DSA 210: Introduction to Data Science

Performance Analysis of Fenerbahçe Football Team Based on Manager Selection

A season-level performance study of head coaches (2005–2025) using EDA, hypothesis testing, and regression modelling

Prepared by: Tibet Aras Tirek – Student ID 30730
Email: tibet.tirek@sabanciuniv.edu
Submission Date: 8 Jan 2025

Table Of Contents:

1. Motivation

 Manager selection is one of the most critical strategic decisions for professional football clubs, as head coaches play a central role in shaping team performance, tactical identity, and long-term success. In the case of Fenerbahçe, a club with a rich history and high competitive expectations, managerial changes have been frequent over the past two decades. These changes have often sparked public and media debates, particularly regarding whether domestic (Turkish) or foreign head coaches deliver superior on-field performance.
Despite strong opinions surrounding this topic, much of the discussion remains anecdotal and subjective. There is limited empirical evidence that systematically evaluates managerial performance using quantitative, season-level data. This gap motivates a data-driven investigation into how different managerial profiles relate to team performance outcomes over time.
 The primary motivation of this study is to analyze Fenerbahçe's seasonal performance between 2005 and 2025 by focusing on head coach characteristics and contextual factors. Specifically, the project aims to examine whether there is a statistically significant difference in team performance—measured by points per game—between Turkish and foreign coaches. Beyond hypothesis testing, the study also seeks to understand how performance indicators and market value dynamics contribute to explaining seasonal success.
By combining exploratory data analysis, statistical hypothesis testing, and regression-based machine learning modelling, this project provides a comprehensive framework for evaluating managerial impact in football. The findings aim to contribute to a more objective understanding of coach selection decisions and demonstrate how data science methodologies can be applied to real-world sports analytics problems.

---

2. Datasets & Data Enrichment

 This study utilizes multiple structured datasets constructed at the season level to analyze Fenerbahçe's managerial performance over time. The core dataset contains information on head coaches and team performance metrics, while an auxiliary dataset provides market value indicators to capture financial and contextual dynamics. All datasets were merged and enriched to create a unified analytical table suitable for exploratory analysis, hypothesis testing, and machine learning modelling.

2.1 Coach Performance Dataset

The primary dataset consists of season-level records for Fenerbahçe head coaches covering the period from 2005 to 2025. Each observation represents a single season under a specific head coach and includes performance-related variables such as:

- Points per game (PPG)

- Win rate

- Average goals scored per match

- Average goals conceded per match

- Total number of matches

- Coach nationality indicator (Turkish vs. foreign)

This dataset enables direct comparison of managerial performance across seasons and provides the foundation for both descriptive and inferential analyses. A binary variable (is_foreign) was used to distinguish between domestic and foreign coaches, allowing for group-based hypothesis testing.

2.2 Market Value Dataset

To incorporate financial context into the analysis, a secondary dataset containing Fenerbahçe's average squad market value per season was used. Market value data reflects the overall quality, depth, and investment level of the squad, which are known to influence on-field performance.

Market value records were aligned with the coach dataset using the season start year as a common temporal key. This ensured a robust and consistent merge across datasets while preserving the chronological structure of the data.

2.3 Data Enrichment and Feature Engineering

Beyond raw performance and market value figures, additional features were engineered to capture temporal dynamics and trends. These enriched variables include:

- Lagged market value (one-season lag)
- Season-over-season percentage change in market value
- Rolling mean and rolling standard deviation of market value over a three-season window

These engineered features were designed to model momentum, volatility, and longer-term trends in squad valuation rather than relying solely on absolute values. To avoid unnecessary

data loss, missing values introduced by lag and rolling calculations were handled using backward filling, ensuring continuity in the time series.

The final enriched dataset integrates performance metrics, coach attributes, and market value dynamics into a single season-level table. This structure supports exploratory data analysis, statistical hypothesis testing, and predictive modelling while minimizing data leakage and preserving temporal integrity.

