucial role in their ability to implement effective obesity prevention measures. Unlike more developed nations, many South American countries lack the financial resources and infrastructure to launch comprehensive government initiatives targeting obesity on a large scale, (Anza-Ramirez et al., 2022) However, as these

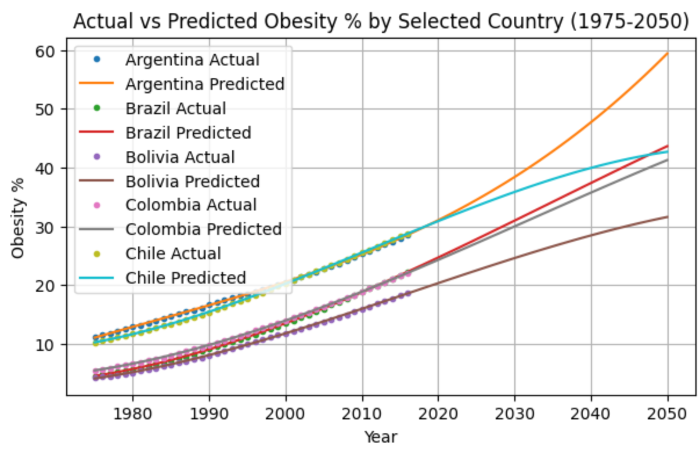

Figure 14: Actual & predicted obesity trends of Argentina, Brazil, Bolivia, Colombia, and Chile from 1975 to 2050.

countries continue to develop economically, there is potential for a shift in these trends. With increased wealth and resources, governments may eventually be able to implement more robust preventative measures and public health initiatives aimed at peaking obesity rates. This could include improvements in urban planning to encourage physical activity, regulations on food marketing, and subsidies for healthier food options. Furthermore, the intelligence and awareness of the health risks associated with obesity will increase as the country develops, making citizens themselves initiate programs to maintain body weight and prevent obesity. As these countries progress, they may follow a similar pattern to more developed nations, potentially seeing a stabilization or even decline in obesity rates in the future.

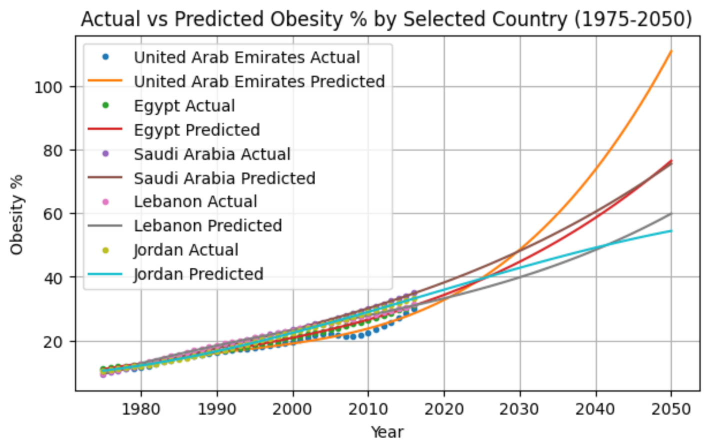

Figure 15: Actual & predicted obesity trends of UAE, Egypt, Saudi Arabia, Lebanon, and Jordan from 1975 to 2050.

As shown in Figure 15, Middle Eastern countries show a consistent upward trend in predicted obesity rates, which are driven by environmental, cultural, and economic factors. The United Arab Emirates stands out as an outlier, 0077ith a decreasing obesity trend from 2005 to 2008 and then a rapidly increasing trend from 2009 to 2016 which is inconsistent with previous years data and potentially leading to inaccurate predictions. The region's hot climate promotes sedentary lifestyles, while traditional diets high in animal-based foods contribute to increased caloric intake. Moreover, the local population is keen to embrace western dietary trends. In Saudi Arabia, the government has introduced 50% excise tax on carbonated drinks in 2017 (Alsukait, et. al., 2020). Other measures on controlling obesity are in the process of implementation in Saudi Arabia and other middle eastern countries. Depending on effectiveness and scale of implementation of these measures, the predictions might change in the coming years.

