Section A Team DS04 - Hanson Chandra (hpc12), Ssu-Ying (Stacy) Chen (sc1014),
Conor Patten (cp420), Peggy Su (ps449), Mercy Zhao (yz1063), Emma Zhu (qz171)

Predicting Weekly Player Availability and Expected Cap Dollars at Risk for the New York Giants

Abstract

This project develops a predictive modeling framework for weekly player availability for the New York Giants, translating probabilities of playing into expected weekly salary cap dollars at risk. We compare a Generalized Linear Model (GLM) and a Random Forest (RF), evaluate discrimination and calibration on a holdout season, and propose a deployment plan that informs roster and financial decisions. Our calibrated RF achieves strong performance (AUC ≈ 0.942 on 2024 holdout) and produces weekly risk metrics that support front office planning. Our proposed deployment reduces unexpected cap exposure and improves weekly roster planning,

Executive Summary

Player availability uncertainty creates costly challenges for NFL teams. Even when injured players do not play, their salaries count against the NFL salary cap. Our model predicts the probability that a player will play in a given week and converts those probabilities into expected cap dollars at risk.

We began with engineering leakage-free lag and rolling features. Subsequently, we trained both GLM and RF models, validated their performance using rolling-origin CV on 2023 data, and finalized the evaluation on a 2024 holdout season. All outputs were then calibrated using Platt scaling.

The Random Forest (RF) model demonstrated excellent discriminative power (AUC: 0.942 vs. GLM: 0.905), and its calibrated outputs accurately reflected observed outcomes. This allowed us to quantify the weekly expected salary cap risk across the roster, which revealed that Wide Receivers and Offensive Linemen drove the largest financial exposure. We then used threshold sweeps to establish clear decision rules for the front office, effectively balancing the minimization of surprise absences against undue cap exposure.

By pinpointing the cap dollars most at risk each week, the deployment of this framework allows the team to make proactive roster moves, expedite decisions regarding IR/waivers, and significantly enhance overall financial planning.

1 Business Understanding

1.1 Problem Framing

In the NFL, uncertainty about player availability is costly: teams must plan practice reps, game lineups, and financial contingencies without knowing whether certain players will be healthy enough to play. For the Giants, this uncertainty matters not only for game outcomes but also because unavailable players still count against the salary cap.

1.2 Objectives and Success Criteria

The primary objective of this project is to predict weekly availability probabilities for each Giants player and translate them into expected weekly salary cap dollars at risk. Model performance will be primarily evaluated using the Area Under the ROC Curve (AUC) on a holdout 2024 season. The secondary KPI is Calibration alignment which is measured through reliability plots. This ensures that the predicted probabilities reflect true observed frequencies. From a business perspective, success is defined by the model’s ability to accurately flag weekly cap dollars at risk while maintaining sensitivity of player to actual availability. To ensure real world applicability, all features are restricted to information available before game day and eliminating target leakage.

1.3 Stakeholders and Actions

The key stakeholders for this project span both the football and financial operations of the New York Giants. The General Manager uses the model to guide roster and IR decisions, while the Cap Manager applies expected cap risk estimates for financial planning and compliance. The Head Coach adjusts depth charts and practice reps based on predicted player's availability and the Training Staff prioritizes medical evaluations for high-risk players. These stakeholders use the model’s insights to make proactive and data-driven decisions that improve roster stability and financial efficiency.

1.4 Value Metric

The key value metric for this project is CapRisk, defined as CapRisk = $(1 - P(\text{play})) \times \text{WeeklyCapHit}$.

Annual player salaries were normalized by dividing total cap hits by the 18 regular-season weeks to estimate weekly exposure. For example, a player earning $18M annually (≈ $1M weekly) with a 60% play probability yields $400K expected risk for that week. This metric translates player availability uncertainty into a financial measure that aligns directly with roster and budget decisions.

2 Data and Scope

2.1 Scope

The analysis utilized New York Giants team data exclusively. For the timeline, 2023 data was selected for modeling and cross-validation (CV), and 2024 data served as the holdout season for final evaluation.

2.2 Sources

The analysis relied on several key data sources, including the master roster and snap files, player salary cap hits, and play-by-play data. This raw data was used to engineer specific features such as touches, tackles, sacks, and quarterback hits. Crucially, all engineered lag and rolling features were designed to ensure no temporal data leakage.

2.3 Target

The target variable did_play is a binary indicator defined as 1 if a player recorded more than zero snaps in a given week, and 0 otherwise.

2.4 Biases and Caveats

Our analysis also has several potential biases and limitations. For instance, the single-team sample may limit the model's ability to generalize to the broader NFL. Additionally, the data may contain injury reporting noise, which could affect the accuracy of the target variable. Finally, the model is susceptible to external factors such as late-breaking information and mid-season role changes that are often not fully captured within the historical dataset.

