**The estimation of obesity levels based on eating habits and physical condition**

A picture containing text

Description automatically generated

|  |
| --- |
| By |
| Mram Al Shehri |
| Nouf Al Ruhaili |
| Noura Al Harthi |
| Nourah Alsaadan |

# Abstract

Obesity has become a global epidemic that has doubled since 1980, with serious consequences for health in children, teenagers and adults. Obesity is a problem has been growing steadily and that is why every day appear new studies involving children obesity, especially those looking for influence factors and how to predict emergence of the condition under these factors [1].

This project presents data for the estimation of obesity levels in individuals from the countries of Mexico, Peru, and Colombia, based on their eating habits and physical condition.

# Table of Contents

[Abstract 2](#_Toc73837913)

[Table of Contents 3](#_Toc73837914)

[List of Figures 4](#_Toc73837915)

[List of Tables 4](#_Toc73837916)

[Chapter 1: Dataset 5](#_Toc73837917)

[Introduction 6](#_Toc73837918)

[1.1Dataset overview 6](#_Toc73837919)

[1.2Survey 7](#_Toc73837920)

[Chapter 2 | Plots and Charts 9](#_Toc73837921)

[Introduction 10](#_Toc73837922)

[2.1 First Plot 10](#_Toc73837923)

[2.2 Second Plot 11](#_Toc73837924)

[2.3 Third Plot 12](#_Toc73837925)

[2.4 Fourth Plot 13](#_Toc73837926)

[2.5 Fifth Plot 14](#_Toc73837927)

[2.6 Sixth Plot 15](#_Toc73837928)

[References 16](#_Toc73837929)

# List of Figures

[Figure 1 7](#_Toc73837650)

[Figure 2 10](#_Toc73837651)

[Figure 3 11](#_Toc73837652)

[Figure 4 12](#_Toc73837653)

[Figure 5 13](#_Toc73837654)

[Figure 6 14](#_Toc73837655)

# List of Tables

[Table 1 7](#_Toc73837695)

# Chapter 1 | Dataset

## **Introduction**

This chapter investigate the dataset and explain the attributes, it also explains the survey questions and answers.

* 1. **Dataset overview**

This Project contains data for the estimation of obesity levels in people from the countries of Mexico, Peru and Colombia, with ages between 14 and 61 and diverse eating habits and physical condition as mentioned by [2], data was collected using a web platform with a survey (table 1), then the information was processed obtaining 17 attributes and 2111 records. With the attributes:

* 1. Frequent consumption of high caloric food (FAVC)
  2. Frequency of consumption of vegetables (FCVC)
  3. Number of main meals (NCP)
  4. Consumption of food between meals (CAEC)
  5. Consumption of water daily (CH20)
  6. Consumption of alcohol (CALC)
  7. Calories consumption monitoring (SCC)
  8. Physical activity frequency (FAF)
  9. Time using technology devices (TUE)
  10. Transportation used (MTRANS)
  11. Smoke
  12. Gender and Age
  13. Height and Weight
  14. NObesity was created by the following equation:

Text

Description automatically generated with medium confidence

Figure 1

With the values:

* + 1. Insufficient Weight less than 18.5
    2. Normal Weight 18.5 to 24.9
    3. Overweight Level I 25 to 27.5
    4. Overweight Level II 27.6 to 29.9
    5. Obesity Type I 30 to 34.9
    6. Obesity Type II 35 to 39.9
    7. Obesity Type III Higher than 40
  1. **Survey**

The table below explains the questions in the survey with their possible Answers

Table 1

| **Questions** | **Possible Answers** |
| --- | --- |
| What is your gender? | * Female * Male |
| What is your age? | * Numeric value |
| what is your height? | * Numeric value in meters |
| what is your weight? | * Numeric value in kilograms |
| Has a family member suffered or suffers from overweight? | * Yes * No |
| Do you eat high caloric food frequently? | * Yes * No |
| Do you usually eat vegetables in your meals? | * Never * Sometimes * Always |
| How many main meals do you have daily? | * Between 1 y 2 * Three * More than three |
| Do you eat any food between meals? | * No * Sometimes * Frequently * Always |
| Do you smoke? | * Yes * No |
| How much water do you drink daily? | * Less than a liter * Between 1 and 2 L * More than 2 L |
| Do you monitor the calories you eat daily? | * Yes * No |
| How often do you have physical activity? | * I do not have * 1 or 2 days * 2 or 4 days * 4 or 5 days |
| How much time do you use technological devices such as cell phone, videogames, television, computer and others? | * 0–2 hours * 3–5 hours * More than 5 hours |
| How often do you drink alcohol? | * I do not drink * Sometimes * Frequently * Always |
| Which transportation do you usually use? | * Automobile * Motorbike * Bike * Public Transportation * Walking |

Chapter 2 | Plots and Charts

## **Introduction**

This chapter contains plots and charts presents the data in chapter 1, these plots are done using python with the pandas, Seaborn, and Matplotlib.