---

3. Data Collection & Preparation

The quality of the data directly influences the validity and reliability of the analysis results. This section outlines the process of transforming raw season-level datasets into a clean, structured, and analysis-ready format. The focus is on ensuring temporal consistency, accurate merging across sources, and minimal information loss during preprocessing.

Initial Steps

- Raw data were loaded into pandas DataFrames from structured Excel files containing coach performance statistics and market value information.
- Season identifiers were standardized by converting them to string format and removing formatting inconsistencies.
- A common temporal key (season_start_year) was extracted from the season variable to ensure consistent alignment across datasets.

Data Aggregation and Merging

- Coach performance data were already structured at the season level, including metrics such as points per game, win rate, goals scored, goals conceded, and total matches.
- Market value data were aggregated at the season level and merged with the coach dataset using the season start year as the join key.
- An inner join strategy was applied to retain only seasons with complete information from both sources.

Preprocessing

- Observations with missing target values (points per game) were removed, as they cannot contribute to statistical testing or predictive modelling.
- No rows were dropped during feature engineering; instead, missing values arising from lagged and rolling calculations were handled using backward filling.
- The dataset was sorted chronologically to preserve the temporal order required for time-based analysis and model evaluation.

Output Dataset

- **Final dataset:** Season-level panel covering the period from 2005 to 2025
- **Unit of observation:** One season
- **Key variables:** Coach performance metrics, coach nationality indicator, and market value features

The resulting dataset is clean, temporally aligned, and enriched with financial context variables. This prepared dataset serves as the foundation for exploratory data analysis, hypothesis testing, and machine learning modelling presented in the subsequent sections.

---

4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to understand the underlying structure, distributional properties, and key patterns within the dataset before proceeding to hypothesis testing and modeling. The analysis focused on evaluating performance differences between Turkish and foreign head coaches across multiple dimensions, including points per game, win rate, goal statistics, trophy achievements, and temporal performance trends..

4.1 Summary Statistics

	points_per_game					win_rate_calc					avg_goals_scored					avg_goals_conceded					has_trophy				
	mean	median	std	min	max	mean	median	std	min	max	mean	median	std	min	max	mean	median	std	min	max	mean	median	std	min	max
ach_type																									
Foreign	1.85	2.00	0.34	0.93	2.09	0.53	0.58	0.13	0.20	0.63	1.70	1.75	0.44	0.80	2.40	1.05	1.00	0.19	0.80	1.50	0.17	0.00	0.39	0	1
Turkish	2.01	2.00	0.29	1.44	2.44	0.59	0.60	0.12	0.33	0.76	1.75	1.70	0.25	1.40	2.30	1.04	1.10	0.18	0.70	1.30	0.23	0.00	0.44	0	1

Initial descriptive statistics were computed for core performance metrics such as points per game (PPG), win rate, average goals scored, average goals conceded, and trophy achievement indicators. These metrics were summarized separately for Turkish and foreign coaches to provide a high-level comparison of central tendency and variability.

Overall, Turkish coaches exhibited slightly higher average points per game and win rates compared to foreign coaches, while goal-related metrics showed relatively similar central values across groups. Standard deviation values indicated moderate variability in performance within both categories, suggesting that coach nationality alone does not fully explain performance dispersion.

4.2 Temporal Performance Trends

To assess how performance evolved over time, season-level trends were examined using line plots and aggregated summaries. Points per game and win rate metrics were tracked across

seasons to observe fluctuations, periods of stability, and potential structural shifts in team performance.

