**EXPLORING BMI FACTORS**

**Abstract**

Abstract This project investigates the potential of unconventional factors for predicting Body Mass Index (BMI) using predictive analytics. Due to the limitations of primary data collection (fewer than 50 responses), the project shifted focus to a secondary dataset sourced from Kaggle.

The main objective was to create a model that uses less-studied factors to predict BMI. The project utilized a diverse set of data analysis techniques, including:

1. **Predictive Modeling:** Building a model to predict BMI based on the dataset.
2. **Hypothesis Testing:** Conducting hypothesis tests to explore relationships within the data.
3. **Clustering Analysis:** Grouping variables to uncover patterns and using these clusters for deeper analysis.
4. **Principal Component Analysis (PCA):** Reducing dimensionality to emphasize the most critical variables.
5. **Scott-Knott Analysis:** Categorizing the data into distinct groups based on statistical significance.
6. **Factor Analysis:** Identifying underlying factors that affect BMI.

By concentrating on unconventional factors and employing a comprehensive data analysis strategy, this project offers a unique contribution to BMI prediction. This approach has the potential to enhance prediction accuracy by incorporating a broader range of variables that may reflect an individual's lifestyle and health habits.

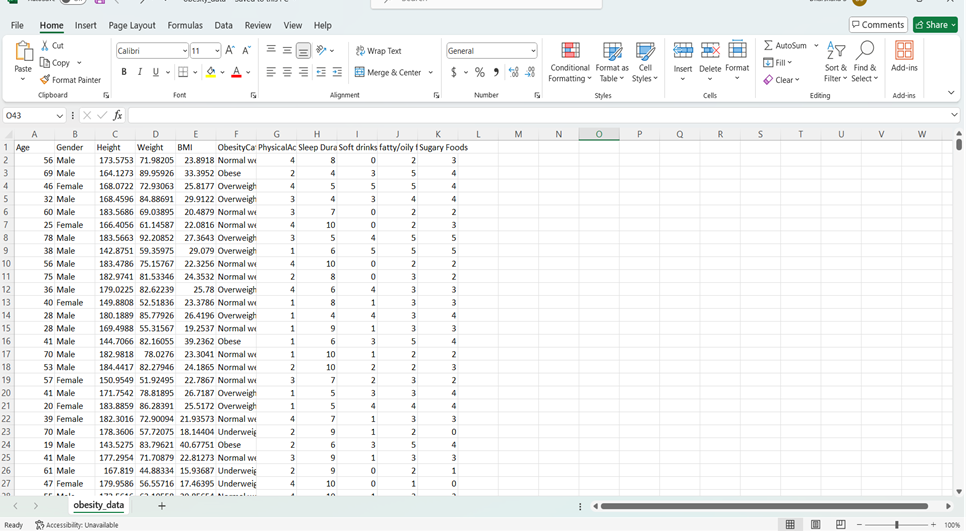
**DATASET:**

**Dependent variable:**

BMI (Body Mass Index) serves as our dependent variable, meaning it's influenced by other factors. BMI is calculated from an individual's height and weight and categorizes them as underweight, normal weight, overweight, or obese. The project considers changes in physical activity level, sleep duration, and dietary habits to affect BMI.

**Independent variables:**

* Physical activity level: This represents how active each person is, ranging from sedentary (low activity) to highly active. An individual's activity level can impact their BMI and overall health.
* Sleep duration: The amount of sleep each person gets on average per night. Sleep duration can influence weight management and metabolism, thereby affecting BMI.
* Soft drinks: The frequency of consuming sugary beverages. High consumption of soft drinks can contribute to weight gain and obesity.
* Fatty/oily foods: How often individuals consume foods high in fats and oils. Regular consumption of fatty foods can affect BMI and overall health.
* Sugary foods: The frequency of consuming foods high in sugar. Regular consumption of sugary foods can contribute to weight gain and influence BMI.



