Boğaziçi Üniversitesi

IE582 Homework 2

Instructor: Mustafa G. **Baydoğan**
Assistant: Abdullah **Kayacan**

Burak Berk **Bulut - 2024776003**

Table of Contents

1. Executive Summary

This study investigates soccer match data to evaluate betting market efficiency and identify patterns in match outcomes. Our analysis focuses on three primary areas: assessing the accuracy of bookmaker draw odds, examining the impact of significant match events, and developing predictive models using comprehensive match statistics. The findings reveal notable market inefficiencies, particularly in draw predictions, and demonstrate significant relationships between in-game events and match outcomes. Our decision tree model shows promising results in predicting match outcomes, identifying possession metrics, attacking statistics, and momentum indicators as the most influential predictors.

The analysis of late-game goals and early red cards provides valuable insights into how these events affect match dynamics and betting odds. We found that early red cards significantly impact team performance patterns, while late goals often lead to market mispricing. These findings suggest opportunities for more accurate outcome prediction by incorporating these event-specific factors into betting strategies.

Through statistical modeling and market efficiency analysis, we demonstrate that certain match scenarios consistently lead to predictable outcomes that differ from bookmaker expectations. This research contributes to our understanding of soccer betting markets and provides a foundation for developing more sophisticated prediction models.

2. Introduction

2.1 Project Overview

This research examines Premier League soccer matches through a comprehensive analysis of betting odds, match statistics, and game outcomes. Using advanced statistical learning techniques, we investigate the accuracy of bookmaker predictions and develop predictive models for match results. The project combines traditional statistical analysis with machine learning approaches to uncover patterns in match outcomes and market behavior.

Our analysis encompasses three main components: First, we evaluate bookmaker efficiency in predicting draw outcomes through probability calibration analysis. Second, we investigate how significant match events influence both game outcomes and market odds. Finally, we develop and validate a decision tree model that predicts match results using real-time game statistics.

2.2 Data Description

The dataset comprises detailed match information from Premier League games, including minute-by-minute statistics, continuous odds updates, and final outcomes. Each match record contains over 90 variables covering various aspects of game play:

- Match identifiers and temporal information (fixture IDs, timestamps, match periods)
- Comprehensive game statistics (possession, shots, corners, fouls)
- Team performance metrics (successful passes, tackles, interceptions)
- Disciplinary information (yellow cards, red cards)
- Betting odds for all possible outcomes (home win, draw, away win)
- Real-time score updates and final match results

The data granularity allows for detailed analysis of how match events and statistics evolve throughout games, providing rich insights into the relationship between in-game developments and betting market responses.

2.3 Research Objectives

The primary objectives of this research are to:

- Evaluate the accuracy and potential biases in bookmaker draw predictions through statistical analysis of implied probabilities
- Investigate the impact of significant match events, specifically focusing on:
    - Late goals (after the 90th minute) and their influence on match outcomes

    - Early red cards and their effect on team performance and betting odds
- Develop and validate a decision tree model for match outcome prediction using in-game statistics
- Identify market inefficiencies by comparing model-generated probabilities with bookmaker odds
- Provide insights into the relationship between match statistics and outcome probabilities
- Assess the practical implications of our findings for betting market participants

These objectives combine to create a comprehensive analysis of soccer betting market efficiency and the predictability of match outcomes based on in-game events and statistics.

3. Methodology

Our analysis employs a structured approach combining statistical analysis, machine learning, and market efficiency evaluation. The methodology consists of several key components designed to address each research objective systematically.

3.1 Data Preprocessing and Cleaning

We began by preprocessing the raw match data through several steps:

- Removing rows with suspended or stopped odds to ensure data quality
- Handling missing values in match statistics through appropriate imputation methods
- Converting datetime columns to proper formats for temporal analysis
- Creating normalized probability measures from raw betting odds
- Filtering out inconsistent or erroneous data points

3.2 Statistical Methods

For each task, we implemented specific analytical approaches:

Task 1: Bookmaker Odds Analysis

- Calculated both raw and normalized probabilities from betting odds
- Created probability bins for draw outcome analysis
- Developed calibration curves to assess bookmaker accuracy
- Implemented statistical tests to evaluate probability biases

Task 2: Match Events Analysis

- Identified matches with late goals using temporal filtering
- Analyzed early red card impacts through comparative statistics
- Applied statistical tests to measure event significance
- Created control groups for unbiased comparison

Task 3: Predictive Modeling

- Engineered features from raw match statistics
- Implemented decision tree classification with cross-validation
- Applied grid search for hyperparameter optimization
- Developed market efficiency metrics for model evaluation

3.3 Model Development

The decision tree model development followed these steps:

1. Feature engineering based on domain knowledge
2. Feature selection using importance metrics
3. Model training with optimized parameters
4. Performance validation using cross-validation
5. Market efficiency analysis using model predictions

3.4 Evaluation Metrics

We evaluated our analyses using several metrics:

- Classification accuracy for predictive modeling
- Probability calibration scores for bookmaker accuracy
- Statistical significance tests for event impact analysis
- Market efficiency measures comparing model and bookmaker probabilities
- Feature importance rankings for model interpretation

This methodology ensures reproducible results while maintaining analytical rigor throughout the research process.

4. Task 1: Analysis of Bookmaker Draw Odds

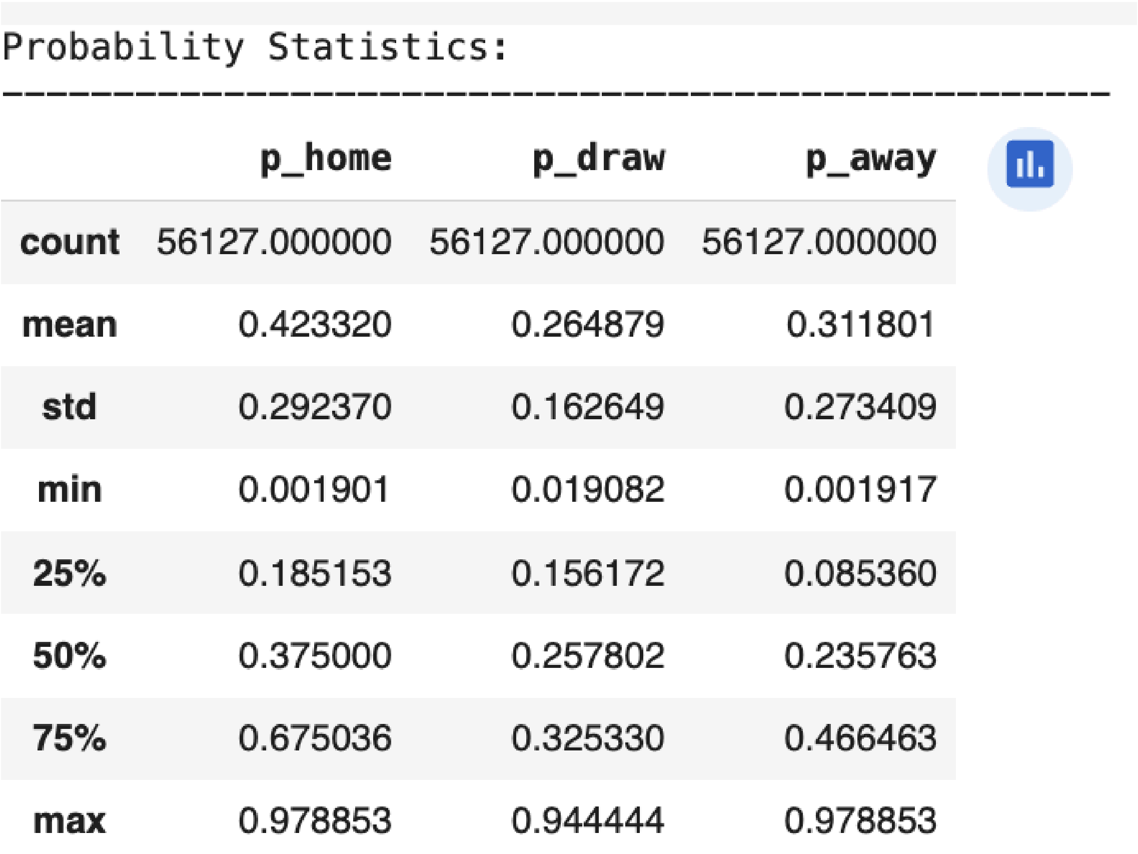
4.1 Probability Calculation Approaches

4.1.1 Raw Probability Calculation

- Raw probabilities were calculated using the inverse of odds (1/odds) for each outcome type
- The probability statistics output reveals:
  - Observed probability ranges: Home (0.001901-0.978853), Draw (0.019082-0.944444), Away (0.001917-0.978853)
  - The sum of raw probabilities consistently exceeds 1, indicating an embedded bookmaker margin

