## Artificial Intelligence And

## Machine Learning

Project Report

Semester-IV (Batch-2022)

**FITPREDICT**

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Faculty Name : Tushar khitoliya Devanshi Dadwal (2210990262)

Dikshant Sohal (2210990283)

Dipankar Sethi (2210990287)

Bhumi Garg (2210990226)

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**INDEX**

|  |  |  |
| --- | --- | --- |
| Sr.No | Content | Page Number |
| 1 | Introduction | 1 |
| 2 | Problem Definition and Requirements | 2-3 |
| 3 | Proposed Design / Methodology | 4-8 |
| 4 | Result | 9 |

**Introduction :**

**Obesity and related health issues have become significant concerns worldwide, with implications for individual health and public healthcare systems. This project aims to analyze and understand factors contributing to obesity, explore body composition metrics, and predict obesity levels and exercise recommendation plans based on various features.**  In this project, we delve into the multifaceted aspects of obesity, aiming to Investigate the various factors that contribute to obesity, including lifestyle, genetics, and environmental influences. Explore dietary habits, physical activity levels, and socioeconomic factors that play a role in obesity prevalence. Examine body composition metrics such as body mass index (BMI), body fat percentage (BFP), and waist-to-hip ratio. Understand how these metrics correlate with obesity risk and overall health. Develop predictive models to estimate obesity levels based on relevant features. Generate exercise recommendation plans tailored to individual characteristics. Based on an individual’s characteristics (e.g., age, gender, BMI), create personalized exercise plans. Consider factors like aerobic exercises, strength training, and flexibility routines. Tailor recommendations to balance weight loss, muscle gain, and overall well-being.

**Problem Definition and Requirements :**

**Problem Statement:**

**Obesity is a significant health concern globally, with implications for individual health outcomes and healthcare systems. Despite its prevalence, understanding the complex interplay of factors contributing to obesity and developing effective interventions remains a challenge. This project aims to address this challenge by leveraging data analytics and predictive modeling techniques.**

**Software Requirements**:

1. Programming Language: Python will serve as the primary programming language for developing the project due to its extensive libraries for data manipulation and machine learning.

2. Machine Learning Libraries: Scikit-learn and Streamline will be utilized for implementing machine learning algorithms and building models for the project.

3. Data Processing Tools: Pandas and NumPy will be used for data preprocessing and manipulation tasks.

4. Model Evaluation: Scikit-learn will provide functions for evaluating the performance of the trained models using metrics such as accuracy, precision, recall, and F1-score.

**Hardware Requirements:**

The hardware requirements for developing and deploying the project are relatively modest:

1. Processor: A multi-core processor (e.g., Intel Core i5 or AMD Ryzen 5) to handle data processing and model training tasks efficiently.

2. Memory (RAM): At least 8GB of RAM to accommodate large datasets and facilitate model training without significant performance degradation.

3. Storage: Sufficient storage capacity (e.g., SSD or HDD) to store datasets, source code, and trained models.

**Data Sets:**

The dataset aims to estimate obesity levels in individuals from Mexico, Peru, and Colombia based on their eating habits and physical condition. The records are labeled with the class variable no obesity , allowing classification into categories such as Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, and Obesity Type III. This dataset provides body composition metrics for individuals. It includes features such as height, weight, body mass index (BMI), body fat percentage, and other related measurements.

Here are some considerations for selecting suitable datasets:

1. Size: The dataset should be sufficiently large to capture diverse patterns and behaviours .

2. Source: This dataset is available on Kaggle.

3. Features: The dataset should include relevant features.

4. Anonymity: Ensure that personally identifiable information (PII) is anonymized or removed from the dataset to comply with data privacy regulations.

5. Source: Utilize reputable sources for obtaining the datasets, such as Kaggle, UCI Machine Learning Repository (with proper permissions).

**Proposed Design / Methodology:**

**Model**

The choice of a predictive model indeed depends on several factors, including the problem’s nature, dataset characteristics, and desired performance metrics. We used linear regression, random forest and decision tree in the model.

**Data Preprocessing**:

**Data Preprocessing**:

The analysis begins with data preprocessing, involving the loading of datasets and examination for missing values. Columns deemed unnecessary for the analysis are removed to streamline the dataset, ensuring data integrity. Additionally, columns are renamed for clarity, enhancing the understandability of the data.

