Estimation of Obesity Levels Based On Eating Habits and Physical Condition

By: Josh Bowser

Brief:

The goal of this project is to develop a comprehensive visual BMI calculator where the individual can input their height and weight. Once the height and weight is entered the user's BMI is calculated, the user is shown where they fall on the World Health Organization's standards of BMI and where they compare to others in their gender from the given dataset and are given helpful tips.

Project Analysis

01

BACKGROUND AND CHALLENGES

Information on the dataset. What was a conflict in the programming?

02

MOTIVATION AND APPROACHES

Why this project? What projects are already out there?

03

RESULTS AND OBSERVATIONS

What results were gathered? Any important observations?

04

CONCLUSION AND FUTURE WORK

Where to go from here?

Problem and Challenges

Project Background

- The dataset selected surveyed 2,112 individuals from around South America and asked them a series of questions such as:
    - What is their age, gender, height, and weight?
    - Do they count calories? How often are they eating high calorie meals?
    - How often are they exercising?
    - What is their water intake?
    - Main mode of transportation
    - Does anyone in their family have a history of high or low BMI?
- Factors such as these have heavy influence on one's BMI

Challenges

- One of the challenges that I faced was understanding if the small sample size was significant enough for the results. As mentioned, there were ~2,000 individuals sampled which in terms of population size is very small.
- Another challenged that I faced was finding adequate approaches. There were several articles that explained how eating alone wasn't the main factor of high BMIs but there didn't seem to be any applications besides BMI calculators. There only seemed to be physical scales that had apps that were able to calculate body fat
- The biggest challenge I also faced was when organizing the dataset, when implementing the machine learning aspect, I had to create a separate pipeline to properly be able to calculate the users BMI and give feedback.

Motivation and Approaches

Motivation

● Complex Factors at Play: Obesity is influenced by a mix of genetics, environment, and societal pressures, making it unfair to place all the blame on individuals who are often navigating these challenges without the right support.

● Lack of Access and Education: Many people face barriers like limited access to healthy food, education, or safe spaces to exercise, which make it much harder to make healthier choices, no matter how hard they try.

Existing Approaches

- Many BMI calculators, like those on NIH, CDC, and WebMD, primarily provide a numerical BMI value, offering limited personalized context or actionable steps, which can reduce their effectiveness in motivating individuals towards positive health changes.
- The oversimplified output of a BMI calculator can be discouraging or misleading, as it doesn't account for factors like muscle mass, body composition, and individual variations, potentially hindering its usefulness as a tool for promoting a holistic understanding of health.

webmd.com/diet/body-bmi-calculator

BMI | Your Waist Size | Healthy Weight Plan

BMI Calculator

*Required - Please complete these fields first

Height

5 ft 4 in

Weight

230 lbs

Calculate

Submit your height and weight to calculate your BMI. Complete optional fields below to learn more about your waist size, health risks, and steps to reach your weight goals.

Your BMI Result

Your Body Mass Index

39.5

Obesity

30+

Underweight
<18.5

Healthy
18.5-24.9

Overweight
25-29.9

Note: BMI is useful for most adults but it has some limitations. Confirm these results with a qualified health professional. Don't use this tool if you are under age 20, pregnant, have an eating disorder, or have a condition that affects your height.

"BMI is just one tool doctors use to understand how your weight affects your health. Your waist size can give a clearer picture of your health risks linked to abdominal fat. Complete the next section below to learn more."

Dr. Brunilda Nazario, MD
Chief Physician Editor, Medical Affairs, at WebMD

Obesity Class 2

"BMI is just one tool doctors use to understand how your weight affects your health. Your waist size can give a clearer picture of your health risks linked to abdominal fat. Complete the next section below to learn more."

Dr. Brunilda Nazario, MD
Chief Physician Editor, Medical Affairs, at WebMD

Obesity Class 2

Your BMI is in the Class 2 Obesity range (between 35 and 39.9).

There are 3 classes of obesity, with Class 3 carrying the greatest health risks.

The higher your obesity class, the greater your risk of health problems like type 2 diabetes, heart disease, stroke, and certain cancers. Taking healthy steps to reduce your weight may help prevent these conditions, stop them from getting worse, or even reverse some of them.

