

Understanding the Drivers of Obesity: A Predictive Analysis



CAL POLY POMONA

Brandan Ly

Gerardo Gutierrez

May 05, 2025

OVERVIEW

1

BACKGROUND

2

HYPOTHESIS

3

EXISTING
LITERATURE

4

FINDINGS & ANALYSIS

5

DATA COLLECTION &
WRANGLING

6

LIMITATIONS

7

DEMO

8

CONCLUSION &
FUTURE WORK

WHAT IS OBESITY?

Obesity is typically defined as a **body mass index (BMI)** of **30 or higher**.

$$\text{BMI} = \frac{\text{Weight (in kilograms)}}{\text{Height}^2 \text{ (in meters)}}$$

Obesity results from a **chronic energy imbalance** between energy intake and expenditure.

Obesity is a **complex** condition with significant social and psychological implications.

Why BMI?

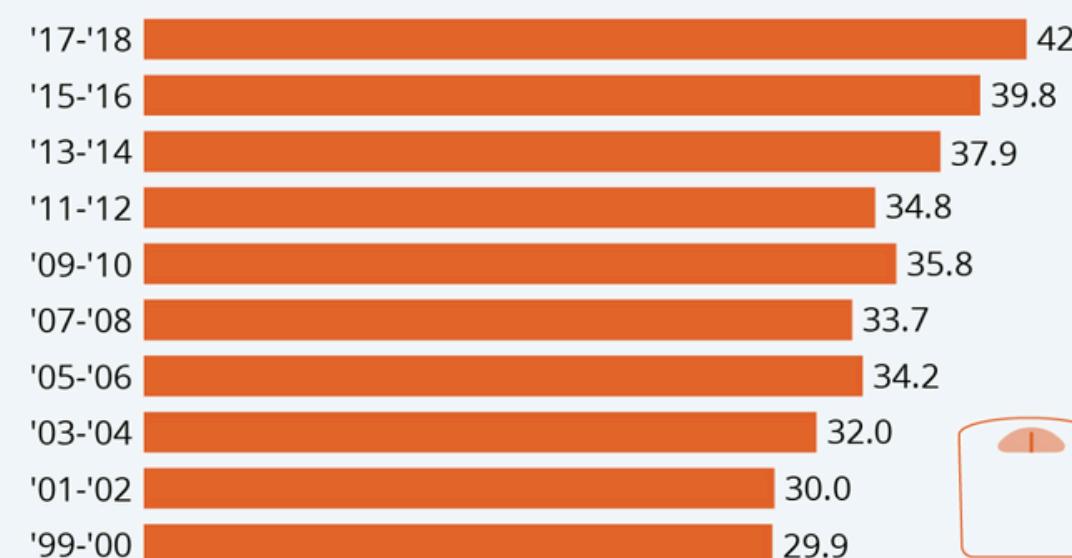
- **Quick** and simple to calculate
- **Inexpensive**
- **Non-invasive**
- **Easily collected** during routine health visits
- Well established population health measure that helps inform public health policy.



PREVALENCE

Obesity Rates Continue to Trend Up in U.S.

Percentage of U.S. adults who are obese based on height and weight survey



Data collected by CDC based on survey of 5,000 U.S. adults

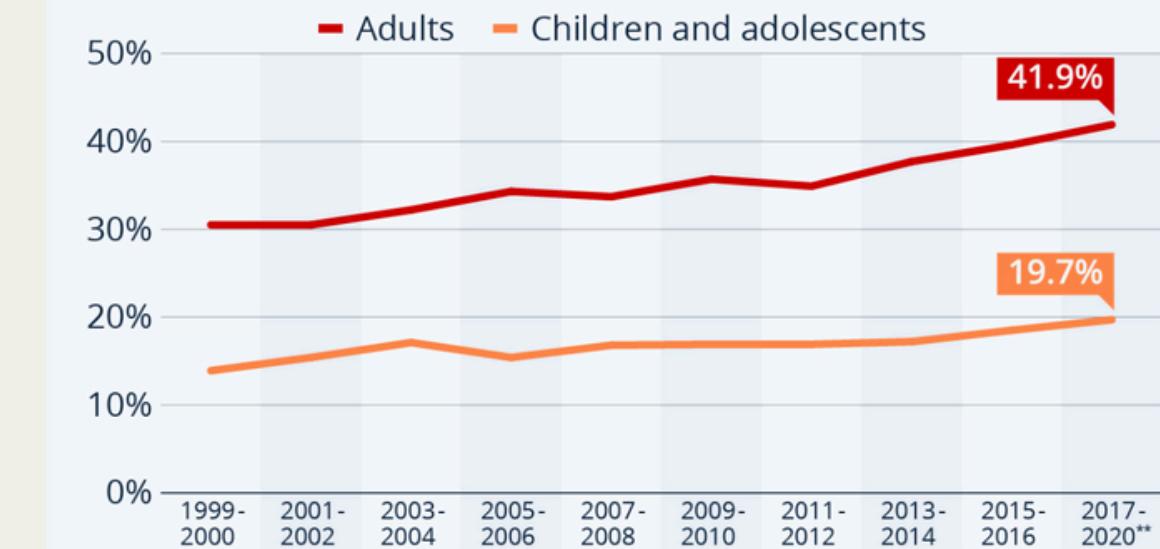
Source: Centers for Disease Control and Prevention



statista

America's Rising Obesity Problem

Prevalence of obesity among adults and children/adolescents in the United States*



* Adults aged 20 and over, children and adolescents aged 2-19

** Partial data collection in 2019-2020 cycle combined with 2017-2018 data for nationally representative estimates

Source: Centers For Disease Control and Prevention



statista

- With obesity rates rising its important to **understand** the behavioral and demographic variables that influence obesity **risk**.
- Awareness** has increased, but obesity epidemic **remains a major concern** globally.

