Obesity Prediction Dashboard Report

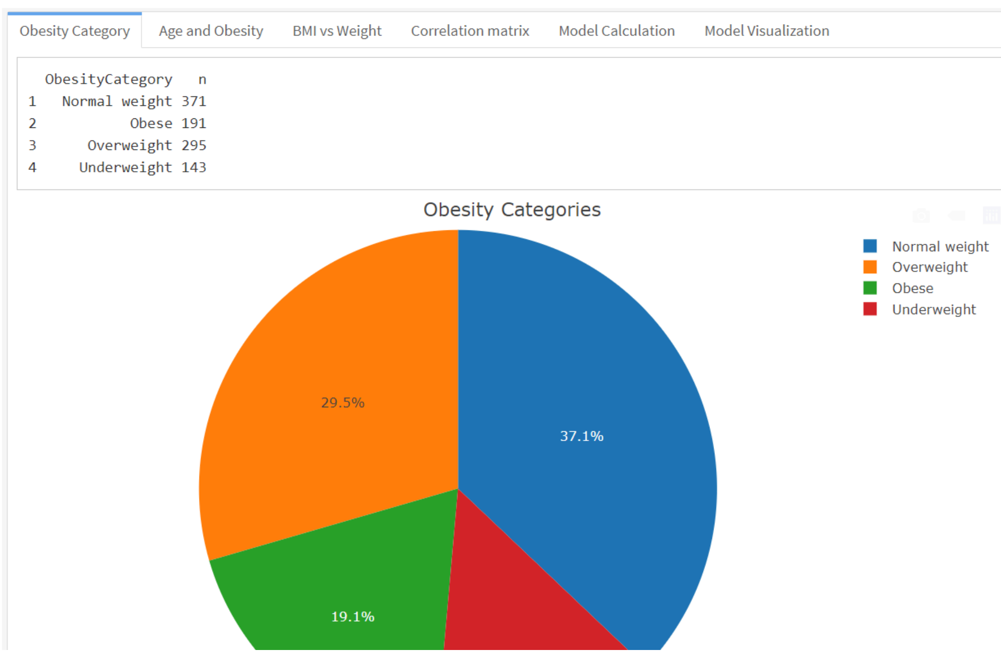
Submitted by:
Disha Patel
10A
MBA(BA)
Indian Institute of Foreign Trade, New Delhi
2023-25

Objective and introduction:

This report aims to provide insights into obesity trends and build a model to predict Body Mass Index (BMI) based on weight and height using a dataset of 1000 individuals. This information can be valuable for healthcare professionals, policymakers, and individuals seeking to understand and manage their weight.

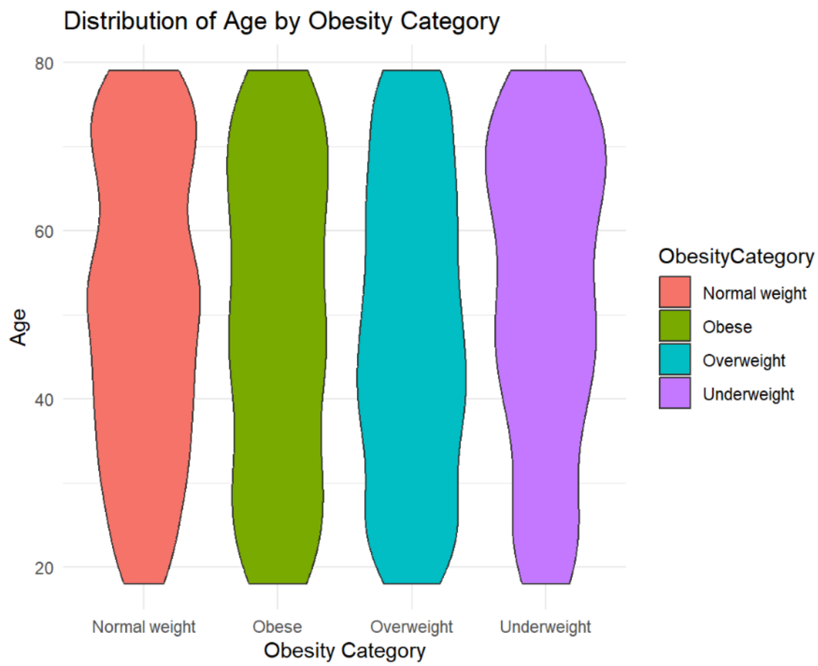
Analysis:

Obesity Prevalence: The initial analysis reveals the distribution of individuals across different obesity categories. A visualization (pie chart) can effectively showcase this distribution, highlighting the prevalence of each category.