3 Feature Engineering and Leakage Control

3.1 Leakage-Free Design

Our feature set relies on lagged variables (such as snaps_lag1, touches_lag1, qb_hits_lag1), rolling windows computed strictly from prior games (such as snaps_roll3_prev), and pre-game contextual factors including rest days, a short-week indicator, home-game status, player age, and BMI. To avoid leakage, cap hits are excluded from model predictors and used solely for downstream business evaluation.

3.2 Final Frame

The finalized modeling dataset, availability_model_frame_clean.csv, contains all engineered features and the target variable used for training and validation. This cleaned frame integrates lagged, rolling, and contextual predictors while adhering to the leakage-free design outlined above, ensuring full reproducibility of the analysis.

4 Modeling Approach

4.1 Models

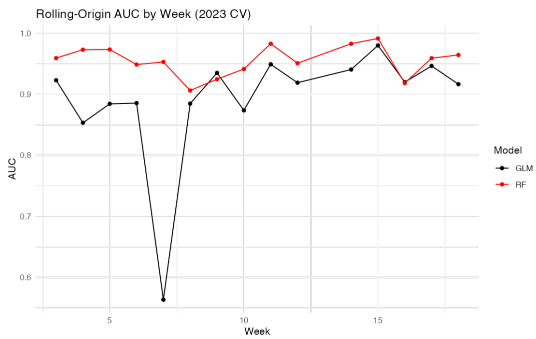
We used two models to predict weekly player availability. A Generalized Linear Model (GLM) with a binomial link function served as an interpretable baseline, providing clear insights into how key features such as injury history and practice participation affect playing probability. A Random Forest model with 600 trees and class weighting was then trained to capture nonlinear interactions between variables, improving predictive accuracy.

4.2 Class Imbalance

Since most players are active each week, the dataset shows class imbalance with fewer unavailability cases. To address this, both models incorporated class weights proportional to class prevalence. This ensures that missed-game events were appropriately emphasized during training and that the model remained sensitive to rare unavailability outcomes.

4.3 Validation

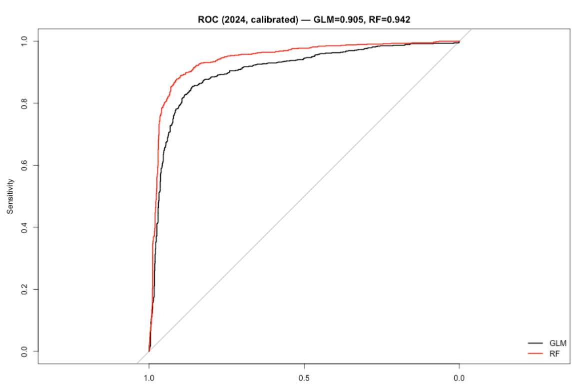
We used rolling-origin cross-validation to evaluate the approach on the 2023 season, where each fold trained on all weeks prior to week w and tested on week w. The final evaluation used a 2023 training set and a 2024 holdout season to assess generalization performance.

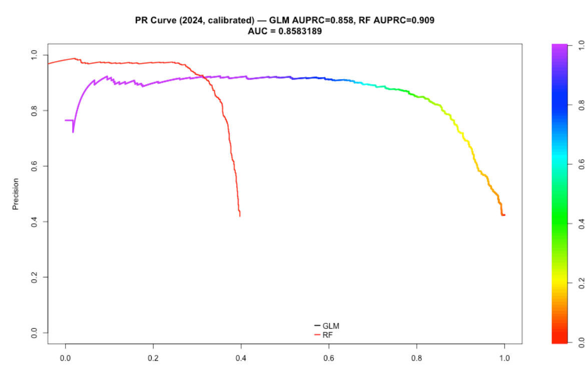


As seen in the graphic, other than the minor fluctuations which reflect natural variation in weekly roster and injury data, the AUC values consistently remain above 0.9. This indicates strong predictive accuracy and stable week-to-week generalization.

4.4 Calibration

The next two plots evaluate model calibration fitted on the 2024 holdout set after applying Platt scaling fitted on 2023 predictions.





The ROC curve graph shows how well the models balance sensitivity and specificity. The RF achieves a higher AUC (red, 0.942) than the GLM (black, 0.905), meaning it better separates available players from unavailable players. While the PR curve focuses on how well the models identify true positives under imbalance conditions.