O B E S I T Y H E A L T H R I S K S

Obesity Health Risks

- Type 2 diabetes (insulin resistance)
- Cardiovascular Disease
- Sleep Apnea
- Osteoarthritis
- Low Back Pain & Mobility Limitations
- Depression, anxiety



You may not be able to change all of your risk factors but **knowing your risk** will help you to take steps to reaching a healthy weight and **lower your risk** for obesity-related health problems.



HYPOTHESIS

Poor nutritional habits are the **key** driver of obesity.



EXISTING LITERATURE



BEHAVIORAL RISK FACTORS OF OBESITY

Cross-sectional, multinational, study of 6,025 children aged 9-11 years

Moderate-to-Vigorous Physical Activity (MVPA):

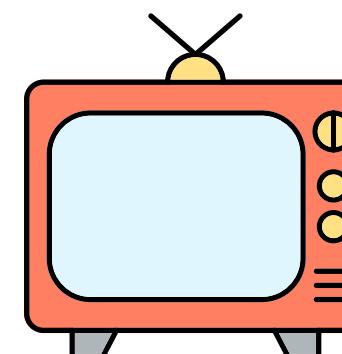
- Lower levels of MVPA were strongly associated with higher odds of obesity.
 - Boys: Odds Ratio (OR) = 0.52
 - Girls: OR = 0.43
 - Children with moderate to high levels of MVPA had about **50 % odds** of being obese.

Sleep Duration:

- Shorter nocturnal sleep duration was also linked to increased obesity risk.
 - Boys: OR = 0.79
 - Girls: OR = 0.71
 - Suggesting that **adequate sleep** is important for preventing obesity.

High TV Time increased obesity odds

- Boys: OR = 1.15
- Girls: OR = 1.07



(Katzmarzyk et al., 2015)

BEHAVIORAL RISK FACTORS OF OBESITY

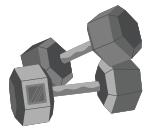
In a cross-sectional study analyzing 48,741 adults across 31 countries :

- Ex-smoking was significantly associated with higher odds of obesity.
- Light/moderate smoking appeared inversely related to obesity, likely due to appetite suppression, though **not recommended** as a preventive measure.
- Daily Fruit and Vegetable Consumption 
 - Men: Significantly protective against both overweight and obesity.
 - Women: No statistically significant association.
- Physical Activity 
 - Positively associated with being overweight, but not significantly associated with obesity.
 - This likely reflects **reverse causation**, where overweight individuals increase activity as a weight-control effort.

(Pengpid & Peltzer, 2017).

BEHAVIORAL RISK FACTORS OF OBESITY

Most Consistently Identified Behavioral Risk Factors



Physical Inactivity – Lack of **moderate to vigorous** physical activity (MVPA) is among the strongest predictors of obesity in both children and adults.



Poor Diet Quality – Diets high in ultra-processed foods, added sugars, and saturated fats contribute significantly to **caloric surplus** and fat accumulation.



High Sedentary Time (e.g., TV viewing) – Excessive screen time is associated with **snacking behavior**, poor sleep, and **reduced energy expenditure**.



Inadequate Sleep Duration – Short sleep (<7 hours/night in adults) disrupts **appetite-regulating hormones** (ghrelin, leptin), promoting overeating.



Smoking Behavior (especially ex-smokers) – Weight gain is common after smoking **cessation**, partly due to increased appetite.



Alcohol Consumption – Moderate to heavy alcohol intake can increase **caloric intake** and alter metabolism.

