Analysis of Obesity Levels Based On Eating Habits and Physical Condition.

Group 3:
Shivani Mangesh Sawant, Urvi Ravindra Dhomne, Sharmitha Chilankuri, Kallamadi Rutva

Introduction

Background:

Obesity is a significant global health issue linked to various chronic diseases, including diabetes, cardiovascular conditions, and hypertension. Understanding the factors contributing to obesity, such as dietary habits and physical activity levels, i intervention strategies.

Dataset:

Title: Estimation of Obesity Levels Based on Eating Habits and Physical Condition
Source: UCI Machine Learning Repository
No of Instances: 2111
No of Variables: 17

Objective:

- To explore the relationship between eating habits, physical activity, and obesity levels, aiming to identify actionable insights for obesity prevention and management.

Dataset Overview

The dataset contains **16 independent variables** and **1 dependent variable**.

Variables:

- **Independent Variables**: Age, Gender, Height, Weight, Dietary Habits (e.g., high-calorie food frequency), Physical Activity, and more.
- **Dependent Variable**: Obesity Level (categorized into seven classes, e.g., Normal Weight, Overweight, Obesity Types I–III)

Sample Size: Diverse individuals aged 14–61 years, capturing a range of dietary and lifestyle behaviors.

Dataset

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Gender	Age	Height	Weight	family_hist	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
Female	21	1.62	64	yes	no	2	3	Sometimes	no	2	no	0	1	no	Public_Transportation	Normal_Weight
Female	21	1.52	56	yes	no	3	3	Sometimes	yes	3	yes	3	0	Sometime	Public_Transportation	Normal_Weight
Male	23	1.8	77	yes	no	2	3	Sometimes	no	2	no	2	1	Frequently	Public_Transportation	Normal_Weight
Male	27	1.8	87	no	no	3	3	Sometimes	no	2	no	2	0	Frequently	Walking	Overweight_Level_I
Male	22	1.78	89.8	no	no	2	1	Sometimes	no	2	no	0	0	Sometime	Public_Transportation	Overweight_Level_II
Male	29	1.62	53	no	yes	2	3	Sometimes	no	2	no	0	0	Sometime	Automobile	Normal_Weight
Female	23	1.5	55	yes	yes	3	3	Sometimes	no	2	no	1	0	Sometime	Motorbike	Normal_Weight
Male	22	1.64	53	no	no	2	3	Sometimes	no	2	no	3	0	Sometime	Public_Transportation	Normal_Weight
Male	24	1.78	64	yes	yes	3	3	Sometimes	no	2	no	1	1	Frequently	Public_Transportation	Normal_Weight
Male	22	1.72	68	yes	yes	2	3	Sometimes	no	2	no	1	1	no	Public_Transportation	Normal_Weight
Male	26	1.85	105	yes	yes	3	3	Frequently	no	3	no	2	2	Sometime	Public_Transportation	Obesity_Type_I
Female	21	1.72	80	yes	yes	2	3	Frequently	no	2	yes	2	1	Sometime	Public_Transportation	Overweight_Level_II
Male	22	1.65	56	no	no	3	3	Sometimes	no	3	no	2	0	Sometime	Public_Transportation	Normal_Weight
Male	41	1.8	99	no	yes	2	3	Sometimes	no	2	no	2	1	Frequently	Automobile	Obesity_Type_I
Male	23	1.77	60	yes	yes	3	1	Sometimes	no	1	no	1	1	Sometime	Public_Transportation	Normal_Weight
Female	22	1.7	66	yes	no	3	3	Always	no	2	yes	2	1	Sometime	Public_Transportation	Normal_Weight
Male	27	1.93	102	yes	yes	2	1	Sometimes	no	1	no	1	0	Sometime	Public_Transportation	Overweight_Level_II
Female	29	1.53	78	no	yes	2	1	Sometimes	no	2	no	0	0	no	Automobile	Obesity_Type_I
Female	30	1.71	82	yes	yes	3	4	Frequently	yes	1	no	0	0	no	Automobile	Overweight_Level_II
Female	23	1.65	70	yes	no	2	1	Sometimes	no	2	no	0	0	Sometime	Public_Transportation	Overweight_Level_I
Male	22	1.65	80	yes	no	2	3	Sometimes	no	2	no	3	2	no	Walking	Overweight_Level_II
Female	52	1.69	87	yes	yes	3	1	Sometimes	yes	2	no	0	0	no	Automobile	Obesity_Type_I
Female	22	1.65	60	yes	yes	3	3	Sometimes	no	2	no	1	0	Sometime	Automobile	Normal_Weight
Female	22	1.6	82	yes	yes	1	1	Sometimes	no	2	no	0	2	Sometime	Public_Transportation	Obesity_Type_I