6. Analysis of Possible Solutions

The analysis of the predicted trends through the python-based prediction model provides a distinct correlation between the economic development of a country and its obesity rates. Developed countries display a peak in their rising obesity levels before 2050. However, developing countries only display exponential or linear increasing trends until 2050. In order to get the peak and thereafter reduction in obesity trends in developing and underdeveloped countries by 2050, multiple solutions need to be effectively implemented on a wide scale. Chapter 6 provides an overview of these solutions.

6.1. Urban planning

Urban planning plays a crucial role in combating obesity by shaping environments that promote healthier lifestyles. Well-designed cities can significantly impact diet, physical activity, and behaviour. For instance, increasing the availability of green spaces, parks, recreational areas and fitness centres encourage outdoor activities and exercise. Implementing pedestrian and cycling infrastructure, such as wide sidewalks and bike lanes in cities and towns, promote active transportation. Mixed-use zoning that combines residential, commercial, and recreational areas reduces reliance on cars and increases daily physical activity. Urban agriculture initiatives, like community gardens and farmers' markets, can improve access to fresh, healthy foods. Additionally, limiting the availability of fast-food outlets and increasing the presence of grocery stores with fresh vegetables and fruits, salad bars can positively influence dietary choices. Smart city designs that prioritize walking, with features like street connectivity and attractive public spaces, naturally encourage more physical activity. By integrating these elements into urban planning, cities can create environments that encourage residents towards healthier behaviours, potentially leading to a peak in obesity rates as these changes take effect over time.

6.2. Policy Interventions

Government policies are powerful tools to prevent the escalation of obesity rates. Implementing strict laws and regulations can lead to significant changes in population health. Nutrition labelling policies, such as mandatory calorie counts on restaurant menus and clear front-of-package labelling, encourage consumers to make informed food choices. Increased taxation on sugar-sweetened beverages and unhealthy foods, can shift consumption patterns towards healthier options. Higher taxation on owning or renting properties for high calorie fast food outlets. Advertising regulations, particularly those targeting children, can reduce exposure to marketing of unhealthy foods. Stricter government regulations enforced by labour unions on work culture, limitations on working hours with an aim to balance work and family life will reduce stress and stress related sleep issues. By implementing and enforcing these policy interventions, governments can create a supportive environment for healthy living.

6.3. Government Programs

Government programs, including the development and promotion of simple apps and devices for tracking health and fitness, can play a significant role in combating obesity. National health initiatives can influence technology to reach wider audiences and provide personalized support. For example, government-sponsored apps could offer features like meal planning, calorie tracking, and physical activity logging, tailored to local dietary guidelines and cultural preferences. These apps could integrate with national health databases, allowing healthcare providers to monitor patients' progress and offer timely interventions. Wearable device programs can encourage increased physical activity by tracking steps, heart rate, and sleep patterns. Government-run websites and social media campaigns can provide reliable, evidence-based information on nutrition and exercise, countering misinformation. Public health agencies could develop AI-powered chatbots to provide support and answer questions about healthy living. By investing in these technological solutions and making them widely available, governments can empower individuals to take control of their health, potentially leading to a peak in obesity rates as more people adopt these tools and make sustainable lifestyle changes.

6.4. Community Programs

Local initiatives and support groups are essential to reduce obesity levels, as they provide personalized, culturally relevant interventions. Community-based weight management programs, led by trained local health workers, can offer tailored advice and support in familiar settings.

Neighbourhood walking groups and fitness or dance classes in community centres can make exercise more accessible and enjoyable. Community cooking or local cooking classes focusing on healthy, affordable meals can improve nutritional knowledge and cooking skills along with social mingling which reduces stress. School-based programs involving parents and children can create a supportive environment for healthy habits at home. Support groups for individuals struggling with obesity can provide emotional support and practical advice. By implementing these community-focused programs, cities and towns can create a network of support that makes healthy living easier and more appealing. These initiatives will slowly take root and spread throughout communities.