FINDINGS & ANALYSIS



CORRELATION

Positive Correlations

1. Family History

2. Age

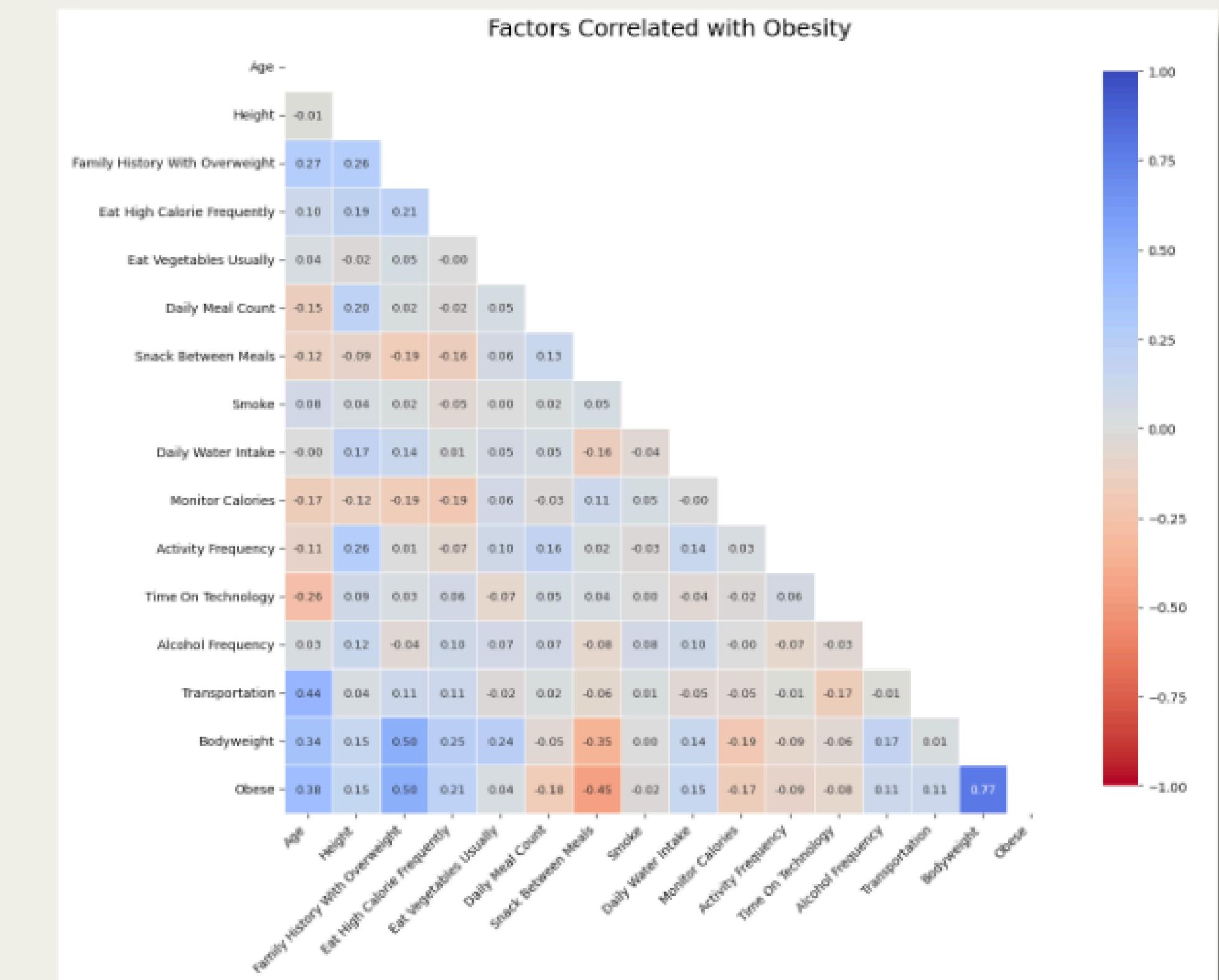
3. High Calorie Meals

4. Eating Vegetables Frequently

Negative Correlations

1. Snack Between Meals

2. Monitor Calories

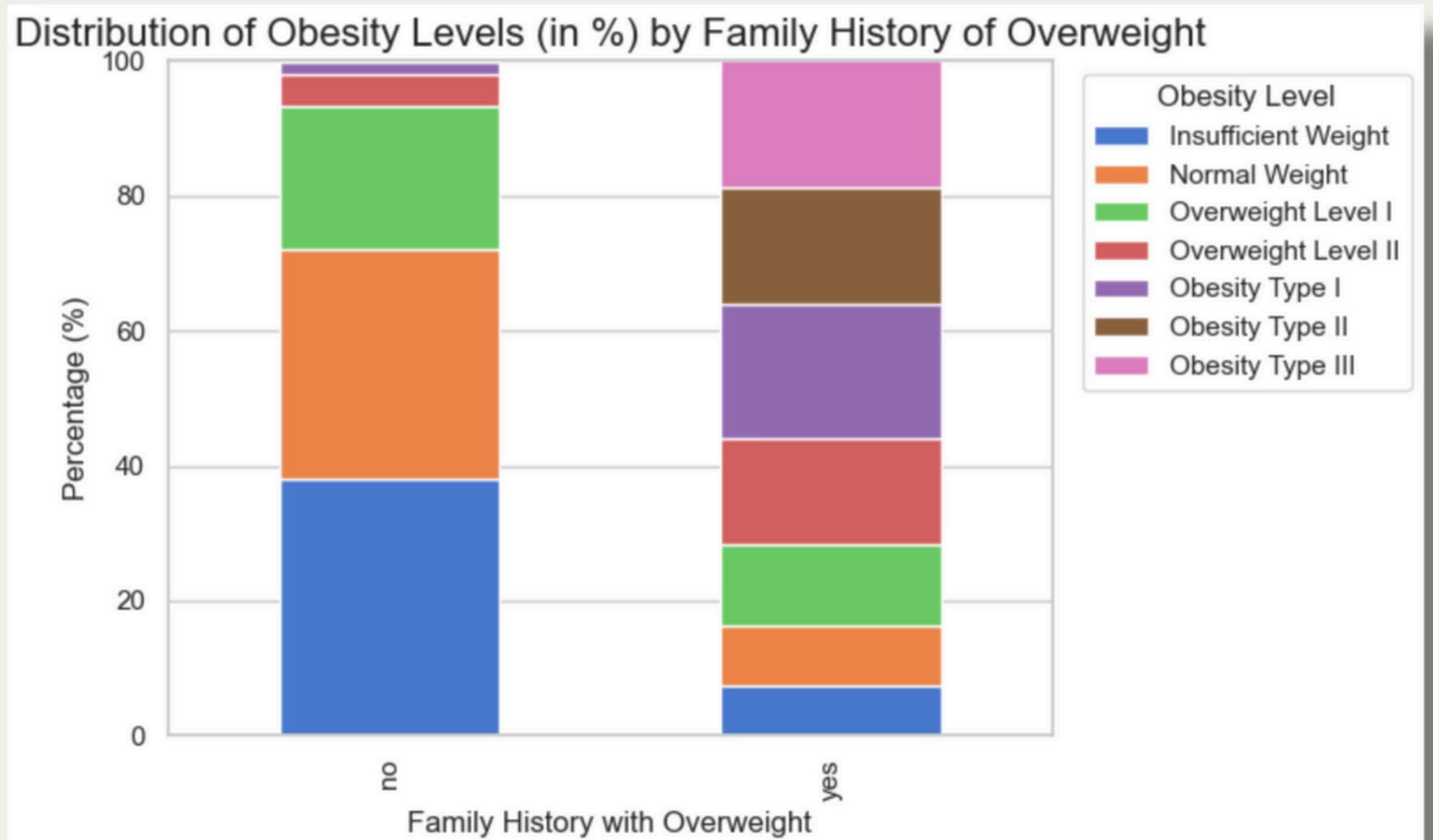


