# Predictive Analysis Report: Powerlifting Total Lift Prediction

## Introduction

This report aims to analyze and predict the total weight lifted by powerlifters based on their age, body weight, and equipment used. The analysis involves exploring the dataset, preprocessing the data, building a predictive model using linear regression, and evaluating the model's performance.

## Dataset Overview

The dataset used for this analysis is the "openpowerlifting-2024-01-06-4c732975.csv" dataset, containing information about powerlifting competitions. It includes features such as age, bodyweight, equipment used, and the total weight lifted (sum of squat, bench press, and deadlift). The dataset was cleaned by removing outliers and unnecessary columns.

## Data Preparation

- The dataset was preprocessed by handling missing values, removing outliers, and encoding categorical variables (equipment).

- Features such as age and body weight were extracted from the dataset to use as predictors for the model.

## Model Building

- A linear regression model was chosen for its simplicity and interpretability.

- The model was trained using the training set, consisting of 80% of the data, and evaluated on the remaining 20% (test set).

## Model Evaluation

- Train R^2 score: 0.4138

- Test R^2 score: 0.4159

- Train RMSE: 111.9793 kg

- Test RMSE: 111.5742 kg

## Summary

- The model explains approximately 41% of the variance in the total weight lifted by powerlifters.

- On average, the model's predictions are approximately 114 kg away from the true values.

- The model's performance is consistent across both the training and test sets, indicating no overfitting or underfitting issues.

## Recommendations for Improvement

- To improve model accuracy, consider building separate models for individual lifts (squat, bench press, deadlift).

- Including interaction variables as predictors could enhance model accuracy to the range of 65 - 70%.

- Further analysis can be done by breaking down age and body weight into multiple groups to understand their effects better.

## Conclusion

In conclusion, the linear regression model provides a reasonable baseline for predicting the total weight lifted by powerlifters based on age, body weight, and equipment used. While the model's accuracy can be improved with additional features and techniques, it serves as a valuable tool for understanding the factors influencing powerlifting performance.

This report provides insights into the predictive analysis process and offers recommendations for further refinement and improvement of the model.