



HEALTH TECH ANALYTICS CENTER

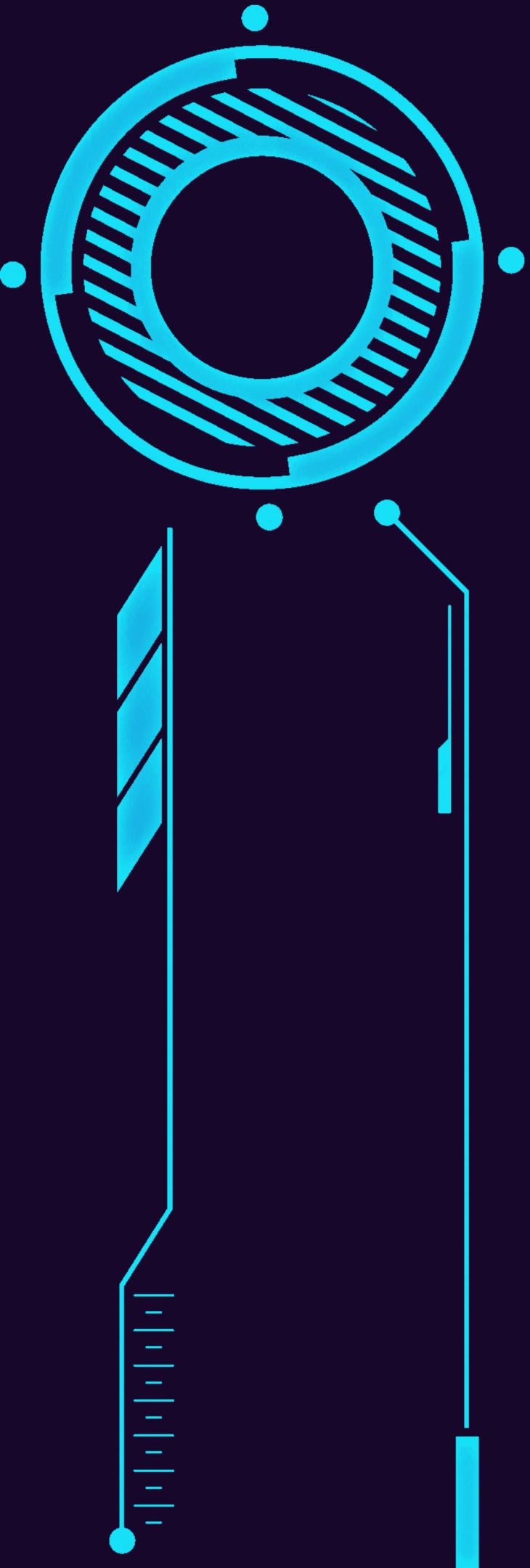
# OBEISITY LEVEL

BASED ON DATA  
ANALYSIS OF PHYSICAL  
ACTIVITY AND  
NUTRITIONAL HABITS

[www.healthtechanalyticscenter.com](http://www.healthtechanalyticscenter.com)

TEAM: ANA MELISSA AND ETTORE CAPONE

FEBRUARY, 2024





# About us

We provide support with advanced analytics, we empower healthcare providers to optimize patient outcomes and improve overall well-being in the fight against obesity.



# Business Problem

Our health center faces challenges in achieving significant weight loss despite tailored nutrition and activity plans. The medical team seeks insights from data analysis to enhance outcomes.





# Hypotheses

## Hypothesis 1



## Hypothesis 2



## Hypothesis 3





# Exploratory Data Analysis (EDA)

01

Understanding data

02

Data cleaning

03

Multivariate analysis

04

Feature engineering

05

Visualization

06

Iterating and refine



# Understanding data



This study surveyed 498 participants aged 14 to 61 via a web-based survey to assess eating habits and physical activity, yielding insights into obesity levels. With 2087 observations and 17 variables, it sheds light on dietary patterns and activity.

77% of the data was generated synthetically using the Weka tool and the SMOTE filter, 23% of the data was collected directly from users through a web platform.



# The data



8 NUMERICAL  
VARIABLES



9 CATEGORICAL  
VARIABLES





# Data cleaning

## VARIABLES UNDERSTANDING

Understanding the variable's meaning and their correlations.

## FUNCTIONS DEFINITION

Defining a function in Python, using libraries such as pandas, matplotlib.pyplot, seaborn and numpy to explore and describe the dataset.

## VISUALIZATION

Creating graphs for exploration, cleaning assurance and data analysis.



# Multivariate analysis

**The multivariate analysis was conducted by simultaneously examining multiple variables within the dataset to uncover relationships and patterns.**

**We analyzed numerical variables like age, weight, height, FCVC, and TUE for relationships and categorical variables like gender for patterns. This comprehensive analysis provided valuable insights for decision-making and strategy development across various domains.**



# Feature Engineering:

- Generate new features or insights to better understand the relationship between obesity levels and genetics.
- For instance, compute BMI using height and weight, or categorize ages into groups.
- In this study classification algorithms such as logistic regression (LR), random forest (RF), and Extreme Gradient Boosting (XGBoost) can be applied.





# Visualization and Summary

- Used Python and Tableau to visualize key findings and insights using appropriate charts, graphs, and plots.
- Summarize the main observations and trends discovered during the EDA process.
- Prepare a report or presentation to communicate the results to stakeholders.