ANALYSIS

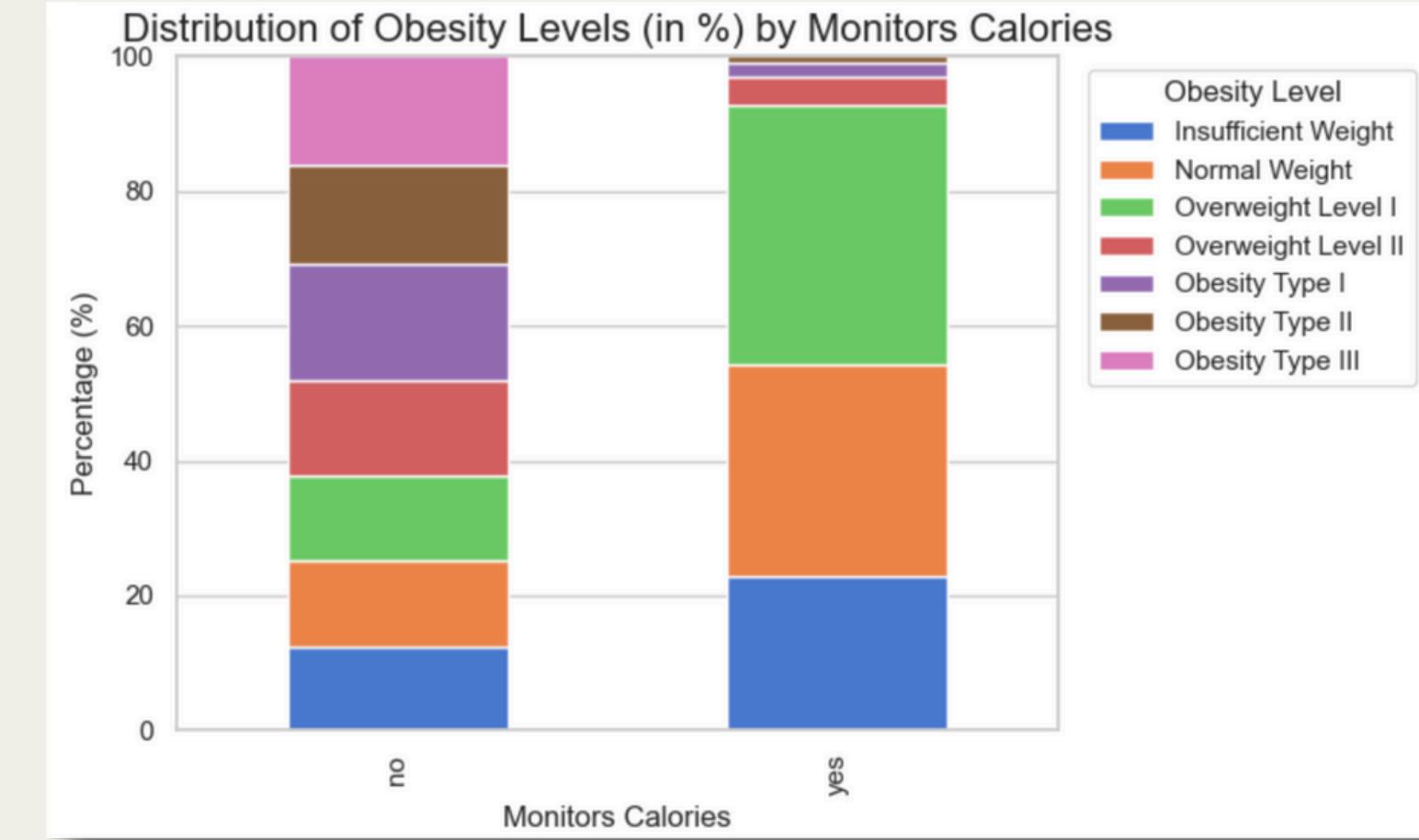
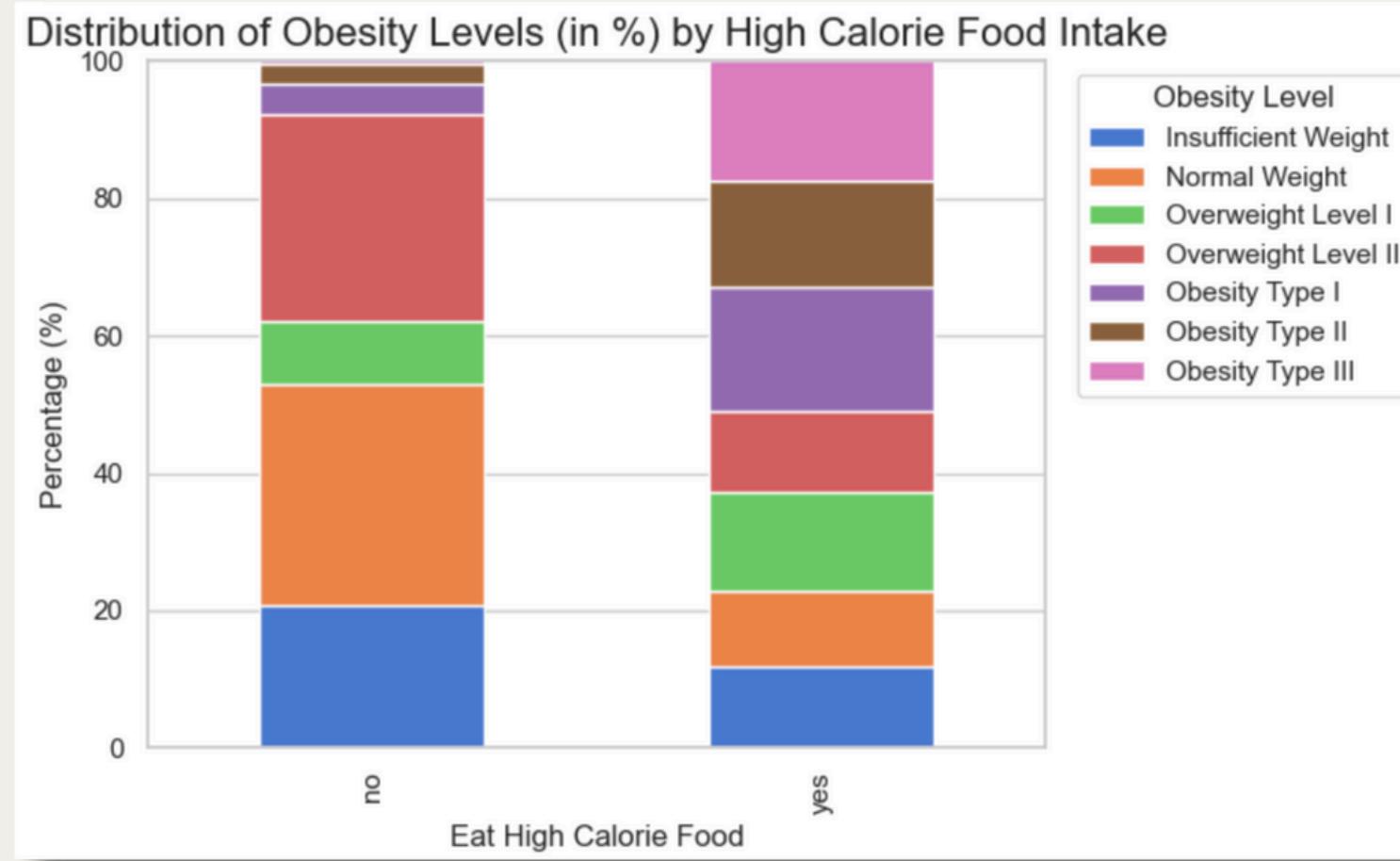
- Over 50% who consume high calorie foods were obese
- Higher obese levels occupy largest portions of bar for yes column
- Normal weight and insufficient weight dominate no column

