Interpretation Of R

Rakshitha Ramachandra K00302101

## Problem Statement

This project studies the relationship between diet and exercise and demographic information of age, weight, height, and gender in terms of their effect on health metrics such as BMI and activity levels. Through this process of addressing inconsistencies, handling missing data, and feature engineering relevant features, it will shed light on insightful insights and patterns that may drive improvement in health and wellness outcomes.

## Solution Summary

1. Data Cleaning and feature engineering : Diet type and exercise type were normalized by standardizing values, removing inconsistencies, and ensuring readability. Missing values in exercise frequency, weight, and height were imputed using contextual rules, including 0.0 for “No Exercise” and age group-based means for height and weight. The calculation of BMI and its categorization into “Low,” “Medium,” and “High” using percentiles was implemented to address skewness and enable meaningful comparisons. Additionally, diet types were grouped into “Plant-Based” and “Non-Plant-Based” categories to simplify analysis and highlight dietary trends.
2. Exploratory Visualization : Extensive visualizations were conducted on the dataset to identify insightful patterns and relationships. Key columns such as type of diet, type of exercise, and BMI were visualized using bar charts, pie charts, and scatterplots. New columns were engineered, including the calculation of BMI and BMI percentile categories for deeper insights into health and activity trends. Data cleaning and imputation were performed to ensure consistency and accuracy for in-depth explorations into demographics, diet preferences, exercise habits, and their interrelations. These visualizations provided valuable perspectives on the dataset structure and key health metrics.
3. Conclusions highlighting the insights you have gained : Exploratory visualizations provided extensive insight into the dataset, showing the interaction between diet and exercise with variations in BMI, considering demographic factors. Conclusion Insights derived from these analyses are outlined herein-present meaningful patterns and trends that address recommendations for action.

## Data Cleaning

The diet\_type column was cleaned to normalize the values to ensure consistency with the removal of special characters, trimming of extra spaces, and converting into title case. Normalization in data means regularizing it in a standard format to make them ready for analysis.

diet\_type  
1 Omnivore  
2 Vegan  
3 Keto  
4 Paleo  
5 Vegetarian

The gender column was standardized to ensure consistency by converting all values to title case and mapping variations like “M” and “Male” to “Male” and “F” and “Female” to “Female.” Additionally, unexpected or missing values were replaced with “Unknown.” This standardization ensures uniformity in the dataset, making it ready for accurate analysis.

gender  
1 Female  
2 Male

The exercise\_type column was cleaned-replacing “None” with “No Exercise” for readability, removing extra spaces, and standardizing text to title case for consistency and better interpretability of the values.

exercise\_type  
1 Pilates  
2 Strength  
3 No Exercise  
4 Cardio  
5 Yoga

The exercise\_frequency column was cleaned, replacing NA with 0.0 when the exercise type is “No Exercise” and with the column median for other NA values. This makes the column consistent without any missing data and thus ready for analysis.

exercise\_frequency  
1 0  
2 1

#### Feature Engineering: Age Grouping and Categorizing

The missing values in the weight and height columns were imputed using the mean values calculated within each age\_group. This group-based approach ensures that the replacements are contextually accurate, reflecting the typical weight and height for individuals in similar age ranges. This method improves data quality while maintaining meaningful variability.

weight height   
 Min. : 61.53 Min. :138.1   
 1st Qu.: 80.55 1st Qu.:163.7   
 Median : 85.04 Median :169.9   
 Mean : 85.00 Mean :169.8   
 3rd Qu.: 89.68 3rd Qu.:176.2   
 Max. :108.24 Max. :202.4

#### Feature Engineering: BMI Calculation and Adjustments

We used a percentile-based categorization of BMI to handle the skewness in the data, since the majority of the persons had a classification of “Overweight.” We divided BMI into “Low,” “Medium,” and “High” categories based on the 33rd and 66th percentiles, which gave us balanced groups. This step will enhance analysis by providing meaningful comparisons between the different BMI groups, as well as uncovering patterns in diet, exercise, and demographics.