6.5. Weight Loss Treatments

Weight loss treatments offer promising solutions to help reduce obesity rates worldwide. Innovative medications like Wegovy (semaglutide) have emerged as powerful tools in the fight against excess weight. These injectable drugs work by sending signals to the brain, suppressing appetite and increasing feelings of fullness, making it easier for people to eat less and lose weight. When combined with healthy eating habits and regular exercise, these treatments can lead to significant and sustainable weight loss, (Egge, 2024). The development of such advanced treatments can considerably improve obesity management. They provide new options for people who have struggled with traditional weight loss methods. As research continues, we can expect even more effective and personalized weight loss solutions to become available. This progress will give doctors more tools to help their patients, allowing them to create tailored treatment plans that work best for each individual. By expanding the range of effective options, we can better address the diverse needs of people dealing with obesity and ultimately contribute to lowering obesity rates on a larger scale.

7. Conclusion

7.1. Summary of Findings

This research has revealed the complex nature of obesity influenced by a range of interconnected factors. Chapter 5 covers discussion on the specific factors that are influencing the predicted obesity trends of different global regions. The fact that most of the countries in one region align with similar predictions confirm that same factors influence countries in that region. If there is a peak and thereafter reduction in obesity levels, of countries in a region, we have seen that the same kind of reforms have been implemented during the same time in those countries. This assures me that the factors influencing obesity, and their relationships are correctly mapped in chapter 2.

The regional analysis discussed in chapter 3 is based on current obesity data. The same regions are discussed in chapter 5 for predicted data. The obesity trends predicted by the python-based prediction model shows similar predictions for countries in a region. These results reaffirm the accuracy and relevance of the model itself as it produces trends that correlate to each other, which can be accurately justified with logical and scientific reasoning. The developed python model is strong enough to predict the future of obesity of 190 countries assuming the influencing factors are unchanged. However, depending on the speed and the scale of implementation of solutions recommended in chapter 6, the predicted trends will also change in future.

7.2. Future Work

Despite performing a thorough analysis, this investigation still has room for potential improvements. The prediction model can be further enhanced if the solutions that contribute to the reduction in obesity trends are also incorporated as features in the model. It will be interesting to see the changes in prediction when certain solutions get implemented. The speed and scale of implementing solutions can be factors which can predict the peak and thereafter reduction in obesity levels. In other words, the model could foresee the impact on obesity by hypothetically implementing certain solutions. For example, if Jordan imposes high taxes based on food sugar and fat content, together with initiating a country wide health awareness program, the model can predict when the obesity peak will be reached for Jordan. The model predictions can help in cost analysis on tax benefit and investment required by the government to implement the solutions until the peak is reached and sustain the changes thereafter versus burden on the health system due to increasing obesity levels. This will provide meaningful insight to the value of the provided solutions in chapter 6 and confirm its effectiveness if implemented, making the results of this report more accurate providing a much more focused discussion on the solutions.

Bibliography

Alsukait, R., et. al., 2020. *Evaluating Saudi Arabia's 50% carbonated drink excise tax: Changes in prices and volume sales.* Economics & Human Biology, [e-journal] 38, p.100868. <https://doi.org/10.1016/j.ehb.2020.100868> [Accessed 28 Sep. 2024].

Anza-Ramirez, C., et. al., 2022. *The urban built environment and adult BMI, obesity, and diabetes in Latin American cities.* Nature Communications, [e-journal] 13(1). <https://doi.org/10.1038/s41467-022-35648-w> [Accessed 15 Sep. 2024].

Australian Chronic Disease Prevention Alliance – ACDPA, 2019. *Obesity and chronic disease - Position statement.* [online] Available at: <https://www.acdpa.org.au/obesity-chronic-disease> [Accessed 29 Aug. 2024].

Bauld, L., 2011. *The impact of smokefree legislation in England: evidence review.* [online] Available at: <https://assets.publishing.service.gov.uk/media/5a74cc7240f0b619c865a867/dh_124959.pdf> [Accessed 1 Sep. 2024].