Why Did We Choose This Project?

- **Global Health Significance**

Obesity is a major public health issue worldwide, linked to chronic diseases like diabetes and cardiovascular conditions. Analyzing its determinants can contribute to developing effective prevention strategies.

- **Comprehensive Dataset**

The chosen dataset provides a rich variety of variables, including dietary habits and physical activity, enabling a holistic analysis of factors influencing obesity levels.

- **Alignment with Academic and Professional Goals**

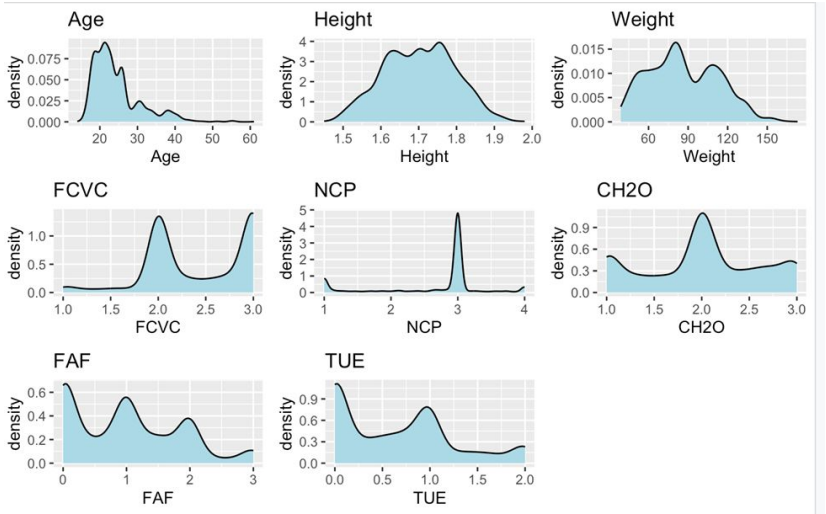
This project combines our academic interests in data analysis with a real-world problem. It also helps develop critical skills in handling complex datasets, formulating hypotheses, and applying analytical techniques relevant to our coursework and future careers.

- **Need for Preventative Measures**

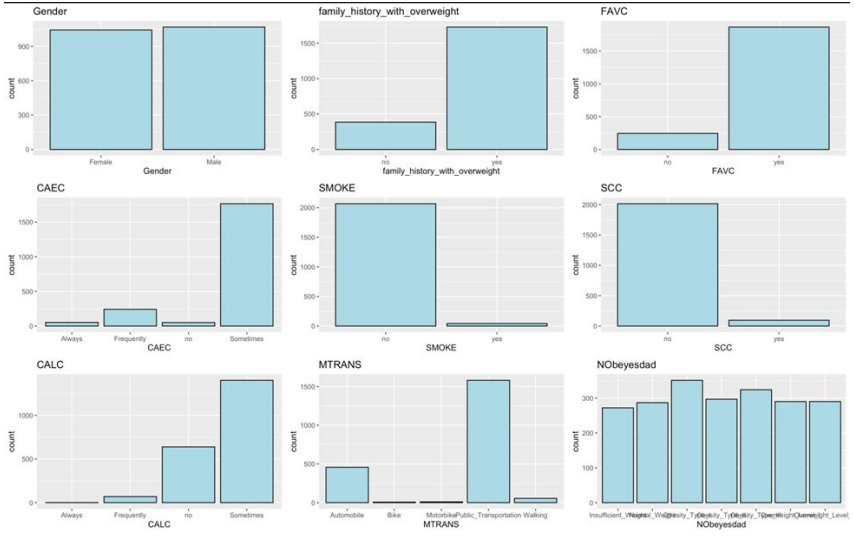
Understanding the relationship between lifestyle factors (e.g., diet, physical activity) and obesity levels can guide the development of effective prevention and management strategies, reducing healthcare burdens.