**Exploratory Data Analysis (EDA):**

Following data preprocessing, exploratory data analysis (EDA) is conducted to gain insights into the datasets. Various visualization techniques, including histograms, box plots, and heatmaps, are utilized to explore data distributions, relationships between variables, and correlations among features. This phase provides valuable insights into the characteristics of the data, facilitating informed decision-making in subsequent steps.

**Feature Engineering and Preprocessing:**

Feature engineering and preprocessing are crucial steps aimed at enhancing the quality of the data for modeling. New features are created based on domain knowledge or feature interactions, enriching the dataset with potentially predictive information. Outliers are detected and handled using appropriate techniques to mitigate their impact on the models. Furthermore, numerical features are normalized to ensure uniformity and improve model performance. The dataset is then split into training and testing sets to facilitate model training and evaluation.

**Model Building**:

With the preprocessed data in hand, model building commences. A Linear Regression model is trained to predict the density variable, leveraging the engineered features. The performance of the model is evaluated using metrics such as R-squared and RMSE, providing insights into its predictive capabilities. Visualization techniques are employed to compare the actual and predicted values, offering further insights into model performance.

**Prediction of Obesity Level and Exercise Recommendation Plan:**

In parallel, prediction tasks are undertaken, including the prediction of obesity levels and exercise recommendation plans. Machine learning models, such as Random Forest Classifier and Decision Tree Classifier, are trained on relevant features to predict these outcomes. The performance of these models is assessed using accuracy metrics, providing an indication of their predictive power and reliability.

**Model Optimization and Fine-Tuning:**

Following initial model training, optimization and fine-tuning steps may be employed to improve model performance further. Techniques such as hyper parameter tuning, feature selection, and ensemble methods can be utilized to enhance model accuracy and generalization. This iterative process aims to refine the models and maximize their predictive capabilities.

**Validation and Cross-Validation:**

Validation and cross-validation techniques are employed to assess the robustness and generalization of the models. Validation sets are used to evaluate model performance on unseen data, while cross-validation provides an estimate of how well the model will perform on new datasets. These techniques help ensure that the models are not overfitting to the training data and can generalize well to new instances.