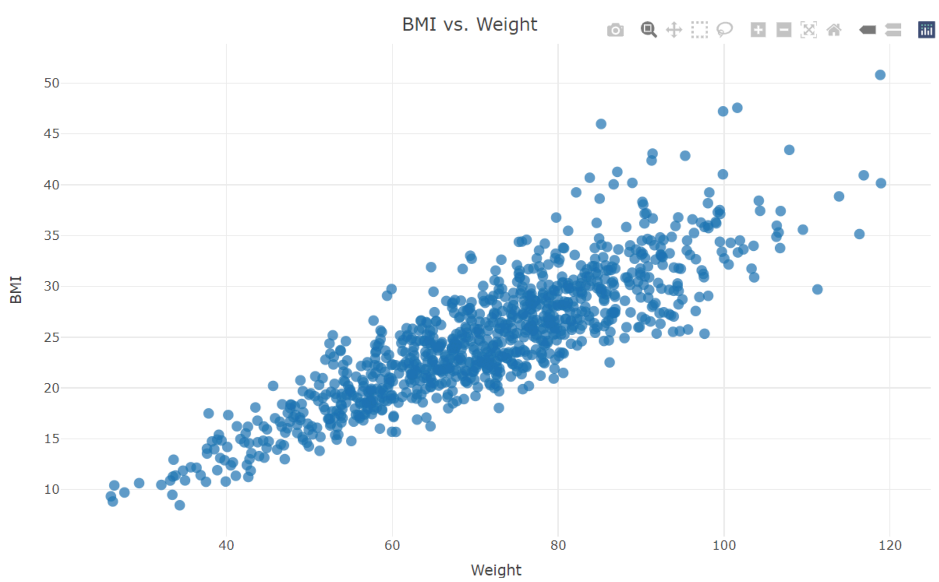


The pie chart visually shows the broad categories the subjects belong to. From the given pie chart 37.1% of the people were normal weight, followed by 29.5% who were overweight with 19.1% of people were obese.

Age and Obesity: A violin plot can illustrate the relationship between age and obesity category. This can reveal potential age groups with higher obesity risk.



Weight and BMI: A scatter plot can explore the relationship between weight and BMI, potentially identifying trends and outliers.

