Weight Change Prediction Model Report

1. Project Overview

The primary objective of this project is to predict weight change based on various factors including caloric intake, activity level, sleep quality, and stress, and to provide relevant advice for weight management strategy. This report outlines the analysis, modeling process, and insights gained from the data, with the aim of understanding the most influential factors in weight management.

2. Data Overview

Dataset Summary:

- **Source:** [Kaggle]
- **Features:** 13 columns, 100 entries
- **Top Features**: Participant ID, Age, Gender, Current Weight (lbs), BMR (Calories), Daily Calories Consumed, Daily Caloric Surplus/Deficit, Weight Change (lbs), Duration (weeks), Physical Activity Level, Sleep Quality, Stress Level, Final Weight (lbs)

Sample Data:

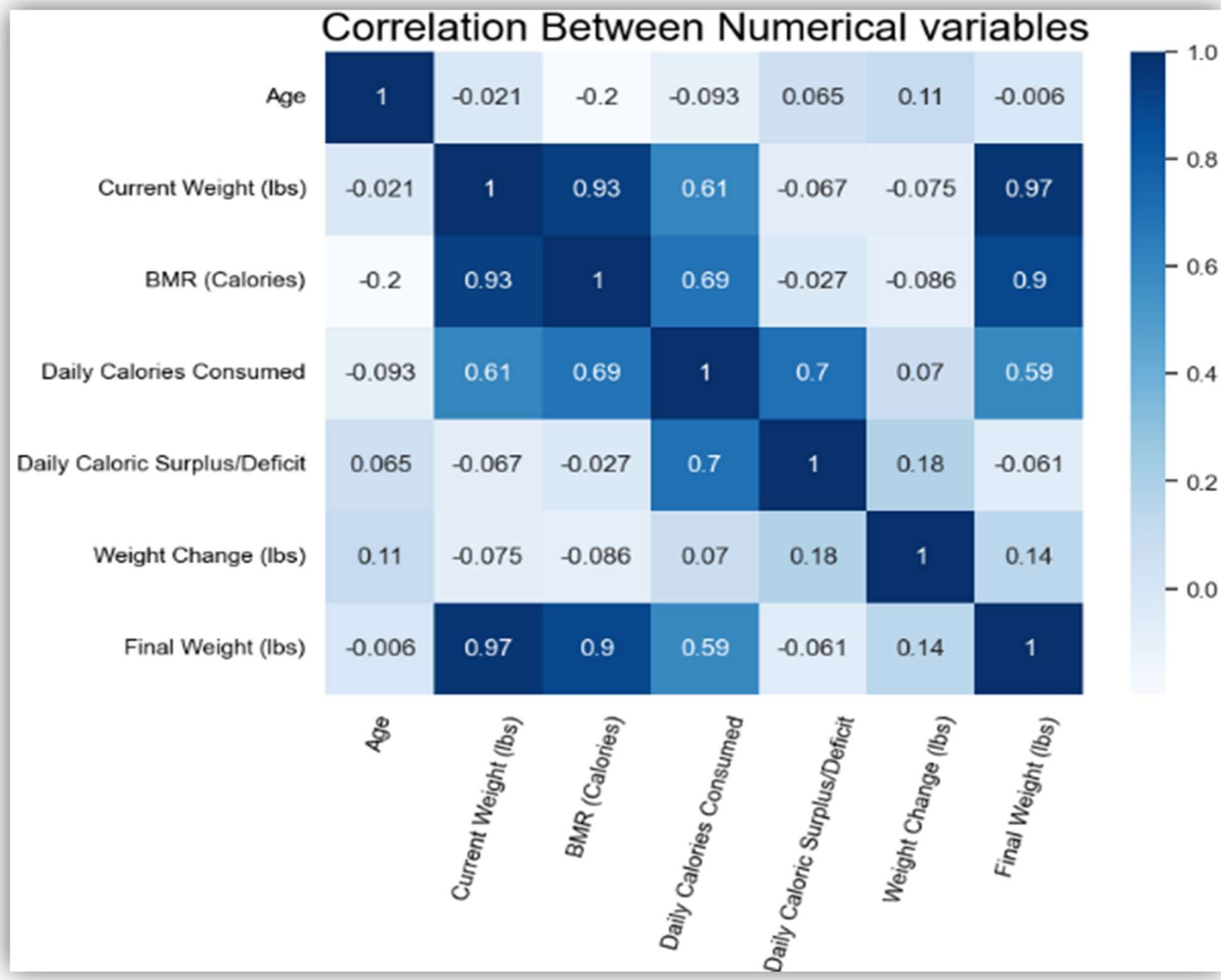
Participant ID	Age	Gender	Current Weight (lbs)	BMR (Calories)	Daily Calories Consumed	Daily Caloric Surplus/Deficit	Weight Change (lbs)	Duration (weeks)	Physical Activity Level	Sleep Quality	Stress Level	Final Weight (lbs)
1	56	M	228.4	3102.3	3916.0	813.7	0.2	1	Sedentary	Excellent	6	228.6
2	46	F	165.4	2275.5	3823.0	1547.5	2.4	6	Very Active	Excellent	6	167.8
...	...	...	...	...	...	...	...	...	...	...	...	...