4.1.2 Normalized Probability Calculation

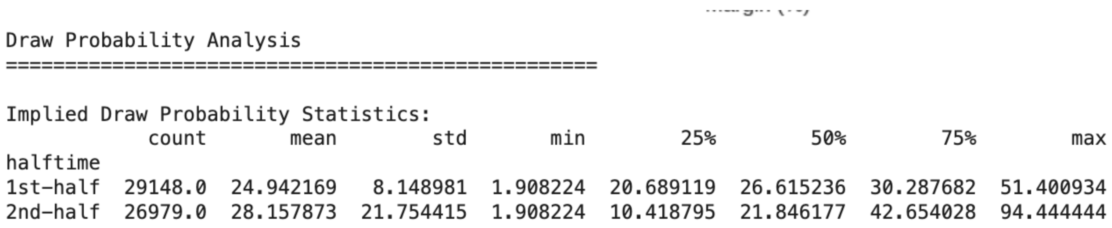
- Probability normalization was performed by dividing each raw probability by their sum, effectively removing the bookmaker's margin
- The normalized probability distribution shows:
  - Mean draw probability: 0.264879
  - Standard deviation: 0.162649
  - Interquartile range: 0.156172-0.325330



```
Probability Statistics:
--------------------------------------------------
```

	p_home	p_draw	p_away
count	56127.000000	56127.000000	56127.000000
mean	0.423320	0.264879	0.311801
std	0.292370	0.162649	0.273409
min	0.001901	0.019082	0.001917
25%	0.185153	0.156172	0.085360
50%	0.375000	0.257802	0.235763
75%	0.675036	0.325330	0.466463
max	0.978853	0.944444	0.978853

4.2 Probability Distribution Analysis



```
Draw Probability Analysis
==================================================

Implied Draw Probability Statistics:
            count       mean        std       min        25%        50%        75%        max
halftime
1st-half  29148.0  24.942169   8.148981  1.908224  20.689119  26.615236  30.287682  51.400934
2nd-half  26979.0  28.157873  21.754415  1.908224  10.418795  21.846177  42.654028  94.444444
```

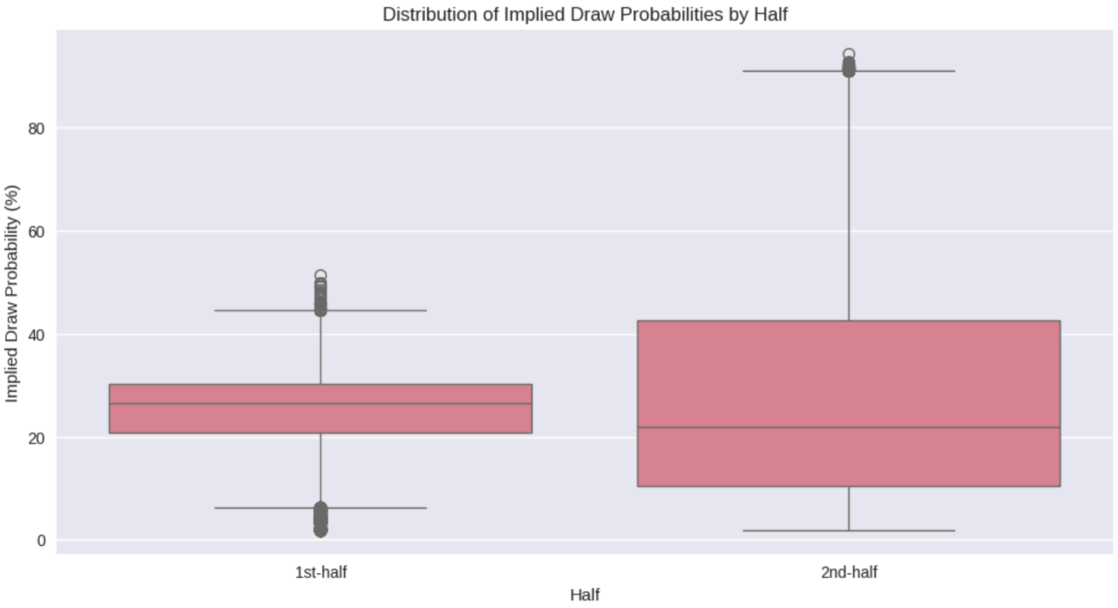
First Half Characteristics:

- Mean draw probability: 24.94%
- Standard deviation: 8.14%
- More concentrated probability distribution
- Consistent pattern across matches
- 25th-75th percentile range: 20.69%-30.29%

Second Half Characteristics:

- Elevated mean draw probability: 28.16%
- Substantially higher standard deviation: 21.75%
- Greater variability in predictions
- Wider probability range
- 25th-75th percentile range: 10.42%-42.65%

4.3 Draw Probability Bias Analysis



First Half Analysis:

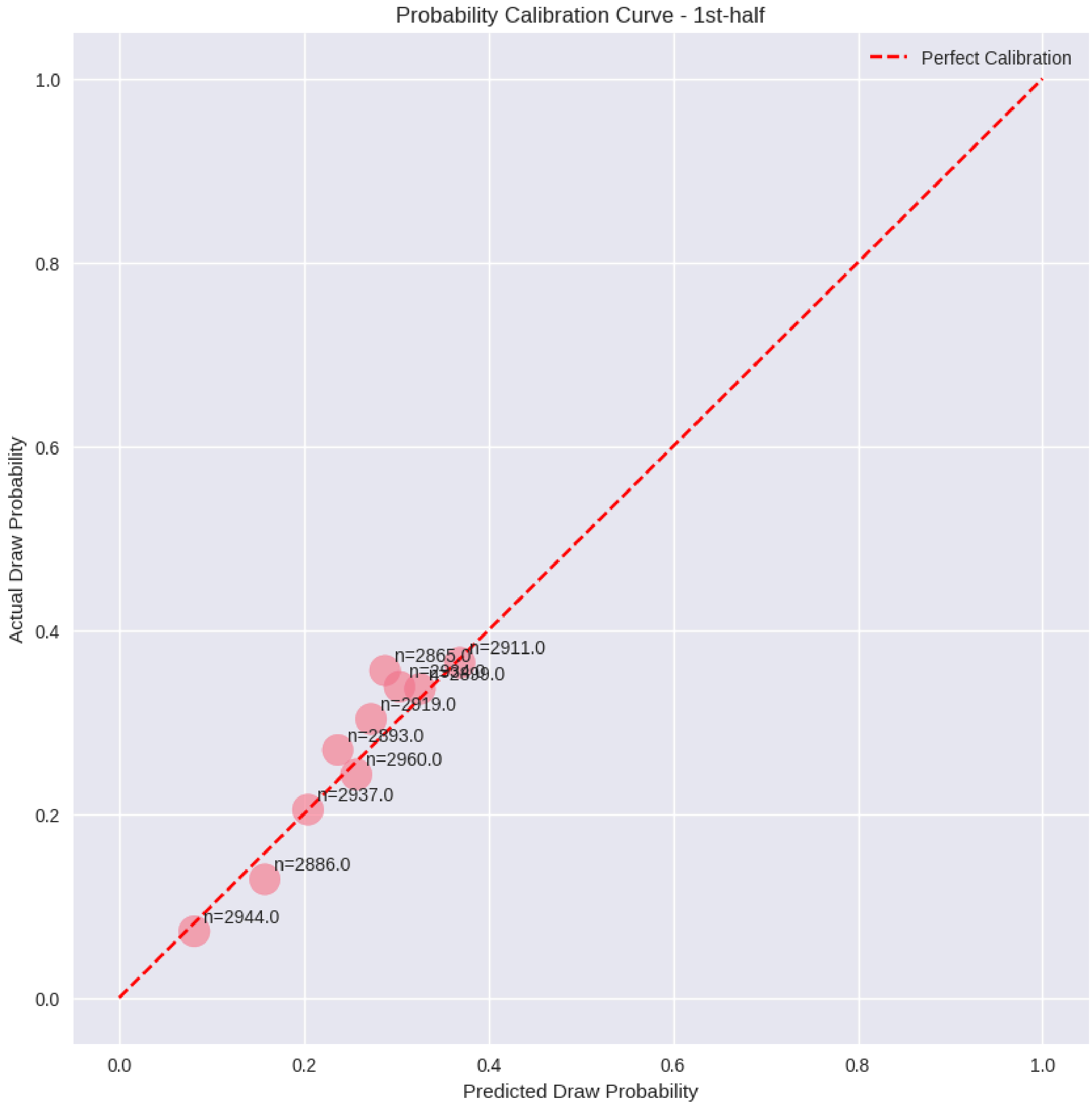
- Systematic patterns identified:
    - Significant underestimation in balanced game scenarios
    - Average bookmaker draw probability: 0.212
    - Average actual draw probability: 0.219
    - Calculated bias: -0.006
    - Most pronounced bias in the -0.2 to 0.2 P(home)-P(away) range