The temporal analysis revealed noticeable season-to-season volatility, particularly during periods with frequent managerial changes. While some high-performing seasons coincided with both Turkish and foreign coaches, no consistent long-term dominance of one group over the other was observed across the full time horizon.
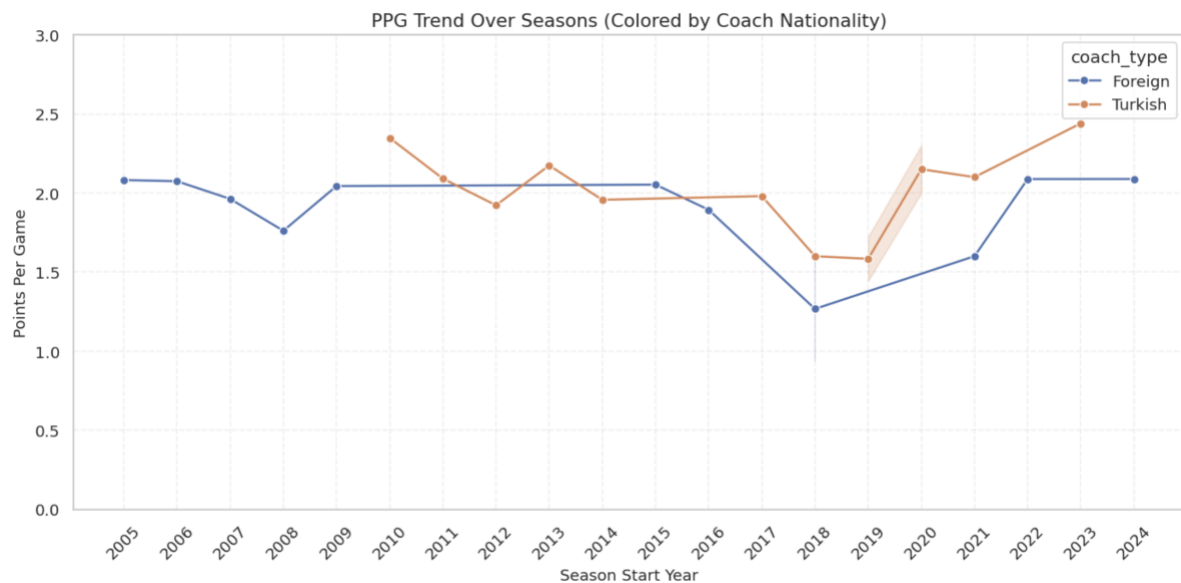
4.3 Distributional Analysis (Goals, Win Rate, PPG)

Distributional properties of key performance indicators were analyzed using boxplots to compare Turkish and foreign coaches. Points per game distributions showed overlapping interquartile ranges, with Turkish coaches having a slightly higher median PPG. However, both groups contained outliers, indicating exceptional high- and low-performance seasons regardless of nationality.

Similarly, win rate distributions demonstrated substantial overlap, reinforcing the idea that while average differences exist, the overall performance ranges of Turkish and foreign coaches are not sharply separated.
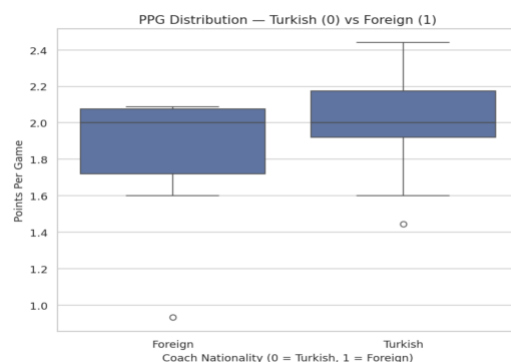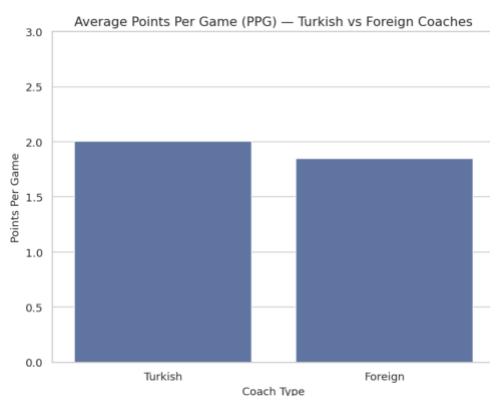
4.4 Goal Performance Analysis

Goal-related metrics were analyzed to evaluate offensive and defensive effectiveness. Average goals scored and conceded per match were compared across coach nationalities. The analysis showed marginal differences between groups, with neither Turkish nor foreign coaches displaying a systematically superior goal-scoring or defensive profile.