3. Exploratory Data Analysis (EDA) and Findings

Data Types and Memory Usage:

- 13 columns: 6 float64, 4 int64, 3 object types
- 100 entries
- We have zero missing values and all data types are already correct.

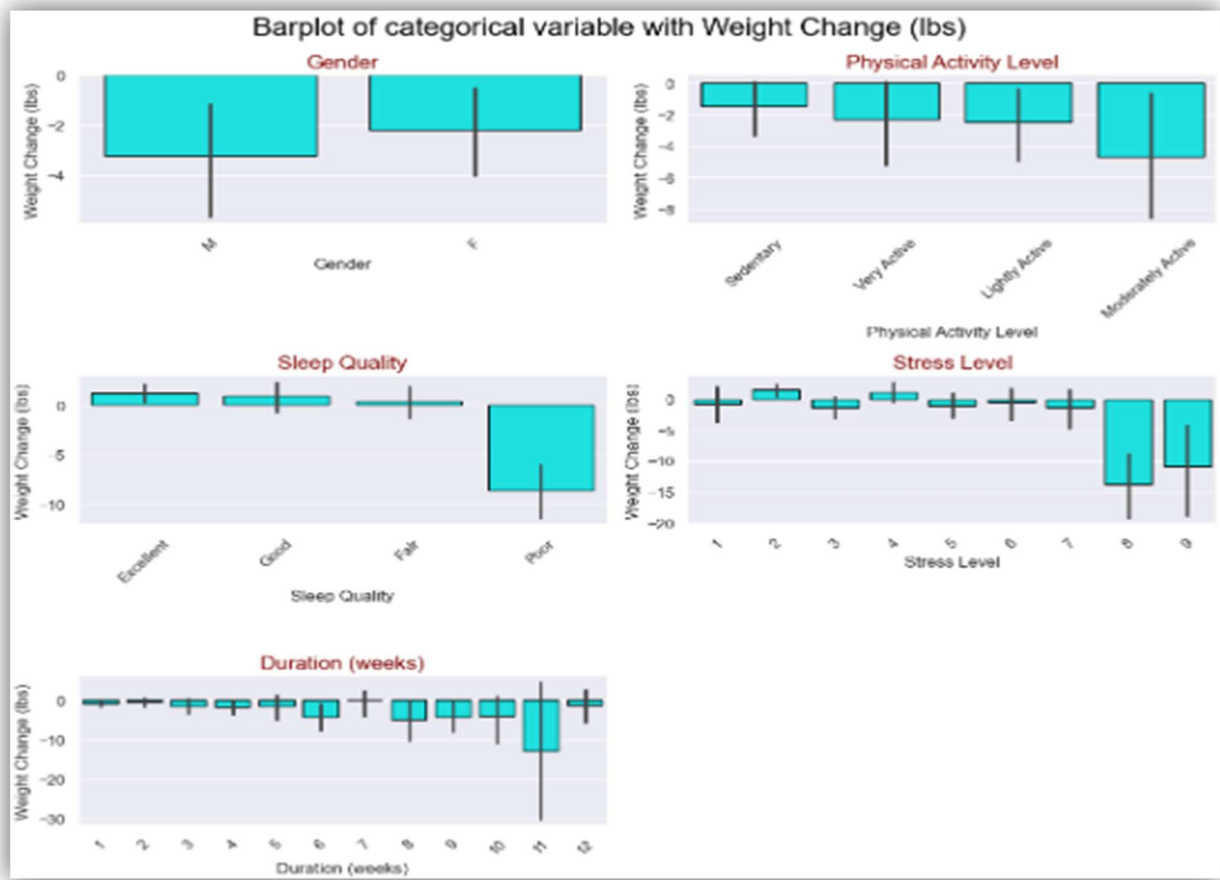
Correlation Analysis:

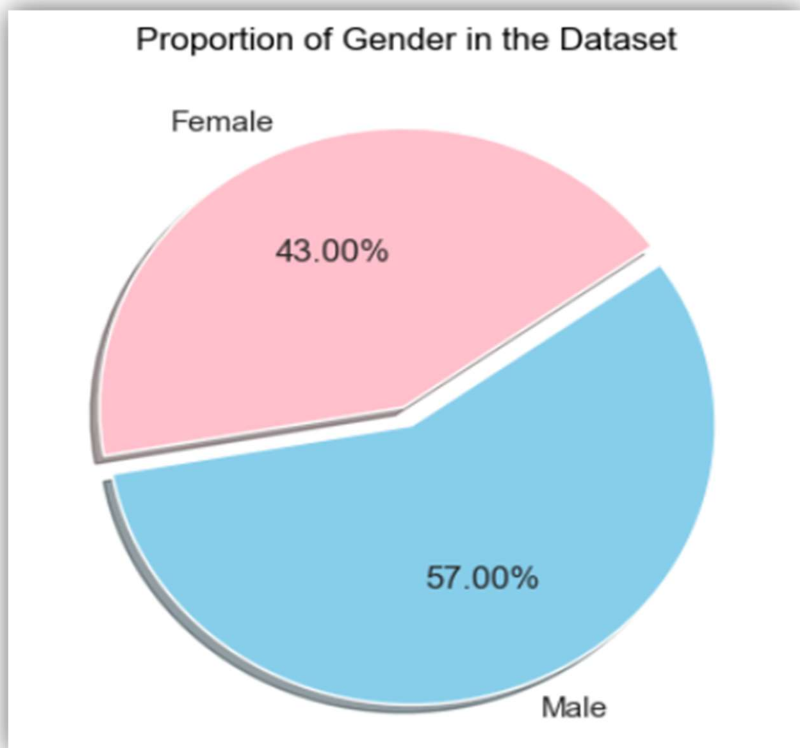


- **Daily Caloric Surplus/Deficit and Weight Change**: Weak positive correlation (0.025).
- **BMR (Calories) and Weight Change**: Weak negative correlation (-0.11).
- Findings suggest non-linear relationships, which may require complex models for better predictions.

Key Insights from EDA:

- Features might interact with each other to influence weight change.
- Multiple regression or machine learning models like Random Forest or Gradient Boosting might better capture these complex interactions.


Barplot of categorical variable with Weight Change (lbs)
Gender
Physical Activity Level
Sleep Quality
Stress Level
Duration (weeks)
Weight Change (lbs)
M
F
Gender
Sedentary
Very Active
Lightly Active
Moderately Active
Physical Activity Level
Excellent
Good
Fair
Poor
Sleep Quality
Stress Level
Duration (weeks)


Proportion of Gender in the Dataset
Female
43.00%
57.00%
Male

4. Data Preprocessing and Feature Engineering

Outlier Treatment:

- Applied Yeo-Johnson transformation to reduce the influence of outliers on 'Weight Change.'

