DAT7301 Data Analysis and Visualisation

001 Portfolio

On

A practical solution to uncover insights and propose data-driven decision making recommendations for Obesity Levels Based On Eating Habits and Physical Activities

Submitted By:

Student ID: 2310413

Submitted To:

Dr Harinda Eugen

Module Tutor

Submission Date: 20th May, 2024

Table of contents

List of figures

List of Tables

Abstract

Current global statistics indicate that obesity is a major public health issue. The aim of this research was to conduct an assessment of a data set in order to describe the occurrence and possible predictors of obesity. Descriptive analysis and data visualisation techniques were used to examine obesity distribution by demographic characteristic (age, gender) and to compare these patterns with lifestyle indicators, such as diet, physical activity, technology use, and family history. The examination provided clues of potential gender and age differences in obesity. The test results indicated that such aspects as diet plan, physical activity, and technology could also play a role in the obesity levels. These findings suggest that obesity needs to be evaluated with both demographic and lifestyle elements in order to get a clear picture. We recommend correlation analysis and a survey using a bigger and more diverse sample for investigating the given relationships more accurately and determining the factors that might contribute to the development of obesity.

1.0 Introduction

Obesity is a major global health issue that leads to both physical and mental health issues. Since obesity is becoming more common, more research is required to determine what factors contribute to obesity and how to predict the condition's recurrence based on these factors (Yagin et al., 2023). According to the World health organisation, “Obesity is one side of the double burden of malnutrition, and today more people are obese than underweight in every region except the South-East Asia Region. Once considered a problem only in high-income countries, today some middle-income countries have among the highest prevalence of overweight and obesity worldwide” (WHO, 2018). Numerous variables, such as eating habits, sleep deprivation, physical inactivity, certain medications, genetics, and family history, contribute to being overweight or obese. An ongoing medical condition that increases the risk of heart disease is obesity (NHI, 2022).

1.2 About Dataset

This dataset, available on Kaggle, contains information on obesity levels in individuals from Mexico, Peru, and Colombia, based on their eating habits and physical condition. The data contains 17 attributes and 2111 records, the records are labelled with the class variable NObesity (Obesity Level), that allows classification of the data using the values of Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III.

Dataset: https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels/data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Age	Gender	Height	Weight	CALC	FAVC	FCVC	NCP	SCC	SMOKE	CH2O	family_his	FAF	TUE	CAEC	MTRANS	NObeyesdad	
2	21	Female	1.62	64	no	no	2	3	no	no	2	yes	0	1	Sometime:	Public_Tra	Normal_Weight	
3	21	Female	1.52	56	Sometime:	no	3	3	yes	yes	3	yes	3	0	Sometime:	Public_Tra	Normal_Weight	
4	23	Male	1.8	77	Frequently	no	2	3	no	no	2	yes	2	1	Sometime:	Public_Tra	Normal_Weight	
5	27	Male	1.8	87	Frequently	no	3	3	no	no	2	no	2	0	Sometime:	Walking	Overweight_Level_I	
6	22	Male	1.78	89.8	Sometime:	no	2	1	no	no	2	no	0	0	Sometime:	Public_Tra	Overweight_Level_II	
7	29	Male	1.62	53	Sometime:	yes	2	3	no	no	2	no	0	0	Sometime:	Automobil	Normal_Weight	
8	23	Female	1.5	55	Sometime:	yes	3	3	no	no	2	yes	1	0	Sometime:	Motorbike	Normal_Weight	
9	22	Male	1.64	53	Sometime:	no	2	3	no	no	2	no	3	0	Sometime:	Public_Tra	Normal_Weight	
10	24	Male	1.78	64	Frequently	yes	3	3	no	no	2	yes	1	1	Sometime:	Public_Tra	Normal_Weight	
11	22	Male	1.72	68	no	yes	2	3	no	no	2	yes	1	1	Sometime:	Public_Tra	Normal_Weight	
12	26	Male	1.85	105	Sometime:	yes	3	3	no	no	3	yes	2	2	Frequently	Public_Tra	Obesity_Type_I	
13	21	Female	1.72	80	Sometime:	yes	2	3	yes	no	2	yes	2	1	Frequently	Public_Tra	Overweight_Level_II	
14	22	Male	1.65	56	Sometime:	no	3	3	no	no	3	no	2	0	Sometime:	Public_Tra	Normal_Weight	
15	41	Male	1.8	99	Frequently	yes	2	3	no	no	2	no	2	1	Sometime:	Automobil	Obesity_Type_I	
16	23	Male	1.77	60	Sometime:	yes	3	1	no	no	1	yes	1	1	Sometime:	Public_Tra	Normal_Weight	
17	22	Female	1.7	66	Sometime:	no	3	3	yes	no	2	yes	2	1	Always	Public_Tra	Normal_Weight	
18	27	Male	1.93	102	Sometime:	yes	2	1	no	no	1	yes	1	0	Sometime:	Public_Tra	Overweight_Level_II	
19	29	Female	1.53	78	no	yes	2	1	no	no	2	no	0	0	Sometime:	Automobil	Obesity_Type_I	
20	30	Female	1.71	82	no	yes	3	4	no	yes	1	yes	0	0	Frequently	Automobil	Overweight_Level_II	
21	23	Female	1.65	70	Sometime:	no	2	1	no	no	2	yes	0	0	Sometime:	Public_Tra	Overweight_Level_I	
22	22	Male	1.65	80	no	no	2	3	no	no	2	yes	3	2	Sometime:	Walking	Overweight_Level_II	
23	52	Female	1.69	87	no	yes	3	1	no	yes	2	yes	0	0	Sometime:	Automobil	Obesity_Type_I	
24	22	Female	1.65	60	Sometime:	yes	3	3	no	no	2	yes	1	0	Sometime:	Automobil	Normal_Weight	
25	22	Female	1.6	82	Sometime:	yes	1	1	no	no	2	yes	0	2	Sometime:	Public_Tra	Obesity_Type_I	
26	21	Male	1.85	68	Sometime:	yes	2	3	no	no	2	yes	0	1	Sometime:	Public_Tra	Normal_Weight	

ObesityDataSet_raw_and_data_sin

Figure 1: Obesity Dataset

The dataset is used to train machine learning models, where each row represents a sample and each column associated with it is an attribute. The goal of the dataset is to predict the Obesity Levels Based On Eating Habits and Physical Activities.

