



# MONTE CARLO SIMULATION AND BETTING SYSTEM FOR PREMIER LEAGUE PREDICTIONS

1

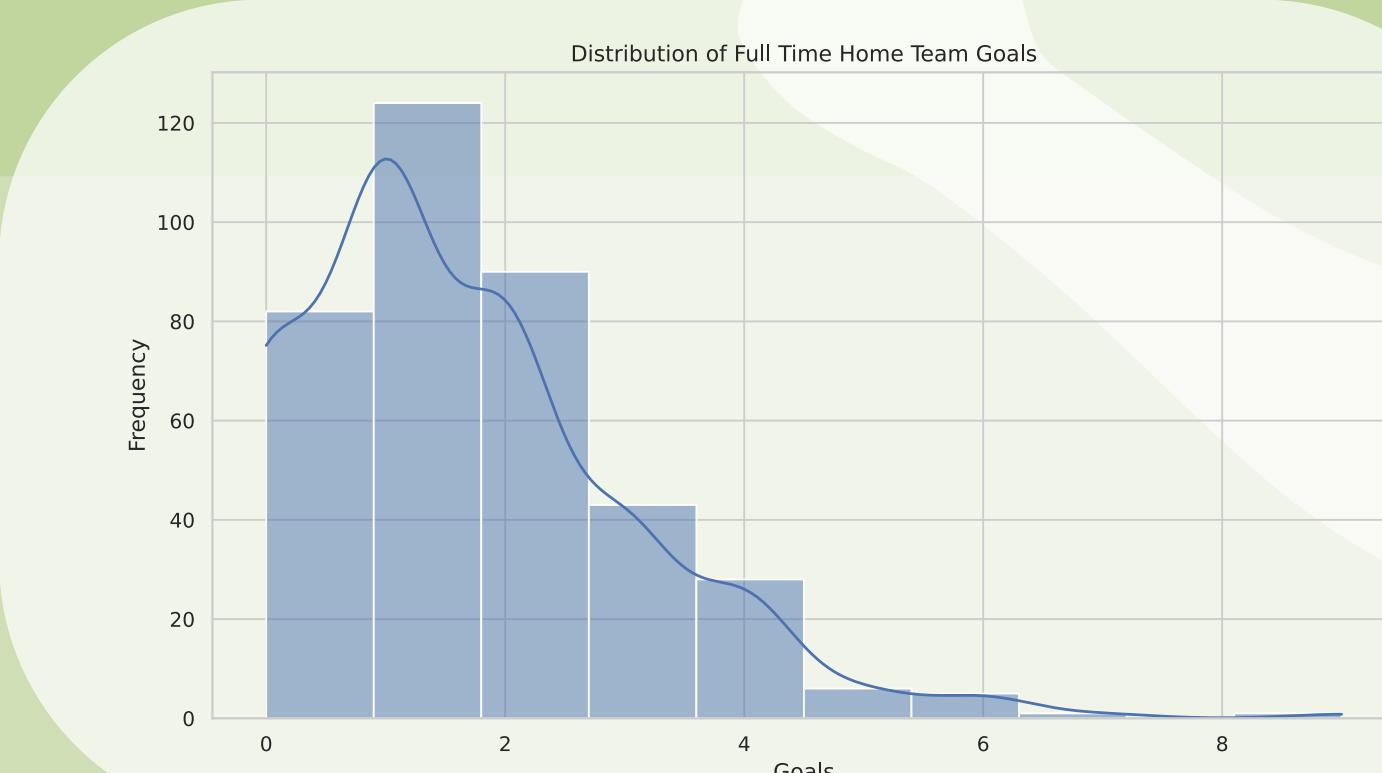
## Introduction

The project aims to leverage machine learning algorithms to predict sports outcomes, project league standings, and provide insights into betting odds. By combining historical sports statistics and real-time odds data, this project enhances the understanding of sports events and their potential outcomes.