These findings suggest that tactical or squad-related factors may play a stronger role in goal outcomes than coach nationality alone.
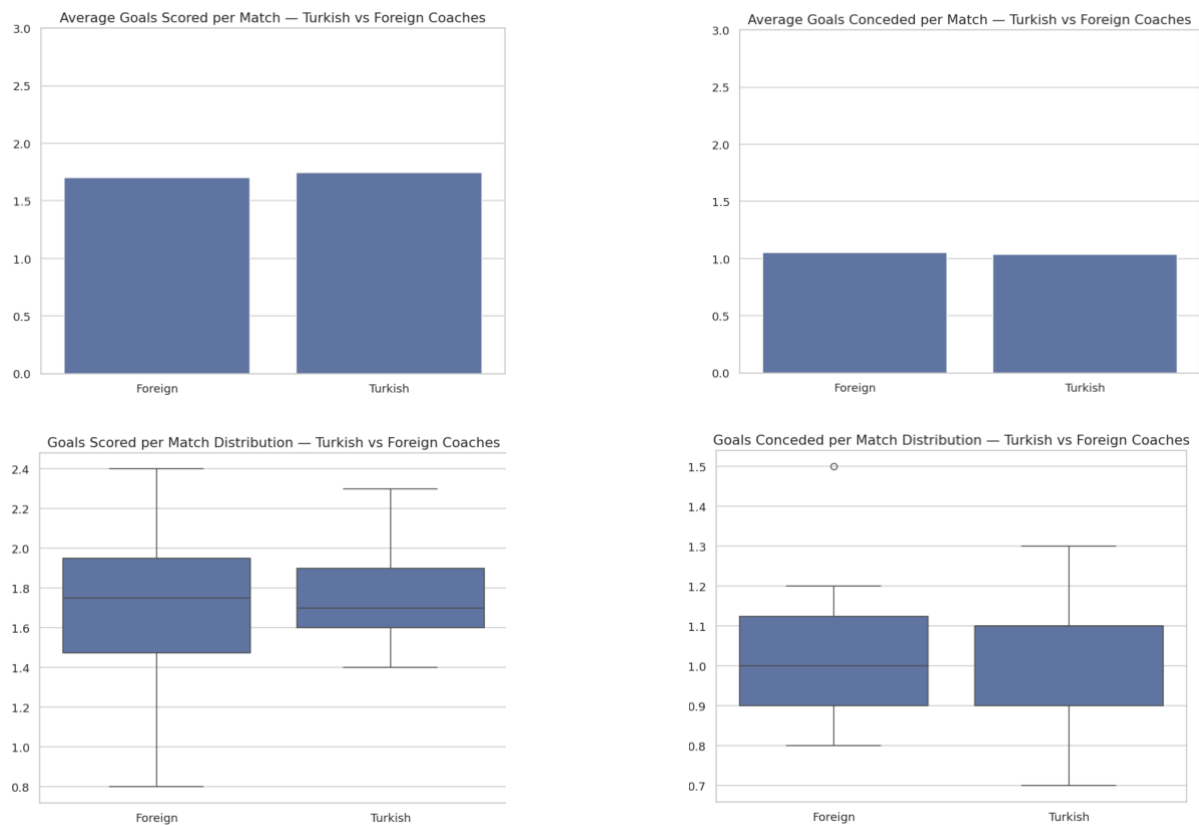
4.5 Trophy Achievement Analysis

To incorporate success beyond match-level performance, a binary trophy achievement feature was analyzed. Trophy presence was compared across Turkish and foreign coaches to assess whether nationality was associated with higher likelihood of silverware.