Statistical Analysis

Density Plot of Numerical Variables:



Bar Chart of Categorical Variables:

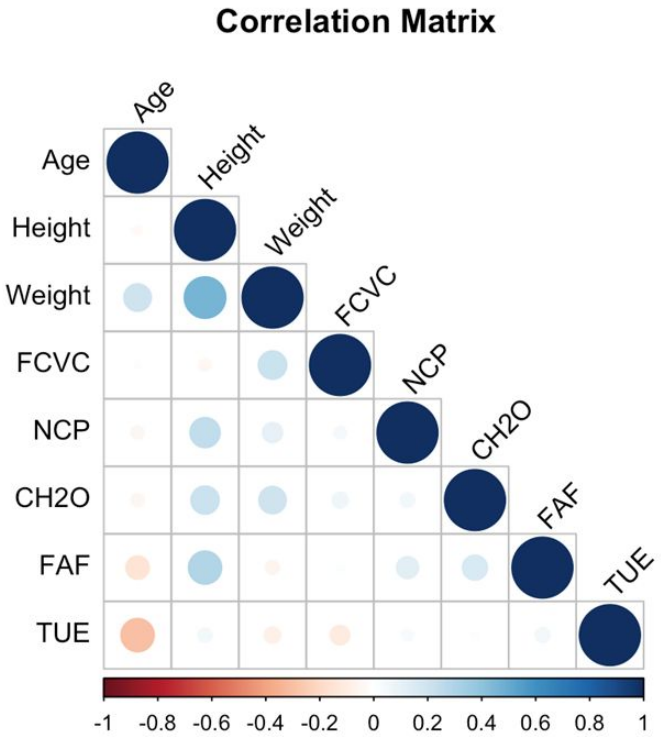


Statistical Analysis

We can see that the following variable pairs have relatively high correlations (above 0.5 or below -0.5):

- Height and Weight (0.84)
- FAVC and FCVC (0.74)
- NCP and CAEC (0.68)
- SMOKE and CH2O (0.67)
- SCC and FAF (0.63)
- FAF and TUE (0.57)
- MTRANS and NObeyesdad (-0.65)

	Age	Height	Weight	FCVC	NCP	CH2O	FAF	TUE
Age	1.00	-0.03	0.20	0.02	-0.04	-0.05	-0.14	-0.30
Height	-0.03	1.00	0.46	-0.04	0.24	0.21	0.29	0.05
Weight	0.20	0.46	1.00	0.22	0.11	0.20	-0.05	-0.07
FCVC	0.02	-0.04	0.22	1.00	0.04	0.07	0.02	-0.10
NCP	-0.04	0.24	0.11	0.04	1.00	0.06	0.13	0.04
CH2O	-0.05	0.21	0.20	0.07	0.06	1.00	0.17	0.01
FAF	-0.14	0.29	-0.05	0.02	0.13	0.17	1.00	0.06
TUE	-0.30	0.05	-0.07	-0.10	0.04	0.01	0.06	1.00