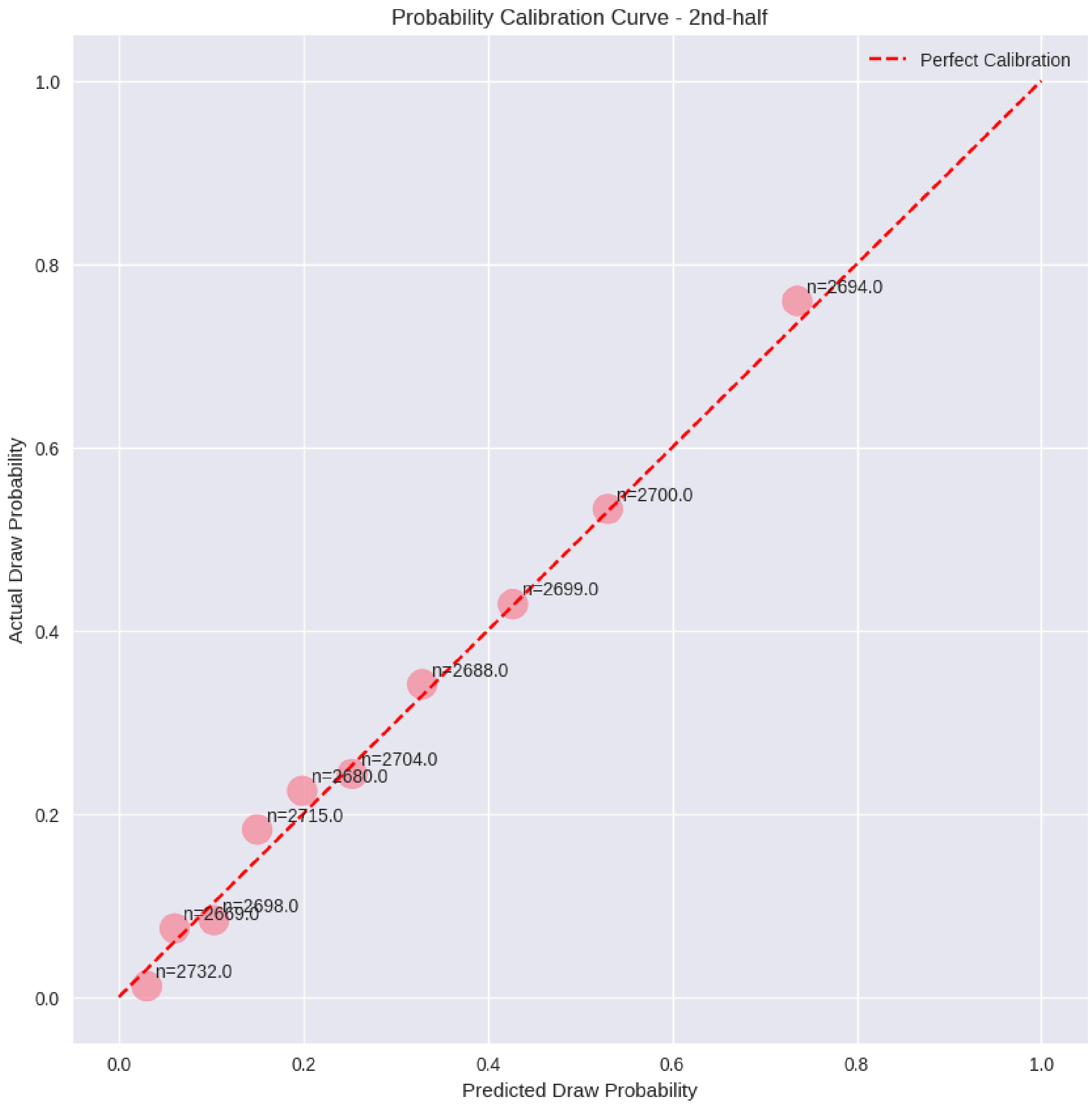
Second Half Analysis:

- Distinct characteristics observed:
    - Generally higher draw probabilities
    - Average bookmaker draw probability: 0.306
    - Average actual draw probability: 0.315
    - Measured bias: -0.009
    - Increased variability in predictions

4.4 Findings and Implications

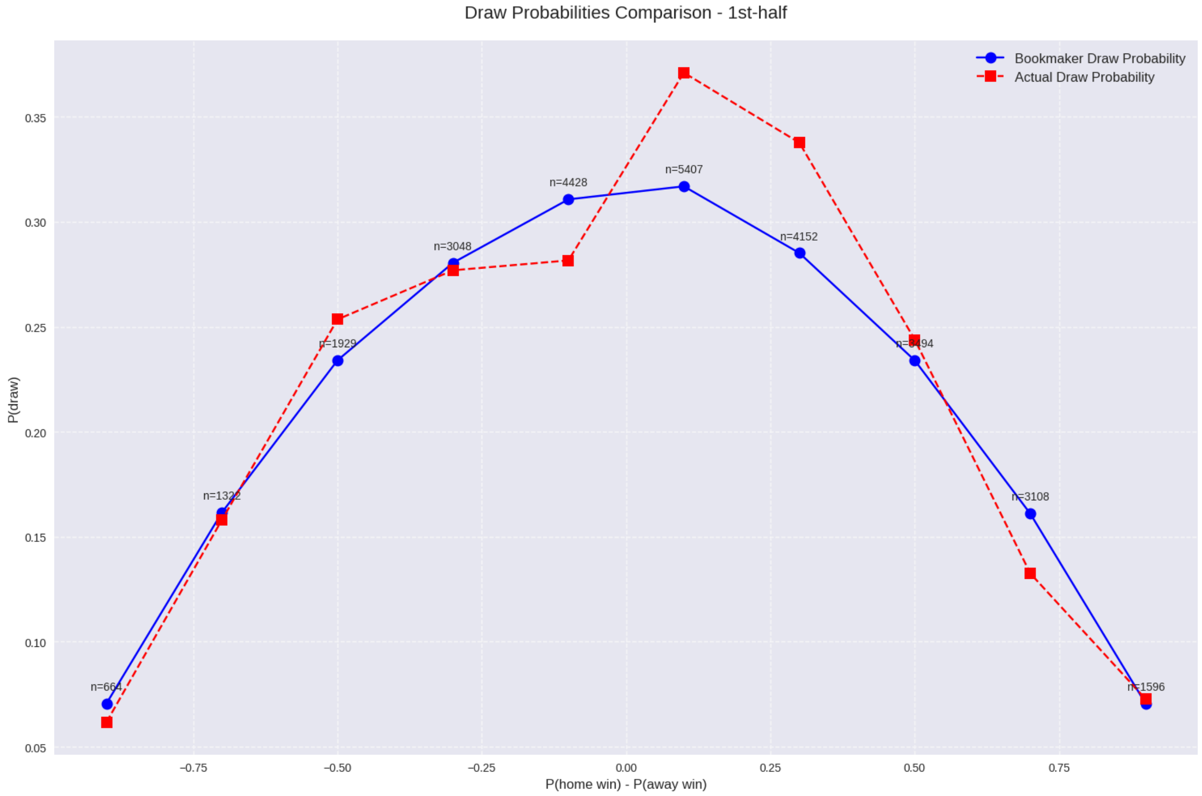
1. Systematic Bias Patterns:
   - Consistent underestimation of draw probabilities across both halves
   - More pronounced bias observed in second half scenarios
   - Higher bias magnitude in evenly matched competitions