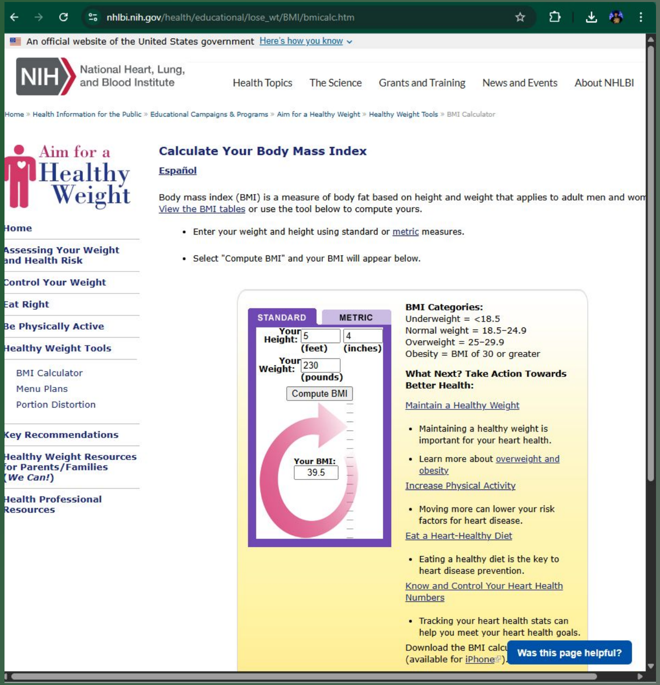
A BMI between 18.5 and 24.9 is considered healthy.

The healthiest weight range for someone your height would be 108 - 145 lbs. But losing even a little weight can be good for your health.

BMI doesn't directly measure body fat or diagnose health conditions. And BMI can sometimes be inaccurate for people with high or low muscle mass. Your doctor can help you understand your health risks and options for a healthy weight management plan.

Existing Approaches and Limitations

- A BMI calculator uses a mathematical formula that takes into account an individual's weight and height to estimate body fat. It's a simple ratio designed to categorize adults as underweight, normal weight, overweight, or obese.
- The standard formula for calculating BMI is:
  BMI = height (in meters)/weight (in kilograms)^2
- By inputting your weight and height (using either metric or standard units), the calculator applies this formula to produce your BMI score. This score is then compared to established ranges to determine your weight category.



Code

```
[495] # Define obesity level categories based on BMI (World Health Organization)
      def categorize_bmi(bmi):
          if bmi < 18.5:
              return 'Underweight'
          elif 18.5 <= bmi < 25:
              return 'Normal weight'
          elif 25 <= bmi < 30:
              return 'Overweight'
          elif 30 <= bmi < 35:
              return 'Obese Class I'
          elif 35 <= bmi < 40:
              return 'Obese Class II'
          else:
              return 'Obese Class III'

      df['Obesity_Category_Calculated'] = df['BMI'].apply(categorize_bmi)
```

```
def get_tips(bmi_category):
    """
    Provides lifestyle tips based on BMI category.

    Args:
        bmi_category (str): BMI category.

    Returns:
        str: Tips for the given BMI category.
    """
    tips = {
        "Underweight": """
            - Focus on nutrient-dense foods: Choose whole grains, lean proteins,
              healthy fats, fruits, and vegetables.
            - Increase calorie intake gradually: Aim for small, frequent meals
              instead of large ones.
            - Include strength training: Build muscle mass to gain weight
              healthily.
            - Consult a healthcare professional:  A doctor or registered dietitian
              can provide personalized advice.
            - Consider adding healthy snacks:  Nuts, seeds, avocados, and full-fat
              dairy can help increase calorie intake.
            - Monitor your progress: Track your weight and adjust your plan as needed.
            """
```

Code

```
[497] # Prepare data for the machine learning model focused on BMI categories
      #I did this because if not it would make me enter all of the questionnaire questions
      X_ml = df[['Gender', 'Height', 'Weight']]
      y_ml = df['Obesity_Category_Calculated']

[498] # Identify categorical and numerical features for the model
      categorical_features_ml = ['Gender']
      numerical_features_ml = ['Height', 'Weight']

[499] # Create a preprocessor for the machine learning model
      preprocessor_ml = ColumnTransformer(
          transformers=[
              ('num', StandardScaler(), numerical_features_ml),
              ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_features_ml)
          ])

[500] # Create a preprocessor for the machine learning model
      preprocessor_ml = ColumnTransformer(
          transformers = [
              ('num', StandardScaler(), numerical_features_ml),
              ('cat', OneHotEncoder(handle_unknown = 'ignore'), categorical_features_ml)
          ])

[501] # I'm retraining the data because when I tried to just use the old training
      # model I kept getting errors about the categories Split data into training and
      # testing sets
      X_train_ml, X_test_ml, y_train_ml, y_test_ml = train_test_split(X_ml, y_ml, test_size = 0.2, random_state = 42)

[502] # Create a pipeline with a classification model (Logistic Regression as a simple starting point)
      pipeline_ml = Pipeline(steps = [('preprocessor', preprocessor_ml),
                                      ('classifier', LogisticRegression(random_state = 42))])

[503] # Train the model
      pipeline_ml.fit(X_train_ml, y_train_ml)
```