Statistical Analysis

Chi Square test for categorical variables associations

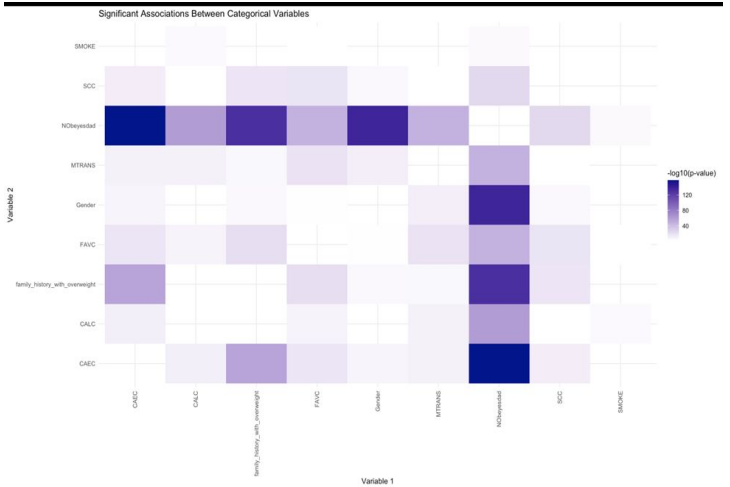
·There appears to be a strong association between Gender and several other variables, including FAVC (Frequent consumption of high caloric food), SMOKE, CH2O, and SCC. This suggests that gender plays a significant role in factors related to obesity & lifestyle.

· The variables NCP (Number of main meals) and CAEC (Consumption of food between meals) also show a notable association, indicating a relationship between the number of meals and the tendency to consume food outside of main meals.

· The variables FAF and TUE (Time using technology) exhibit a significant association, implying a link between body fat levels and the amount of time spent using technology.

· The variable MTRANS (Transportation method) shows a strong negative association with NObeyesdad (Obesity level), suggesting that the mode of transportation used is related to the obesity level.

· Other significant associations exist between variables like SMOKE and CH2O, SCC and FAF, and FAVC and FCVC (Frequency of consumption of vegetables), indicating complex relationships between lifestyle, dietary, and physiological factors.

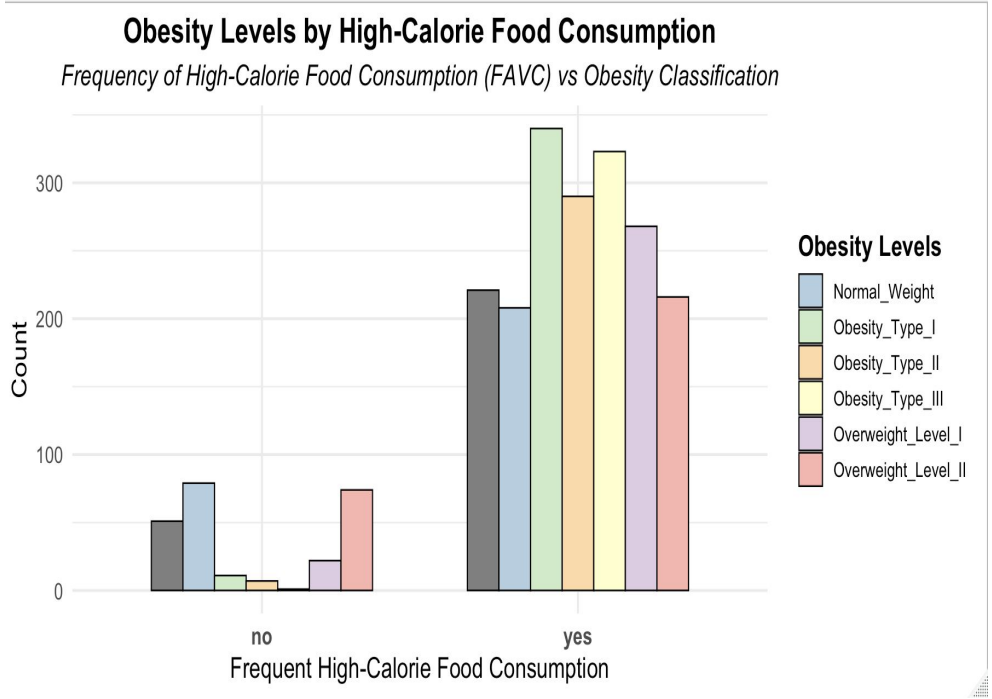


Research Hypothesis 1

Individuals with frequent consumption of high-calorie foods have a higher likelihood of obesity than those with infrequent consumption.

```
        Pearson's Chi-squared test

data:  contingency_table
X-squared = 233.34, df = 6, p-value < 2.2e-16
```



Research Hypothesis 1

Individuals with frequent consumption of high-calorie foods have a higher likelihood of obesity than those with infrequent consumption.