While some successful trophy-winning seasons were observed under both groups, the distribution did not indicate a strong or exclusive association between coach nationality and trophy attainment.

Total Trophies Won by Each Coach
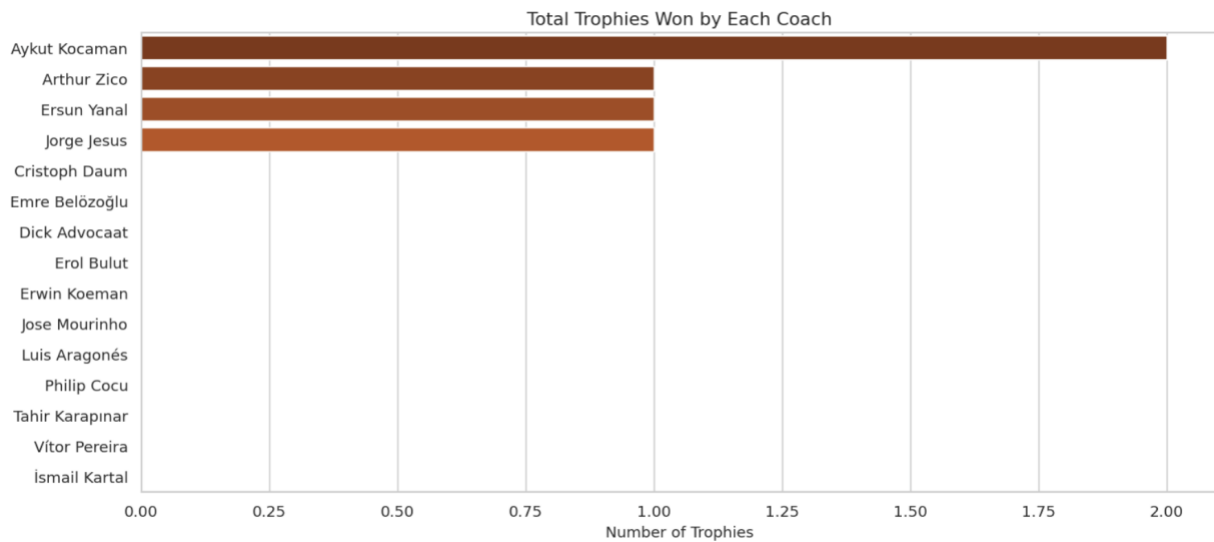
4.6 Coach-Level Performance Comparison

Finally, coach-level summaries were constructed to compare individual managers across multiple performance dimensions simultaneously. This analysis highlighted that performance outcomes varied significantly within each nationality group, emphasizing that individual coach characteristics and contextual factors likely dominate over nationality-based effects.

	coach_name	seasons	avg_ppg	avg_win_rate	avg_goals_scored	avg_goals_conceded	total_trophies
0	Arthur Zico	2	2.02	0.59	2.05	1.05	1
1	Aykut Kocaman	4	2.08	0.62	1.80	1.02	2
2	Cristoph Daum	2	2.06	0.62	2.10	1.05	0
3	Dick Advocaat	1	1.89	0.55	1.70	1.10	0
4	Emre Belözoğlu	1	2.30	0.70	1.80	1.10	0
5	Erol Bulut	1	2.00	0.62	1.80	1.10	0
6	Ersun Yanal	3	1.83	0.53	1.67	1.10	1
7	Erwin Koeman	1	1.60	0.40	1.20	1.20	0
8	Jorge Jesus	1	2.09	0.63	2.10	0.90	1
9	Jose Mourinho	1	2.09	0.62	1.80	0.90	0
10	Luis Aragonés	1	1.76	0.50	1.50	1.00	0
11	Philip Cocu	1	0.93	0.20	0.80	1.50	0
12	Tahir Karapınar	1	1.44	0.33	1.40	1.30	0
13	Vítor Pereira	2	1.83	0.52	1.50	0.90	0
14	İsmail Kartal	3	2.17	0.64	1.83	0.87	0

4.7 Key EDA Insights

The EDA phase revealed that:

- Turkish coaches showed slightly higher average performance metrics, but

- Performance distributions largely overlapped across nationalities,
- Significant variability existed within both groups,
- No strong visual separation emerged based solely on coach nationality.