Code

```
# Get user input
print("Please enter your details to determine your BMI category:")
gender_user = input("Gender (Male/Female): ").capitalize()
try:
    height_user = float(input("Height (in meters): "))
    weight_user = float(input("Weight (in kilograms): "))
except ValueError:
    print("Invalid input for height or weight. Please enter numeric values.")
    exit()
```

```
Please enter your details to determine your BMI category:
Gender (Male/Female): Male
Height (in meters): 1.63
Weight (in kilograms): 104
```

```
[549] # Create a DataFrame for the user input
      user_data = pd.DataFrame({'Gender': [gender_user], 'Height': [height_user], 'Weight': [weight_user]})
```

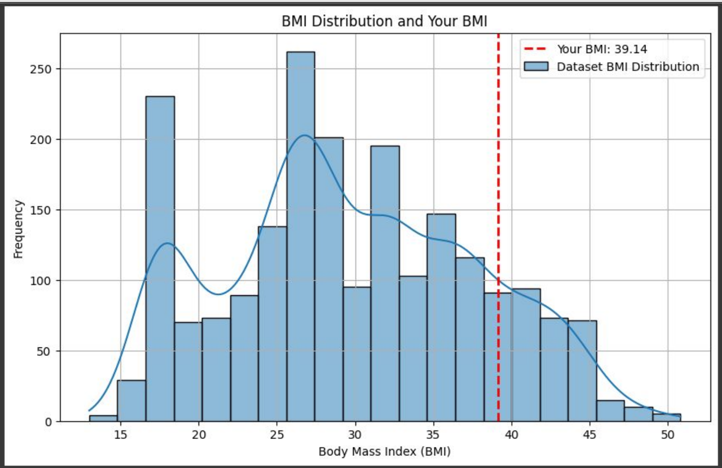
```
[550] # Predict the BMI category for the user
      predicted_category_user = pipeline_ml.predict(user_data)[0]
      print(f"\nBased on your input, your BMI category is predicted to be: {predicted_category_user}")
```

```
Based on your input, your BMI category is predicted to be: Obese Class II
```

```
[551] # Calculate user's BMI for visualization
      bmi_user = weight_user / (height_user ** 2)
```

```
[552] # Visualization: Compare user's BMI to the distribution of BMIs in the dataset
      plt.figure(figsize = (10, 6))
      sns.histplot(df['BMI'], kde=True, label = 'Dataset BMI Distribution')
      plt.axvline(bmi_user, color = 'red', linestyle = 'dashed', linewidth = 2, label = f'Your BMI: {bmi_user:.2f}')
      plt.title('BMI Distribution and Your BMI')
      plt.xlabel('Body Mass Index (BMI)')
      plt.ylabel('Frequency')
      plt.legend()
      plt.grid(True)
      plt.show()
```