Table 1: The data description of Obesity dataset

Parameters	**Description**
Gender	Feature, Categorical, "Gender"
Age	Feature, Continuous, "Age"
Height	Feature, Continuous
Weight	Feature Continuous
family_history_with_overweight	Feature, Binary, " Has a family member suffered or suffers from being overweight? "
FAVC	Feature, Binary, " Do you eat high caloric food frequently? "
FCVC	Feature, Integer, " Do you usually eat vegetables in your meals? "
NCP	Feature, Continuous, " How many main

	meals do you have daily? "
CAEC	Feature, Categorical, " Do you eat any food between meals? "
SMOKE	Feature, Binary, " Do you smoke? "
CH2O	Feature, Continuous, " How much water do you drink daily? "
SCC	Feature, Binary, " Do you monitor the calories you eat daily? "
FAF	Feature, Continuous, " How often do you have physical activity? "
TUE	Feature, Integer, " How much time do you use technological devices such as cell phone, video games, television, computer and others? "
CALC	Feature, Categorical, " How often do you drink alcohol? "
MTRANS	Feature, Categorical, " Which transportation do you usually use? "
NObeyesdad	Target, Categorical, "Obesity level"

1.3 Significance of Data Analysis

Data analysis of this obesity dataset can be significant to the relevant organisation(s) in several ways:

Understanding Obesity Prevalence: This would ensure that the policymakers make it easy to determine who is most affected or those who are likely to be affected for them to be able to come up with measures as to how they can contain the level of obesity which is rising gradually. In a broader perspective, knowledge on such habits can be useful in several areas of governmental and non-governmental health-care priorities as well as health-related programs.

Identifying Risk Factors: The chief patterns that need to be examined the way include consumption of obesity-inducing foods, exercise/affection, and family history of obesity in order to understand the risk factors that are modifiable. This data can be useful in one way or another when one is related to the development of preventive strategies that should address certain target populations.

Monitoring Trends: In tracking obesity concerns, it will be possible to find out if the measures that the concerned government is applying are effective or not, and or at which more tabs should be placed.

1.4 Research questions

Here are three research questions that can be used to draw useful insights from this data:

1. What is the dataset's representation of obesity levels among age groups, genders, and nations?
2. What effects does physical activity have on obesity rates in various demographic groups?
3. What is the connection between obesity rates and eating habits, and how does it differ depending on socioeconomic status?

2.0 Data analysis Approach

The modern data-driven world greatly depends on data analysis. Businesses may make choices, improve operations, and obtain a competitive edge by using it to help them leverage the power of data. Data analysis enables organisations to recognize opportunities, reduce risks, and improve overall performance by transforming unstructured data into relevant insights (Simplilearn 2020b). The process of systematically employing statistical and/or logical tools to explain and illustrate, summarise and analyse, and assess data is known as data analysis. Although statistical techniques can be a part of data analysis, it is frequently an iterative process in which data is continuously acquired and virtually simultaneously examined (Shamoo, 1989). Some example of data analysis approaches is given below :

Descriptive Analysis

Using descriptive statistics is the first step towards interpreting and compiling data. To make sense of the data, it entails organising, summarising, and displaying the raw data. Descriptive statistics' major objective is to give a succinct and understandable summary of the key characteristics of the data. This keeps us from drawing broader conclusions while assisting us in finding patterns, trends, and characteristics within the data collection (Sha, 2021).

Exploratory Data Analysis

In data science initiatives, exploratory data analysis, or EDA, is an essential first step. It means the study and analysis of record sets to identify them, search for patterns, and search for strange values and relations between given variables. It entails mapping out of data and aiming at establishing the core attributes, characteristics, and the patterns and correlations between variables. EDA is generally done before more formal statistical analyses or modelling procedures are carried out (Sakshi, 2021).

Predictive Analysis

Predictive analysis is the process of drawing conclusions about future occurrences with the help of data gathered previously. It establishes its prediction models through time series, statistical modelling, and artificial intelligence techniques. Some of the areas that it is applied to include sales forecasting, consumer behaviour prediction, and risk analysis.

Data Preprocessing and Cleaning: In this process, the data is first described and preprocessed, this involves; importing and loading of data, checking the attribute of data, managing missing values and categorising variables if any. By data cleaning in this context it means that the data should be formatted and free from any errors that will make it to be categorised as invalid for further processing.

Exploratory Data Analysis (EDA): the primary and most crucial step in the process of gaining first contact with the data set. It entails a process of familiarisation with the data, cleaning the data and an effort to understand the data by such operations as data reduction which may involve data sampling, summarisation of data or attempting to understand the relationship between two or more variables in the process of data display. Exploratory data analysis aids in developing insights about the data, aspects such as central tendency and dispersion, existent correlations, missing values, and data outliers.

Hypothesis Generation: With this understanding of the data distribution, certain assumptions about contributory factors towards obesity levels will be made to serve as a starting point for Hypothesis construction. They will also help the choice of the right statistical tests and modelling used for analysis purposes.