These findings motivated the use of formal hypothesis testing to statistically evaluate whether observed differences were significant, as well as regression-based modeling to control for additional explanatory variables.

---

5. Hypothesis Testing

This section investigates whether there is a statistically significant difference in overall managerial performance between **Turkish and foreign head coaches** at Fenerbahçe. Performance is primarily measured using **points per game (PPG)**, which captures match outcomes while accounting for differences in total matches coached.

5.1 Research Hypotheses

- **H_0 (Null Hypothesis):**
  There is **no statistically significant difference** in overall performance between Turkish and foreign coaches.

- **H_1 (Alternative Hypothesis):**
  There **is a statistically significant difference** in overall performance between Turkish and foreign coaches.

5.2 Statistical Methodology

Before selecting the appropriate statistical test, standard assumption checks were performed:

- **Normality** of the points per game distribution for each group was assessed using the **Shapiro–Wilk test**.
- **Homogeneity of variances** between groups was examined using **Levene's test**.

The normality tests indicated that while the distribution of PPG for Turkish coaches did not significantly deviate from normality ($p > 0.05$), the distribution for foreign coaches violated the normality assumption ($p < 0.05$). Levene's test suggested no significant difference in variances between the two groups.

Given the violation of the normality assumption in at least one group, a **non-parametric Mann–Whitney U test** was selected to compare the two independent samples. This approach does not rely on distributional assumptions and is suitable for small sample sizes.

5.3 Test Results

The Mann–Whitney U test yielded the following results:

- **U statistic:** 100.0
- **p-value:** 0.2419

Since the p-value exceeds the predefined significance level ($p > 0.05$), the null hypothesis cannot be rejected.

5.4 Interpretation of Findings

The results indicate that there is **no statistically significant difference** in points per game between Turkish and foreign coaches during the analyzed seasons. Although descriptive statistics suggest slightly higher average performance for Turkish coaches, this difference is not strong enough to be considered statistically meaningful.

This finding implies that, based solely on points per game, coach nationality does not appear to be a decisive factor in Fenerbahçe's seasonal performance outcomes.

5.5 Visual Confirmation

To complement the statistical tests, boxplots and bar charts were used to visually compare the distribution and central tendency of points per game across coach nationality groups. These visualizations support the statistical findings by showing substantial overlap between the distributions of Turkish and foreign coaches.
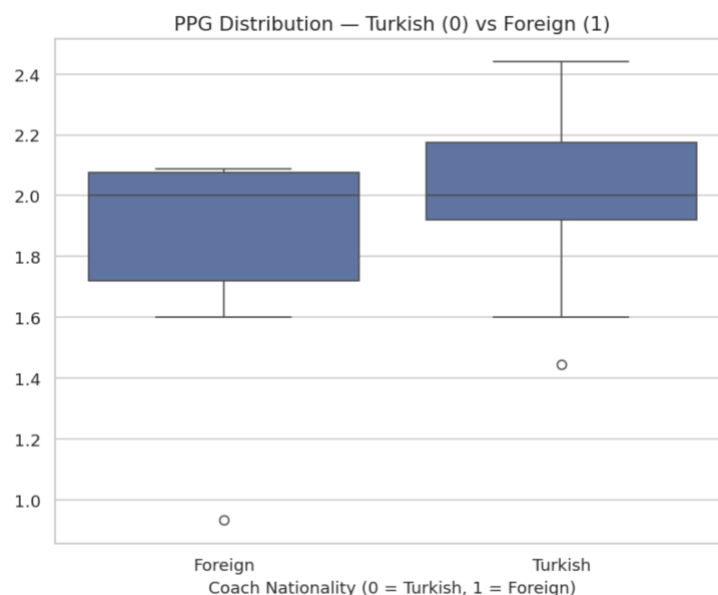
6. Machine Learning Methodology