Code

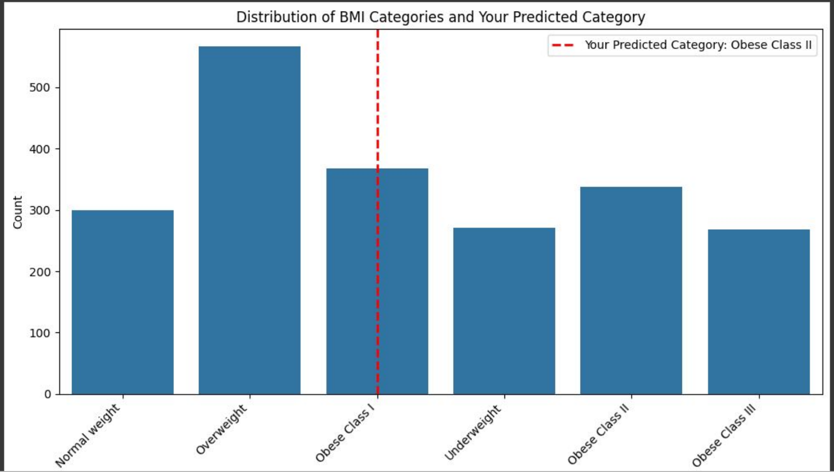


[553]

```
# Visualization: Compare user's BMI category to the distribution of BMI categories in the dataset
plt.figure(figsize = (10, 6))
sns.countplot(data = df, x = 'Obesity_Category_Calculated', order = df['Obesity_Category_Calculated'].value_counts().index, palette = 'viridis')
plt.axvline(x = df['Obesity_Category_Calculated'].value_counts().index.get_loc(predicted_category_user), color = 'red', linestyle = 'dashed', linewidth = 2, label = f'Your Predicted Category: {predicted_category_user}')
plt.title('Distribution of BMI Categories and Your Predicted Category')
plt.xlabel('BMI Category')
plt.ylabel('Count')
plt.xticks(rotation = 45, ha = 'right')
plt.legend()
plt.tight_layout()
plt.show()
```

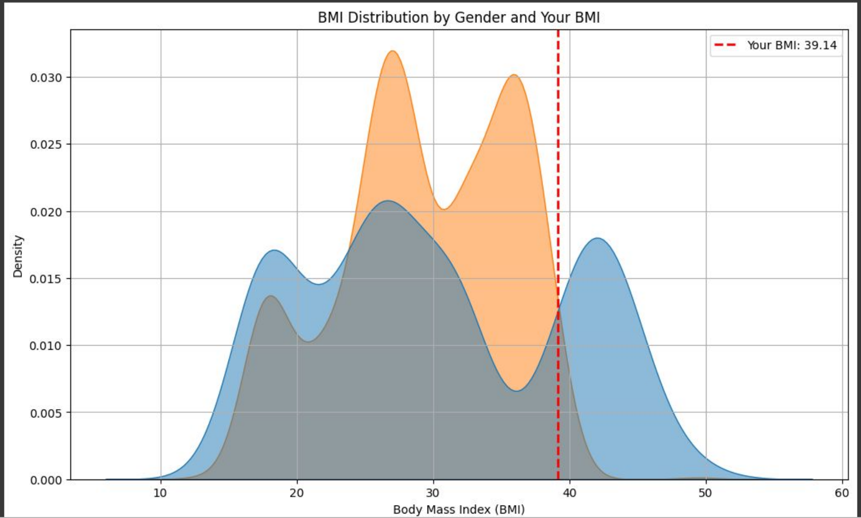
Code

```
# Visualization: Compare user's BMI category to the distribution of BMI categories in the dataset
plt.figure(figsize = (10, 6))
sns.countplot(data = df, x = 'Obesity_Category_Calculated')
plt.axvline(x = df['Obesity_Category_Calculated'].value_counts().index.get_loc(predicted_category_user), color = 'red', linestyle = 'dashed', linewidth = 2, label = f'Your Predicted Category: {predicted_category_user}')
plt.title('Distribution of BMI Categories and Your Predicted Category')
plt.xlabel('BMI Category')
plt.ylabel('Count')
plt.xticks(rotation = 45, ha = 'right')
plt.legend()
plt.tight_layout()
plt.show()
```



Code

```
# Visualization: BMI distribution by Gender with user's BMI highlighted
plt.figure(figsize = (12, 7))
sns.kdeplot(data = df, x = 'BMI', hue = 'Gender', fill = True, alpha = 0.5)
plt.axvline(bmi_user, color = 'red', linestyle = '--', linewidth=2, label = f'Your BMI: {bmi_user:.2f}')
plt.title('BMI Distribution by Gender and Your BMI')
plt.xlabel('Body Mass Index (BMI)')
plt.ylabel('Density')
plt.legend()
plt.grid(True)
plt.show()
```



Code

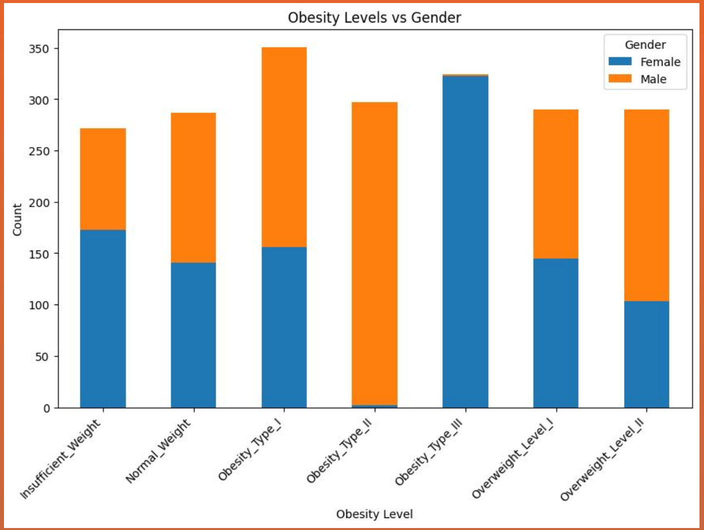
```
[555] tips = get_tips(predicted_category_user)
      print(f"Tips for BMI: {predicted_category_user} ")
      print(tips)
```