Chi-Square Test Results:

The Chi-Square test results show a statistically significant association between the frequency of high-calorie food consumption (FAVC) and the obesity level (NObeyesdad), with a p-value < 0.05.

Conclusion:

These results support the research hypothesis that individuals with frequent consumption of high-calorie foods have a higher likelihood of obesity.

Research Question 1

How do dietary habits (e.g., consumption frequency of high-calorie foods) influence obesity levels among individuals?

1. **Findings from the Chi-Square Test**:

There is a **statistically significant association** between the frequency of high-calorie food consumption (FAVC) and obesity levels (NObeyesdad), with a p-value < 0.05.

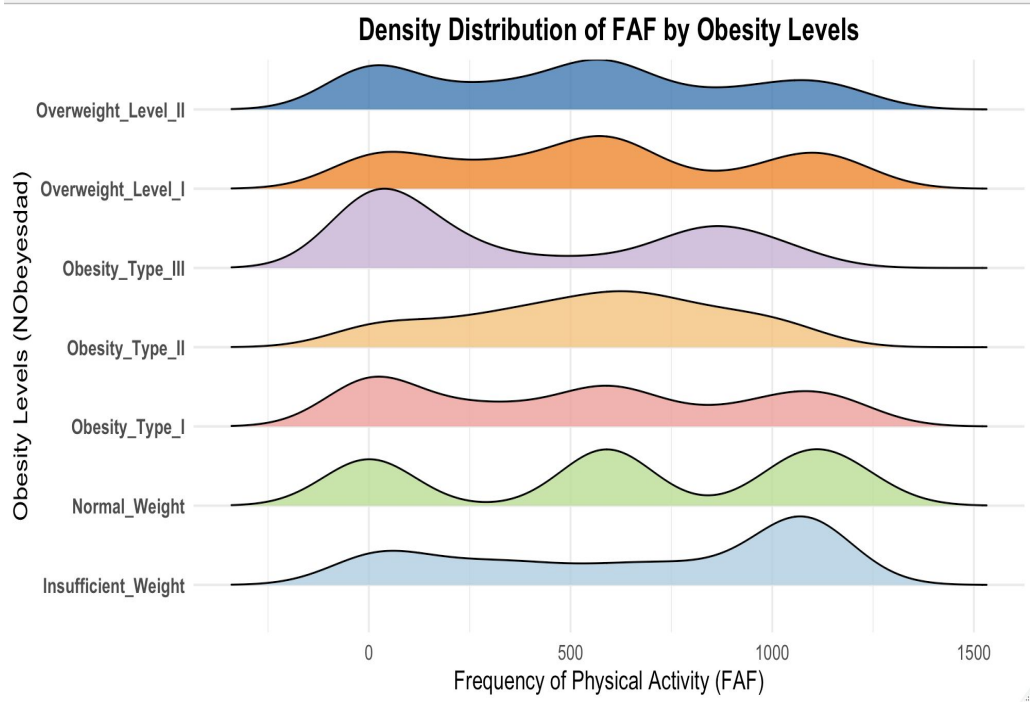
2. **Correlation Analysis**:
   - A strong correlation (0.74) exists between high-calorie food frequency (FAVC) and vegetable consumption frequency (FCVC), showing how dietary habits interrelate.
3. **Odds Ratio Analysis**:
   - Individuals with frequent consumption of high-calorie foods (FAVC = Yes) are more likely to fall into higher obesity categories compared to those with infrequent consumption (FAVC = No).
4. **Conclusion**:
   - The results confirm the hypothesis that frequent high-calorie food consumption significantly increases the likelihood of obesity. This highlights the need to address high-calorie diets in obesity interventions.