3

## Exploratory Data Analysis

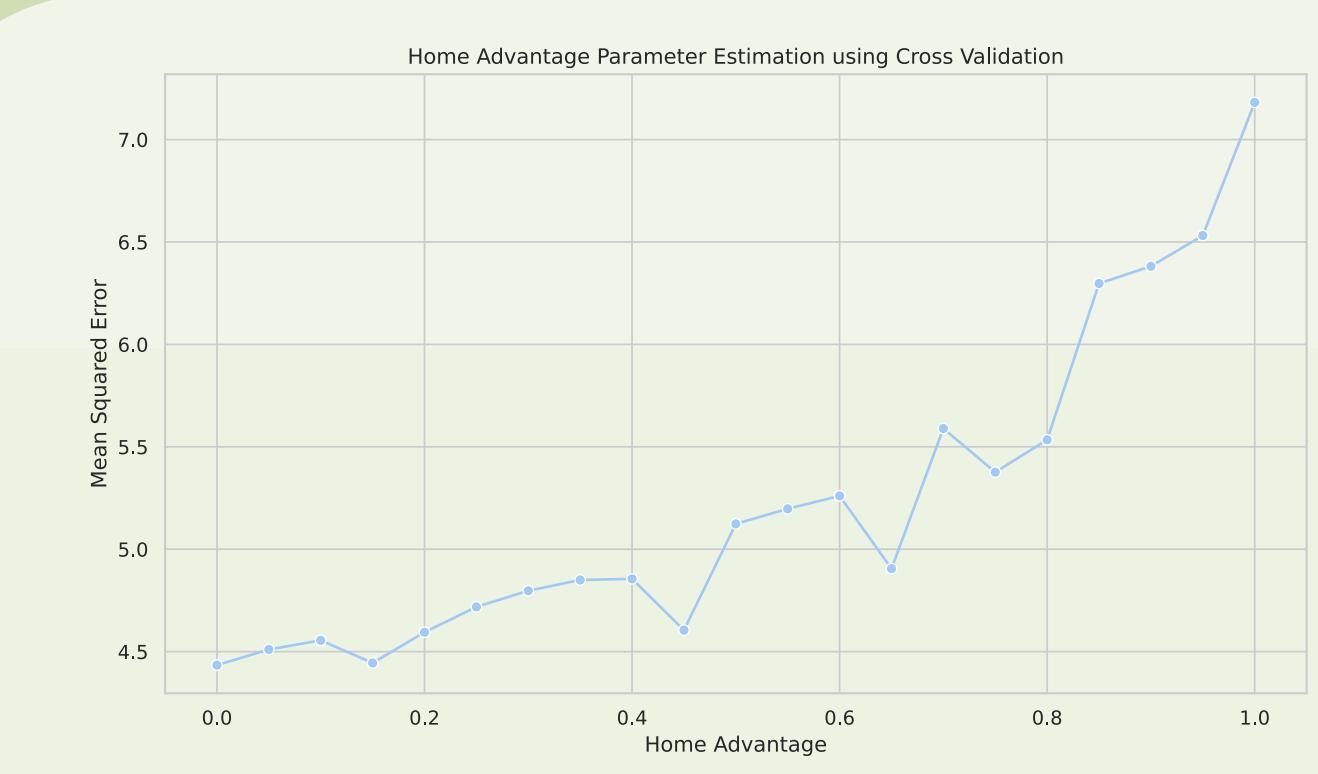
The dataset offers a comprehensive view of English Premier League matches for the 2022–2023 season. It includes match results, stats, and diverse betting odds. This resource is valuable for analyzing match outcomes, betting trends, and statistical modeling in sports analytics.



4

## Methodology

- Optimal split value derived from MAE.
- Simulate EPL table using various approaches:
  - GLM and Poisson-based Monte Carlo: Utilize Poisson regression through R's 'glm' function to simulate standings, calculating attack, defense strengths, and home advantage.
  - Manual team data reduction for strengths: Compute strength based on average goals, ignoring nuances like opponent quality.
  - PCA for team strength and table prediction: PCA captures key statistics using two components, representing over 95% variability, and integrates home advantage.
  - Determine home advantage through cross-validation
  - Factor Analysis gave inconclusive results due to inconducive data dimensions.
- Procrustes Algorithm compares non-metric MDS values from models.
- Choose best model and predict odds/probabilities.
- Compare predicted odds with actual organization-provided odds.
- Translate odds to monetary terms, ensuring the house profits.



8

## References

- James Adam Gardner, *Modeling and Simulating Football Results*, 2011.
- J. F. C. Kingman. *Poisson Processes*. Oxford University Press, 1993.
- Rue H, Salvesen O. 2000 Prediction and retrospective analysis of soccer matches in a league. *J. R. Stat. Soc. Ser. D (Stat.)* 49, 399–418.
- Source: UCD 2023 Spring Multivariate Analysis coursework
- Source: "English Premier League and Championship Soccer Data" by Football-Data.co.uk.