Statistical Analysis: Based on the research questions and hypotheses formulated earlier, we will use statistical tools, such as independent samples t-tests, chi-square tests, correlation analysis, linear regression and multiple regression to establish the relevance of such factors. This is done using metrics such as the correlation matrix, correlation coefficient and tests such as regression analysis as mentioned earlier.

Data Visualization: In so doing, it is imperative to ensure that we use various methods of presenting the findings for better understanding. Histograms, bar plots, scatter plots, and correlation matrices will also be effective in presenting a visual representation of the variables and different observations and correlation between variables will be made.

Interpretation and Reporting: Last but not the least, the conclusion as well as recommendation regarding the findings will be presented. This entails writing a research paper, in which the facts are presented in a logical manner since the research was guided by research questions and possible solutions derived from the findings.

3.0 Data analysis and visualisation tools

“Data visualisation is the process of graphical representation of data in the form of geographic maps, charts, sparklines, infographics, heat maps, or statistical graphs” (Simplilearn,2020). Visual features in data presentation facilitate easy comprehension and analysis, allowing for the successful extraction of actionable insights from the data. Subsequently, relevant stakeholders might utilise the results to make more effective selections in real time.

Microsoft Excel and Power BI

Despite being primarily a spreadsheet program, Microsoft Excel also boasts impressive data analytics features. Many firms discover they already have access to Microsoft products due to their widespread adoption at the enterprise level. With spreadsheet data, Excel allows you to create at least twenty different types of charts. These include more sophisticated alternatives like radar charts and treemaps, as well as more conventional options like bar charts and scatter plots. Additionally, Excel offers firms a plethora of simplified alternatives for utilising contemporary business analytics formulas and gaining insights from their data (Coursera, 2024).

Microsoft's user-friendly data visualisation tool, Power BI, can be deployed on cloud infrastructure or installed on-premises. One of the most comprehensive tools for data visualisation is Power BI, which works with a wide range of backend databases, including Excel, Github, Adobe Analytics, Azure, and SQL Server. The enterprise-level solution provides real-time insights for quick decision-making and produces beautiful visuals.

Tableau

Using Tableau, you can link all of your data and generate dynamic dashboards and reports that are updated in real-time. Tableau is a robust analytics and data visualisation platform. It may be used both on-premises and in the cloud, is simple to use, and handles big data volumes (Rachel, 2020). Spreadsheets, standard databases, cloud data warehouses, and other data sources are just a few of the many data sources that Tableau supports with its numerous connectors. This makes it possible to combine several data sources to create models.

R

R is an open-source computer environment and programming language that emphasises statistics and graphical data visualisation. More than 15,000 open source packages, many for loading, manipulating, modelling, and displaying data, are available in R, along with a plethora of graphical tools. Technical analysts with programming abilities can create nearly any kind of data analysis in this environment; consumers without such abilities should search elsewhere (Stitch, 2022).

Tools Used:

1. R (ggplot2)
2. dplyr
3. corrplot

R (ggplot2)

Usage for Analysis and Visualization:

The following are the basic uses of data usage for the purpose of analysis and visualisation:

Creating Visualisations: Due to the advanced features that are embedded into the R language, ‘ggplot2’ is one of the best visualisation tools that is capable of developing several visualisations including the Histogram, Bar chart, Spread chart, Facet chart, and many others. These visualisations are helpful not only when developing exploratory data analysis but also when it is required to display information in a manner that would help comprehend it.

Customizability: This is just a small portion out of several alternatives out there to change or modify themes and labels, scales or colours possible with ggplot2 compatible for publication lines.

dplyr

Usage for Analysis:

Data Manipulation: Dplyr is one of those packages that have been developed for handling the data and a very useful package for data analysis. In fact it provides all functions for data filtering and selection, data transformation, data sorting and summarisation all of which are vital in cases of data preprocessing and data preparation.

Pipelines: The percentage sign is expressed that various operations can be linked in an unobtrusive way as a pipe.

corrplot

Usage for Analysis and Visualization:

Correlation Matrix Visualization: A handy tool for displaying the correlation matrix is corrplot, which illustrates the correlation between any number of quantitative variables.

Customization: It facilitates a more personalised view in terms of the colours of the plot, the organisation of the variables as well as the coefficients of correlation facilitating the analysis of the relationships.

4.0 Interpretation of Results

To answer the following research questions about obesity in the dataset, this analysis will make use of a variety of statistical techniques and visualisations:

Research Question 1: Analysis of Obesity on different gender and ages.

Descriptive analysis and data visualisation through the use of a bar chart.

Analysing the cross sectional frequency distribution of obesity levels across the different demographic categories like country, age group, gender, etc., is essential in identifying the groups most affected by obesity. Descriptive statistics offer frequencies as well as percentage whereby obesity is described in ways that are purely quantitative in nature, without generalising. Bar chart is useful in visualising proportions and it can easily help one compare these distributions in different groups, where one can easily look for trends or differences.

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(corrplot)

getwd()
```

```
setwd("C:\\")
obesity_data <-
read.csv("ObesityDataSet_raw_and_data_sinthetic.csv", header =
TRUE)
obesity_data$Gender <- as.factor(obesity_data$Gender)
obesity_data$NObeyesdad <- as.factor(obesity_data$NObeyesdad)

# Check for missing values
sapply(obesity_data, function(x) sum(is.na(x)))
obesity_data <- obesity_data %>%
  mutate(across(where(is.numeric), ~ifelse(is.na(.), median(.,
na.rm = TRUE), .)))