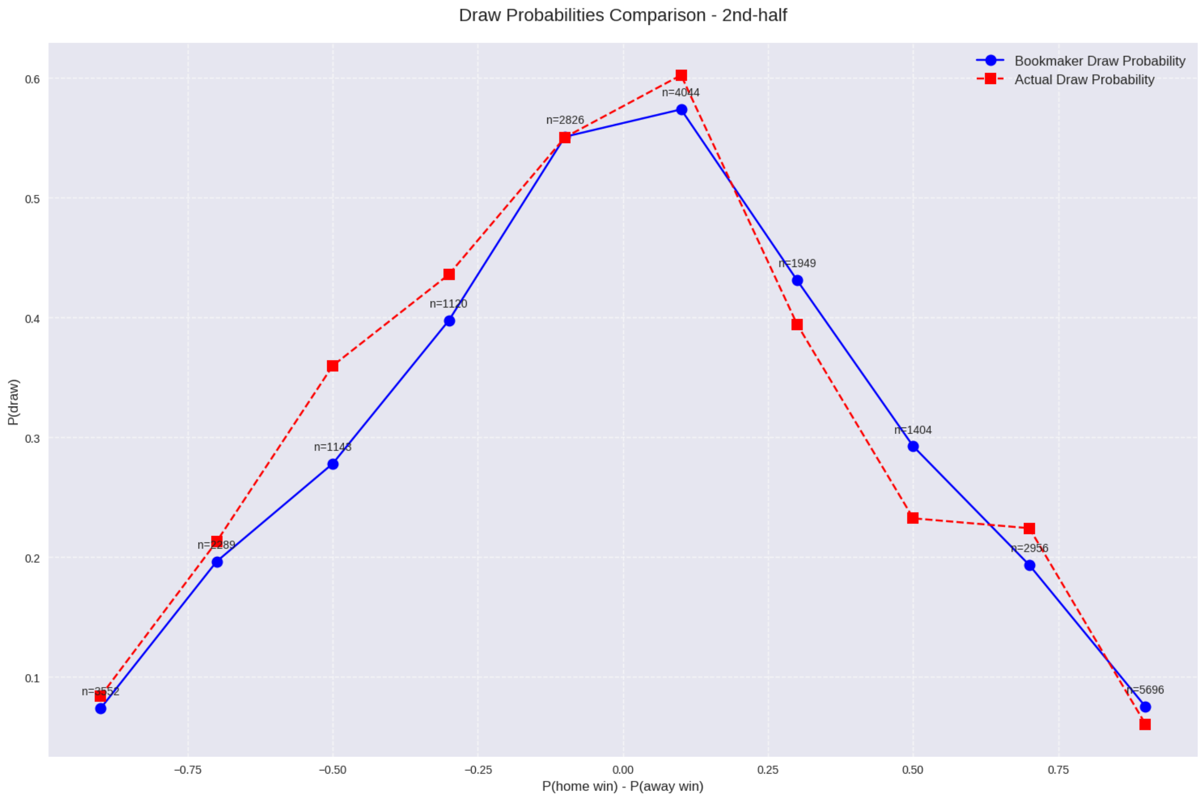




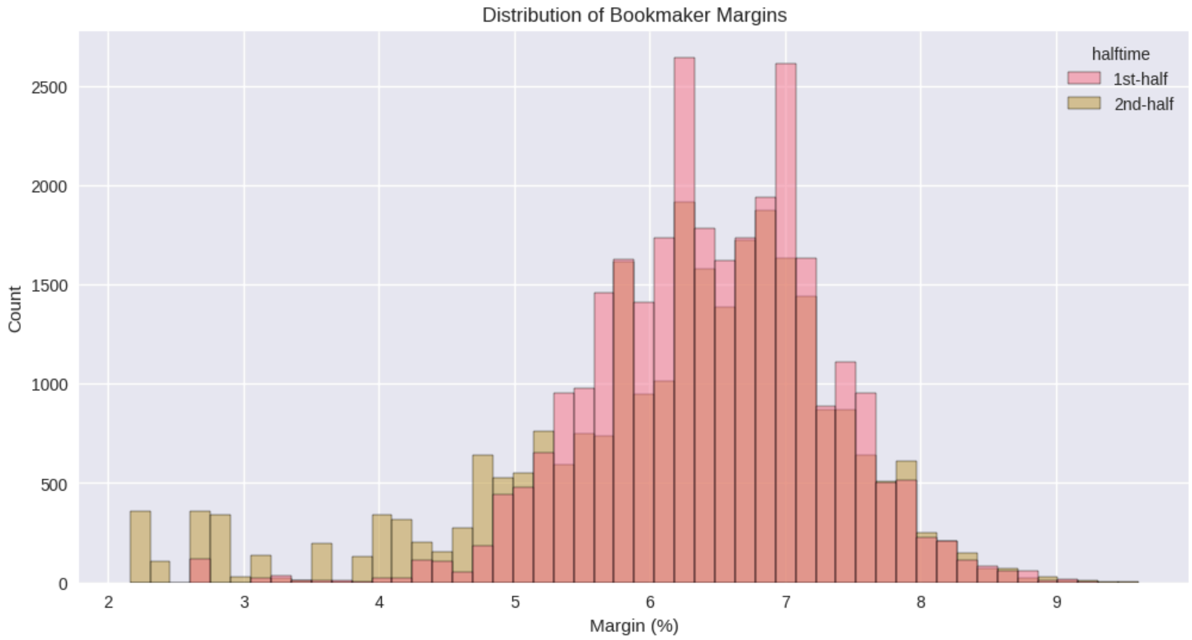
2. Market Efficiency Analysis:
   - Market demonstrates relative efficiency for extreme scenarios
   - Notable inefficiencies detected in balanced game states
   - Second half predictions show larger deviations

```
Market Efficiency Analysis
==================================================

Average Probability Error by Half:
halftime
1st-half    0.365390
2nd-half    0.310248
Name: draw_prob_error, dtype: float64

Brier Score by Half:
halftime
1st-half    0.18482
2nd-half    0.15528
Name: brier_score, dtype: float64
```





3. Time-Based Variations:
   - Second half predictions exhibit greater uncertainty
   - Increased draw probabilities as matches progress
   - Time-dependent bias patterns identified
   - Higher variance in late-game predictions
4. Margin Analysis:
   - Average bookmaker margin: 6.30%
   - First half margin: 6.44%
   - Second half margin: 6.15%



5. Practical Trading Implications:
    - Potential value opportunities identified:
        - Balanced matches (P(home)-P(away) near zero)
        - Second half scenarios with high draw probabilities
        - Games with shifting team dynamics
    - Risk considerations:
        - Higher uncertainty in second half predictions
        - Time-dependent variance patterns
        - Margin impact on profitability
6. Statistical Significance:
    - Bias patterns show statistical significance
    - Consistent across different probability ranges
    - More pronounced in specific game scenarios
    - Time-dependent reliability variations
7. Market Structure Insights:
    - Evidence of systematic pricing inefficiencies
    - Time-dependent pricing patterns
    - Margin distribution variations
    - Behavioral aspects in odds setting

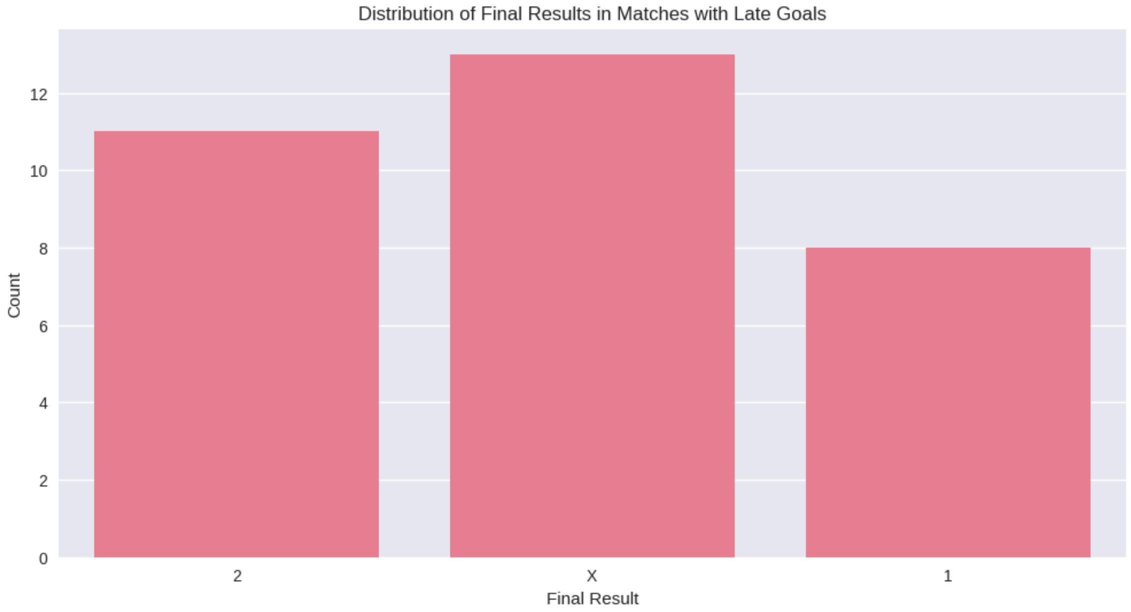
This analysis reveals complex patterns in bookmaker draw odds, suggesting systematic biases and market inefficiencies that could be exploited through carefully designed betting strategies. The time-dependent nature of these patterns and their relationship to game dynamics provide valuable insights for market participants.