**REGRESSION:**

**Define:**

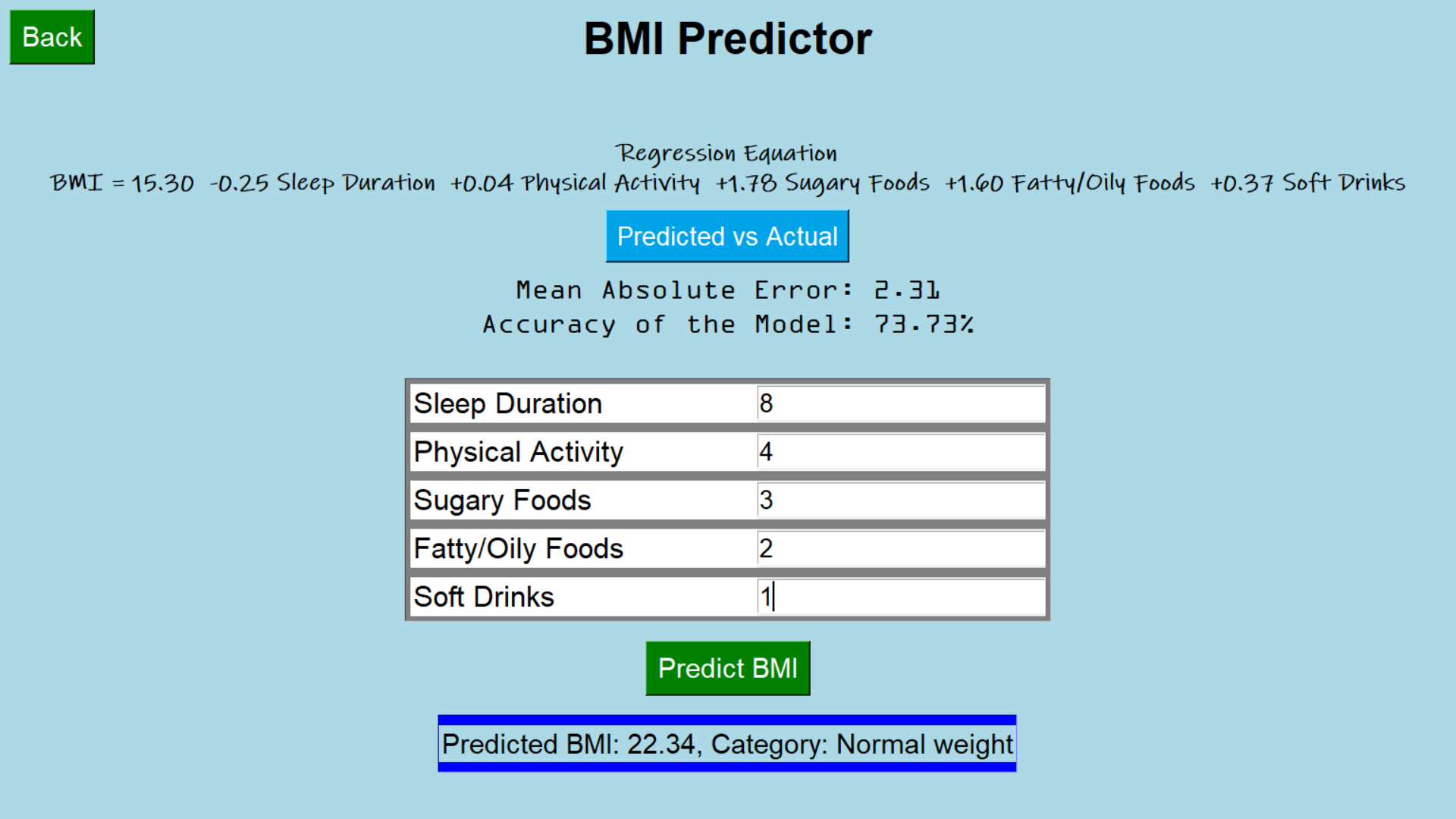
Regression is a statistical method used to analyze the relationship between one dependent variable and one or more independent variables. It aims to understand how the independent variables affect the dependent variable and to predict the value of the dependent variable based on the values of the independent variables.

**What we have done:**

Several functions were defined:

* predict\_bmi(entries, result\_label): This function predicts BMI based on user input and updates the result label with the prediction.
* Labels displaying Mean Absolute Error (MAE) and Model Accuracy.
* User input frame with entry fields for factors like sleep duration, physical activity, etc.
* Predict BMI button to trigger BMI prediction based on user input.
* Result frame to display the predicted BMI and category.
* Close button to exit the application.

#### **Screenshot of Output:**



**INFERENCE:**

This script builds a graphical user interface (GUI) application for predicting BMI (Body Mass Index) based on user input. Users can input their sleep duration, physical activity level, and consumption habits of sugary foods, fatty/oily foods, and soft drinks. The application utilizes a regression model trained on a dataset to make predictions. Additionally, it provides information about the regression equation used for prediction, the mean absolute error, and the accuracy of the model. Users can utilize this tool to get an estimate of their BMI category and assess their health status.

### **Factor Analysis:**

### **Define:**

#### Factor Analysis (FA) is a statistical method used to identify latent factors or constructs that explain correlations among observed variables. It aims to reduce the dimensionality of the data while retaining as much information as possible by uncovering the underlying structure or patterns within the variables.