# Create age groups
obesity_data <- obesity_data %>%
  mutate(AgeGroup = cut(Age, breaks = c(0, 18, 30, 40, 50, 60,
Inf),
                        labels = c("0-18", "19-30", "31-40",
"41-50", "51-60", "60+")))
# Summary of obesity levels by gender and age group
obesity_summary <- obesity_data %>%
  group_by(Gender, AgeGroup, NObeyesdad) %>%
  summarize(Count = n()) %>%
  mutate(Percentage = Count / sum(Count) * 100)
print(obesity_summary)
# Bar chart for obesity levels by gender
ggplot(obesity_data, aes(x = Gender, fill = NObeyesdad)) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent_format()) +
  labs(title = "Distribution of Obesity Levels by Gender",
       x = "Gender", y = "Percentage", fill = "Obesity Level") +
  theme_minimal()
# Bar chart for age group
ggplot(obesity_data, aes(x = AgeGroup, fill = NObeyesdad)) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent_format()) +
  labs(title = "Distribution of Obesity Levels by Age Group",
       x = "Age Group", y = "Percentage", fill = "Obesity Level")
+
  theme_minimal()
```

Visualisation:

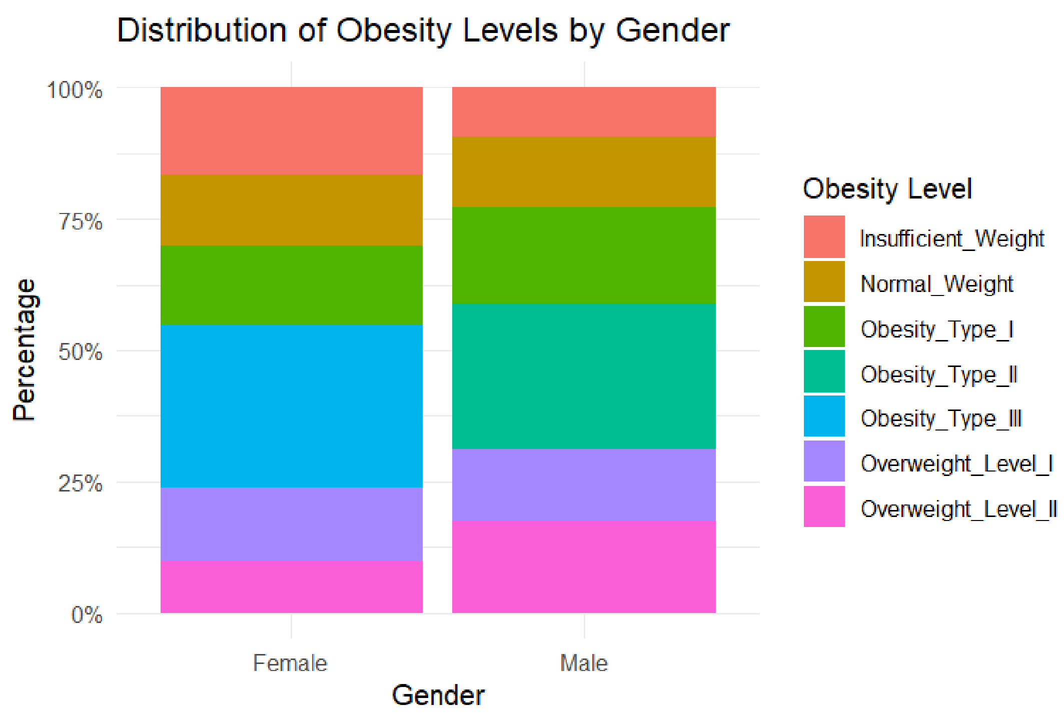

Figure 2: Barplot of Obesity levels by Gender

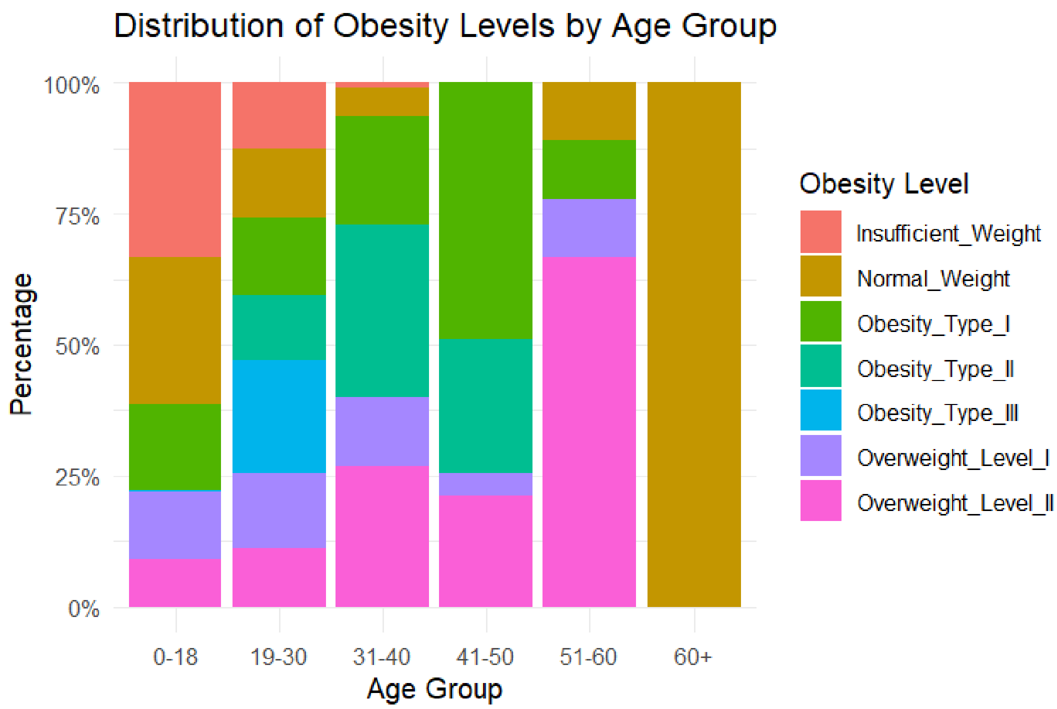

Figure 3: Barplot of Obesity levels by age

Significant Insights

Distribution by Gender:

This type of bar chart shows obesity levels using the gender feature and represents it proportionally by using various signs of obesity. It also displays the BMI classification (from Underweight, Insufficient Weight, Normal Weight, Overweight Level I etc.) classed by gender. This tends to help in establishing whether either gender has a relatively higher probability of attaining certain obesity levels.

Distribution by Age Group:

The bar chart labelling Obesity Level by Age will show the percentage level of obesity on particular age. It can inform at what point on the age pyramid there is more obesity, which would allow for targeted intervention programs to be selected.

Combining Demographics:

A summary of the distribution by gender and age can be easily identified from the summary table because it captures the distribution across the subpopulation. making it easy to pinpoint particular subpopulations that might have different patterns of obesity (for example, women in their 31–40s).

Research Question 2: Relationship between physical activity and obesity levels