2

## Features

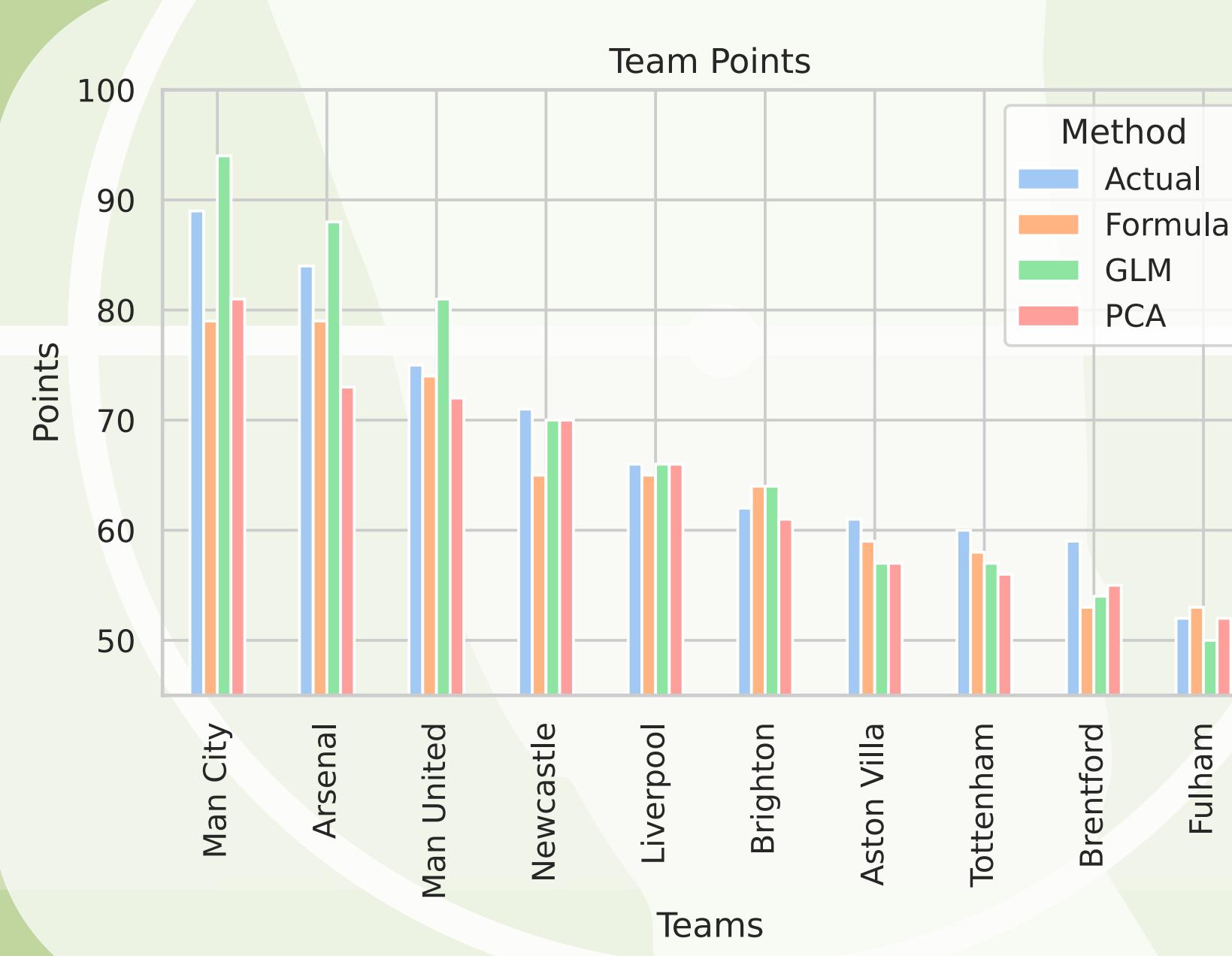
- Outcome Predictions:** Utilize multiple trained machine learning models to predict the outcomes of upcoming sports matches and league standings.
- Betting Insights:** Compare real-time betting odds with predicted outcomes, offering insights into potentially profitable betting opportunities.
- Monetary Conversion:** Convert betting odds into monetary terms, allowing users to understand potential returns on their bets.

The Poisson distribution is used in simulating football matches via MCMC due to its fit for modeling event counts i.e., goals scored. The derived Poisson regression model effectively captures this pattern. Goal distribution graph supports its relevance, confirming its role in football match simulation.

Football analysis reveals a home team advantage, reflected in a Poisson regression parameter (home ground advantage). The 22-23 season graph confirms more home team goals, enhancing realism and reducing outcome differences, measured by MSE.