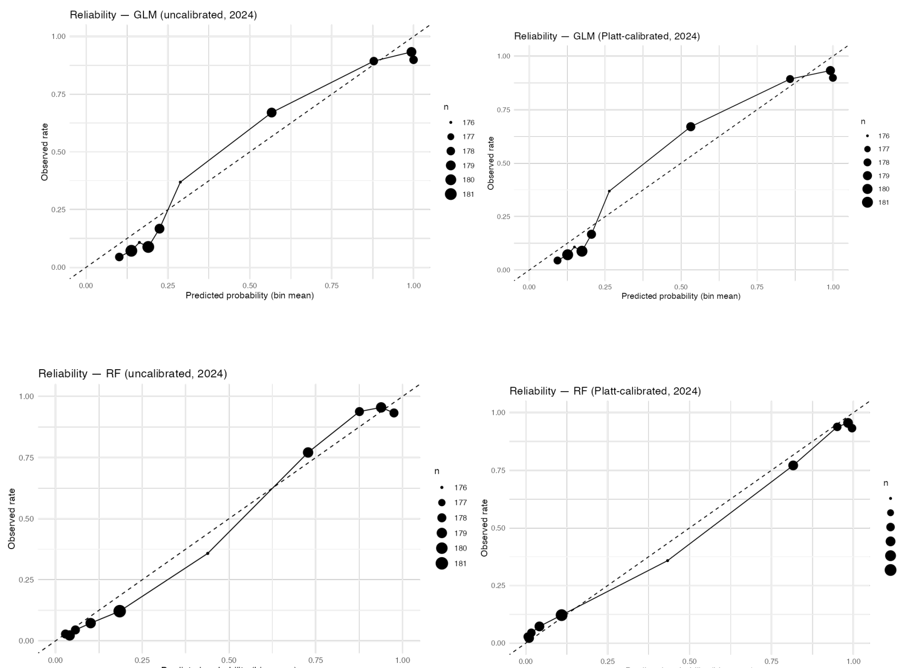
The RF again outperforms the GLM, with AUPRC of 0.909 compared to 0.858, demonstrating higher precision in predicting actual availabilities while maintaining recall.

5 Evaluation Results

5.1 Discrimination

On the 2024 holdout set, the RF model achieved an AUC of 0.942, outperforming the GLM at 0.905. This indicates that the RF model can correctly rank a randomly chosen player-week who played above one who did not by approximately 94% of the time, demonstrating strong discriminative power in predicting player availability.

5.2 Calibration

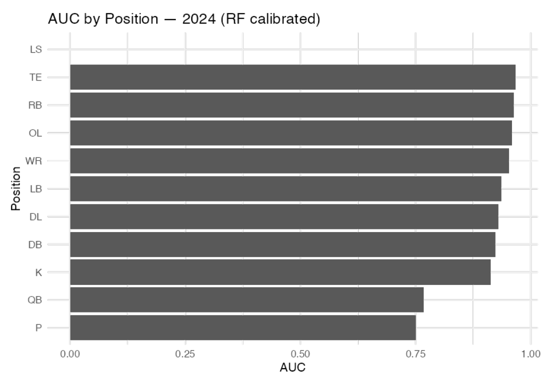


After calibration, the reliability curve tracks the 45° line more closely, shrinking bin gaps and aligning predicted with observed play rates.

5.3 By-Position AUC

We evaluated the Random Forest model's predictive performance by player position using the 2024 holdout season. Wide receivers (WR) with AUC values of 0.963, running backs (RB) with AUC values of 0.955, and linebackers (LB) with AUC values of 0.950, indicates that the model was highly effective at distinguishing available from unavailable players in these positions. Offensive linemen (OL) with AUC values of 0.812, defensive linemen (DL) with AUC values of 0.835, and tight ends (TE) with AUC values of 0.846, although lower but still reflects complexity in predicting availability for these roles. These results highlight that model accuracy varies by position, which can guide position-specific risk management and roster decisions.

From 2024 RF: WR = 0.963, RB = 0.955, LB = 0.950; OL = 0.812, DL = 0.835, TE =0.846.



5.4 Error Analysis

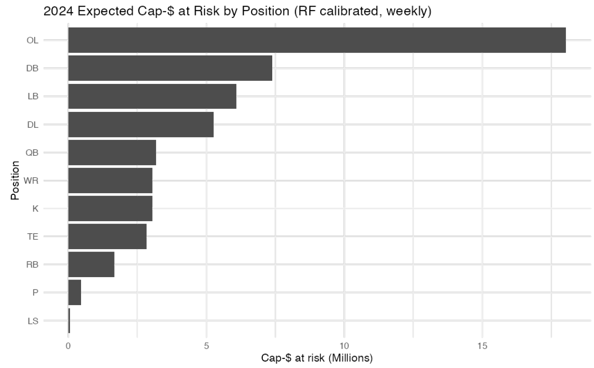
Error analysis revealed that positions with fewer samples, such as specialists, shows unstable calibration bins. This means that predicted probabilities for these players are less reliable and we should be cautious when interpreting their availability predictions.