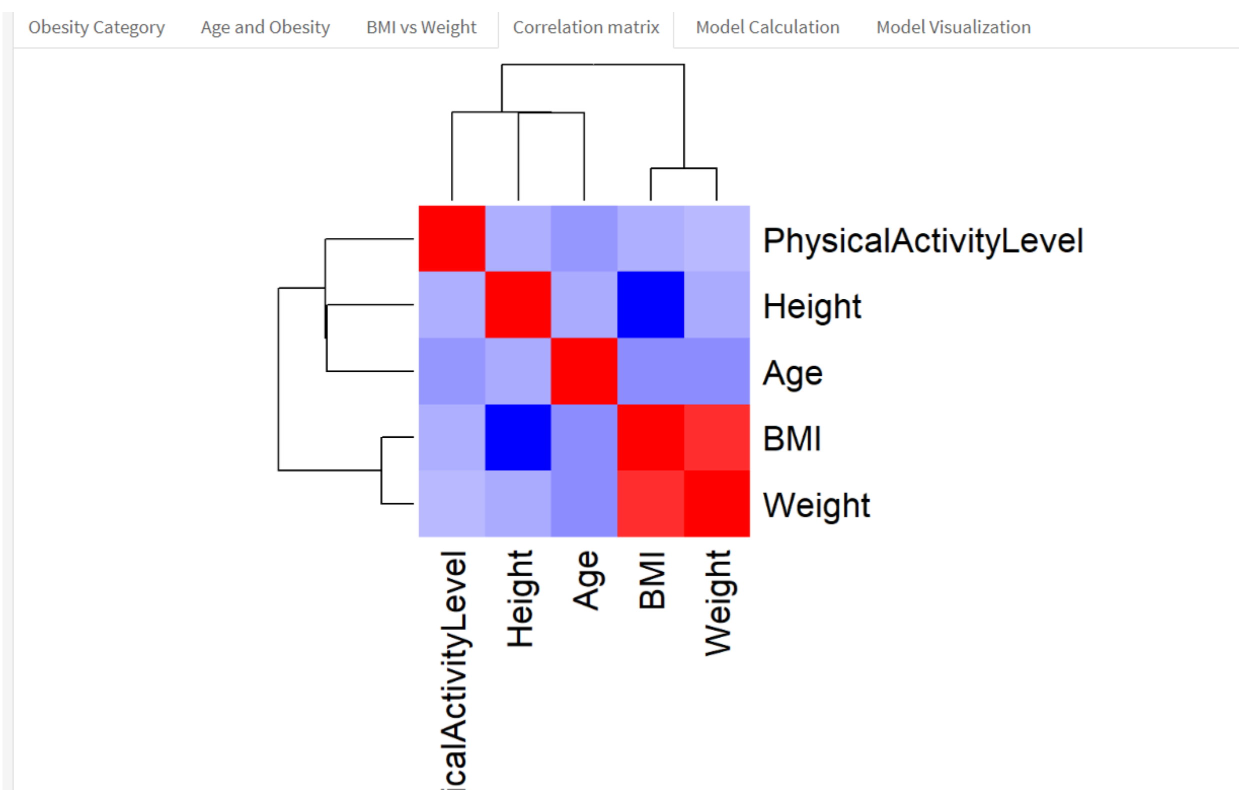


Positive Correlation: The plot shows a positive correlation between weight and BMI. This means that as weight increases, BMI tends to increase as well. This aligns with understanding how BMI is calculated, as it directly incorporates weight and height.

The outliers you mentioned could be due to various reasons, such as:

1. **Measurement errors**: Inaccuracies in measuring weight or height can lead to data points falling outside the expected range.
2. **Extreme cases:** Individuals with very high or low muscle mass, or those with conditions affecting body composition, might deviate from the typical trend.

Variable Correlations: A heatmap can visualize the correlations between age, weight, height, BMI, and physical activity level. This can uncover potential associations and guide further investigation.



- The matrix is square and symmetrical, indicating that the correlation between each pair of variables is listed twice (e.g., the correlation between Age and BMI is the same as the correlation between BMI and Age).
- The diagonal elements are all 1.0, which is expected as a variable perfectly correlates with itself.

- The color gradient ranges from blue (negative correlation) to red (positive correlation), with white indicating no correlation.
- The intensity of the color reflects the strength of the correlation.

Specific Correlations:

- There is a strong positive correlation between BMI and Weight (red), indicating that as weight increases, BMI also tends to increase.
- There is a moderate positive correlation between Age and Weight (orange), suggesting that older individuals tend to weigh more on average.
- There is a weak negative correlation between Age and Height (light blue), implying that older individuals may be slightly shorter on average.
- There is a weak positive correlation between Height and BMI (light orange), suggesting that taller individuals may have slightly higher BMIs on average.

Predicting BMI: A linear regression model is trained to predict BMI based on weight and height. The model's performance is evaluated using metrics like R-squared and RMSE. The coefficients of the model explain the impact of each variable (weight and height) on BMI prediction.

Key Findings:

- **High Model Accuracy:** The model explains a significant proportion of the variance in BMI, with an **adjusted R-squared of 0.9822**. This indicates that weight and height have a strong influence on BMI and the model can accurately predict it based on these factors.
- **Weight and BMI Relationship:** Each unit increase in weight is associated with an average increase of 0.347 units in BMI, holding height constant. This highlights the significant positive impact of weight on BMI.
- **Height and BMI Relationship:** Each unit increase in height leads to an average decrease of 0.297 units in BMI, holding weight constant. This suggests that taller individuals tend to have lower BMIs compared to shorter individuals with the same weight, likely due to a larger frame size distributing weight more evenly.
- **Model Limitations:** It's important to remember that the model is based on a specific dataset and may not generalize perfectly to other populations. Additionally, it doesn't account for other factors that can influence BMI, such as diet, physical activity, and genetics.

Obesity Category | Age and Obesity | BMI vs Weight | Correlation matrix | Model Calculation | Model Visualization

```
Call:
lm(formula = BMI ~ Weight + Height, data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.1319 -0.3752 -0.1873  0.1692  6.1484 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 50.654948   0.497732   101.8   <2e-16 ***
Weight       0.346935   0.001895   183.1   <2e-16 ***
Height      -0.296681   0.002819  -105.3   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8428 on 797 degrees of freedom
Multiple R-squared:  0.9823,	Adjusted R-squared:  0.9822 
F-statistic: 2.207e+04 on 2 and 797 DF,  p-value: < 2.2e-16
```

Obesity Category | Age and Obesity | BMI vs Weight | Correlation matrix | Model Calculation | Model Visualization