

# Beating the Odds: Football Data Insights

"Some people believe football is a matter of life and death - I am very disappointed with that attitude. I can assure you it is much, much more important than that."

- Bill Shankly (1913-1981)

This project began with a familiar dream among football fans - that, with enough logic, statistics, and late-night spreadsheets, one might find the hidden pattern that beats the bookmakers. What started as curiosity soon turned into a structured experiment in probability and data analysis. Using Power BI, a Poisson-based model, and bookmaker odds, I explored the thin line between mathematical prediction and market reality. At some point, the goal shifted from profit to understanding. The deeper I went, the clearer it became that while football outcomes can be modeled, market efficiency always absorbs any edge. The illusion of beating the odds turned into a case study in how logic and transparency can thrive - even when luck wins.

## Key Insights

- Predictability exists - profitability does not. Statistical models outperform randomness, but bookmaker margins erase every gain.
- The paradox: football is chaotic in single matches yet orderly in mass - predictable in pattern, not in outcome.
- Margins neutralize advantage. Temporary edges appear in small clusters but vanish as markets self-correct.

## Objective

To test whether a statistically grounded football prediction model can achieve consistent profitability after adjusting for bookmaker margins - and, more importantly, to show how data analytics can expose order within apparent chaos.

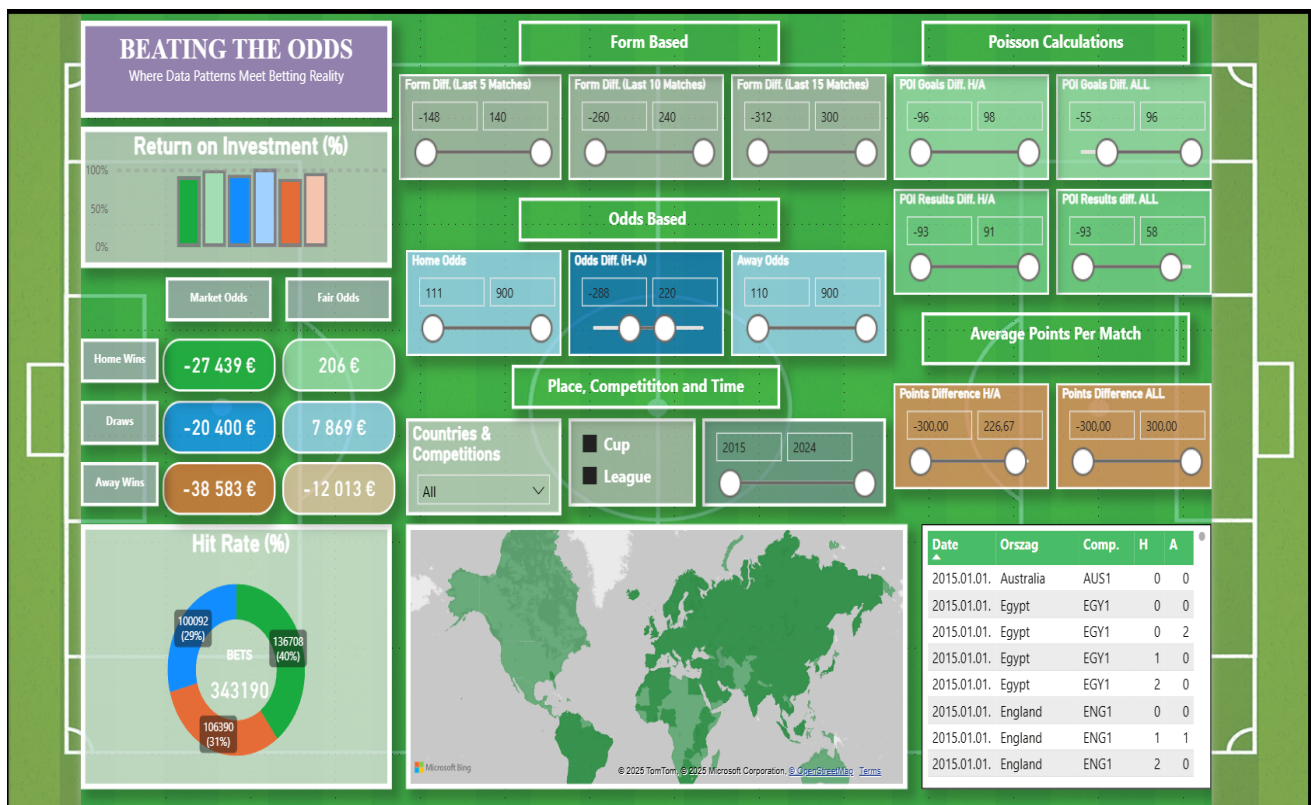
## Dataset and Methodology

- Scope: 494,274 matches (2019-2024) from nearly 300 competitions.
- Source: Primarily OddsPortal.com, cross-checked with open football databases.
- Cleaning: The heaviest phase - unifying team names, fixing missing fixtures, resolving duplicates, and aligning multilingual naming.
- Normalization: Bookmaker odds were adjusted to Fair Odds by removing overround margins so that total probabilities summed to 100%.

With these steps, the dataset became large-scale, consistent, and transparent - a rare quality in betting data projects.

## Modeling Approach

The analytical core was the Poisson distribution, a classic in football modeling that estimates goal probabilities from each team's average performance. It allows for calculating expected outcomes (Home, Draw, Away) and comparing them with bookmaker odds. I tested Bayesian refinements, Bale's expected-goal ratios, and Elo-style form factors, but Poisson proved the most stable and interpretable. Ultimately, it was not about finding the best model - but about revealing how markets behave once mathematics meets emotion.



## Dashboard Overview

The Power BI dashboard, Beating the Odds, visualizes how mathematics, form, and market pricing interact across nearly half a million matches. Main features include ROI and Hit Rate for Home, Draw, Away outcomes, Market vs Fair ROI comparisons, and interactive filters for odds range, competition, and season. In essence, it is a probability laboratory - not a betting tool, but a transparent simulator showing how statistical logic behaves in an irrational system.

## Conclusion

If betting markets operated on perfectly fair odds, the model would yield consistent returns. In the real world, margins quietly absorb every advantage. The project's true value lies not in profit but in perspective - in transforming a futile chase for winnings into a structured study of probability, logic, and market behavior. You cannot beat the odds, but you can understand them - and sometimes, that is the more meaningful victory.

## Selected Evaluations

Gemini (2025): Transforms a futile chase for profit into a structured exploration of probability and market behavior. Exemplary for scope and clarity.

Claude (2025): A rare case where failure becomes success - honest, disciplined, and insightful.

ChatGPT (2025): A blend of rigor and humility. Demonstrates real analytical maturity - understanding over prediction.

This work is free to use and share.

If so, please mention me in goodwill - like buying me a virtual beer.