## **2.1 First Plot**

Chart

Description automatically generated

Figure 2

This is a mixed of violin and swarm plots that compares the Obesity level and age based on gender, as it shows most of the people who has participated in the survey in age range between 17 and early 40’s.

In the obesity type 3 most of them are women with age range between 18 and younger than 30, and people with obesity type 2 most of them are men with age range between 20 and early 40's,

People with age +30 and overweighed most of them are men, no men with age +30 are Insufficient weight (underweight)

## 

## **2.2 Second Plot**

Chart, bar chart

Description automatically generated

Figure 3

This is a histogram plot that compares the Obesity level with the transportation used, as it shows people with overweight levels do not walk, bike, or use motorbike as a transportation method, maybe because these transportation methods require a high fitness level.

We cannot consider the public transportation or the Automobile as an effective measure for the obesity level because they don’t show the activity levels.

## **2.3 Third Plot**

Chart

Description automatically generated

Figure 4

This is a violin plot that compares the age, gender and the family history with overweight, as it shows for the ‘Yes’ case, the highest point for females is in the early 20’s to 30, and then it spread a little under 40’s, for males the highest point is in the early 20’s as well and then it spread in the early 30’s

For the ‘No’ case, they are less than the ‘Yes case’ and for both genders they start to spread in around 20’s, and the females are spreading more than males.

## **2.4 Fourth Plot**

Chart, bar chart

Description automatically generated

Figure 5

This is a histogram plot that compares the obesity level with the family history with overweight, as it shows that people with a family history with overweight’s have a higher level of overweight’s, on the contrary, people with no family history with overweight’s have lower levels.

## 

## **2.5 Fifth Plot**

Chart, bar chart

Description automatically generated

Figure 6

This figure shows bar plots that compares the good habits with the obesity level, and as it shows:

The first one shows the consumption of water daily, and people with obesity type 3 are drinking more than the others, which is obvious that people with higher weights needs more water than the others.

The second one shows the physical activity frequency, and people with normal wight and Insufficient weight do more exercises than the others, and the obesity type 3 people does less than the other, which means the higher the weight goes the less activity level they got.

The third one shows the frequency of consumption of vegetables, and people with obesity type 3 eats more than the rest.

The last one shows the number of main meals, and people with obesity type 3 eats more than the rest.

In conclusion, good or healthy habits cannot be considered as an effective measure of the obesity level, because people with higher obesity levels must eat and drink more than people with lower levels, except for the physical activity frequency.

## **2.6 Sixth Plot**

Chart, box and whisker chart

Description automatically generated

Figure 7

This figure shows box plots that compares the bad habits with the weight, and ages in grouped by youth, adult, and seniors, and as it shows:

The first one shows the smoke, and smokers youth have lower weights than the nonsmokers, but for the adult smokers the have higher weights than nonsmokers, study’s shows smokers have lower wights levels than nonsmokers.

The second one shows the consumption of food between meals, and seniors tend to not eat between meals, for the adults and youth who tend to eat between meals have lower weight range.

The third one shows the consumption of alcohol, people who don’t drink have lower wight range than the others.

The last one shows the frequent consumption of high caloric food, and as its obvious people who eat high calories have higher weight ranges for all the age groups.

In conclusion, most of the bad habits affect the weight level and can be consider as an effective measure for the obesity levels.

# References

**[1] Obesity Level Estimation Software based on Decision Trees | Article *in* Journal of Computer Science · January 2019 https://www.sciencedirect.com/science/article/pii/S2352340919306985**

**[2] M.V. Olmedo, La obesidad: un problema de salud pública. Revista de divulgació científica y tecnológica de la Universidad Veracruzana (2011) Recuperado de:** [**https://www.uv.mx/cienciahombre/revistae/vol24num3/articulos/obesidad/**](https://www.uv.mx/cienciahombre/revistae/vol24num3/articulos/obesidad/)