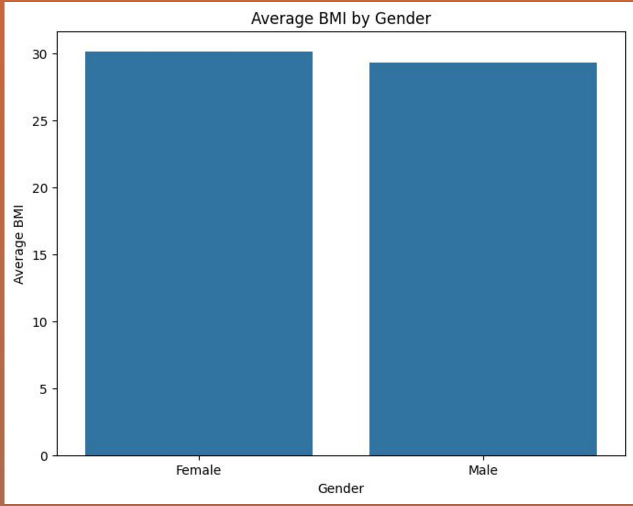
```
Tips for BMI: Obese Class II

            - Adopt a comprehensive approach: Combine intensive lifestyle
              modifications with medical supervision.
            - Work with a multidisciplinary team:  Include a physician,
              dietitian, psychologist, and exercise specialist.
            - Consider medical interventions:  Discuss options like weight loss
              medications or bariatric surgery with your doctor.
            - Address underlying issues:  Explore any emotional or psychological
              factors that may be contributing to obesity.
            - Focus on long-term management:  Obesity is a chronic condition that
              requires ongoing care.
            -  Get support:  Join a support group to connect with others facing
               similar challenges.
```

Results and Observations

Observations

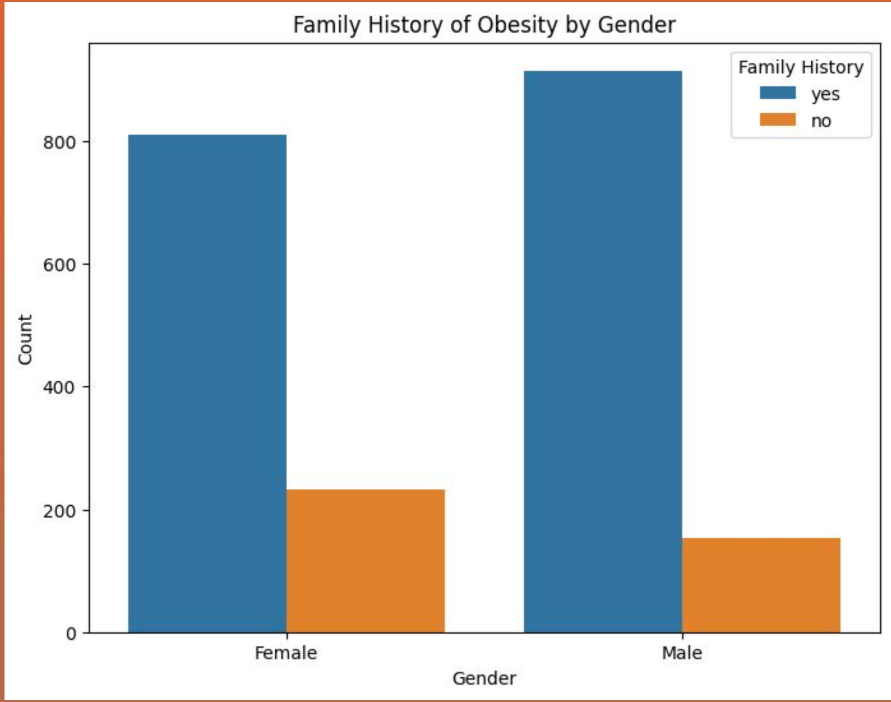
- BMI between Males and Females
- The first image shows the distribution of males and females across different obesity levels, while the second image indicates that females have a slightly higher average BMI than males.
- It's possible that the slightly higher average BMI in females (as seen in the second image) contributes to the observed distribution of females in the various obesity level categories shown in the first image.