Interpretation

- Individuals w/ family history overweight likely to fall into moderate to severe obesity



ANALYSIS



- **90% of individuals who do not consume high calories foods were not obese**
- **Over 50% who consume high calorie foods were obese**

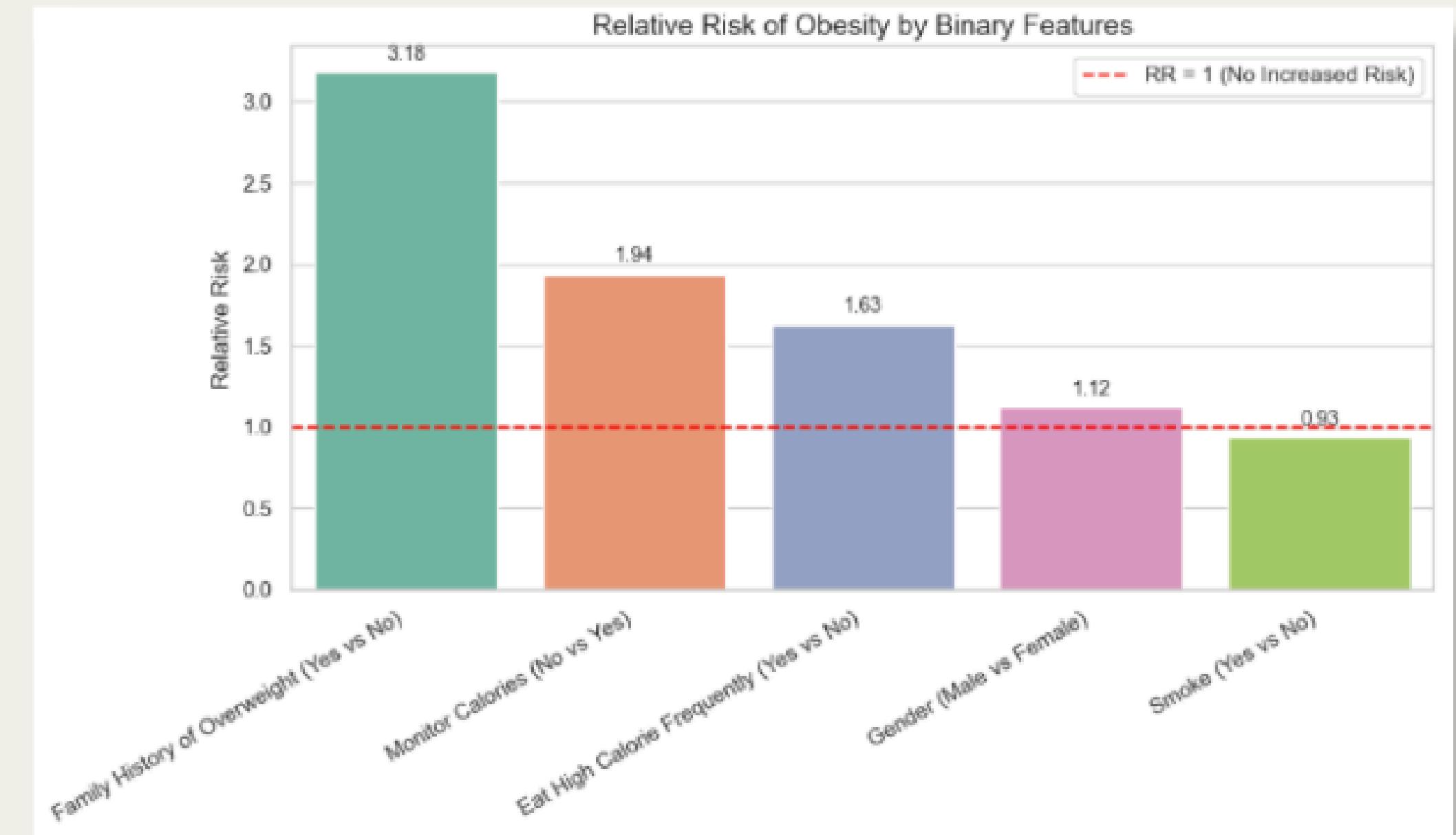
- **Over 90% of individuals who monitor calories were not obese.**
- **Over 40% of individuals who did not monitor calories were obese.**