Feature Engineering:

- **New Features Created**:
    - **Caloric Intake Per Weight**: Dropped due to redundancy.
    - **Physical Activity MET Value**: Used to estimate energy expenditure.
    - **Activity Weighted Calories**: Combines BMR and MET value to calculate calories burned based on activity level.

Feature Selection and Redundancy Handling:

- Dropped ‘Final Weight’ to avoid target leakage and redundancy with ‘Weight Change.’
- Removed 'Caloric Intake Per Weight' and 'Physical Activity Level' as their effects are covered in ‘Activity Weighted Calories.’

---

5. Modeling Process and Evaluation

Models Tested and Performance Metrics:

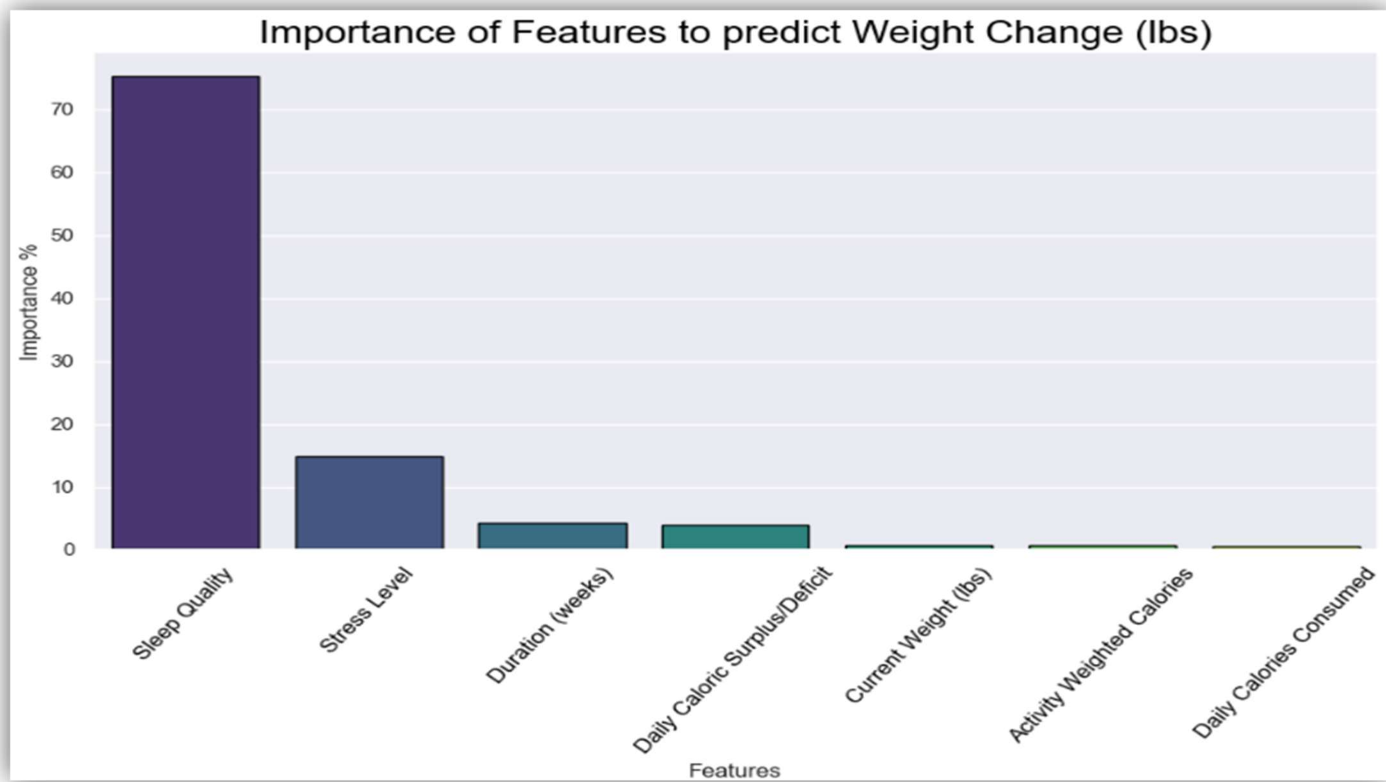
Model	R^2 Score	Adjusted R^2 Score	Mean Absolute Error	Mean Squared Error	Root Mean Squared Error
Multi-Linear Regressor	0.242713	-0.438845	0.481131	0.473570	0.688164
Decision Tree Regressor	0.898584	0.807311	0.195354	0.063420	0.251834
Support Vector Regressor	0.041605	-0.820951	0.629212	0.599334	0.774167
Random Forest Regressor	0.903197	0.816075	0.198354	0.060536	0.246040
XGB Regressor	0.906328	0.822022	0.182008	0.058578	0.242029
KNN Regressor	-0.300782	-1.471486	0.797495	0.813446	0.901912
Gradient Boosting Regressor	0.874872	0.762257	0.215300	0.078249	0.279730
AdaBoost Regressor	0.840218	0.696415	0.257434	0.099920	0.316101