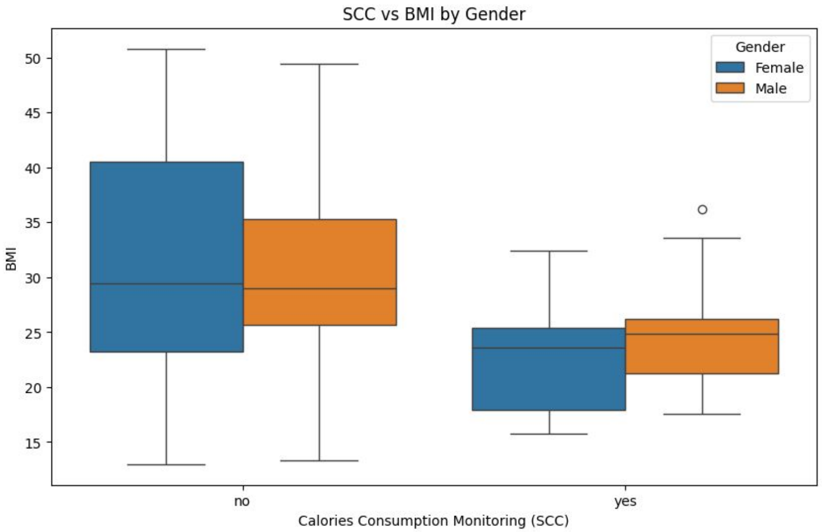
Observations

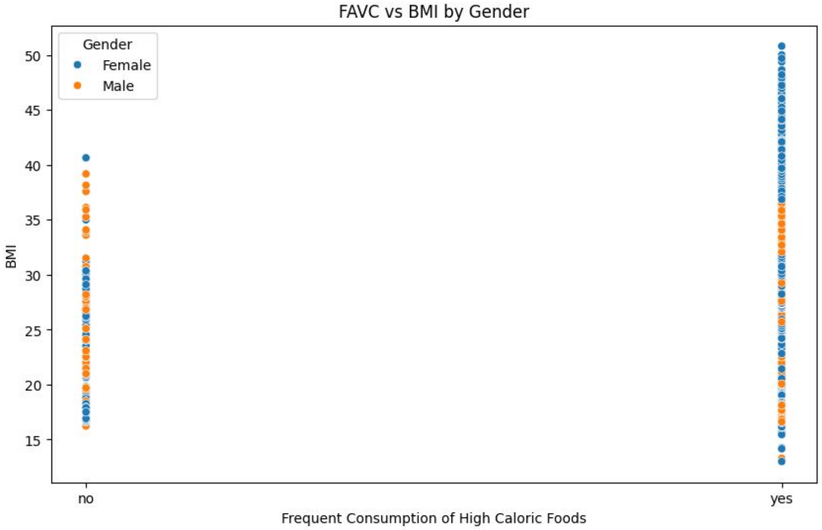
- Within both the female and male groups, a larger number of individuals report having a family history of obesity compared to those who do not.
- The proportion of individuals with a family history of obesity appears to be somewhat higher among males than females in this dataset.



Observations

- Individuals who report monitoring their calorie consumption ("yes") tend to have lower BMIs on average compared to those who do not ("no"), for both males and females.
- Among those who do not monitor their calorie consumption ("no"), females appear to have a slightly higher median BMI and a wider range of BMI values compared to males.
- Even among those who do monitor their calorie consumption ("yes"), there's still a considerable spread in BMI values for both genders, suggesting that calorie monitoring is one factor among others influencing BMI.