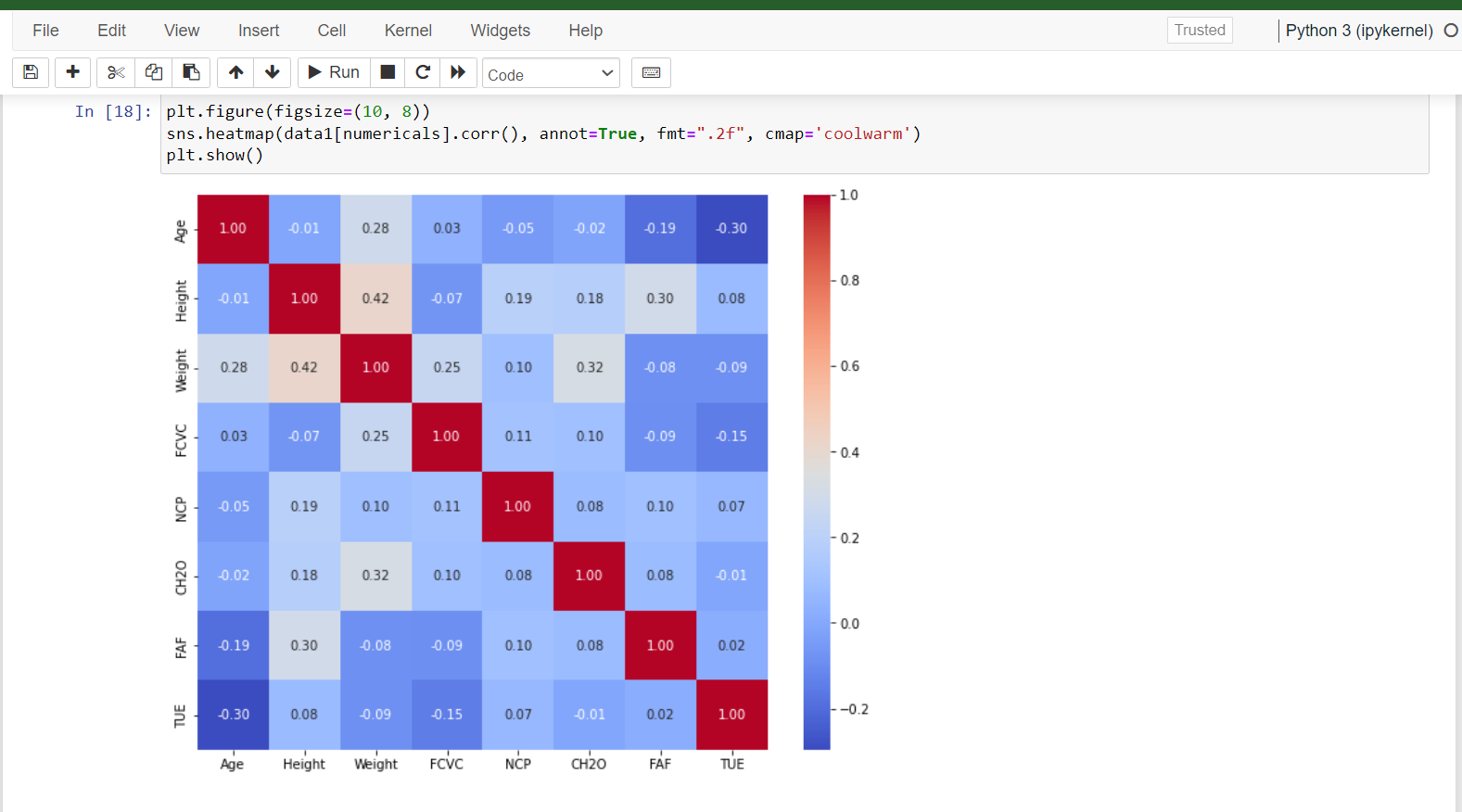
**Deployment and Monitoring:**

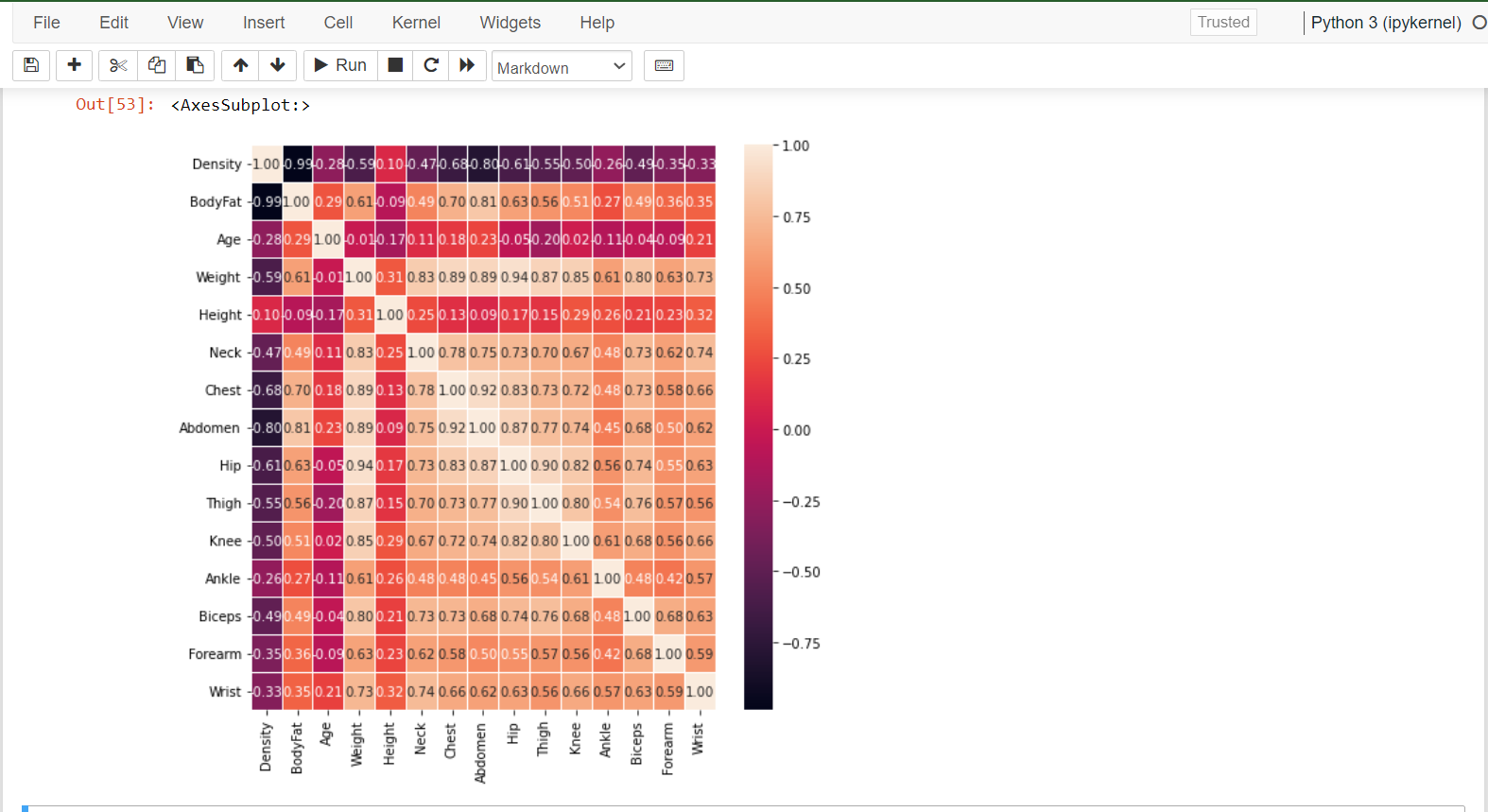
Once satisfactory model performance is achieved, the models can be deployed into production environments for real-world applications. Continuous monitoring and evaluation of model performance are essential to ensure that the models remain effective over time. Feedback loops may be established to incorporate new data and update the models periodically, maintaining their relevance and accuracy in dynamic environments.