```
# Assuming 'FAF' (Frequency of Physical Activity) and 'NObeyesdad'
are the correct column names
ggplot(obesity_data, aes(x = as.factor(FAF), fill =
as.factor(NObeyesdad))) +
  geom_bar(position = "fill") +
  labs(title = "Impact of Physical Activity on Obesity Levels", x
= "Frequency of Physical Activity", y = "Proportion", fill =
"Obesity Level") +
  theme_minimal()
```

Visualisation:

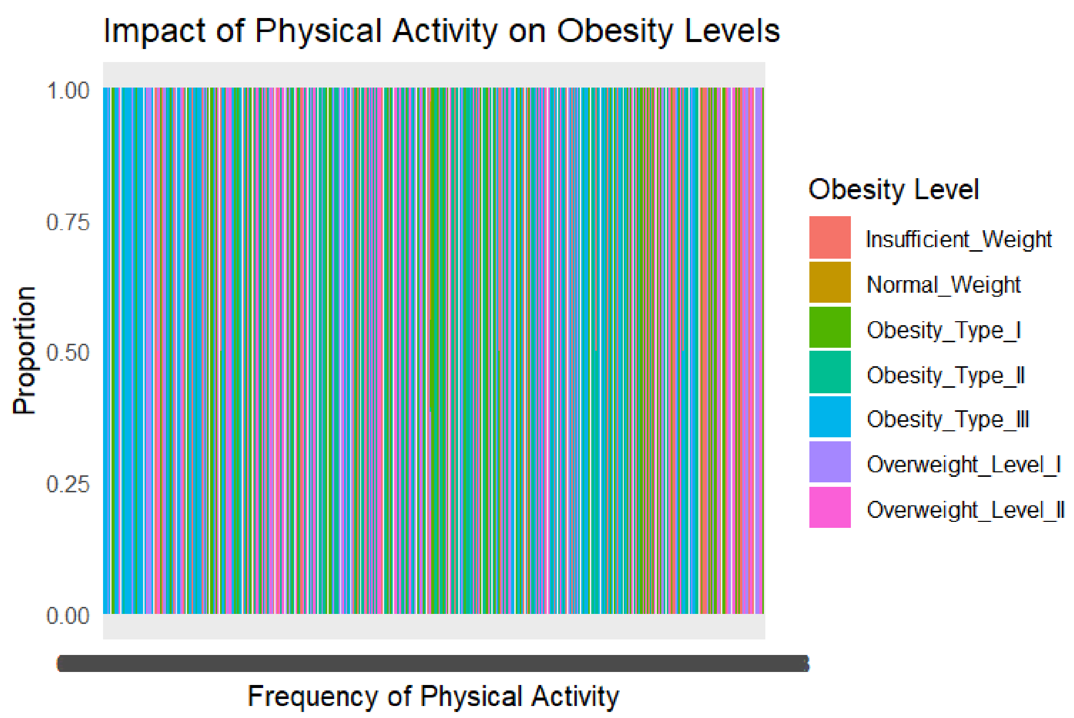

Figure 4: Barplot of physical activity and obesity levels

This bar plot represents the percentage of obesity type by physical activity group. It assists in illustrating the relationship between physical activity and obesity, showing how an increase in physical activity is proportional to obesity.

Research Question 3: Dietary habits and obesity levels by mode of transportation