5

## Visualizing Predictions



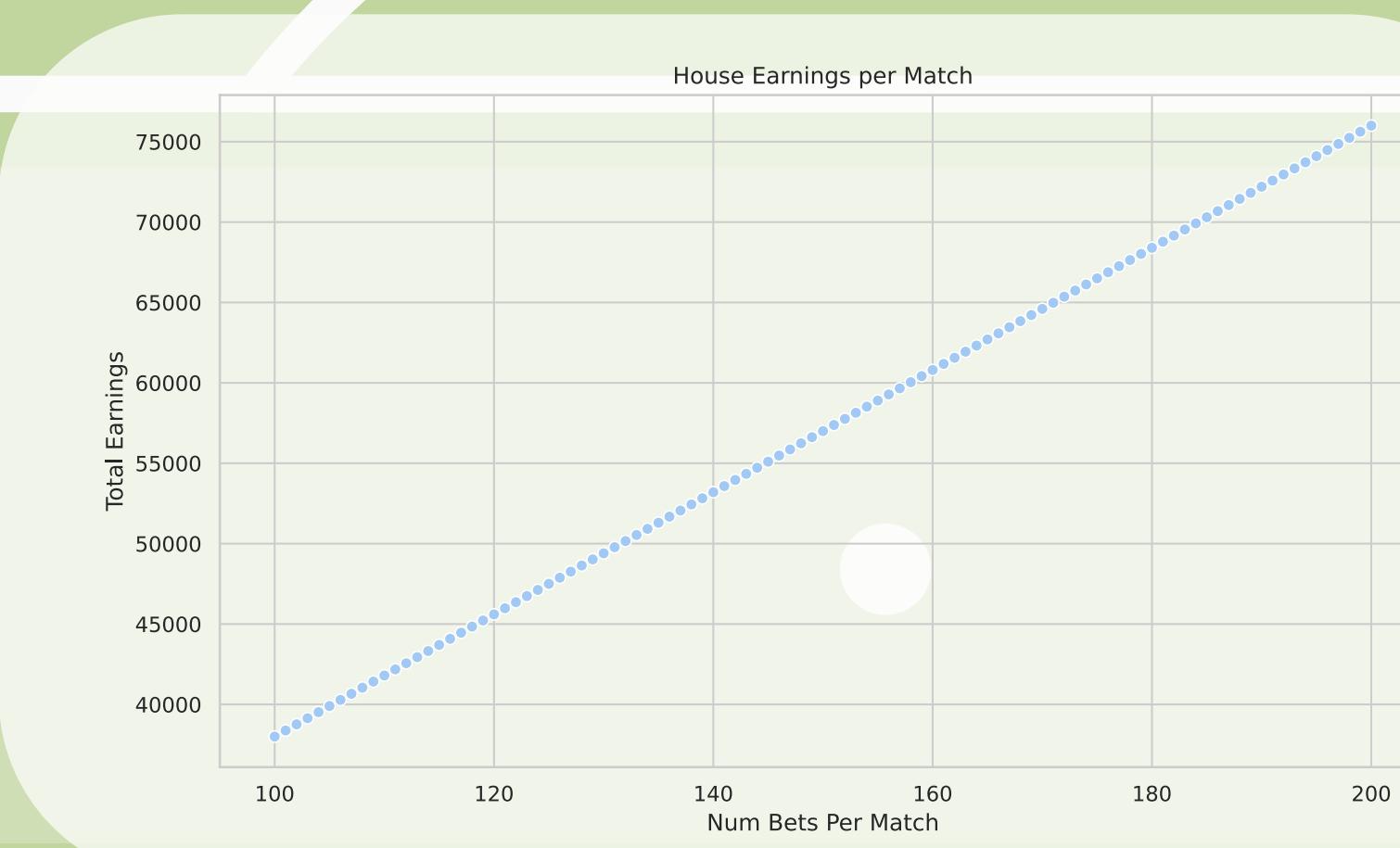
7

## Convert Odds to Monetary Terms

A simulated English Premier League (EPL) betting algorithm employs team parameters, odds, and randomness to predict the house's earnings from bets. It explores the impact of the number of bets per match on earnings, providing insights into EPL betting trends, while recognizing real-world complexity beyond the model's scope.