Research Hypothesis 2

Individuals with low physical activity levels have a higher likelihood of obesity than those with moderate to high activity levels.

```
        Pearson's Chi-squared test

data:  table_faf_obesity
X-squared = 7678.3, df = 7134, p-value = 4.275e-06
```



Research Hypothesis 2

Individuals with low physical activity levels have a higher likelihood of obesity than those with moderate to high activity levels.

1. **Findings from Physical Activity Frequency (FAF)**:
   - Individuals with **low physical activity frequency (FAF = 0)** are significantly more likely to fall into higher obesity categories
2. **Statistical Evidence**:
   - A significant negative correlation between **Physical Activity Frequency (FAF)** and obesity levels (from correlation analysis and regression models) indicates that as physical activity decreases, the likelihood of higher obesity levels increases.
   - The multinomial logistic regression further confirms that individuals with **low activity levels** are more likely to belong to obesity categories compared to those with moderate or high physical activity.
3. **Mode of Transportation Analysis**:
   - Non-active transportation modes (e.g., automobiles) are predominantly linked to higher obesity levels.
   - Active modes (e.g., walking) are associated with **Normal Weight** or **Overweight_Level_I**.
4. **Conclusion**:
   - The hypothesis is supported: **Low physical activity levels significantly increase the likelihood of obesity**, highlighting the importance of promoting moderate to high physical activity as part of public health strategies to combat obesity.

Research Question 2

What is the relationship between physical activity levels and obesity?

1. **Correlation Analysis**:

   The correlation matrix shows a significant negative correlation (-0.65) between the mode of transportation (MTRANS) and obesity levels (NObeyesdad). This suggests that active transportation methods (e.g., walking, biking) are associated with lower obesity levels.

2. **Chi-Square Test for Associations**:

   Variables related to physical activity, such as FAF (Physical Activity Frequency), are significantly associated with obesity levels. The relationship implies that reduced physical activity correlates with higher obesity levels.

3. **Regression Analysis**:

   Multinomial logistic regression results suggest that lower levels of physical activity and sedentary transportation choices are strong predictors of higher obesity categories.

4. **Conclusion**:

   Physical activity levels have a substantial inverse relationship with obesity. Encouraging more frequent physical activity and promoting active transportation can effectively mitigate obesity risks.

Prediction Model

Multinomial Logistic Regression to Predict Obesity Category:

Since the output of our prediction model consists of categorical data with 7 categories, we opted to use logistic regression.
We selected multinomial logistic regression because the output involves more than two categories.
We evaluated the model using a subset of the data and achieved an accuracy of 97.91%.

predictions	Insufficient_Weight	Normal_Weight	Obesity_Type_I	Obesity_Type_II
Insufficient_Weight	267	3	0	0
Normal_Weight	5	278	0	0
Obesity_Type_I	0	0	350	2
Obesity_Type_II	0	0	1	295
Obesity_Type_III	0	0	0	0
Overweight_Level_I	0	6	0	0
Overweight_Level_II	0	0	0	0

predictions	Obesity_Type_III	Overweight_Level_I	Overweight_Level_II
Insufficient_Weight	0	0	0
Normal_Weight	0	6	0
Obesity_Type_I	0	0	2
Obesity_Type_II	0	0	0
Obesity_Type_III	324	0	0
Overweight_Level_I	0	275	10
Overweight_Level_II	0	9	278

Conclusion

The study on the **Estimation of Obesity Levels Based on Eating Habits and Physical Conditions** provides critical insights into the relationships between lifestyle factors and obesity. Using the dataset, we examined dietary habits, physical activity, and other behavioral factors to identify their influence on obesity levels. Here are the key takeaways:

1. **Dietary Habits and Obesity**
2. **Physical Activity and Obesity**
3. **Statistical Insights**

Future Work:

- Further analysis could explore the combined impact of psychological factors (e.g., stress coping capacity) and demographic variations (e.g., age, gender) on obesity.
- Machine learning models could be applied for more precise obesity risk predictions.