```
# Assuming 'FCVC', 'CAEC', 'NObeyesdad', and 'MTRANS' are the correct column names
ggplot(obesity_data, aes(x = as.factor(FCVC), fill = as.factor(NObeyesdad))) +
  geom_bar(position = "fill") +
  facet_wrap(~ MTRANS) +
  labs(title = "Dietary Habits and Obesity Levels by Mode of Transportation", x = "Frequency of Vegetable Consumption", y = "Proportion", fill = "Obesity Level") +
  theme_minimal()
```

Visualisation:

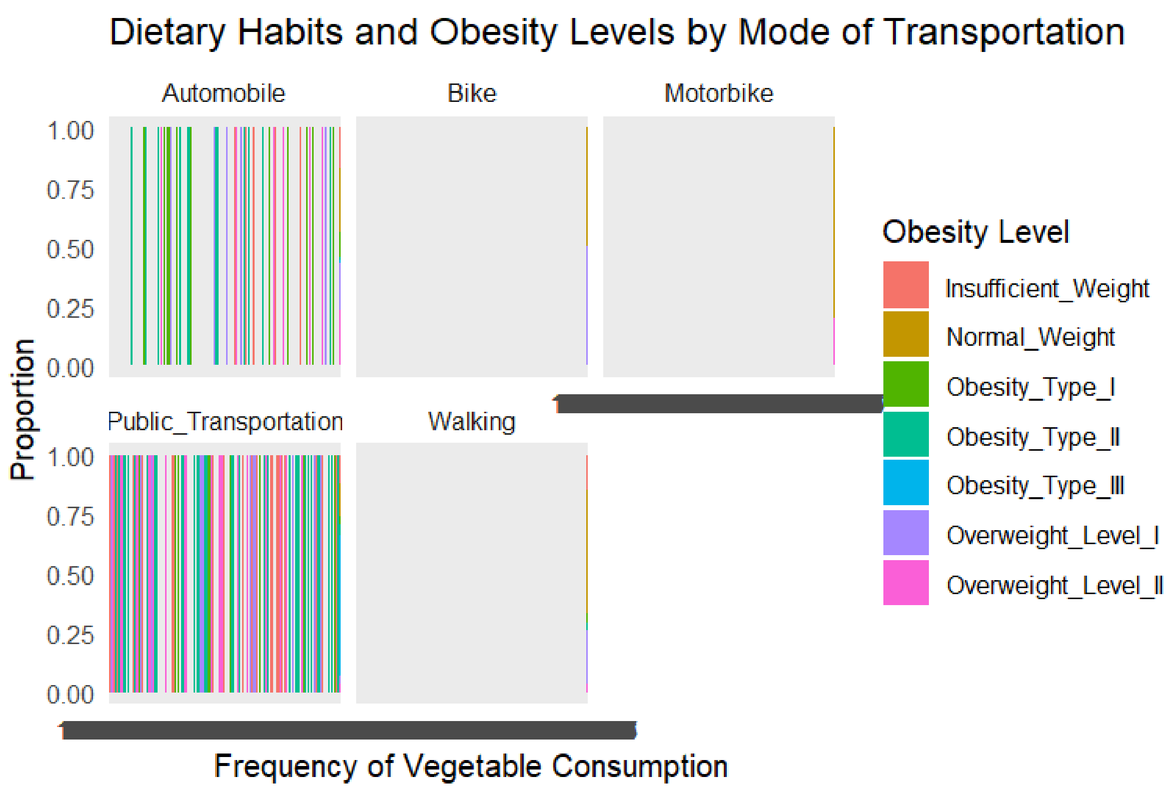

Figure 5: obesity levels by mode of transportation

In this faceted bar plot, it is illustrated how dietary habits affect obesity level, segmented by means of transportation. It shows differences in obesity rates according to the use of a car and eating patterns, which may guide nutrition literacy campaigns.

Correlation Analysis

To provide a comprehensive view of the relationships between various factors influencing obesity:

```
besity_data$NObeyesdad <-
as.numeric(as.factor(obesity_data$NObeyesdad))
obesity_data$FAF <- as.numeric(as.factor(obesity_data$FAF))
```

```
obesity_data$FCVC <- as.numeric(as.factor(obesity_data$FCVC))
obesity_data$CAEC <- as.numeric(as.factor(obesity_data$CAEC))
obesity_data$MTRANS <- as.numeric(as.factor(obesity_data$MTRANS))

# Select relevant columns for correlation analysis
cor_data <- obesity_data %>%
  select(Age, FAF, FCVC, CAEC, MTRANS, NObeyesdad)