	Bet365	IW
Home Win	0.4804063	0.4752186
Draw	0.01951954	0.03103072
Away win	0.4203848	0.4144028

Table 1: Correlation between the predicted odds and the odds set by the organisations which is seen to be positively correlated



6

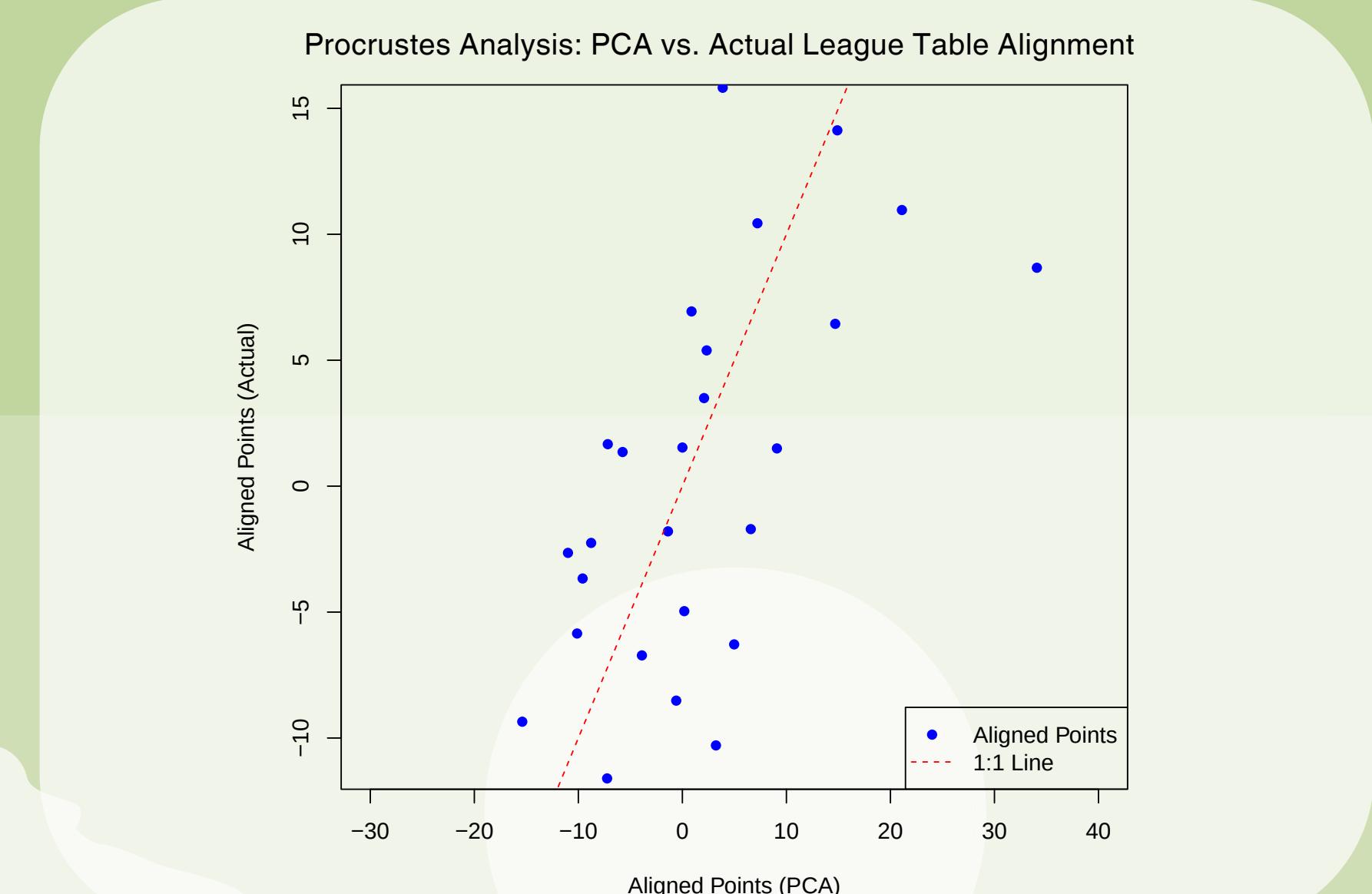
## Comparison of Models

By employing GLM, PCA, and a formulated prediction method to evaluate English Premier League standings, PCA emerges superior in MAE, MPAE, and Procrustes Sum of Squares assessments, highlighting its superior performance as seen in the next graph.

	GLM	PCA	Formula
MAE Score	2.950000	3.800000	2.750000
MAPE Score	6.471286	8.003657	5.434667
Procrustes SS	4014.528028	3086.628006	5705.070739

Table 2: Evaluation matrix comparison

The scatter plot visualizes the results of a Procrustes analysis performed on two sets of Non-metric Multidimensional scaled data of the actual standing and pca standing table. Procrustes analysis is a technique used to align two sets of data points in a way that optimally matches their shapes while preserving their relative distances. The blue points correspond to the aligned data points from the "Actual" set, while the red dashed line indicates the 1:1 reference line, highlighting how well the alignment matches the original data.



10

## Authors

- Vishalkrishna Bhosle - 22205276
- Shubham Sharma - 22201541