High BMI Low BMI Medium BMI   
 340 330 330

#### Feature Engineering: Diet Category

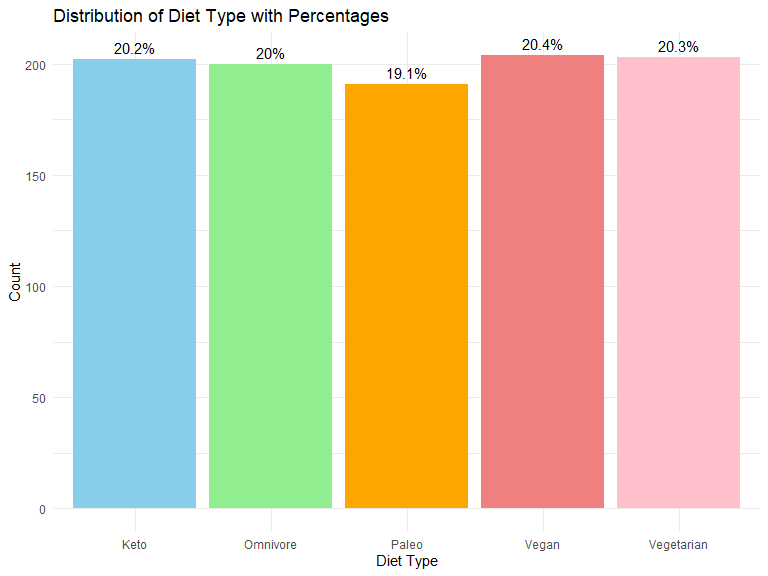
We simplified the diets into two categories: “Plant-Based” and “Non-Plant-Based.” For comparison clarity and to showcase dietary trends, “Plant-Based” will consist of Vegan and Vegetarian diets, while “Non-Plant-Based” includes Omnivore, Keto, and Paleo diets. This way, comparisons between these diet types and health metrics, exercise, and demographics would make more sense.

# A tibble: 5 × 2  
 diet\_type diet\_category   
 <chr> <chr>   
1 Omnivore Non-Plant-Based  
2 Vegan Plant-Based   
3 Keto Non-Plant-Based  
4 Paleo Non-Plant-Based  
5 Vegetarian Plant-Based

## Exploratory visualizations

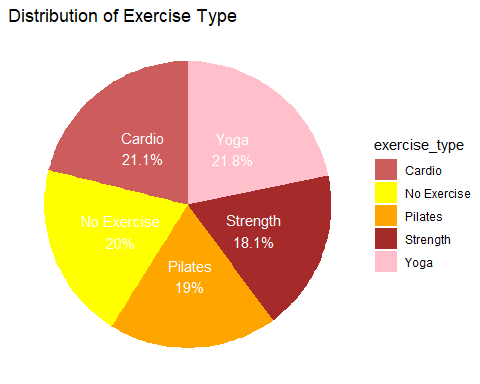
### Distribution of Diet Types in the Dataset

This pie chart shows the proportion of people following each diet type. Vegan is the most popular diet type at 20.4%, while Paleo is the least popular, with only 19.1% of people following it.



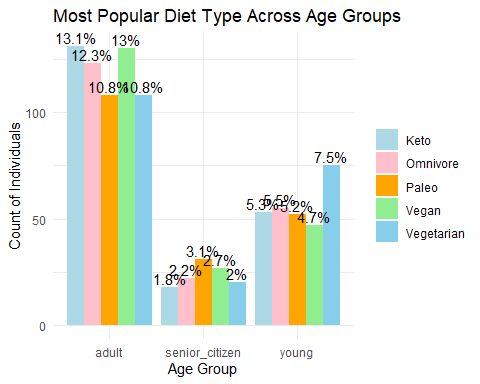
### Distribution of Excercise in the dataset

This pie illustrates how, out of all kinds, yoga is represented by 21.8%, followed closely by cardio at 21.1%, and in that order comes Pilates by 19%, Strength by 18.1%, and No Exercise by 20%. In this representation, it is somewhat equitably divided, though yoga and cardio are most preferred activities.



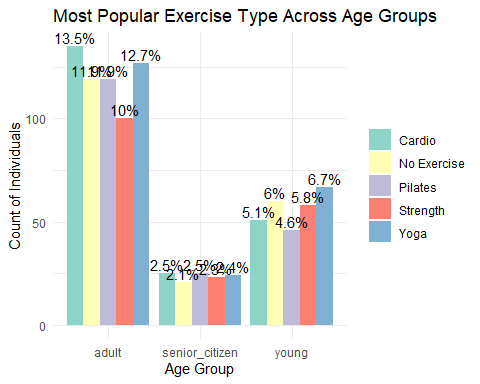
### Most Popular Diet Type Across Age Groups

This bar chart visualizes the distribution of diet types across different age groups, using the new age\_group variable created earlier. The chart reveals that among adults, Keto is the most popular diet type, while among senior citizens, Paleo dominates. For young individuals, Vegetarian is the most popular diet type. The legend shows the diet types, and the percentage labels indicate the proportion of each diet type within each age group.