RELATIVE RISK

$$RR = \frac{p_1}{p_2}$$

Individuals with **family history** are more than **3 times** as likely to be obese.

Individuals who do **NOT monitor calories** are almost **2 times** as likely to be obese.



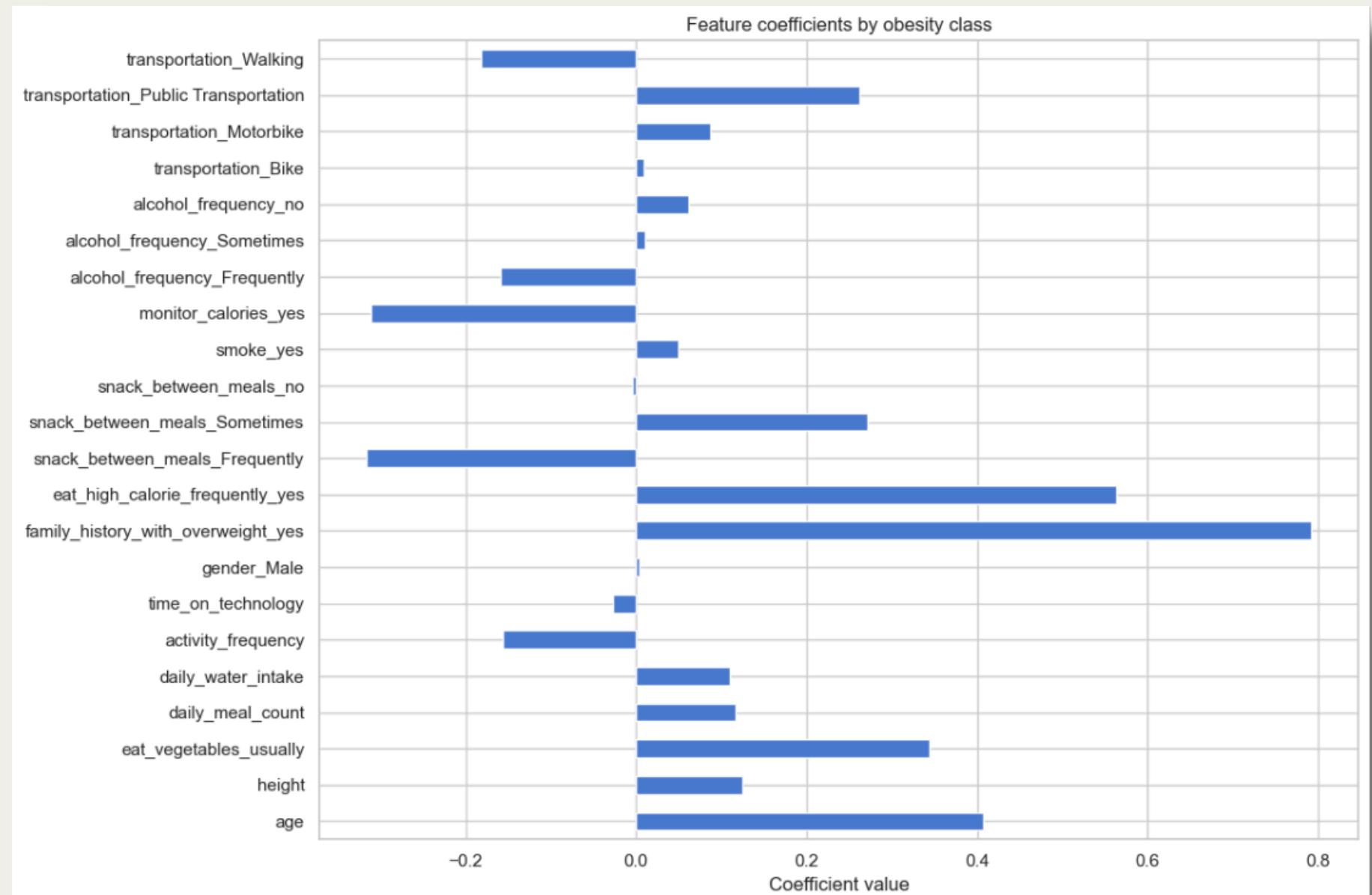
PREDICTION

Top Predictors

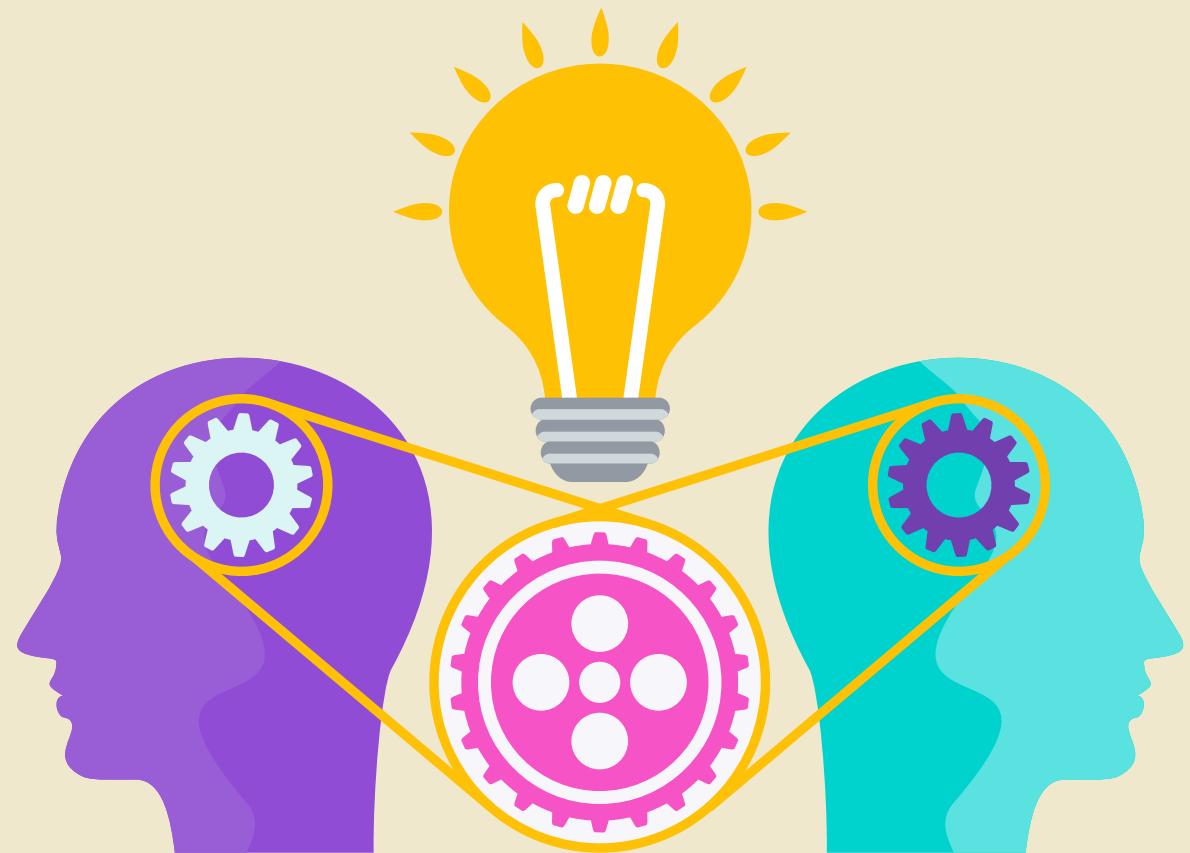
- Family History with Overweight
- Eat High Calorie Foods
- Public Transportation
- Age

Analysis

- Family history & high calorie foods increase likelihood of obesity
- Inverse for monitor calories



METHODS



METHODS

Dataset Collection

- UC Irvine Machine Learning Repo
- Survey Data from Mexico, Peru, Colombia
- 2111 Instances, 77% synthetic.
- 16 Attributes



Data Wrangling

- No missing variables
- Changing column names
- Reformatting categorical data
- Hot encoding for logistic regression
- Adding BMI feature to dataset
- Adding obese binary feature



LIMITATIONS

- 77% of 2111 records were generated with **SMOTE**
- **Lack of public data.**
 - There are very few large-scale obesity datasets in the public domain due to **privacy laws** (e.g., HIPAA, GDPR).

Inappropriate Uses

- Clinical research and medical decisions
- Public Health Policy
- Generalization to other populations