Center for Science in the Public Interest, 2020. *10-Year Anniversary of the Healthy, Hunger-Free Kids Act of 2010 (HHFKA)*. [pdf]. Available at: <https://www.cspinet.org/sites/default/files/media/documents/resource/HHFKA_10Anniv_factSheet_2.pdf> [Accessed 15 Sep. 2024].

Egge, J., H., 2024. *Ny Medisin Mot Fedme Skal Være Mer Effektiv Enn de vi Har I Dag.* <www.nrk.no/trondelag/danske-forskere-utvikler-ny-medisin-mot-fedme-_-mener-den-er-mer-effektiv-enn-wegovy-og-andre-1.16890414> [Accessed 16 Sept. 2024].

Harvey, C., and James N., 2023. *What is BMI? - Definition, Formula & Calculation - Video & Lesson Transcript*. [online] Available at: <https://study.com/academy/lesson/what-is-bmi-definition-formula-calculation.html> [Accessed 8 Aug. 2024].

Diabetes.org., 2024. *Genetics of Diabetes | American Diabetes Association.* [online] Available at: <https://diabetes.org/about-diabetes/genetics-diabetes#:~:text=Type%202%20diabetes%20has%20a,can%20also%20play%20a%20role> [Accessed 5 Aug. 2024].

Doran, L., 2023. *Healthiest Diets in the World (and how they help you live longer).* [online] Available at: <https://www.william-russell.com/blog/worlds-healthiest-diets/> [Accessed 1 Sep. 2024].

Ferrari, G., et. al., 2022. *Perceived urban environment attributes and obesity indices in adults: an 8-Nation study from Latin America.* Scientific Reports, [e-journal] 12(1). <https://doi.org/10.1038/s41598-022-24209-2> [Accessed 15 Sep. 2024].

Foster, N., et al., 2018. *Regulatory measures to fight obesity in Small Island Developing States of the Caribbean and Pacific, 2015 – 2017.* Revista Panamericana de Salud Pública, [e-journal] 42, pp.1–7. <https://doi.org/10.26633/rpsp.2018.191> [Accessed 15 Sep. 2024].

Gabriel, A. S., Ninomiya, K. and Uneyama, H., 2018. *The Role of the Japanese Traditional Diet in Healthy and Sustainable Dietary Patterns around the World.* Nutrients, [e-journal] 10(2), pp.173–173. <https://doi.org/10.3390/nu10020173> [Accessed 11 Aug. 2024].

Goh, P., Gan, L. and Pang, J., 2012. *OBESITY: PREVENTION & MANAGEMENT UNIT NO. 1 OBESITY IN SINGAPORE, PREVENTION AND CONTROL*. [online] Available at: <https://www.cfps.org.sg/publications/the-singapore-family-physician/article/71_pdf> [Accessed 9 Aug. 2024].

Hurt, R.T., et. al., 2011. *Obesity Epidemic.* Journal of Parenteral and Enteral Nutrition, [e-journal] 35(5S). <https://doi.org/10.1177/0148607111415110> [Accessed 4 Aug. 2024].

Institute of Business Forecasting & Planning, 2024. *MAPE (Mean Absolute Percentage Error) | IBF.org.* [online] Available at: <https://ibf.org/knowledge/glossary/mape-mean-absolute-percentage-error-174> [Accessed 13 Sep. 2024].

Knutson, K.L. and Van Cauter, E., 2008. *Associations between Sleep Loss and Increased Risk of Obesity and Diabetes.* Annals of the New York Academy of Sciences, [e-journal] 1129(1), pp.287–304. <https://doi.org/10.1196/annals.1417.033> [Accessed 1 Sep. 2024].

Mayo Clinic, 2021. *Metabolic syndrome - Symptoms and causes.* [online] Available at: <https://www.mayoclinic.org/diseases-conditions/metabolic-syndrome/symptoms-causes/syc-20351916> [Accessed 1 Sep. 2024].

Mayo Clinic, 2024. *Diabetes - Symptoms and causes.* [online] Available at: <https://www.mayoclinic.org/diseases-conditions/diabetes/symptoms-causes/syc-20371444> [Accessed 5 Aug. 2024].