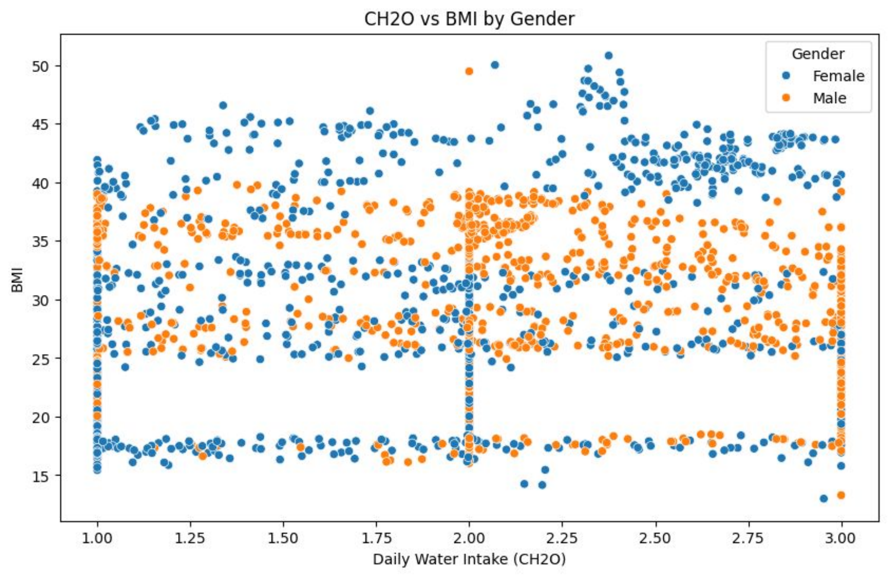




Observations

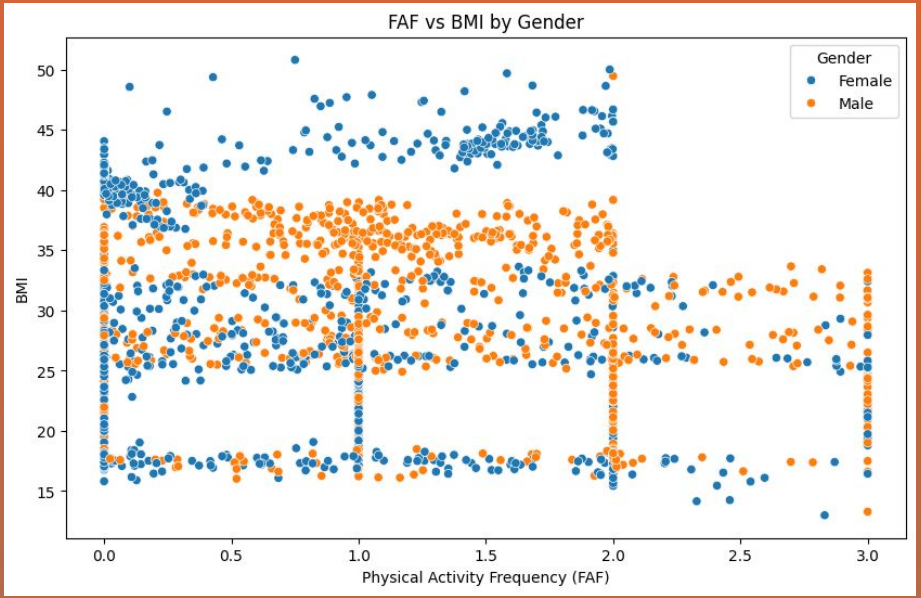
- There doesn't appear to be a strong linear correlation between daily water intake (CH2O) and BMI for either males or females in this data. Individuals across a wide range of BMIs are observed at various levels of water intake.

- There might be a slight visual tendency for individuals with very low BMIs (around 15-20) to be clustered at lower and higher ends of the water intake spectrum, while those with higher BMIs are spread across all levels of reported water intake. However, this observation needs further statistical analysis to confirm.



