Oilers Data

Sam Kluthe DSC 680

Introduction

This project analyzes historical game-level data for the Edmonton Oilers to identify patterns in team performance and to develop a predictive model for game outcomes. By incorporating contextual factors such as home versus away games, travel distance, rest between games, and recent team performance, the analysis demonstrates how sports analytics can support performance evaluation and decision-making in professional hockey.

Business Problem

Professional sports organizations operate in highly competitive environments where small performance advantages can meaningfully influence season outcomes. For NHL teams such as the Edmonton Oilers, understanding how factors like travel fatigue, scheduling, and recent performance trends relate to wins and losses can inform coaching decisions, roster management, and strategic planning. This analysis explores whether measurable contextual variables are associated with game outcomes and assesses their usefulness in a predictive modeling framework.

Background and History

The Edmonton Oilers are one of the NHL's most established franchises, competing across multiple eras of league expansion, rule changes, and scheduling formats. Over time, the league has evolved to include longer travel distances, more back-to-back games, and increased emphasis on data-driven decision-making. As a result, historical game data provides an opportunity to examine how contextual factors have influenced performance across seasons and to evaluate whether modern analytical techniques can uncover consistent patterns in wins and losses.

Data Explanation

The dataset used in this analysis consists of historical Edmonton Oilers game results compiled from multiple season-level tables. Prior to analysis, extensive data cleaning and standardization were performed. Game dates were converted to datetime format, team names were standardized to account for franchise relocations and renamings, and outcome indicators were recalculated based on goals for and goals against.

Additional features were engineered to support analysis. These include indicators for home versus away games, binary win outcomes, days of rest between games, back-to-back games,

travel distance between consecutive games using arena latitude and longitude, and accumulated travel distance during road trips. Performance-based features such as rolling ten-game win percentage and season-to-date win percentage were also derived. These steps resulted in a structured dataset suitable for exploratory analysis and modeling.

Methods

This study employed a combination of descriptive analysis and statistical modeling. Exploratory analysis was conducted using summary statistics, correlation analysis, and visualizations to examine relationships between game outcomes and contextual factors such as travel distance and rest. Win rates were evaluated across binned travel and rest categories to identify potential nonlinear patterns.

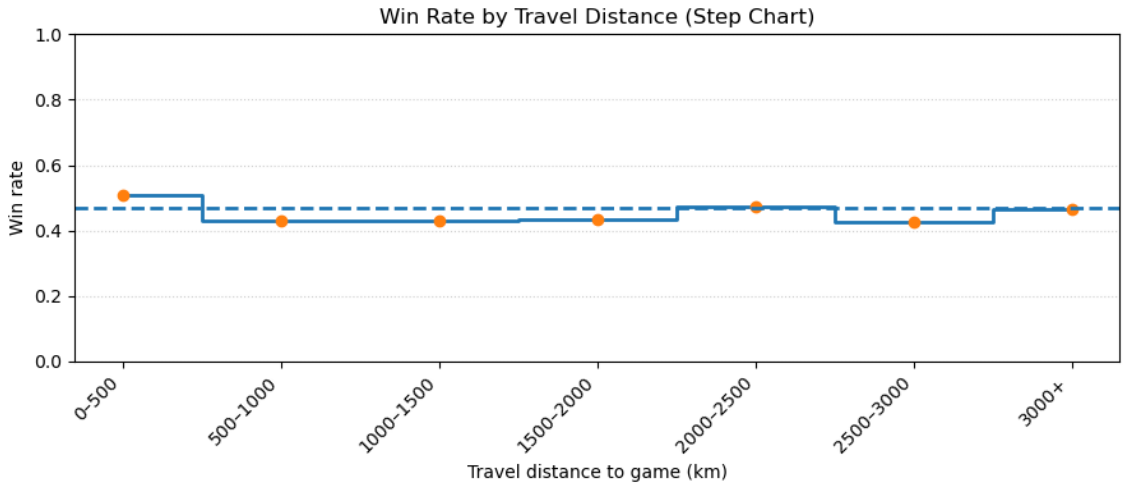
For predictive modeling, a logistic regression model was developed to estimate the probability of winning a game. The model included key features identified during exploratory analysis: home versus away status, accumulated travel distance, back-to-back games, and rolling ten-game win percentage. Model coefficients were interpreted to assess the direction and relative strength of each factor. Model fit was evaluated using pseudo R-squared and likelihood-based statistics.