Appropriate Uses

- Training and benchmarking machine learning models.
- Learning data science lifecycle.



SMOTE

What is SMOTE?

- SMOTE stands for **Synthetic Minority Oversampling Technique**.
- It **generates synthetic data** points for the minority class by interpolating between existing samples.
- Commonly used to **balance imbalanced** datasets where one class is **underrepresented**.



Why SMOTE?

- The dataset had a **class imbalance**, with far **fewer obese** individuals compared to non-obese.
- SMOTE is applied to prevent model bias toward the majority class.
- It helps the model learn patterns from both classes more effectively, improving prediction on minority cases.

SAMPLE MINORITY OVERSAMPLING TECHNIQUE

- **SMOTE** helps by generating **synthetic** data points for the **minority class** (e.g., obese individuals), instead of just duplicating them.
- It works by creating **new points between** existing minority class **neighbors**.

Example

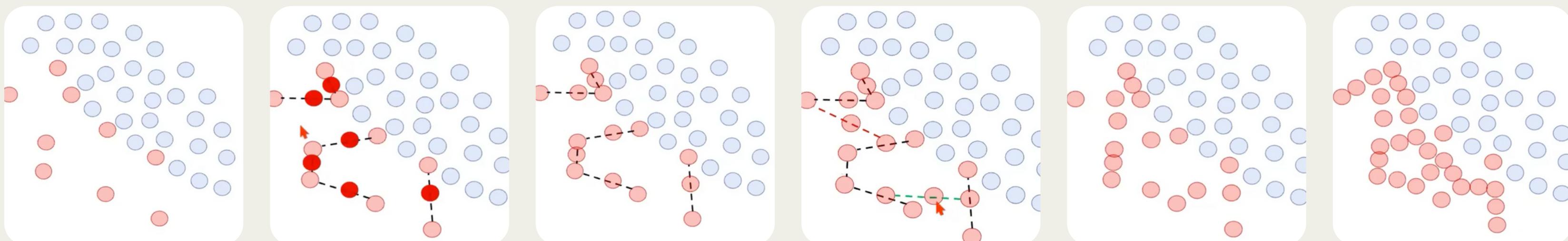
Person A: BMI = 33, meals/day = 3

Person B: BMI = 35, meals/day = 4

→ SMOTE might create a new point:

BMI = 34, meals/day = 3.5

- This technique **helps balance the dataset** and improve model performance.