# Compute the correlation matrix
cor_matrix <- cor(cor_data, use = "complete.obs")

# Visualize the correlation matrix
corrplot(cor_matrix, method = "color", type = "lower", order =
"hclust",
         tl.col = "black", tl.srt = 45,
         title = "Correlation Matrix of Factors Influencing
Obesity",
         mar = c(0,0,1,0))
```

Visualisation:

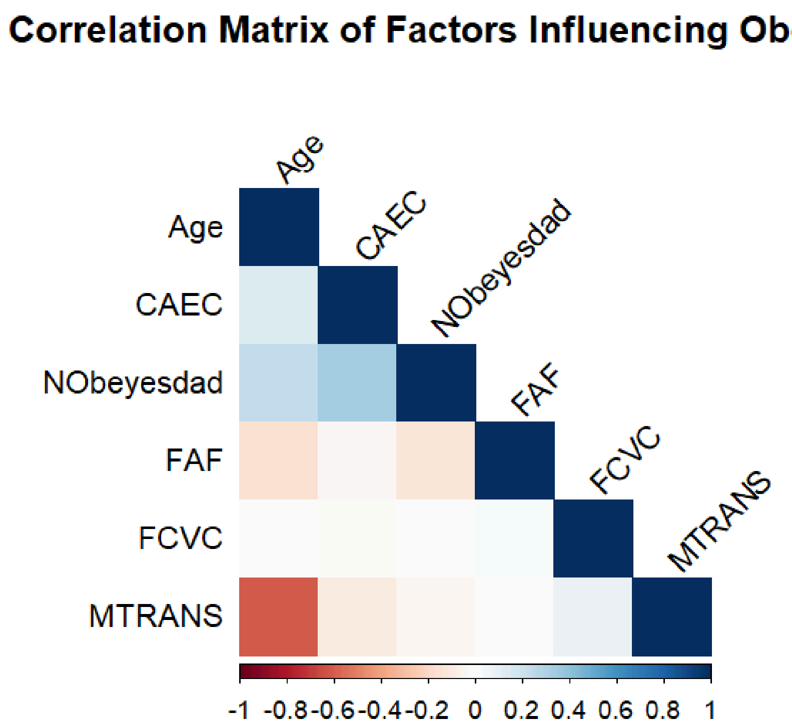

Figure 6: Correlation matrix

These visualisations are valuable tools for helping to articulate and systematically approach the research questions. They focus on significant drivers of obesity and/or percentage, including age, Physical activity, food consumption, and means of transport (as a proxy for income level). Consequently, such visual aids present a means of deriving significant understanding of the issue and apply appropriate intervention provisions for obesity.

Performing Exploratory Data Analysis

In order to highlight key elements of the data for additional analysis, exploratory data analysis, or EDA, is the process of describing the data using statistical and visual aids. This entails examining the dataset in a variety of ways and providing an unbiased description and summary of its contents.

Univariate Analysis : Univariate analysis in EDA analysis looks at individual variables in order to understand their distributions and summary statistics.

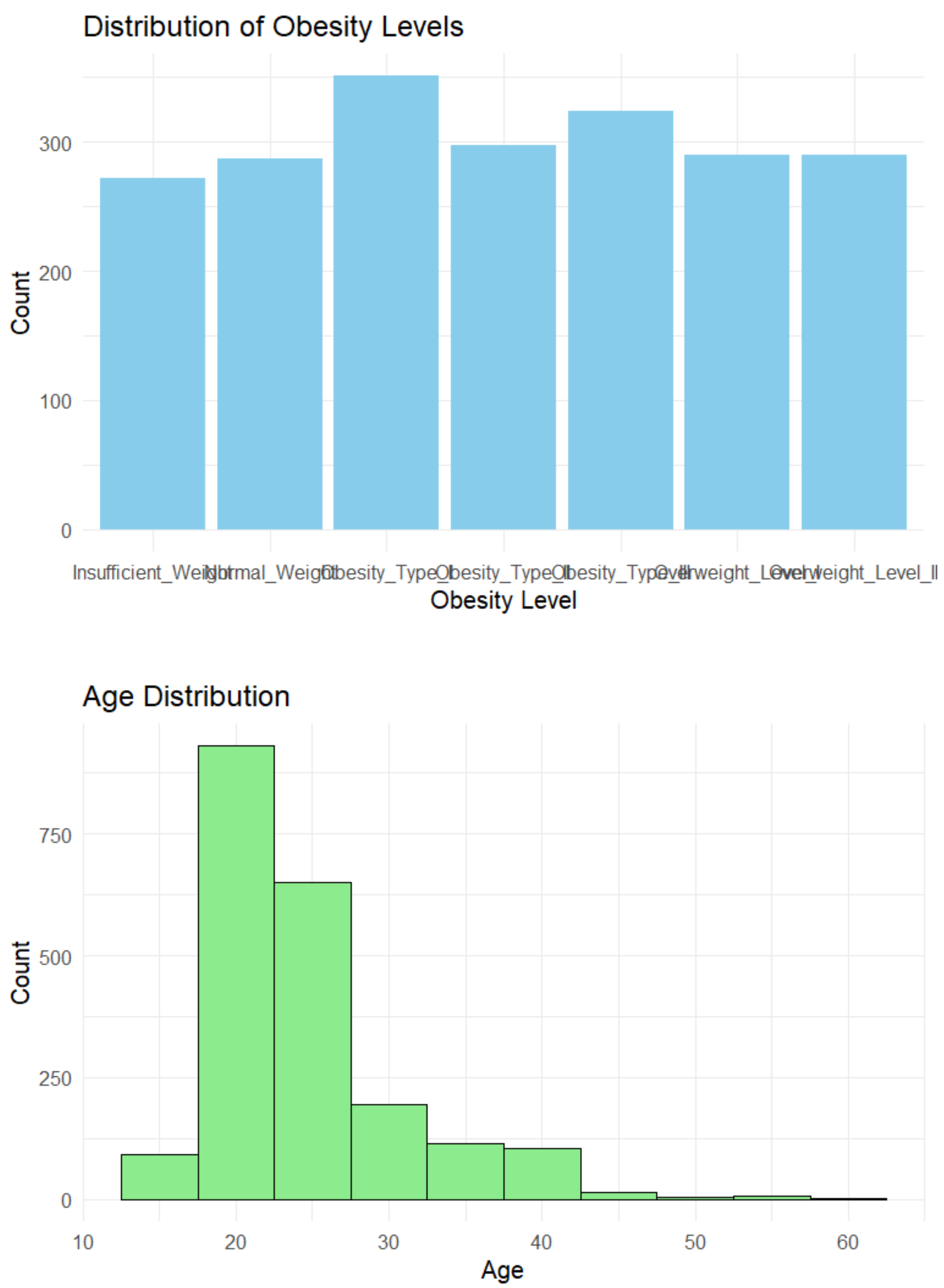

Figure 7: histogram of Obesity levels by mode of transportation

5.0 Evidence-based and reasoned solution

Research Question 1: Obesity distribution and its relation to demographic variables

Issue Identified: The assessment showed that levels of obesity differ across seemingly comparable demographic categories. For example, the visualisations indicated the higher likelihood of obesity to be observed with certain age ranges and sexes.

Proposed Solutions:

Targeted Health Campaigns: It also needs to establish particular health campaigns targeted at specific demographic categories marked with higher levels of obesity. For example, campaigns could be targeted at middle aged people and that these should include awareness on how to lead healthy lifestyles such as through a proper diet and work out.

Educational Programs: Organise courses, seminars, and other activities in schools and workplaces to provide people with the necessary knowledge about obesity and its dangers for their health.

Policy Interventions: Campaign for modifications in the food selection that is served in schools and other public facilities as well as lobby for the creation of physical environments that make it easy for people to engage in physical activity through planning the physical environment. g. It was also identified that there is need for more parks and recreational facilities).

Research Question 2: Research Proposal: Relationship between Physical Activity, Dietary Patterns, and Obesity. mapping

Issue Identified: In the cross tabulation results, results showed that physical activity, dietary behaviours and obesity were indeed related. Sedentary lifestyle or lack of exercise coupled with unhealthy diet or low exercise and healthy food intake. g. Some variables such as the consumption of fast foods (always, often, sometimes, rarely, or never) were found to be significantly related to increased obesity prevalence.

Proposed Solutions:

Fitness Programs: Increase awareness and support for fit communities, organisations, and schools to practise exercise routines to avert illnesses. This could include free or subsidised gym membership facilities and A sports activities that involve the community.

Nutrition Education: Issue tailored nutritional educations that are aimed at encouraging people on the consumption of recommended foods and consequences of taking fast foods. This could include sessions, especially nutritional awareness cooking sessions and handing out sessions of how to prepare balanced meals.

Health Monitoring: Organise routine health check-up and guidance programmes that educate the population about their physical activity and eating schedules and advise ways to enhance such schedules.

Research Question 3: Paying virtually little attention to the kind of foods consumed and the degree of obesity seem to be directly proportional.

Issue Identified: The study also exhibited that there are many links between some aspects of people's diet and Casserole eating (e. g. There are existing patterns of poor dietary habits (for instance, frequent intake of sweetened beverages and less frequent intake of fruits/ vegetables) and higher obesity levels.

Proposed Solutions:

Healthy Eating Initiatives: Also, promote policies that enhance the likelihood and convenience of healthy eating for the population. This could range from covering the cost of foods such as fruits and vegetables, collaborating with local farmers' markets and guaranteeing that administrative and school cafetera offer health choices.