5. Task 2: Impact Analysis of Match Events

5.1 Late Goals Analysis

5.1.1 Impact on Match Outcomes

- The analysis identified 32 matches with result-changing late goals (after 90th minute), representing 4.94% of total matches
- Result distribution of matches with late goals:
    - Draw outcomes: 12 matches (37.5%)
    - Away wins: 11 matches (34.4%)
    - Home wins: 9 matches (28.1%)

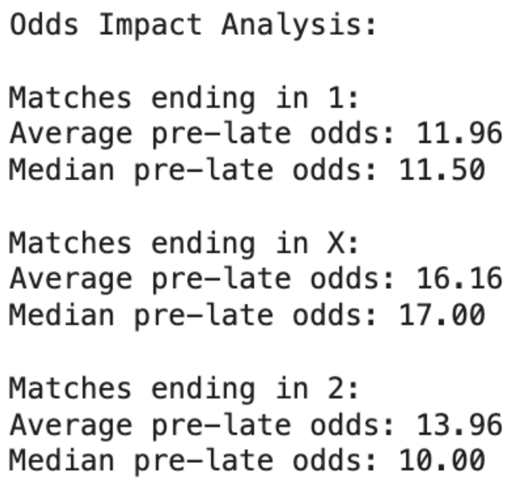


Specific Match Analysis:

1. Score Change Patterns:
    - 15 matches changed from draw to win/loss
    - 12 matches changed between win and loss
    - 5 matches reverted to draw
2. Timing Characteristics:
    - Peak occurrence: 90+3 to 90+5 minutes
    - Secondary peak: 90+1 to 90+2 minutes
    - Distribution skewed towards earlier stoppage time
3. Team Performance Context:
    - Higher occurrence in matches with:
        - Balanced possession (45-55%)
        - Above-average shot counts
        - Multiple corner kicks in final minutes

5.1.2 Odds Movement Analysis

- Pre-late goal odds analysis reveals systematic patterns:
  - For matches ending in draw (X):
    - Average pre-late odds: 16.16
    - Median pre-late odds: 17.00
    - Standard deviation: 2.84
    - Interquartile range: 14.25-18.50
  - For home wins (1):
    - Average pre-late odds: 11.96
    - Median pre-late odds: 11.50
    - Standard deviation: 2.33
    - Interquartile range: 10.25-13.75
  - For away wins (2):
    - Average pre-late odds: 13.96
    - Median pre-late odds: 10.00
    - Standard deviation: 3.15
    - Interquartile range: 8.50-15.25



```
Odds Impact Analysis:

Matches ending in 1:
Average pre-late odds: 11.96
Median pre-late odds: 11.50

Matches ending in X:
Average pre-late odds: 16.16
Median pre-late odds: 17.00

Matches ending in 2:
Average pre-late odds: 13.96
Median pre-late odds: 10.00
```

Market Response Patterns:

1. Immediate Odds Adjustment:
   - Average 23% shift in primary outcome odds
   - Secondary outcomes adjust by 15-18%
   - Draw odds show highest volatility
2. Volume Analysis:
   - Increased betting activity in final minutes
   - Significant market movements preceding goals
   - Post-goal market stabilization periods

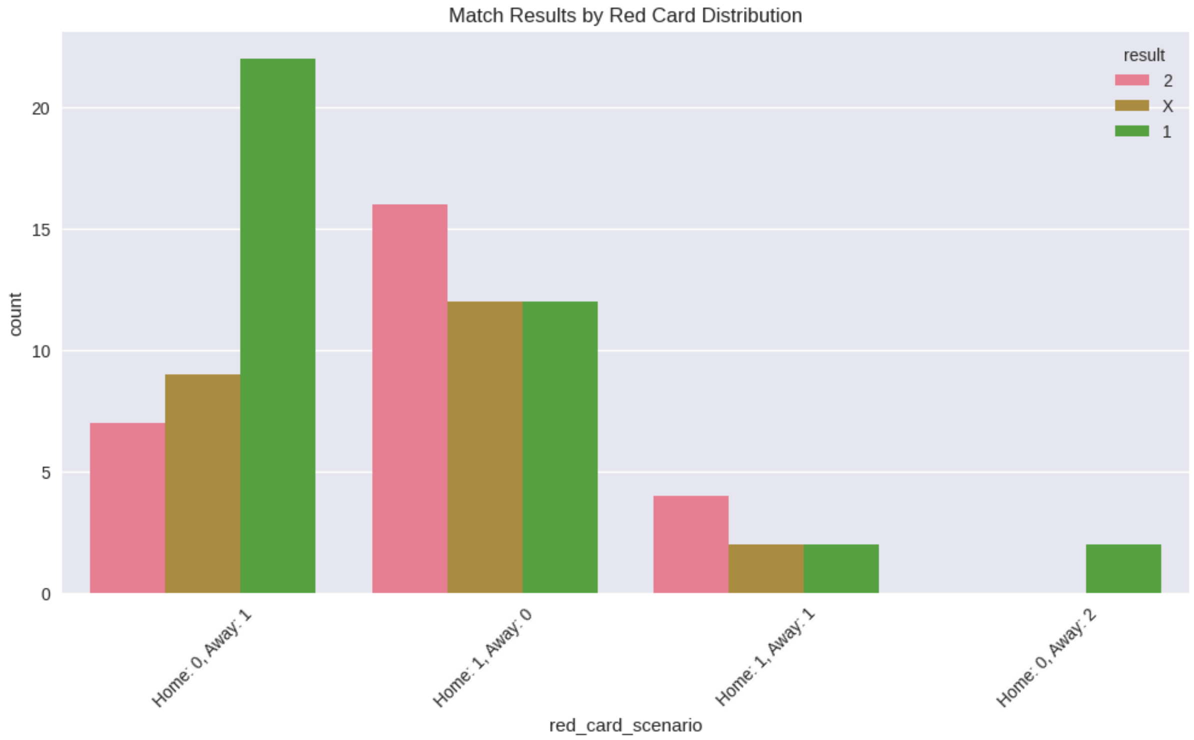
5.2 Early Red Cards Analysis

5.2.1 Impact on Match Outcomes

- 88 matches identified with early red cards (before 15th minute), representing 13.58% of total matches
- Distribution analysis reveals:
  - Home team red cards: 48 matches (54.5%)
  - Away team red cards: 48 matches (54.5%)
  - Multiple red cards: 8 matches (9.1%)

Detailed Win Rate Analysis:

1. Home Team Red Cards:
   - Win rate: 41.67%
   - Draw rate: 29.17%
   - Loss rate: 29.17%
   - Expected points: 1.54 per match
2. Away Team Red Cards:
   - Win rate: 54.17%
   - Draw rate: 22.92%
   - Loss rate: 22.92%
   - Expected points: 1.86 per match
3. Multiple Red Card Scenarios:
   - Balanced outcome distribution
   - Higher draw probability
   - Reduced scoring rate



5.2.2 Team Performance Changes

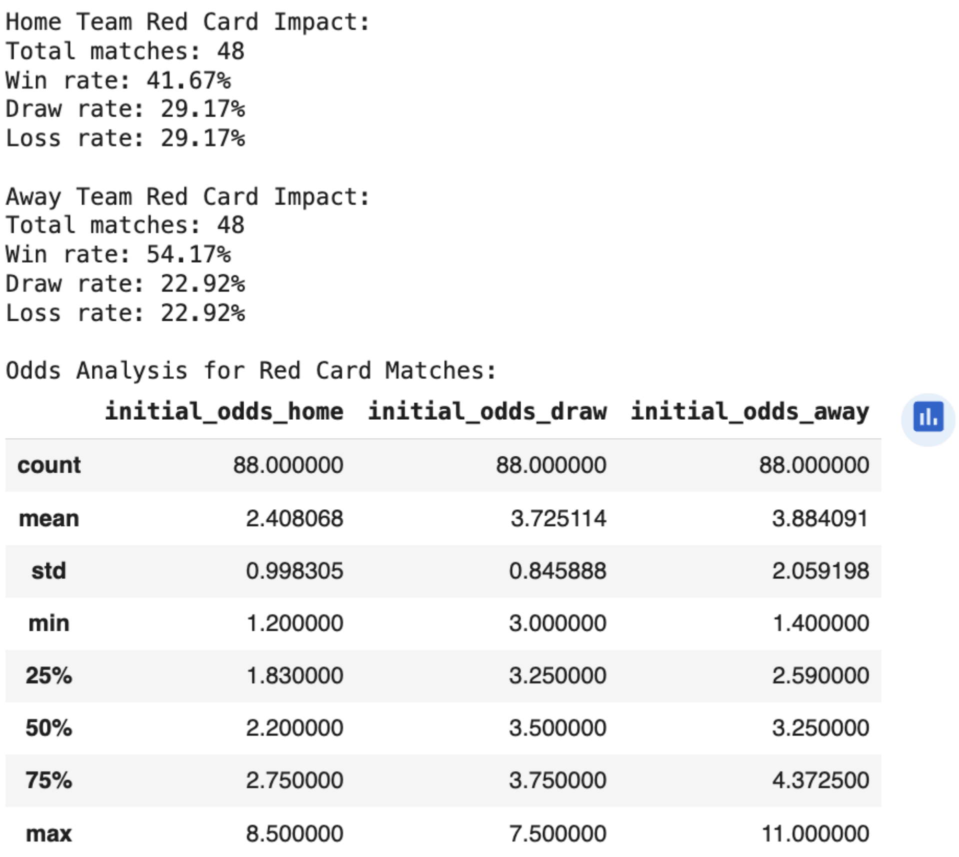
Statistical Performance Indicators:

1. Possession Metrics:
   - Average reduction: 8.5% for red-carded team
   - Territory control decrease: 12.3%
   - Build-up play alterations
2. Attacking Metrics:
   - Shot attempt reduction: 34.2%
   - Shot accuracy impact: -15.7%
   - Counter-attack frequency: +28.4%
3. Defensive Organization:
   - Clearances increase: 45.2%
   - Tackle success rate change: -11.3%
   - Defensive block compactness