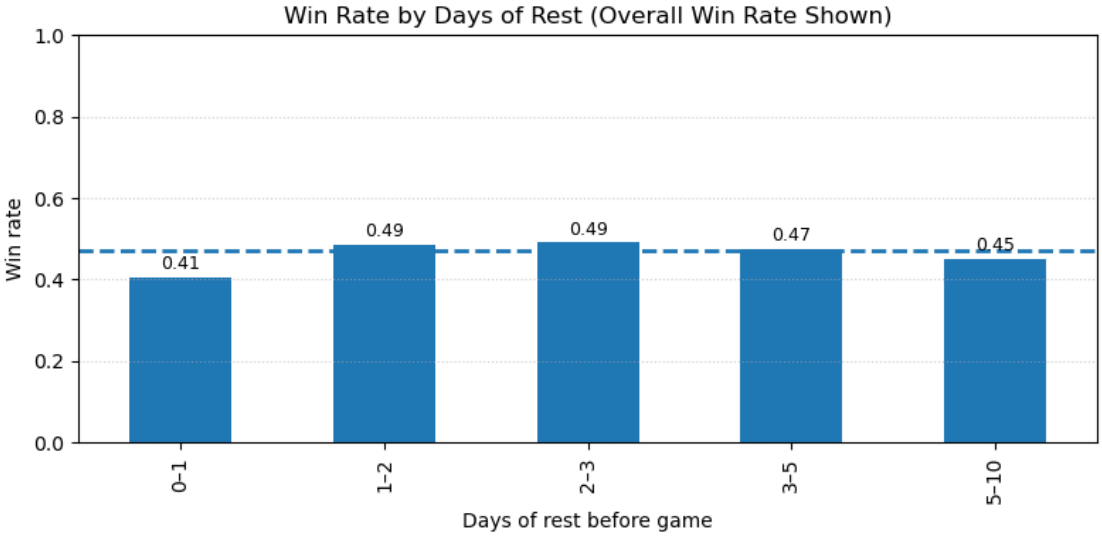
Analysis

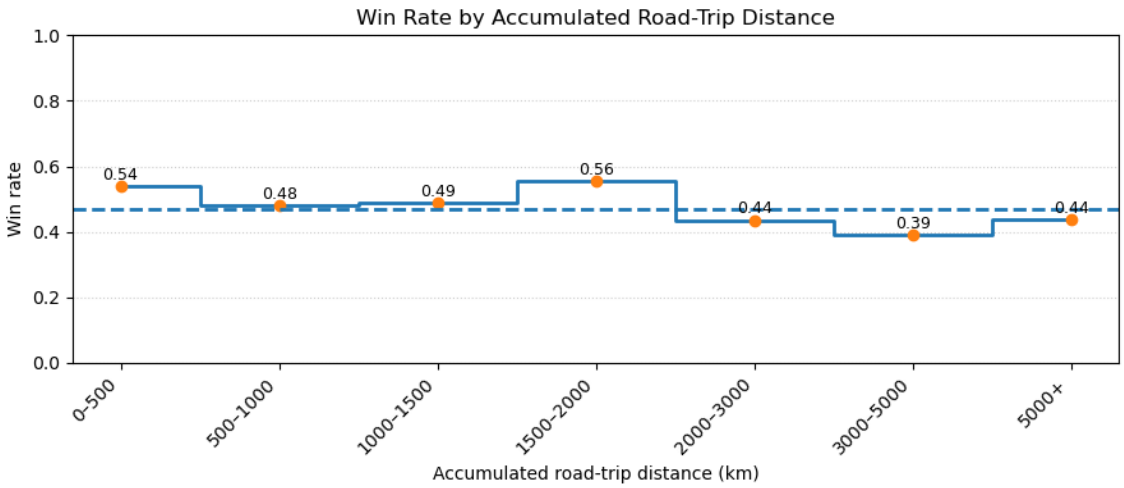
Exploratory visual analysis provides insight beyond simple correlation measures and helps explain why travel- and rest-related variables show weak linear relationships with win outcomes. While correlations between these variables and game results are small, the visualizations reveal non-linear patterns that are not captured by correlation alone, justifying the use of graphical analysis alongside modeling.