#### **What we have done:**

#### **Data Preparation:**

#### Loaded the obesity-related data from a CSV file.

#### Encoded categorical variables and selected relevant features for analysis.

#### **Kaiser criterion:**

#### The Kaiser criterion suggests retaining factors with eigenvalues greater than 1. This guideline helps identify the optimal number of latent factors to extract from the data. Essentially, it prioritizes components that explain more variance than a single original variable would.

#### **Factor Analysis (FA):**

#### Applied FA to uncover the latent factors contributing to the observed variables.

#### Utilized the varimax rotation method to improve the interpretability of factor loadings by maximizing their variance.

#### **Visualization:**

#### Created a graphical user interface (GUI) with buttons to facilitate the visualization of results.

#### Implemented functionalities to display:

#### Scree plot showing eigenvalues to aid in determining the number of factors to retain.

#### Heatmap of factor loadings to visualize the relationships between variables and factors.

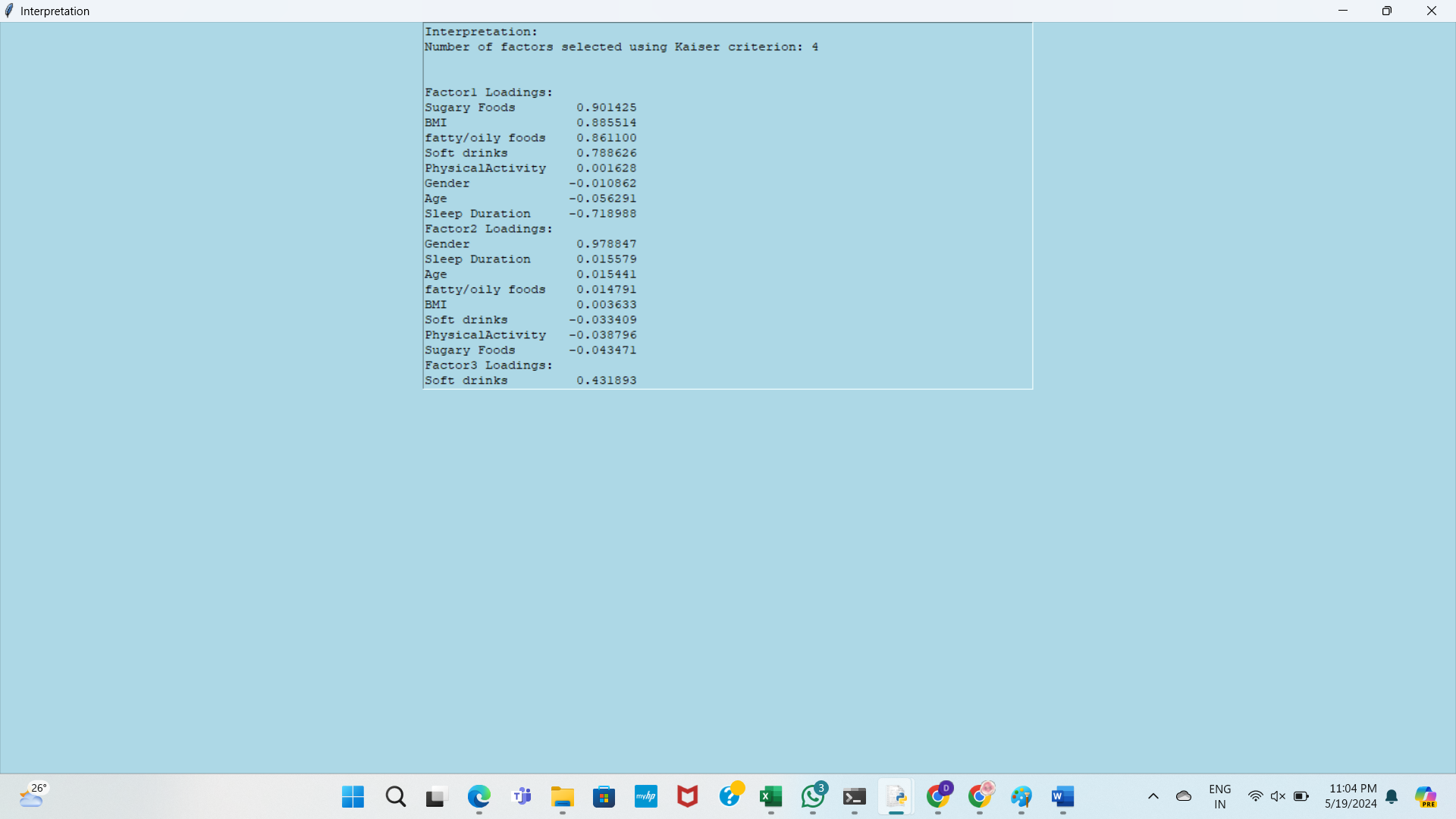
#### Interpretation of the analysis results, including the number of factors selected and factor loadings.