6.1 Model Selection and Objective

To model seasonal team performance, a supervised regression approach was adopted. The primary objective was to predict **points per game (PPG)** at the season level using a combination of **market value dynamics** and **on-field performance indicators**. Given the relatively small dataset size and the presence of correlated explanatory variables, **Ridge Regression** was selected due to its ability to mitigate multicollinearity through L2 regularization and to provide stable coefficient estimates.

6.2 Feature Set and Market Value Dynamics

The final feature set was designed to capture both contemporaneous performance indicators and the temporal dynamics of squad market value. Market value information was incorporated using multiple time-aware transformations to reflect both level and momentum effects. Specifically, the following features were used:

- Average squad market value (million euros)
- One-season lagged market value
- Season-over-season percentage change in market value
- Rolling mean and rolling standard deviation of market value (3-season window)
- Average goals scored per match
- Average goals conceded per match
- Win rate
- Total matches played
- Coach nationality indicator (Turkish vs. foreign)

After feature construction and merging, the final dataset consisted of **25 observations and 16 variables**, covering the period **2005–2024**.

6.3 Time-Based Train–Test Split

To prevent data leakage and preserve the temporal structure of the data, a **time-based train–test split** was applied instead of random sampling. Seasons from **2005 to 2019** were used for training (n = 18), while more recent seasons from **2020 to 2024** were reserved for testing (n = 7). This approach ensures that the model is evaluated on genuinely unseen future data, closely mimicking real-world forecasting conditions.

6.4 Model Pipeline and Training

All preprocessing and modelling steps were implemented within a unified **machine learning pipeline** to ensure reproducibility and consistency. The pipeline included feature scaling followed by Ridge Regression estimation. This design choice allowed preprocessing steps to be learned exclusively from the training data and subsequently applied to the test set without information leakage.

6.5 Model Evaluation and Baseline Comparison

Model performance was evaluated using **Mean Absolute Error (MAE)**, **Root Mean Squared Error (RMSE)**, and **R² score**. As a benchmark, a baseline model predicting the historical mean PPG from the training set was constructed.
Ridge Regression Test Performance:

- MAE: 0.094
- RMSE: 0.105
- R²: 0.81

Baseline (Train Mean) Performance:

- MAE: 0.296
- RMSE: 0.328

The Ridge Regression model substantially outperformed the baseline across all evaluation metrics. An R² value of 0.81 indicates that a large proportion of the variation in seasonal points per game can be explained by the selected market value dynamics and performance-related features.

6.6 Prediction Analysis

Predicted and actual PPG values for the test seasons were closely aligned, with the model successfully capturing both high-performing and lower-performing seasons. While minor deviations were observed, the predictions remained within a narrow error range, further supporting the explanatory strength of the model.
It is important to note that the primary objective of this modelling exercise is not to achieve maximal predictive accuracy, but rather to demonstrate the applicability of supervised learning methods to football performance analysis within a limited historical dataset.

7. Key Findings

The main insights uncovered throughout the project combine findings from exploratory data analysis, statistical hypothesis testing, and regression-based machine learning to evaluate the impact of head coach nationality on Fenerbahçe's seasonal performance.