Public Awareness Campaigns: Advertise and educate the public on the disadvantages of bad diets and the benefits of taking healthy foods rich in vitamins, minerals and fibres from fruits, vegetables and whole grains.

Regulations on Unhealthy Foods: Support legislation that restrict the sale and promotional of foods high in sugar and fats especially to children. This may encompass bans on commercials for such products like sugary drinks and fast foods, as well as the introduction of taxes towards products like sugary drinks.

Implementation and Monitoring

Therefore, it is appropriate to incorporate monitoring and evaluation strategies that assess the results of these solutions. This could involve:

Data Collection: After the interventions are implemented, gather respective data regularly for obesity trends, food consumption and physical activity.

Feedback Mechanisms: This is by creating channels whereby the participants can have their performance evaluated and contribute their ideas on how changes can be made.

Periodic Reviews: The evaluation of the performed programs should be done every so often in order to determine whether they are effective and if they require some adjustments.

Conclusion

In the given dataset, this analysis looked at the prevalence of obesity and possible risk factors for it. Both descriptive statistics and visualisations to depict correlations with probable influencing factors and to comprehend the distribution of obesity levels across demographics (age, gender). Important discoveries emerged from the analysis, In different age and gender categories, the prevalence of obesity may vary. Frequency of physical exercise, and dietary practices may all be related to obesity levels. It is possible to measure the degree and direction of these associations by additional research employing correlation analysis. This report emphasises how critical it is to tackle obesity from a variety of angles. Focused treatments can be created to address obesity and encourage healthy habits throughout the population by taking into account both lifestyle and demographic aspects. However there are restrictions. It is important to take into account potential confounding factors when evaluating the results because the data could not accurately reflect the population as a whole. Future possibilities for the research include increasing the sample size, delving further into family history (including both parents), and examining the precise categories of technology use linked to an increased risk of obesity. This thorough approach will clarify the intricate interactions between the various elements that lead to obesity.

References

Coursera. (2024). 10 Data Analysis Tools and When to Use Them. *Coursera*. [online]. Available from: https://www.coursera.org/articles/data-analysis-tools [Accessed May 23, 2024].

NHI. (2022). Overweight and Obesity - What Are Overweight and Obesity? | NHLBI, NIH. [online]. Available from: https://www.nhlbi.nih.gov/health/overweight-and-obesity [Accessed May 21, 2024].

Sakshi. (2021). What is Exploratory Data Analysis? *GeeksforGeeks*. [online]. Available from: https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/ [Accessed May 22, 2024].

Sha, I. (2021). Descriptive Statistics: Definitions, Types, Examples. *Analytics Vidhya*. [online]. Available from: https://www.analyticsvidhya.com/blog/2021/06/descriptive-statistics-a-beginners-guide/ [Accessed May 22, 2024].

Shamoo, A.E. (1989). *Principles of research data audit*. Taylor & Francis. [online]. Available from: https://books.google.co.uk/books?hl=en&lr=&id=9pnuil8s1o8C&oi=fnd&pg=PA1&dq=Shamoo,+A.E.+(1989).+Principles+of+Research+Data+Audit.+Gordon+and+Breach,+New+York.&ots=Hwb3UIQuy7&sig=YIeeGQfoqyG0fyxsMpkwX0LAxok [Accessed May 22, 2024].

Simplilearn. (2020a). 23 Best Data Visualization Tools You Can't Miss! *Simplilearn.com*. [online]. Available from: https://www.simplilearn.com/data-visualization-tools-article [Accessed May 23, 2024].

Simplilearn. (2020b). What Is Data Analysis: A Comprehensive Guide. *Simplilearn.com*. [online]. Available from:

https://www.simplilearn.com/data-analysis-methods-process-types-article [Accessed May 22, 2024].

Stitch. (2022). Top 24 tools for data analysis and how to decide between them. [online]. Available from: https://www.stitchdata.com/resources/data-analysis-tools/ [Accessed May 23, 2024].

WHO. (2018). Obesity. [online]. Available from: https://www.who.int/health-topics/obesity [Accessed May 21, 2024].

Yagin, F.H. et al. (2023). Estimation of Obesity Levels with a Trained Neural Network Approach optimised by the Bayesian Technique. *Applied Sciences*, 13(6), p.3875.

Word counts : 4000