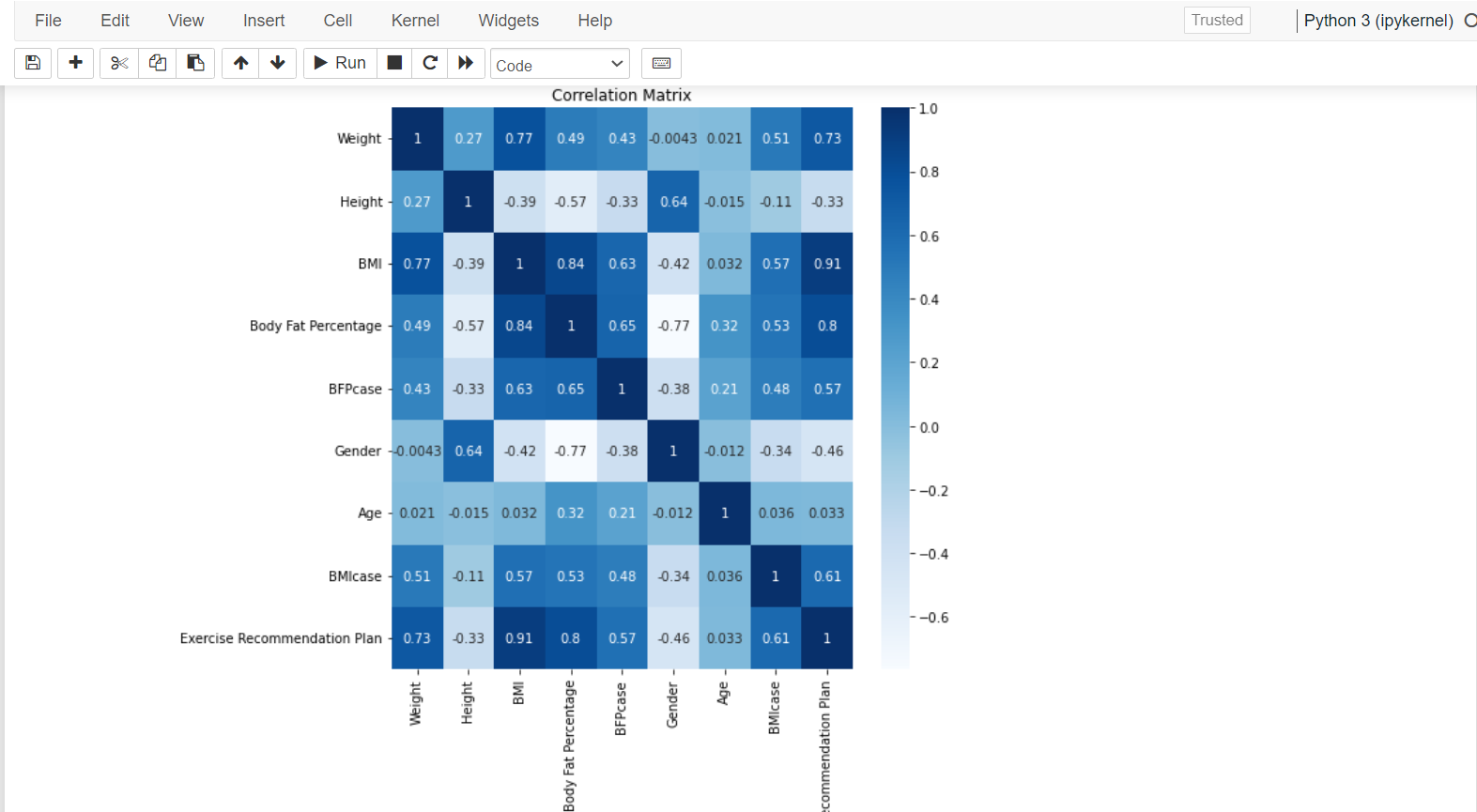
**Conclusion:**

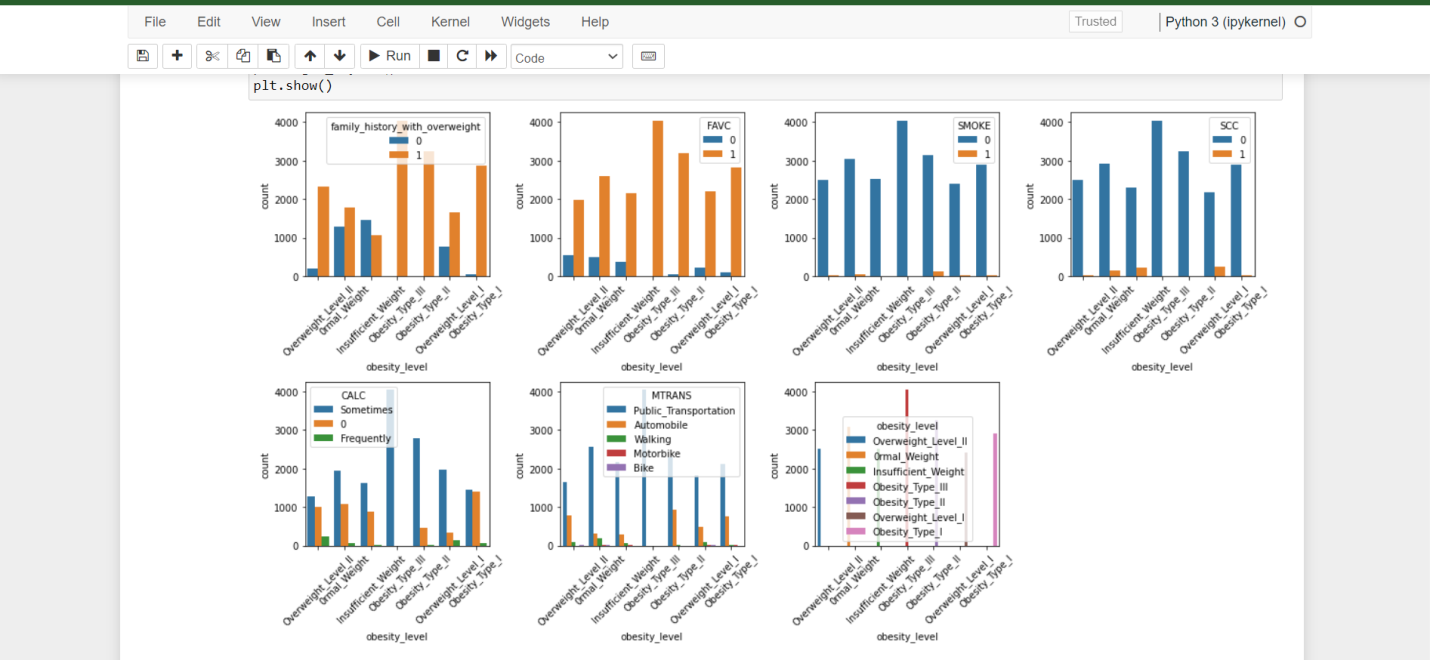
Overall, the analysis follows a structured approach encompassing data preprocessing, exploratory data analysis, feature engineering, model building, optimization, validation, deployment, and monitoring. By iteratively refining the models and leveraging advanced techniques, the analysis aims to develop robust predictive models capable of making accurate predictions and driving actionable insights in various domains.