Win rate by days of rest demonstrates a clear non-linear relationship. Games played on zero or one day of rest are associated with lower win rates, while performance improves with moderate rest before leveling off or declining slightly with extended rest. This pattern explains the near-zero correlation between rest days and wins and supports the inclusion of a back-to-back indicator in the regression model. In contrast, win rates by single-game travel distance remain relatively stable across distance categories, suggesting that isolated long trips do not consistently reduce win probability.

Analysis of accumulated road-trip distance indicates that travel fatigue may develop over the course of extended road trips rather than from individual games. Win rates remain near the overall average early in road trips but decline as cumulative distance increases. Although accumulated travel distance does not emerge as a statistically significant standalone predictor in the logistic regression, the visual evidence suggests that travel effects are context-dependent. Together with the regression results, these findings highlight the value of combining exploratory visualization and multivariate modeling to identify performance patterns that would be obscured by correlation analysis alone.







Conclusion

This analysis demonstrates that while contextual factors such as travel and rest contribute marginally to game outcomes, recent team performance and home-ice advantage play a much larger role in predicting wins for the Edmonton Oilers. The logistic regression model confirms that short-term performance momentum is a strong and statistically significant predictor of success. These findings highlight the importance of combining contextual and performance-based variables when modeling outcomes in professional sports.

Assumptions, Limitations, Recommendations and Challenges

This analysis assumes that historical game records are accurate and consistently reported across seasons; however, changes in league structure, travel logistics, and rules over time may limit comparability across eras. The model is further constrained by the absence of player-level variables such as injuries, lineup changes, and goaltender selection, all of which can meaningfully influence individual game outcomes. Future applications of this work could incorporate opponent strength metrics, player-level statistics, or advanced performance indicators such as expected goals, as well as explore more flexible modeling approaches, including tree-based or ensemble methods, to better capture nonlinear relationships. Despite these limitations, the results support the continued use of recent performance indicators and schedule-related context in pre-game evaluation processes, particularly awareness of back-to-back games and cumulative fatigue. The analytical workflow developed in Python can be readily reused and automated for ongoing season analysis, with updated game data enabling regular recalculation of rolling metrics and

generation of model-based insights to support game preparation and strategic decision-making.

Ethical Assessment

This analysis relies exclusively on publicly available, non-personal data and therefore poses minimal ethical risk. Ethical considerations include responsible interpretation of results, transparency in modeling assumptions, and avoiding overstatement of predictive capabilities, particularly when communicating findings to non-technical audiences.

10 Questions

What was the main goal of this analysis?

Why look at travel and rest instead of just wins and losses?

Why didn't the correlation numbers look very strong?

Why did you focus on total road-trip travel instead of one long trip?

What does "recent performance" mean in this project?

Why do back-to-back games matter?

What factor seemed to matter most for winning?

Why didn't travel distance alone have a big impact in the model?

What are the limits of using data like this to explain sports performance?

Are there any ethical concerns when using sports data to make decisions

Appendix

Hockey-Reference.com. (n.d.). *1980–81 Edmonton Oilers schedule and results*. Sports Reference LLC. Retrieved January 13, 2026, from https://www.hockey-reference.com/teams/EDM/1981_games.html