#### **Screenshot of Output:**

#### 



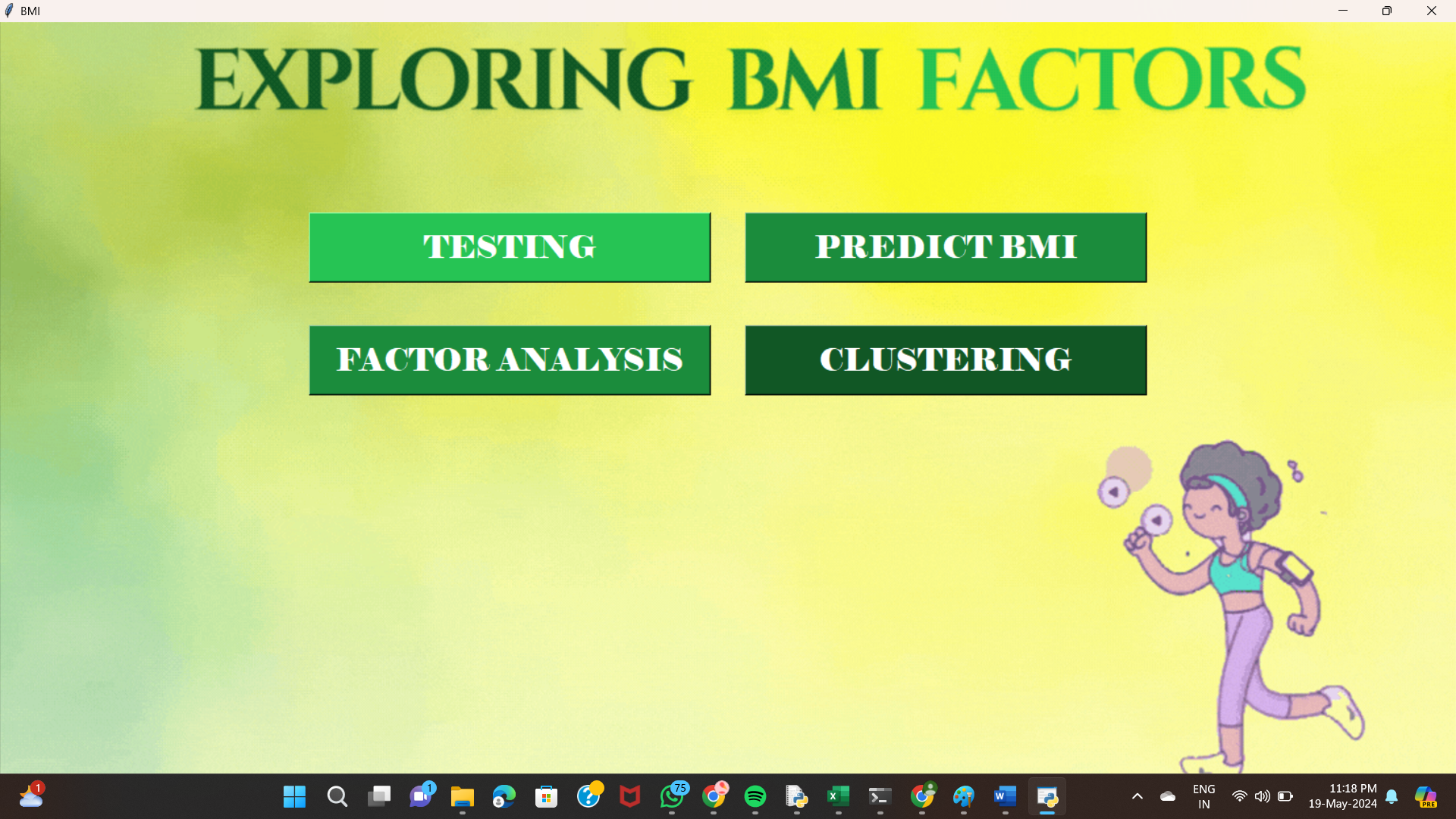




#### **INFERENCE:**

* Factor Analysis helps in uncovering the latent factors underlying the observed variables, providing insights into the underlying structure of the data.
* By examining factor loadings, we can understand how each variable contributes to the identified factors, facilitating interpretation and decision-making in obesity-related research or interventions.

**Gui Screenshots:**



****

****