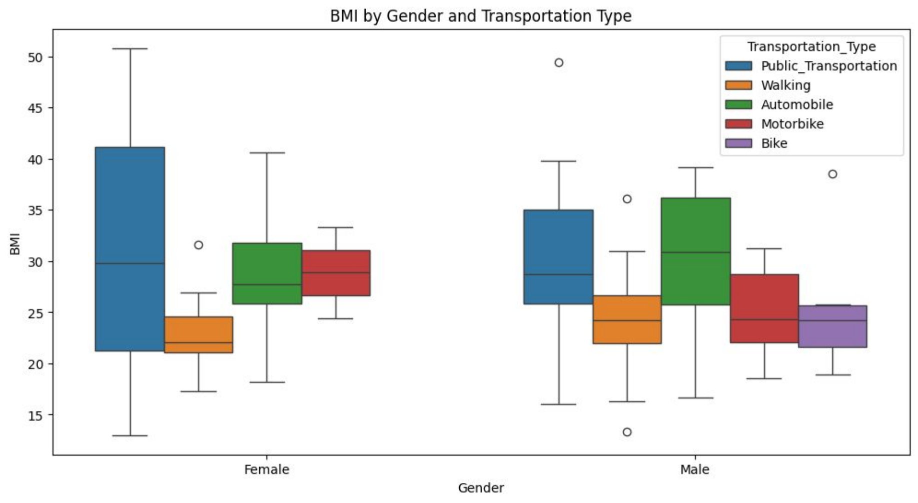
Observations

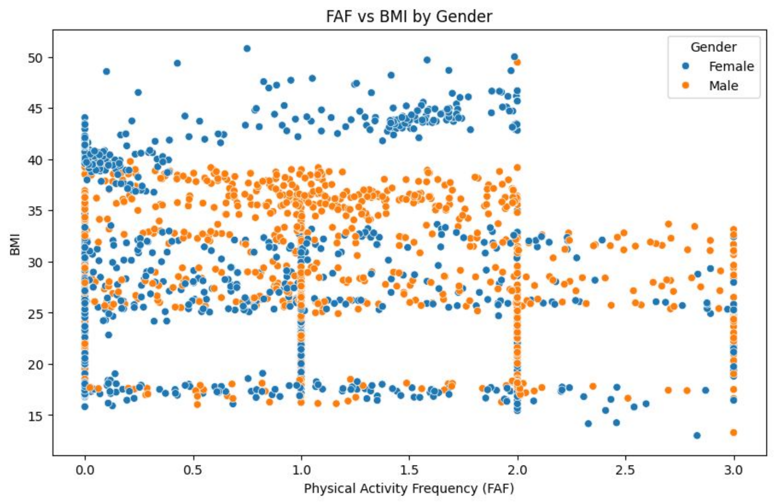
- Generally, as Physical Activity Frequency (FAF) increases, there appears to be a trend toward lower BMIs for both males and females, although the relationship isn't strictly linear.
- At lower levels of physical activity, females seem to exhibit a wider range of BMIs, including some higher values, compared to males.
- There are clusters of data points suggesting that even with similar levels of physical activity, females might have, on average, slightly higher BMIs than males in this dataset, particularly in the mid-range of FAF.



Observations

- For both genders, individuals who primarily use active transportation methods like walking or biking tend to have lower median BMIs compared to those who mainly use passive methods like automobiles.
- Females using public transportation appear to have a higher median BMI compared to females using other transportation types, while males using public transportation show a more varied range of BMIs.
- Among those who use automobiles, males tend to have a slightly higher median BMI compared to females in this dataset.





Results

- BMI and Gender: Females in this dataset tend to have a slightly higher average BMI compared to males.
- Obesity Levels and Gender: There are different distributions of males and females across various obesity levels.
- Physical Activity and BMI: Generally, higher physical activity frequency seems to be associated with lower BMIs for both genders.
- Calorie Monitoring and BMI: Individuals who monitor their calorie intake tend to have lower BMIs, regardless of gender.
- High Caloric Food Consumption and BMI: Frequent consumption of high-caloric foods appears to be linked to higher BMIs for both males and females.
- Water Intake and BMI: There isn't a clear linear relationship between reported daily water intake and BMI.
- Transportation Type and BMI: Active transportation methods (walking, biking) are generally associated with lower BMIs in both genders.
- Family History of Obesity: A significant portion of both males and females report a family history of obesity, with a slightly higher proportion among males in this specific dataset.

Conclusion and Future Work

Conclusion

This project highlights the complex interplay between BMI, gender, and various lifestyle factors, underscoring the need for a more nuanced approach to understanding and addressing weight-related health. By examining these relationships, we can identify key areas for targeted interventions and further research.

Future Work

- Machine learning algorithms could be employed to predict individual BMI trajectories based on a wider range of variables, including genetic predispositions, detailed dietary habits, and longitudinal physical activity patterns.
- Advanced techniques like deep learning can be used to identify complex, non-linear relationships between various factors and BMI, potentially uncovering novel insights that traditional statistical methods might miss.
- Personalized interventions could be developed using machine learning to tailor recommendations for weight management based on an individual's unique risk profile and predicted BMI trajectory.

References

Estimation of Obesity Levels Based On Eating Habits and Physical Condition [Dataset]. (2019). UCI Machine Learning Repository. https://doi.org/10.24432/C5H31Z.

Prevalence of Obesity and Severe Obesity Among Adults: United States, 2017–2018 Authors: Centers for Disease Control and Prevention (CDC) Year: 2020 Journal/Conference: National Center for Health Statistics Data Brief, No. 508 https://www.cdc.gov/nchs/products/databriefs/db508.htm

Overweight & Obesity Statistics Authors: National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Year: 2022 Journal/Conference: National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity

THANK YOU