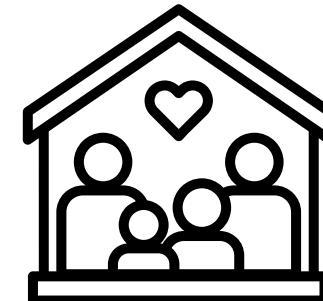
DEMO



CONCLUSIONS

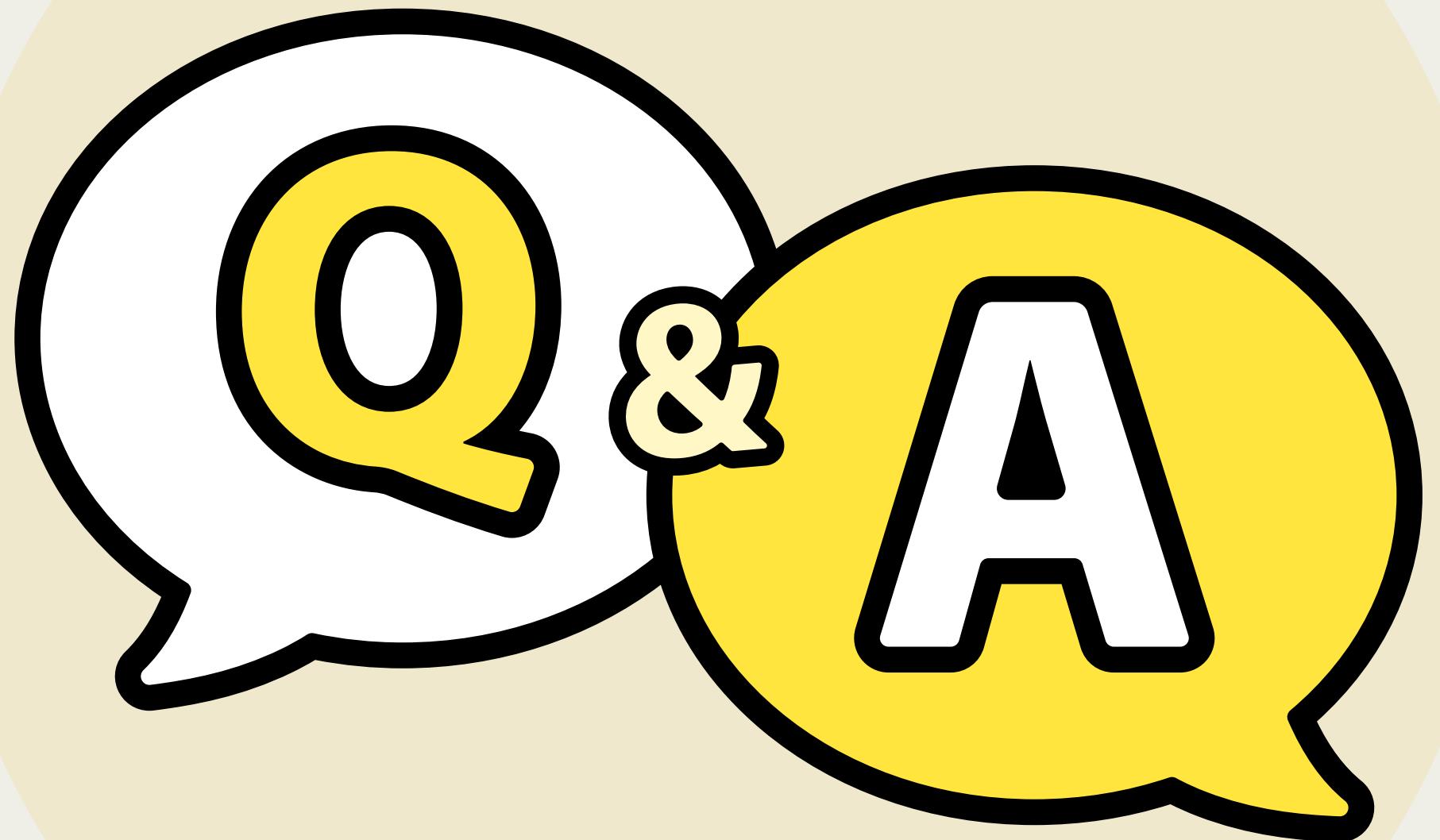
Conclusion

- **Dietary factors** are, indeed, amongst the greatest risk factors of obesity.
- **Family history** appears to have the greatest influence on obesity.
 - However, this likely stems from inherited behaviors and lack of nutritional knowledge



Further Work

- Further investigate how **marketing** and **public education** influences dietary decisions.
- Investigate **socioeconomic** factors and **geographic** prevalence to have a more targeted approach.
- We can apply these methods and models to a **non-synthetic** dataset to draw more accurate insights and inform public policy.

The logo features the letters "Q" and "&" on a white speech bubble, and the letter "A" on a yellow speech bubble, all set against a large light beige circle.

Q & A

REFERENCES

- Center for Disease Control
- Pi-Sunyer, F. X. (2002). The medical risks of obesity. *Obesity Surgery*, 12(S1), S6–S11. <https://doi.org/10.1007/BF03342140>
- Stewart, M. W. (2009). Obesity and the perianesthesia patient. *Journal of PeriAnesthesia Nursing*, 24(5), 332–334. <https://doi.org/10.1016/J.JOPAN.2009.07.006>