McLaren, M., 2022). *Cutting the Fat: Why is Japan's Obesity Rate So Low?* [online] New Thinking. Available at: <https://www.newthinking.com/health/cutting-fat-why-japans-obesity-rate-low> [Accessed 9 Aug. 2024].

McLennan, A.K. and Ulijaszek, S.J., 2014. *Obesity emergence in the Pacific islands: why understanding colonial history and social change is important.* Public Health Nutrition, [e-journal] 18(8), pp.1499–1505. <https://doi.org/10.1017/s136898001400175x> [Accessed 15 Sep. 2024].

Moyo, J., 2021. *Obesity emerges as health hazard in cash-strapped Zimbabwe*. [online]. Available at: <https://www.aa.com.tr/en/africa/obesity-emerges-as-health-hazard-in-cash-strapped-zimbabwe/2315231> [Accessed 9 Aug. 2024].

National Heart, Lung, and Blood Institute, National Institute of Health, 2022. *Causes and Risk Factors.* [online] Available at: <https://www.nhlbi.nih.gov/health/overweight-and-obesity/causes>[Accessed 5 Aug. 2024].

National Institute of Diabetes and Digestive and Kidney Diseases, 2018. *Insulin resistance & prediabetes.* [online] Available at: <https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/prediabetes-insulin-resistance> [Accessed 1 Sep. 2024].

National Library of Medicine (NLM), U.S. Department of Health and Human Services, 2017. *Type 2 diabetes: MedlinePlus Genetics.* [online] Available at: <https://medlineplus.gov/genetics/condition/type-2-diabetes/#inheritance> [Accessed 1 Sep. 2024].

Norwegian Institute of Public Health, 2024. *FHI.no*. [online] Available at: <https://www.fhi.no/en/he/hin/health-disease/overweight-and-obesity-in-norway---/?term=> [Accessed 9 Aug. 2024].

Okati-Aliabad, H., et. al., 2022. *Prevalence of Obesity and Overweight among Adults in the Middle East Countries from 2000 to 2020: A Systematic Review and Meta-Analysis.* Journal of Obesity, 2022, [e-journal] pp.1–18. <https://doi.org/10.1155/2022/8074837> [Accessed 11 Aug. 2024].

Oslo Kommune, 2023. *The Norwegian Oil Fund - Cultural building blocks.* [online]. Available at: <https://www.oslo.kommune.no/english/welcome-to-oslo/norwegian-society/cultural-building-blocks/the-norwegian-oil-fund/#gref> [Accessed 8 Sep. 2024].

Panicker, V., 2011. *Genetics of Thyroid Function and Disease.* The Clinical Biochemist Reviews, [e-journal] 32(4), p.165. Available through: National Library of Medicine <ncbi.nlm.nih.gov> [Accessed 1 Sep. 2024].

Pico, R. C., 2024. *How leisure and sports are shaping urbanism: the case of Singapore*. [online] Available at: <https://www.tomorrow.city/how-leisure-and-sports-are-shaping-urbanism-the-case-of-singapore/> [Accessed 9 Aug. 2024].

Popkin, B.M. and Reardon, T., 2018. *Obesity and the food system transformation in Latin America.* Obesity Reviews, [e-journal] 19(8), pp.1028–1064. <https://doi.org/10.1111/obr.12694> [Accessed 11 Aug. 2024].

Ramsay Health Care, 2024. *BMI Formula | Body Mass Index | Ramsay Health Care UK.* [online] Available at: <https://www.ramsayhealth.co.uk/treatments/weight-loss-surgery/bmi/bmi-formula> [Accessed 8 Aug. 2024].

Ratey, J., 2008. *SPARK: THE REVOLUTIONARY NEW SCIENCE OF EXERCISE & THE BRAIN.* [online] Available at: <https://v.fastcdn.co/u/3a1b1cdf/28217742-0-Spark-by-John-Ratey-.pdf> [Accessed 6 Aug. 2024].

Ritchie, H. and Roser, M., 2024. *Obesity*. [online] Available at: <https://ourworldindata.org/obesity> [Accessed 6 Aug. 2024].