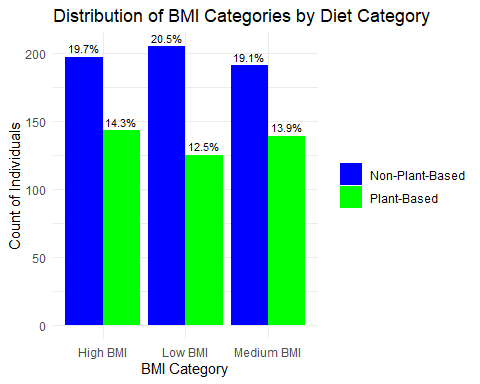
### Most Popular Exercise Across Age Groups

The bar chart below shows the distribution of exercise type by age group. It can be seen that Yoga is the most favourite among the young and adults, while Cardio also had a very strong participation in adults. In senior citizens, the majority had No Exercise, seconded by Yoga and Strength, respectively. The chart also includes labels showing the percentages of each type of exercise in each of the different age groups.



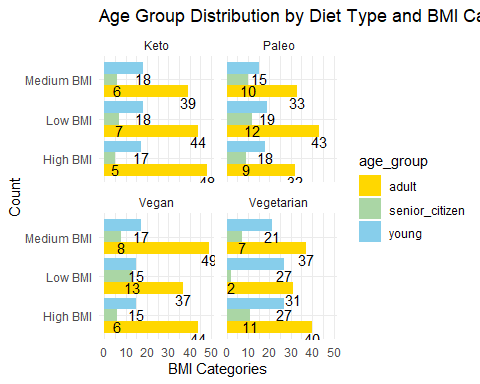
### Diet Category vs. BMI Category Distribution

The chart shows the distribution of BMI by diet type. The non-plant-based diets have higher proportions in high and medium categories of BMI, while plant-based diets have a little more in the low category of BMI. This reflects that plant-based diets are associated with lower levels of BMI.



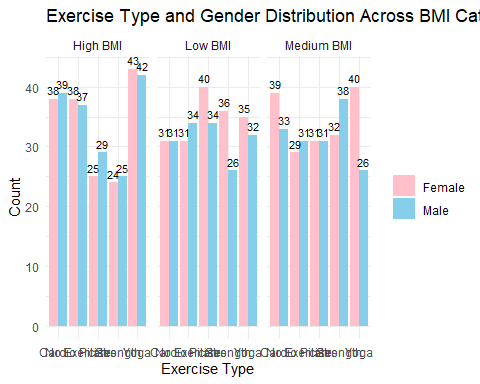
### Diet Type and Age Group by BMI Percentile

This facet-wrapped bar chart shows the distribution of age groups (young, adult, senior citizen) across BMI categories (Low, Medium, High) for different diet types (Keto, Paleo, Vegan, Vegetarian). Adults dominate all the categories of BMI throughout the diet types, especially in the High and Medium categories of BMI. Young individuals are more represented in Medium BMI, while senior citizens are the least represented. The chart shows how different the age group distributions are by diet and BMI.



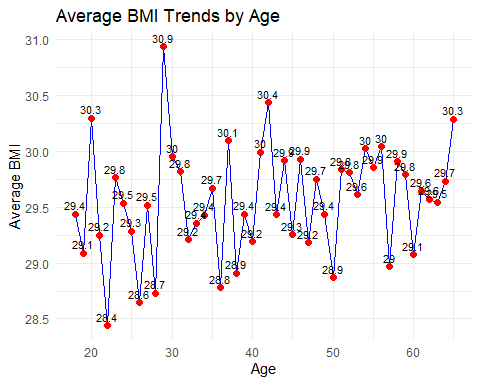
### **Exercise Type and Gender Distribution Across BMI Categories**

This bar chart represents the distribution of various exercise types, both male and female, in relation to the BMI categories (High, Low, and Medium). Each panel displays a BMI category, and the bars within a specific panel count the number of people for each exercise type, segregated by gender. Yoga records the highest number of participants across all BMI categories. It is mainly done by females with high BMI. The exercises of Cardio and No Exercise have been selected by the two genders quite evenly. Pilates and Strength have slightly lower participation overall. The chart brings out different preferences for exercise based on gender and BMI. It also demonstrates how the activities differ according to the demographics of the population.