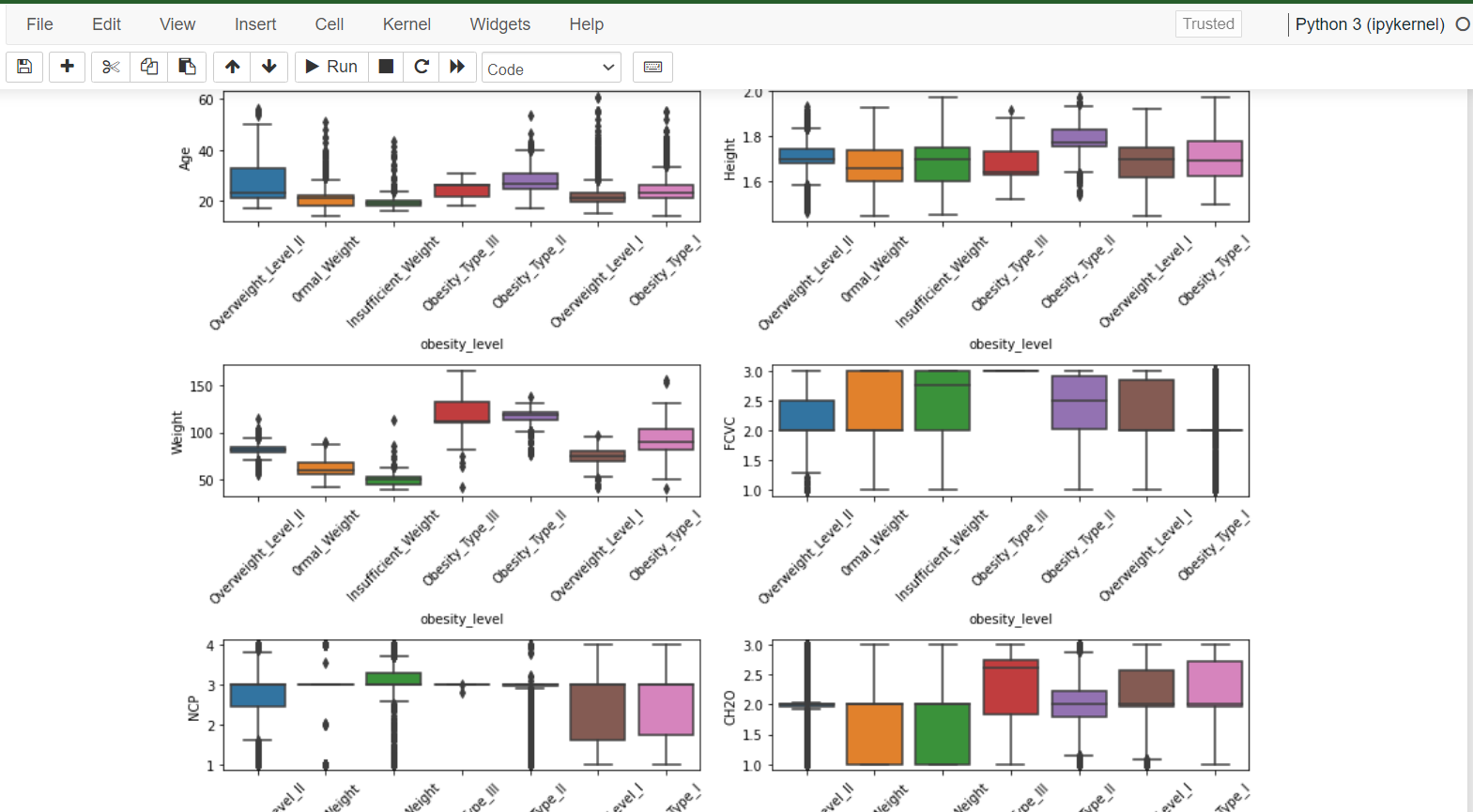
**Visuals :**











**User Interface :**