# Iterate and refine

To iterate and refine is a process commonly used in problem-solving and project development, where successive iterations or cycles of improvement are made to refine a solution or product

It's a flexible and adaptive method that enables teams to respond to changing requirements, address issues, and optimize outcomes over time.

- **Iterate...**

...through the EDA process as needed, exploring different variables, visualizations, or analyses to gain deeper insights.

- **Refine...**

...the analysis based on feedback from stakeholders or additional data exploration.



# Conclusions

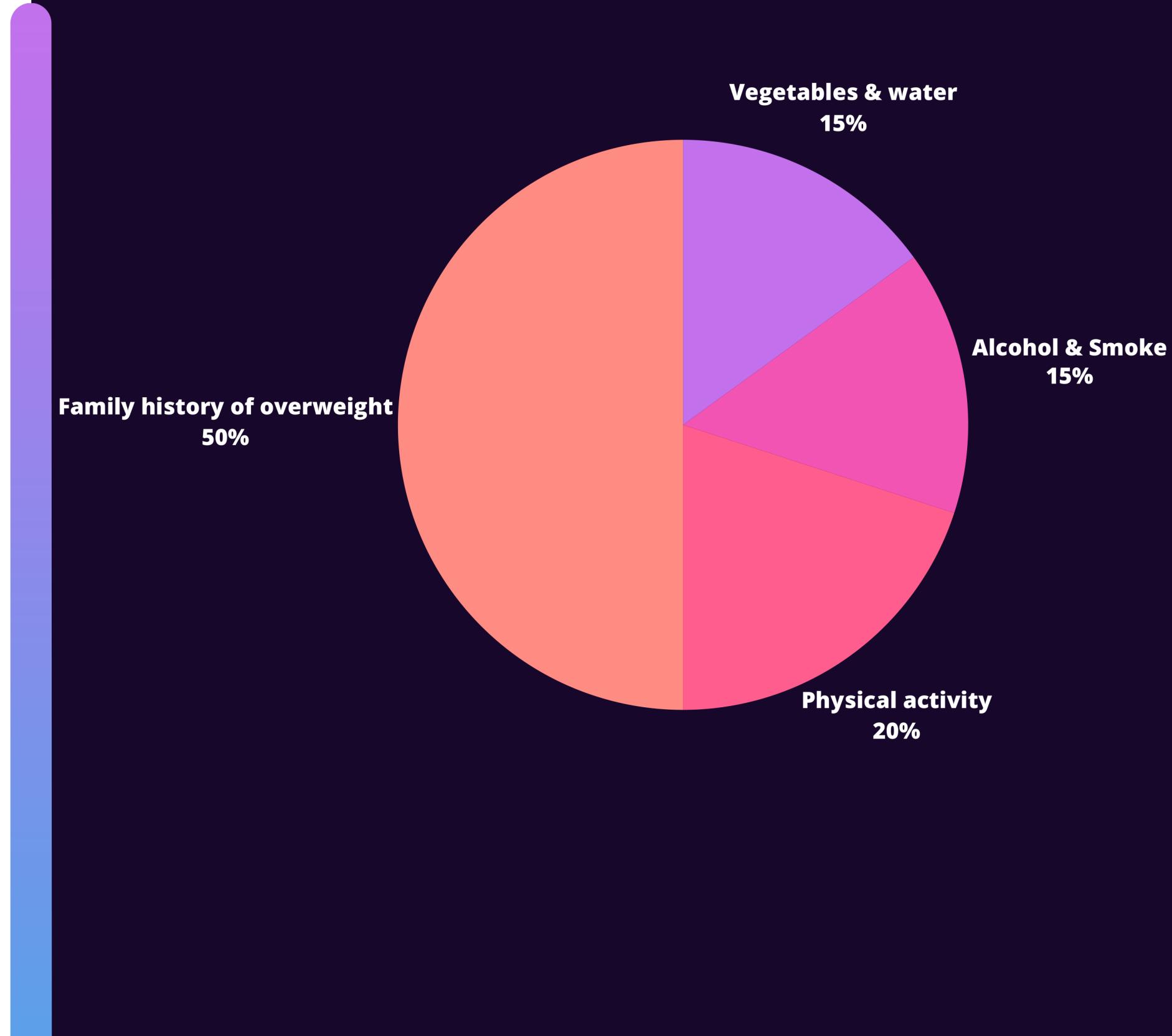
H1: We could not find a significant relation between vegetables and water consumption with obesity levels.

H2: We could not find a significant relation between alcohol consumption and smoking with obesity levels.

H3: We found some trends between frequency of physical activity with obesity levels.

We found a strong relation between Family history of overweight and obesity levels.

Further data collection, research, and collaboration with geneticists are essential to advance our understanding and address the complexities of the problem effectively.



# CONTACT

[www.healthtechanalyticscenter.com](http://www.healthtechanalyticscenter.com)

[analyzeme@healthtechanalytic  
scenter.com](mailto:analyzeme@healthtechanalytic<br/>scenter.com)

+123-456-7890

Storkower Straße 132, 10407  
Berlin



[www.healthtechanalyticscenter.com](http://www.healthtechanalyticscenter.com)

**THANK  
YOU**