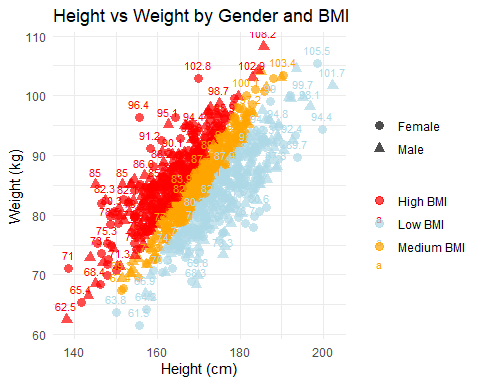
### Average BMI Trends by Age

The chart demonstrates the average BMI trends at different ages. It shows that the BMI levels keep changing, with some obvious peaks and troughs. Younger and middle-aged groups appear more varied, perhaps because of the lifestyle or physical activity they engage in. At the ages of 30 and 40, BMI spires, thus showing the possibility of weight gain in these two periods. After the age of 50, BMI remains stable, thus indicating that a certain weight pattern is the case with the aged. This visual representation of BMI change over time and the identification of the main points for possible health interventions are the core elements of the chart.



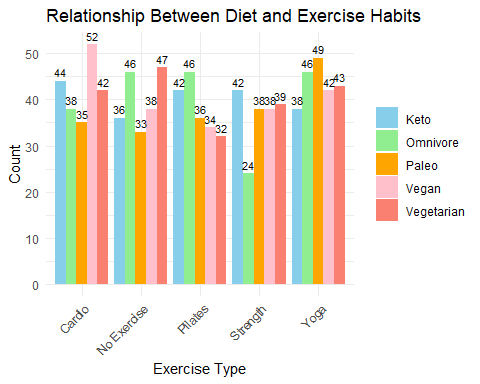
### Weight vs Height by Gender and BMI

The scatter plot is not only the visual representation of height and weight but also the association of these two variables categorized by gender and BMI levels. One point is a person whose color says BMI category (Low, Medium, or High) and shape shows gender (circle for females, triangle for males). The persons with a high BMI are clustered at higher weights, on the other hand, the latter category of BMI persons are grouped at the lower weight segment. BMI categories and gender are the specific issues in which the plot shows that the height-weight relationship is influenced.



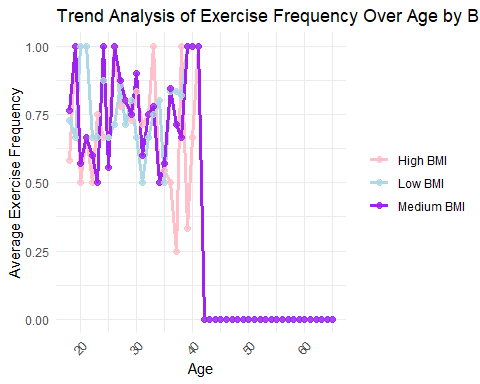
### Relationship Between Diet and Exercise Habits

The bar chart shows the link between the types of diets and the habits of doing exercise. Yoga is the favorite sports type among most of those who are vegan (49) and get in the diet of Keto (46). Regular involvement and participation of Cardio and No Exercise are the patterns that Vegan followers prefer, and (52) them show a higher degree of inclination to No Exercise. As far as strength training is concerned, it has a minimal number of participants in every diet, especially among Vegetarians (24). The graph presents differences in the preferences of a person for a sport that are related to their diet and show that in most diet categories Yoga and Pilates in most diet categories.