Ruth J. F. L., and Giles S. H. Y. *"The Genetics of Obesity: From Discovery to Biology."* Nature Reviews Genetics, [e-journal] vol. 23, no. 23, 23 Sept. 2021, pp. 1–14, www.nature.com/articles/s41576-021-00414-z, https://doi.org/10.1038/s41576-021-00414-z. [Accessed 1 Sept. 2024].

Saltiel, A.R. and Olefsky, J.M., 2017. *Inflammatory mechanisms linking obesity and metabolic disease.* Journal of Clinical Investigation, [e-journal] 127(1), pp.1–4. <https://doi.org/10.1172/jci92035> [Accessed 5 Aug. 2024].

Sanyal, D. and Raychaudhuri, M., 2016. *Hypothyroidism and obesity: an Intriguing Link.* Indian Journal of Endocrinology and Metabolism, [e-journal] 20(4), p.554. <https://doi.org/10.4103/2230-8210.183454> [Accessed 15 Sep. 2024].

Senthilingam M., 2015. *How paradise became the fattest place in the world*. [online] Available at: <https://edition.cnn.com/2015/05/01/health/pacific-islands-obesity/index.html> [Accessed 11 Aug. 2024].

Shubrook, N., 2022. *Why is the Japanese diet so healthy?* [online] Available at: <https://www.bbcgoodfood.com/health/healthy-food-guides/why-japanese-diet-so-healthy> [Accessed 1 Sep. 2024].

Sidani, J.E., et. al., 2016. *The Association between Social Media Use and Eating Concerns among US Young Adults.* Journal of the Academy of Nutrition and Dietetics, [e-journal] 116(9), pp.1465–1472. <https://doi.org/10.1016/j.jand.2016.03.021> [Accessed 7 Aug. 2024].

Torres, S.J. and Nowson, C.A., 2007. *Relationship between stress, eating behavior, and obesity.* Nutrition, [e-journal] 23(11-12), pp.887–894. <https://doi.org/10.1016/j.nut.2007.08.008> [Accessed 1 Sep. 2024]

Verhaegen, A.A. and Van Gaal, L.F., 2019. *Drugs That Affect Body Weight, Body Fat Distribution, and Metabolism.* [online] Available at: <https://www.ncbi.nlm.nih.gov/books/NBK537590/> [Accessed 1 Sep. 2024].

Vilmate LLC , 2019. *Advantages of Python over Other Programming Languages.* [online] Available at: <https://vilmate.com/blog/python-vs-other-programming-languages/> [Accessed 11 Sep. 2024].

Waidmann, T., et. al., 2022. *Obesity across America Geographic Variation in Disease Prevalence and Treatment Options*. [online] Available at: <https://www.urban.org/sites/default/files/2022-02/obesity-across-america.pdf> [Accessed 1 Sep. 2024].

Wang, M.L., et. al., 2019. *BMI and related risk factors among U.S. Marshallese with diabetes and their families.* [e-journal] 26(8), pp.1196–1208. <https ://doi.org/10.1080/13557858.2019.1640351>

Watson, B. (2021). *How Genes Influence Your Resting Metabolic Rate (RMR)?* [online] Available at: <https://www.xcode.life/genes-and-sports/how-genes-influence-your-resting-metabolic-rate-rmr-2/> [Accessed 6 Aug. 2024].

WHO, 2022*. Obesity rising in Africa, WHO analysis finds.* [online] Available at: <https://www.afro.who.int/news/obesity-rising-africa-who-analysis-finds> [Accessed 28 Sep. 2024].

World Health Organisation, 2020. Obesity. [online] Available at: <https://www.who.int/health-topics/obesity#tab=tab_1> [Accessed 4 Aug. 2024].

World Health Organisation, 2024. *Obesity and overweight.* [online] Available at: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight#:~:text=Worldwide%20adult%20obesity%20has%20more,16%25%20were%20living%20with%20obesity> [Accessed 4 Aug. 2024].

Yale University, 2019. *Linear Regression.* [online] Available at: <http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm> [Accessed 11 Sep. 2024].