Best Model: XGB Regressor with R^2 score of 0.906, suggesting strong predictive power before optimization.

---

6. Feature Importance and Optimization

Final Feature Importance:



- **Sleep Quality (75.2%)) is the most influential factor in weight change, suggesting that good sleep is a crucial element for weight management.**
- **Stress Level (14.8%)also plays a significant role in weight fluctuations, indicating that managing stress is important.**
- **Duration (weeks) (4.3%) and Daily Caloric Surplus/Deficit (3.9%) also have importance in predicting weight change.**
- **Current Weight (lbs) and Activity Weighted Calories hold** minimal **importance, suggesting their relevance in predicting weight change.**

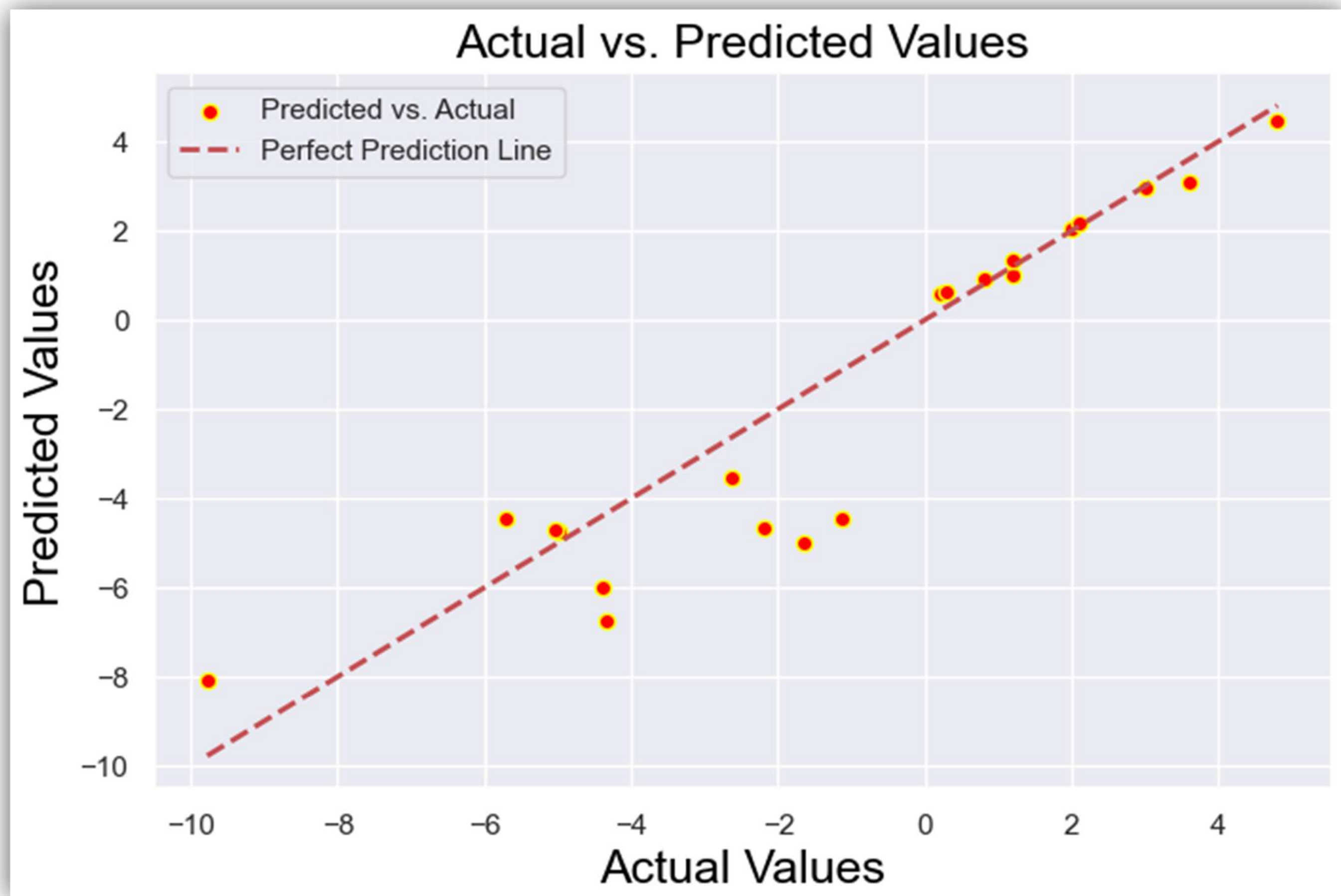
Hyperparameter Tuning Results:

- **Best Parameters**: 'learning_rate'=0.2, 'max_depth'=4, 'min_samples_leaf'=1, 'min_samples_split'=2, 'n_estimators'=100, 'subsample'=0.9
- **After Tuning Performance:**
    - **Training R^2 Score**: 0.998
    - **Testing R^2 Score**: 0.944
    - Improved generalization, indicating a well-optimized model with minimal overfitting.

---

7. Results and Analysis

- **R^2 Improvement**: The optimized model explains 94.4% of the variance in 'Weight Change' on testing data, showing a significant improvement.
- **Adjusted R^2**: Improvement to 0.911, supporting generalization capability.
- **MAE, MSE, RMSE**: All metrics decreased after tuning, indicating better prediction accuracy.



8. Key Insights & Advice for Weight Management Strategy

- **Prioritize Sleep Quality:** As the most significant predictor, improving sleep quality can be a powerful strategy for controlling weight change. Aim for 7-9 hours of quality sleep per night, as poor sleep can disrupt metabolic and hormonal processes that regulate appetite and fat storage.
- **Stress Management:** With stress levels playing a significant role (14-15%), managing stress through mindfulness, exercise, or relaxation techniques can support weight management efforts. Chronic stress leads to higher cortisol levels, which can promote fat retention, especially around the abdomen.
- **Monitor Caloric Surplus/Deficit:** The increased importance of Daily Caloric Surplus/Deficit suggests that tracking daily caloric intake and expenditure is crucial for weight control. Consuming fewer calories than the body burns will lead to weight loss, while a surplus leads to weight gain.
- **Physical Activity:** Since Activity Weighted Calories (related to Met values and BMR) has some predictive importance, maintaining regular physical activity can help burn calories and contribute to weight maintenance. Aim for a mix of aerobic exercises and strength training for optimal results.
- **Consider Duration (Weeks):** Weight change is a gradual process, and the duration over which the changes occur matters. Long-term habits tend to have more lasting effects on weight compared to short-term efforts.

- **Individual Weight Management:** As Current Weight (lbs) has some influence, weight management strategies should be personalized. Those with higher starting weights may need more significant changes in diet and activity levels to see the same results as those starting at a lower weight.

---

9. Conclusion & Strategy

To optimize weight change and management, focus on improving sleep quality and managing stress. Regular exercise and maintaining a healthy caloric balance will also be crucial for achieving sustainable weight loss or maintenance. Regularly track progress over weeks and adjust habits as needed for optimal results.