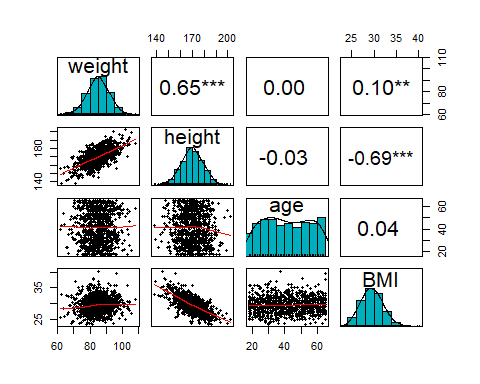
### Trend Analysis of Exercise Frequency Over Age

The line chart depicts the average levels of exercise in different age groups according to BMI (High, Medium, Low). Patients having a variable BMI of Low, Medium, and High, within the age range of 20-40 experience significantly deviant exercise patterns, Some of the low BMI people are engaging in very active exercises thus showing slightly better results. Outside the 40 age barrier, all the BMI groups experience physical activity regression which in the end comes as a steady lower level. The chart clearly shows the drop in exercise time with age and this could be due to the people not feeling like doing it any longer. Thus, older adults should be the main target for exercise programs.



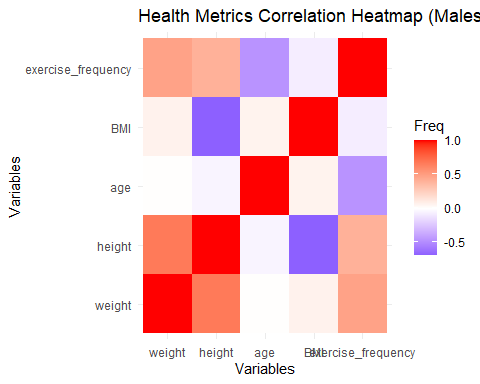
### Correlation Analysis and Relationships Between Health Metrics

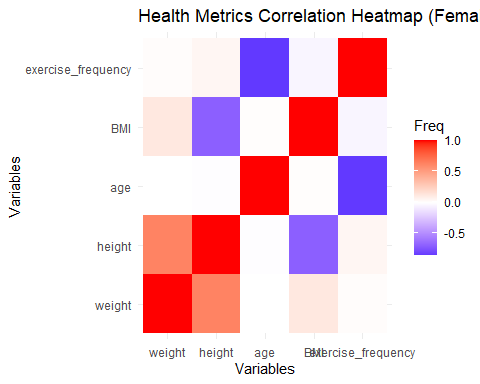
This correlation matrix provides a visual examination of the connections between weight, height, age, and BMI. The diagonal contains the histograms of each variable, which provides the information about the data distribution. The upper triangle part of the matrix displays the correlation coefficients, where the greater the values, the stronger the associations. For instance, there is a high positive correlation between weight and height which means that taller people are often heavier. On the other hand, BMI and height have a high negative correlation, meaning that taller people tend to have relatively low BMI. The lower triangle includes scatter plots accompanied by trend lines, which further reveal the relationships between the variables. In conclusion, this matrix shows the particular interactions between these health metrics.



### Gender-Based Heatmap for Health Metrics

The heatmaps highlight correlations between health metrics for males and females. All of them exhibit an evident positive attachment between body mass index and height together with a negative one between BMI and height as well. In males, BMI has a more positive connection with age as opposed to females. In the female, exercise frequency slightly increases age. Therefore, these patterns ultimately show that there exist gender-specific differences in health metric interactions.





## Conclusion

The results from this analysis provide important insights into the connections between health parameters, diet preferences, exercise behaviors, and demographic characteristics:

1. Diet and Exercise Patterns: Regardless of the diet followed, yoga emerges as the most popular exercise, particularly among Vegan and Paleo diet followers, while strength training sees the least participation.
2. Body Mass Index (BMI) and Demographics: Over the past three years, adults have had the highest BMI representation, with medium and high BMI categories being the most prevalent. Young individuals, however, dominate the low BMI group. This variation suggests differences in lifestyle and metabolic factors across age groups.
3. Age and Activity: A decline in exercise frequency begins around age 40, highlighting a growing sedentary trend among middle-aged and older individuals. Targeted interventions for physical activity are necessary to address this issue and improve health outcomes in these populations.
4. Correlations Between BMI, Height, and Weight: BMI correlates negatively with height, indicating that taller individuals tend to have a lower BMI, while weight is positively associated with higher BMI levels, particularly in people engaging in less physical exercise.
5. Sex-Specific Trends: BMI varies significantly with age in males, whereas females show a positive relationship between exercise frequency and age. These differences point to the need for gender-specific wellness initiatives to maximize health benefits for each group.

The insights from this analysis underscore the importance of designing health programs tailored to individual dietary preferences, exercise patterns, and demographic attributes. Developing personalized health and wellness interventions can significantly contribute to improved health outcomes and enhanced quality of life for diverse populations.