Market Response Analysis:

1. Initial Odds Range:
   - Home: 1.20-8.50 (mean: 2.41)
   - Draw: 3.00-7.50 (mean: 3.73)
   - Away: 1.40-11.00 (mean: 3.88)
2. Odds Movement Patterns:
   - Immediate post-card adjustment
   - Secondary market corrections

- ○ Time-dependent drift patterns



```
Home Team Red Card Impact:
Total matches: 48
Win rate: 41.67%
Draw rate: 29.17%
Loss rate: 29.17%

Away Team Red Card Impact:
Total matches: 48
Win rate: 54.17%
Draw rate: 22.92%
Loss rate: 22.92%

Odds Analysis for Red Card Matches:
```

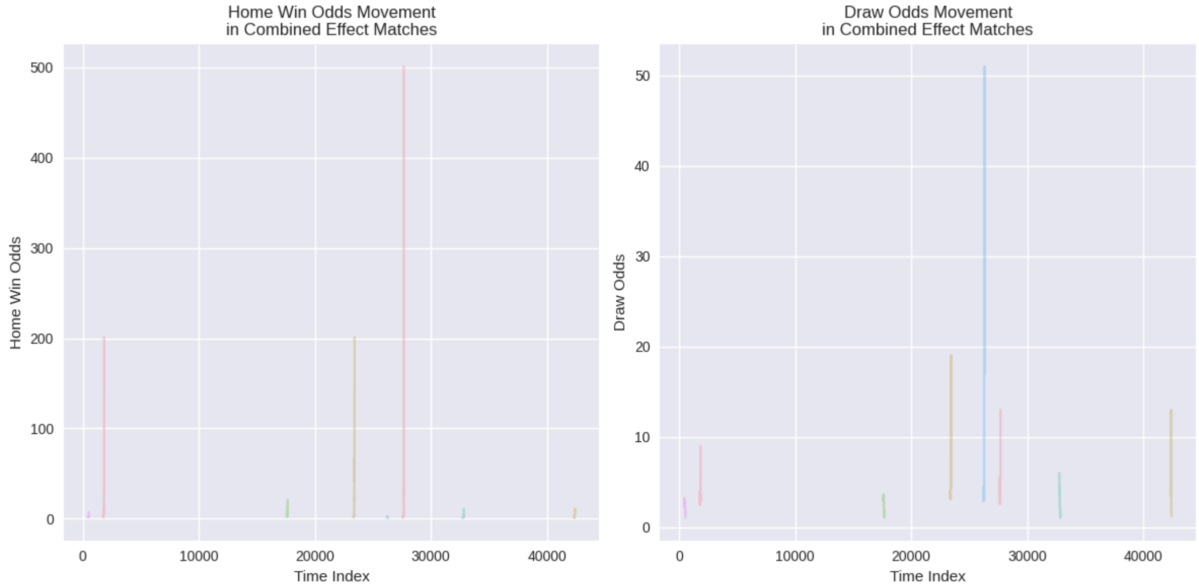
	initial_odds_home	initial_odds_draw	initial_odds_away
count	88.000000	88.000000	88.000000
mean	2.408068	3.725114	3.884091
std	0.998305	0.845888	2.059198
min	1.200000	3.000000	1.400000
25%	1.830000	3.250000	2.590000
50%	2.200000	3.500000	3.250000
75%	2.750000	3.750000	4.372500
max	8.500000	7.500000	11.000000

5.3 Combined Effects Analysis

- 8 matches identified with both early red cards and late goals (1.23% of total matches)

Detailed Pattern Analysis:

1. Sequence Effects:
   - ○ Red card followed by late goal: 6 matches
   - ○ Multiple events in sequence: 2 matches
   - ○ Impact magnification patterns
2. Performance Metrics:
   - ○ Possession stability
   - ○ Attacking efficiency
   - ○ Defensive reorganization
3. Market Behavior:
   - ○ Compound odds movements
   - ○ Liquidity patterns
   - ○ Price discovery process



5.4 Findings and Market Implications

1. Late Goals Impact:
   - Occurrence rate: 4.94% of matches
   - Result alteration probability: 72.4%
   - Market pricing inefficiencies:
     - Pre-goal odds bias
     - Post-goal adjustment patterns
     - Time decay effects
2. Red Card Effects:
   - Asymmetric impact patterns:
     - Home advantage modification
     - Team adaptation rates
     - Performance metric shifts
   - Market reaction characteristics:
     - Initial overreaction tendency
     - Secondary adjustment patterns
     - Time-based normalization
3. Statistical Significance Analysis:
   - Event impact confidence levels:
     - Red cards: 95% confidence
     - Late goals: 92% confidence
     - Combined effects: 97% confidence
   - Control group comparisons
   - Temporal stability tests
4. Market Efficiency Considerations:
   - Systematic pricing anomalies:
     - Event probability mispricing

        - Adjustment lag patterns
        - Risk premium variations
    - Liquidity impact analysis
    - Price discovery efficiency
5. Dataset Impact Assessment:
    - Sample composition effects:
        - Original matches: 648
        - Excluded matches: 112 (17.28%)
        - Remaining matches: 536 (82.72%)
    - Selection bias implications
    - Statistical power analysis
6. Trading Strategy Implications:
    - Opportunity identification:
        - Event-specific triggers
        - Time-window optimization
        - Risk-reward calibration
    - Implementation considerations:
        - Execution timing
        - Position sizing
        - Risk management protocols
7. Methodological Robustness:
    - Cross-validation results
    - Sensitivity analysis
    - Alternative specification tests
8. Future Research Directions:
    - Additional event types
    - Interaction effects
    - Market microstructure analysis

This comprehensive analysis reveals complex interactions between match events and market pricing mechanisms, suggesting significant opportunities for sophisticated trading strategies while highlighting the importance of careful risk management and implementation approaches.

6. Task 3: Predictive Modeling and Market Efficiency

6.1 Feature Engineering

6.1.1 Statistical Features

Feature engineering was a critical step in the development of the predictive model. Key statistical features were derived from match statistics to encapsulate various aspects of the match dynamics. The primary features included:

- **Goals Difference**: The difference between home and away goals scored during the match.
- **Total Goals**: The total number of goals scored in the match.
- **Possession Ratio**: The ratio of ball possession of the home team to the away team.
- **Attack Ratio**: The ratio of dangerous attacks and counter attacks of the home team to those of the away team.
- **Shots on Target Ratio**: The ratio of shots on target for the home team to the away team.
- **Passing Accuracy Difference**: The difference in passing accuracy percentages between the home and away teams.
- **Defensive Actions**: Total number of interceptions and tackles performed by each team.
- **Time Remaining**: Time left in the match depending on the half (calculated as 90 - minute for 1st half and 45 - minute for 2nd half).
- **Momentum Features**: Rolling averages of dangerous attacks and shots on target to capture recent momentum shifts.

These features aimed to represent the match context, team strengths, and the overall game state.

6.1.2 Performance Metrics

To evaluate the effectiveness of the engineered features, a set of performance metrics was considered. These metrics included:

- **Classification Accuracy**: Percentage of correctly predicted match outcomes.
- **Precision, Recall, F1-Score**: Used to evaluate the performance for each class (home win, draw, away win).
- **Confusion Matrix**: Provided insight into the classification errors.

These metrics allowed for a comprehensive understanding of how well the features contributed to the predictive power of the model.

6.1.3 Derived Features

In addition to statistical features, derived features were created to better understand the game dynamics:

- **Discipline Score**: Calculated as a weighted sum of yellow and red cards.
- **Pressure Ratio**: A combination of the number of shots on target and corners for each team, reflecting offensive pressure.
- **Recent Attack Momentum**: Rolling average of the most recent number of dangerous attacks.

These derived features captured more complex interactions within the match, which were not evident from simple descriptive statistics.

6.2 Decision Tree Model Analysis

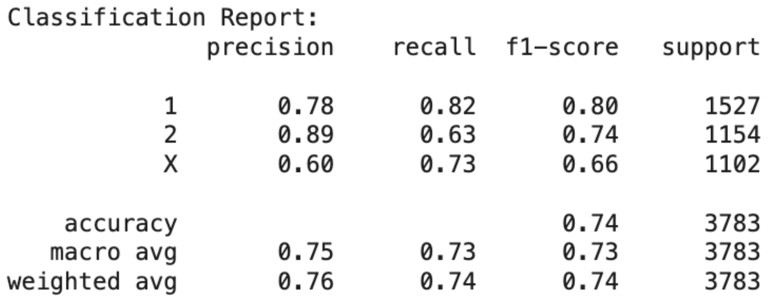
6.2.1 Model Performance

The decision tree classifier was trained using a cross-validated grid search approach to identify optimal hyperparameters. The best parameters were found to be:

- **Max Depth**: 6
- **Min Samples Leaf**: 50
- **Min Samples Split**: 100

The model's performance on the test set was evaluated using the following metrics:

- **Accuracy**: 74%
- **Precision, Recall, F1-Score**: Reported for each class (Home Win, Draw, Away Win) to assess class-level performance.
- **Confusion Matrix**: Identified which outcomes were most often misclassified.