- Exploratory data analysis showed that Turkish head coaches exhibited slightly higher average points per game and win rates compared to foreign coaches. However, performance distributions largely overlapped, and substantial variability was observed within both groups, suggesting that nationality alone does not strongly differentiate performance outcomes.
- Statistical hypothesis testing using the Mann–Whitney U test found no significant difference in points per game between Turkish and foreign coaches ($p = 0.2419$). This result indicates that the observed differences in average performance are not statistically meaningful and should not be interpreted as evidence of a systematic nationality-based advantage.
- Time-based performance trends further supported this finding, as high-performing and low-performing seasons occurred under both Turkish and foreign coaches, with no consistent long-term dominance of either group across the analyzed period.
- The Ridge Regression model demonstrated strong explanatory power, achieving an R^2 score of 0.81 on the test set and substantially outperforming a baseline model based on the historical mean. This suggests that seasonal performance is better explained by a combination of market value dynamics and performance indicators rather than coach nationality alone.
- Prediction results showed close alignment between actual and estimated points per game in recent seasons, reinforcing the model's ability to capture meaningful relationships between squad value trends, on-field metrics, and overall performance.

Together, these findings indicate that while coach nationality is often emphasized in public discourse, it is not a statistically decisive factor in Fenerbahçe's seasonal success. Instead, team performance appears to be driven by a broader set of contextual and performance-related variables, highlighting the importance of data-driven evaluation in managerial decision-making.

8. Limitations & Future Work

While this project provides a structured and data-driven evaluation of managerial performance at Fenerbahçe, several limitations should be acknowledged when interpreting the results.

Limitations

- The analysis is conducted at the **season level**, resulting in a relatively small sample size. This limits the statistical power of hypothesis testing and restricts the complexity of machine learning models that can be reliably applied.
- Coach performance is evaluated using aggregate seasonal metrics such as points per game and win rate. These measures do not fully capture contextual factors such as injuries, fixture congestion, squad rotation, or in-season managerial changes, which may influence performance outcomes.
- Some explanatory variables used in the regression model, including win rate and goal-related metrics, are inherently linked to season outcomes. As a result, the machine learning model should be interpreted primarily as **explanatory rather than strictly predictive**.
- The study focuses on a single club, which limits the generalizability of the findings. Results observed for Fenerbahçe may not directly extend to other clubs or leagues with different competitive and organizational structures.

Future Work

- Future studies could expand the dataset by including multiple clubs from the same league or across different leagues, enabling more robust cross-sectional comparisons and improving statistical power.
- Incorporating **match-level data** would allow for more granular analysis of managerial impact and facilitate the use of advanced time-series or panel data methods.
- Additional contextual variables such as player injuries, transfer expenditures, wage budgets, or tactical indicators could further enrich the analysis and improve model interpretability.
- More advanced causal inference techniques could be applied to better isolate the effect of coach nationality from confounding factors, providing deeper insights into managerial decision-making in football.

Overall, addressing these limitations would allow future research to build more comprehensive, generalizable, and causally robust models of football performance analysis.

---

9. Technology Stack

The project was implemented using Python and widely adopted data science libraries. Data collection, preprocessing, exploratory analysis, hypothesis testing, and machine learning modelling were conducted in **Jupyter Notebook**.

- **Programming Language:** Python
- **Data Manipulation:** pandas, numpy
- **Visualization:** matplotlib, seaborn

- **Statistical Analysis:** scipy
- **Machine Learning:** scikit-learn
- **Development Environment:** Jupyter Notebook

This technology stack enabled efficient data handling, reproducible analysis, and clear presentation of results throughout the project.

---

10. Project Timeline

This section outlines the project's lifecycle from the initial proposal to final submission, detailing the temporal distribution of tasks. It highlights the iterative nature of the research, demonstrating how foundational work paved the way for subsequent discoveries.

Phase	Description	Deadline
Phase 1	Project proposal	Oct 31, 2025
Phase 2	Data collection & EDA & Hypothesis Testing	Nov 28, 2025
Phase 3	ML implementation	Jan 2, 2026
Phase 4	Final report & submission	Jan 9, 2026