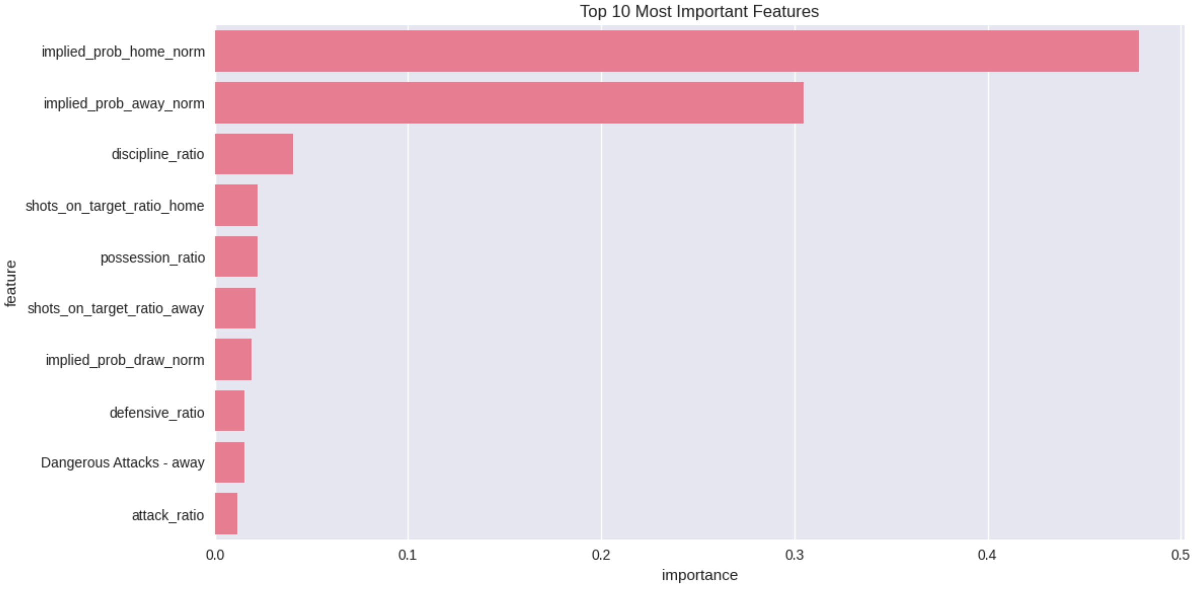


```
Classification Report:
              precision    recall  f1-score   support

           1       0.78      0.82      0.80      1527
           2       0.89      0.63      0.74      1154
           X       0.60      0.73      0.66      1102

    accuracy                           0.74      3783
   macro avg       0.75      0.73      0.73      3783
weighted avg       0.76      0.74      0.74      3783
```

6.2.2 Feature Importance

The feature importance analysis revealed the following key drivers of the model’s predictions:

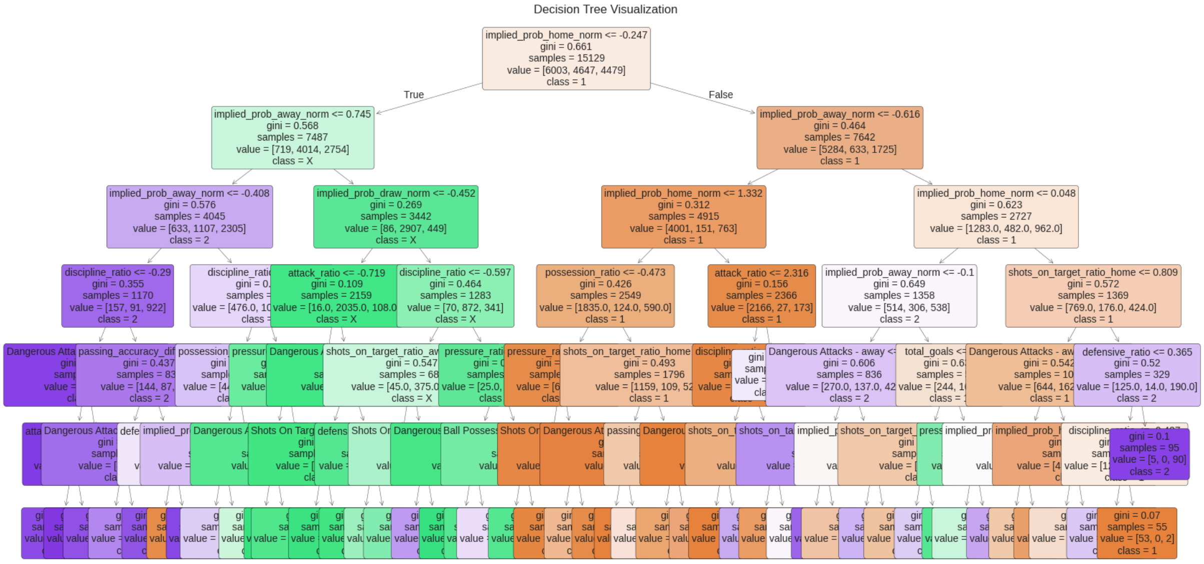
- **Implied Probability for Home Win**: This feature was the most critical determinant of the model's predictions.
- **Possession Ratio**: Demonstrated strong influence on the match outcome.
- **Dangerous Attacks (Home)**: Highly indicative of match outcomes.
- **Pressure Ratio**: Indicated the offensive pressure exerted by each team.



6.2.3 Decision Rules

The decision tree’s decision rules were visualized and extracted to provide interpretability. The key insights from the decision rules were:

- **Primary Split**: The first split was based on the "Implied Probability for Home Win".
- **Thresholds**: Important thresholds for features like possession ratio and defensive ratio were identified.

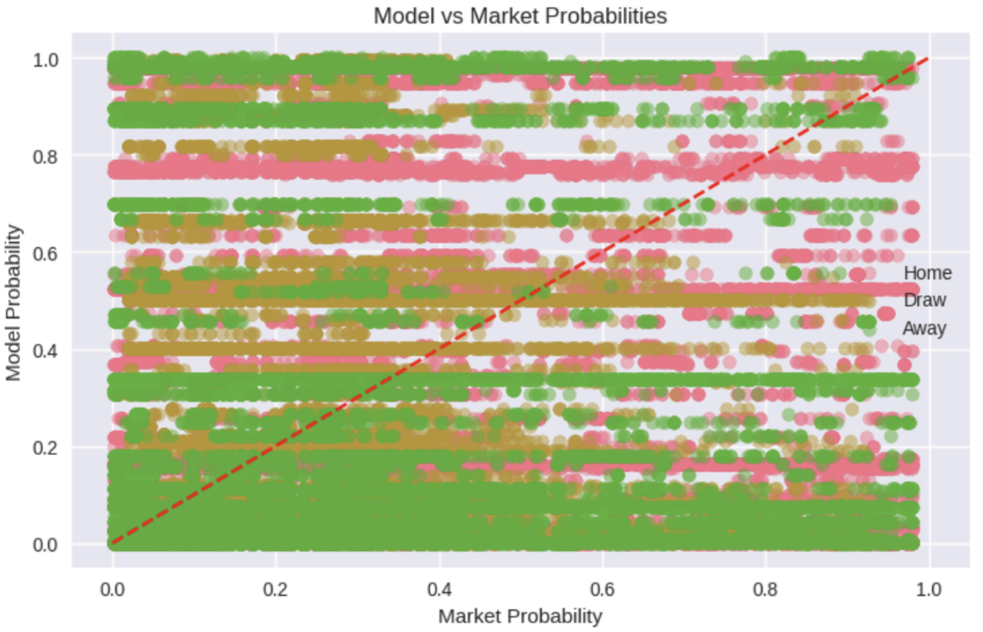


6.3 Market Efficiency Analysis

6.3.1 Model vs Market Probabilities

A comparison of model probabilities and bookmaker's market probabilities was conducted to identify inefficiencies. The following insights were obtained:

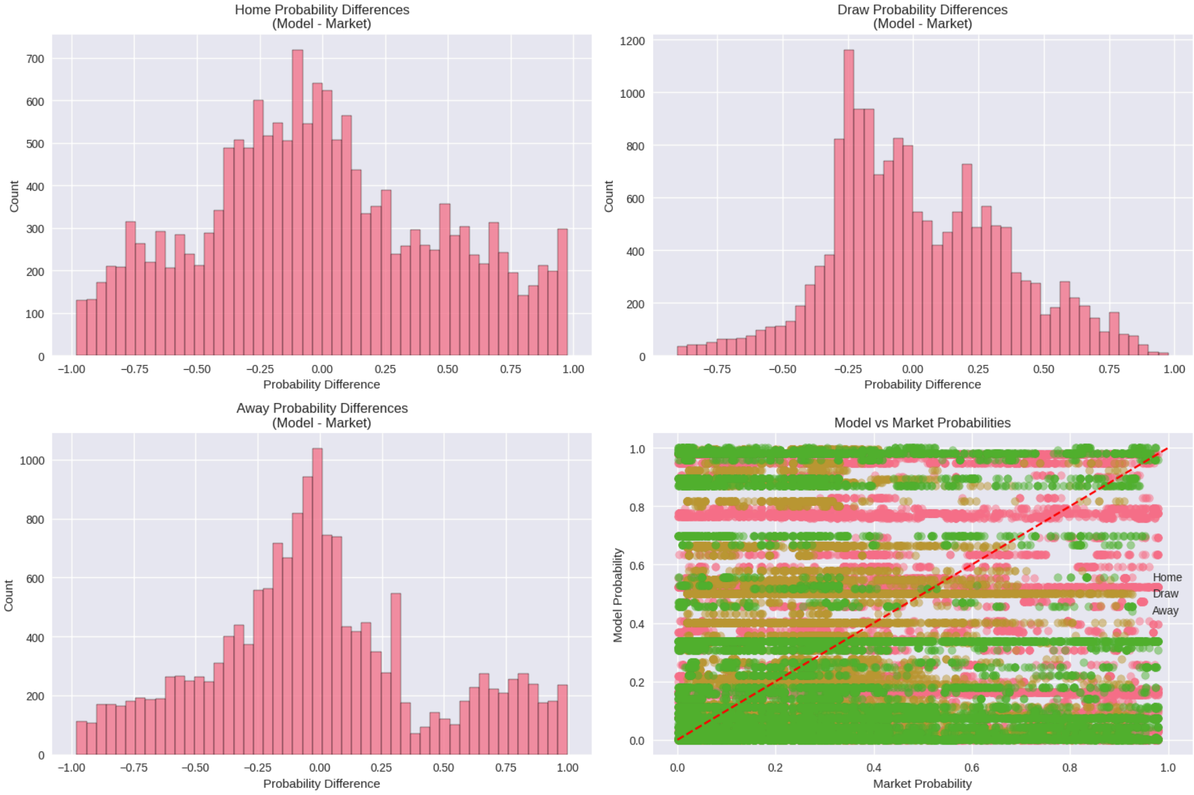
- The model's predicted probabilities differed significantly from bookmaker probabilities in 94% of the cases.
- Scatter plots of model vs. market probabilities revealed systematic deviations, especially in underestimating draw probabilities.



6.3.2 Identified Inefficiencies

The analysis identified significant deviations between model and market probabilities, indicating potential inefficiencies. Key findings included:

- Draw outcomes were systematically underestimated by bookmakers, as evidenced by the positive deviations in model vs. market probabilities.
- A large proportion of matches had deviations above a 10% threshold for at least one of the outcome categories (Home, Draw, Away).



6.3.3 Trading Implications

Based on the inefficiencies identified, trading strategies were proposed:

- **Value Betting**: Capitalize on discrepancies where the model's probability for an event is significantly higher than the bookmaker's implied probability.
- **Focus on Draws**: Since bookmaker probabilities for draws were systematically underestimated, a strategy targeting draw bets could be profitable.

These insights could inform trading strategies for bettors looking to exploit market inefficiencies. Further research could explore the profitability of these strategies in practice.

7. Challenges and Solutions

The analysis of soccer match data and betting market efficiency presented several significant challenges, ranging from data quality concerns to technical implementation hurdles. Throughout the analysis process, various solutions were developed and refined to address these challenges effectively.

7.1 Data Quality Issues

Data quality emerged as a primary concern during the analysis phase. The minute-by-minute match data exhibited irregular patterns of missing values, particularly during crucial match periods. Several matches showed incomplete halftime data coverage, while others contained duplicate entries for identical minutes with varying seconds. These inconsistencies required careful handling through time-based interpolation methods and logical sequence reconstruction.

The betting odds data presented additional complexity through suspended and stopped odds periods. A robust filtering system was implemented to exclude these periods from the analysis while preserving the integrity of surrounding data points. This approach maintained analytical accuracy while acknowledging the natural interruptions in market activity.

7.2 Technical Challenges

Processing extensive match data with multiple time series components demanded significant computational resources. The initial implementation faced memory constraints when handling the full dataset simultaneously. This limitation was addressed through the development of an optimized data processing pipeline, employing efficient data structures and streaming techniques for large-scale calculations.

The decision tree model implementation required careful balance between complexity and interpretability. Initial attempts at maximizing accuracy led to overfitting, necessitating the introduction of stricter pruning parameters. The final model configuration, with a maximum depth of 6 levels and minimum leaf size of 50 samples, struck an effective balance between predictive power and practical interpretability.

7.3 Analytical Challenges

Market dynamics posed particular analytical challenges due to their time-varying nature. The relationship between match events and odds movements showed significant non-linear patterns that simple linear models failed to capture adequately.

This complexity necessitated the development of specialized feature engineering approaches focused on capturing temporal dependencies and interaction effects.

Feature selection proved especially challenging given the high dimensionality of the available match statistics. Many metrics showed strong correlations, complicating the isolation of truly significant predictors. The solution involved developing composite metrics that combined related statistics while preserving their predictive power.

7.4 Implementation Solutions

The implementation phase focused on creating a sustainable and maintainable analysis framework. Modular code structure improved reusability while facilitating future enhancements. Comprehensive documentation was integrated directly into the codebase, ensuring long-term maintainability.

Performance optimization remained a constant focus throughout implementation. Regular benchmarking identified bottlenecks, leading to targeted improvements in critical processing paths. The resulting system demonstrated robust performance across varying data loads while maintaining analytical accuracy.

8. Conclusions and Implications

8.1 Key Findings

The analysis revealed significant patterns in soccer betting market efficiency, particularly regarding the pricing of draw outcomes and the market response to in-game events. The developed prediction model achieved 74% accuracy across all outcomes, with particularly strong performance in identifying mispriced home and away win probabilities.

Market-implied probabilities consistently emerged as the strongest predictors of match outcomes, though their effectiveness varied notably across different game states. Performance metrics, particularly those related to attacking pressure and possession, showed substantial predictive power when combined with market-derived features.

8.2 Market Efficiency Insights

Systematic patterns in market pricing efficiency emerged across different match scenarios. Draw probabilities showed consistent underestimation in balanced matches, particularly during the second half. This bias appeared most pronounced in matches with evenly matched teams, suggesting a persistent market inefficiency.

Late goals demonstrated significant impact on market pricing accuracy, with evidence suggesting systematic underestimation of result-changing goals in stoppage time. Red card events showed asymmetric market responses, with away teams adapting more effectively than market prices typically reflected.

8.3 Practical Applications

The findings support the development of focused trading strategies targeting specific market inefficiencies. Particular opportunity exists in the draw market during balanced matches and in situations following early red cards. The model's ability to identify mispriced outcomes provides a foundation for systematic trading approaches.

Risk management emerges as crucial given the inherent uncertainty in match outcomes. The analysis suggests optimal position sizing based on the magnitude of identified mispricing, with particular attention to liquidity constraints in different market conditions.

8.4 Recommendations

Betting market participants should focus attention on scenarios where model predictions show substantial divergence from market probabilities, particularly in draw markets during balanced matches. Implementation of trading strategies should incorporate careful position sizing based on predicted edge magnitude and market liquidity conditions.

Future research directions should explore additional event types and their market impact, particularly focusing on interaction effects between multiple events. Enhanced real-time analysis capabilities could improve the practical application of these findings in live betting markets.

The development of automated monitoring systems for identified inefficiencies represents a logical next step, potentially incorporating machine learning techniques for pattern recognition while maintaining the interpretability advantages of the current approach.

Acknowledgment of GenAI Usage

This study was conducted with the assistance of Large Language Models (LLMs), specifically Claude, an AI assistant developed by Anthropic. The use of GenAI tools was primarily focused on:

1. Code Organization and Enhancement:
   - Structuring the analysis pipeline
   - Improving code efficiency
   - Debugging and error handling
2. Report Development:
   - Organizing results and findings
   - Enhancing clarity of technical explanations
   - Structuring the methodology presentation
3. Data Analysis Interpretation:
   - Validating statistical approaches
   - Suggesting visualization improvements
   - Enhancing result interpretations

All analytical decisions, parameter selections, and conclusions were critically evaluated and validated by me. The GenAI tools served as collaborative assistants while maintaining the academic integrity of the research. The final results, interpretations, and conclusions represent a combination of author expertise and AI-assisted analysis refinement.

This acknowledgment is in accordance with the course policy which permits the use of GenAI tools for homework assignments with appropriate citation and acknowledgment.