6 Business Evaluation and Scenario Analysis

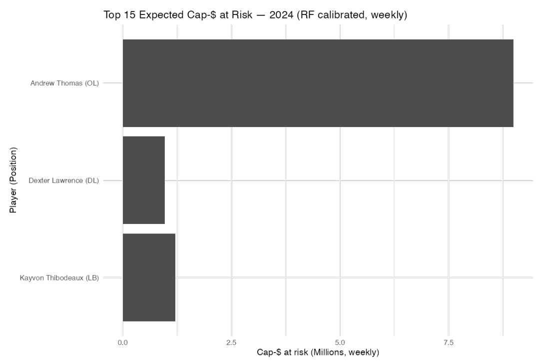
6.1 Weekly Cap Normalization

We normalized the annual cap hits by dividing them across 18 weeks to derive the precise weekly cap hit. This step was crucial for accurately setting the weekly risk metrics and preventing inflation to annualized figures.

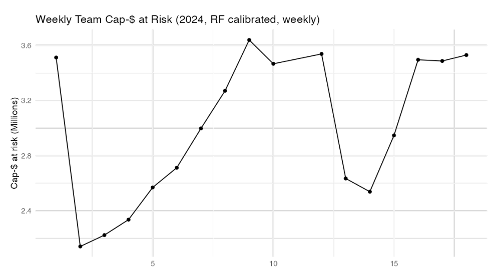
6.2 Cap Dollars at Risk



The expected cap graph shows that OL leads weekly cap at risk by a wide margin, follow by DB, LB and DL. It suggests that we should put priority on OL depth and vigilance for back-seven volatility.

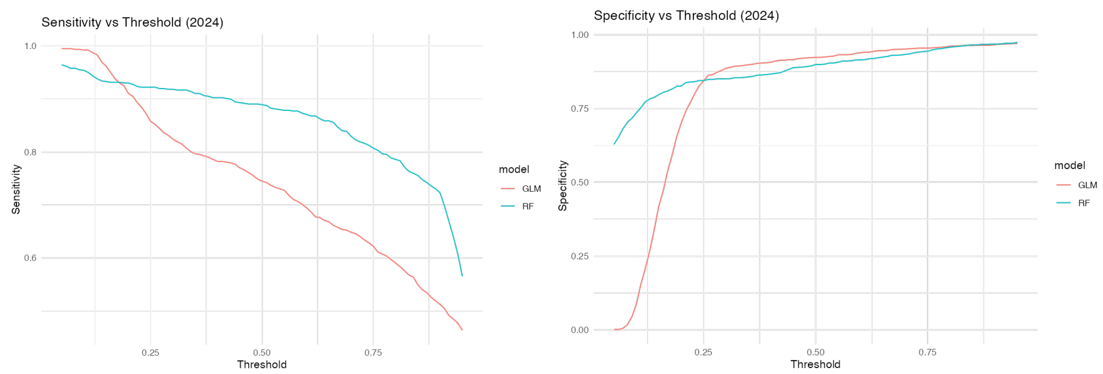


The top 15 cap risk players graph highlights specific high-risk player-weeks, with Andrew Thomas (OL) standing out with by far the largest weekly exposure.



The weekly cap risk graph shows team-level cap-at-risk rising into week 6, then peaking again around week 12 before leveling off. Therefore, we can come to the conclusion that mid-season pressure points where depth planning and workload management matter most.

6.3 Threshold Sweeps



The analysis curves illustrate the classic threshold trade-off: lowering the probability threshold (τ) increases sensitivity while decreasing specificity. We identified an optimal operating point at τ≈0.35, where the model achieved a sensitivity of approximately 0.91 and a specificity of 0.72. This threshold flags about $5.1million weekly cap at risk, effectively capturing the majority of financial risk with only moderate false positives.

6.4 Operating Policy

The final deployment involves the Random Forest model, calibrated with a threshold of 0.35, which yields a high sensitivity of 0.91 and a specificity of 0.72. This operating point results in approximately $5.1 million of weekly caps being flagged. The resulting policy dictates that any player falling below this threshold is flagged for immediate staff review every Tuesday.

7 Deployment Plan

7.1 Workflow

The deployment workflow is structured to provide timely actionable insights each week. On Monday, we ingest the prior week's data on snaps, injuries, and roster updates. On Tuesday morning, the model is re-trained and used to generate updated predictions. By Tuesday midday, the Cap Risk Report is distributed to the front office, enabling informed roster and financial decisions for the upcoming week.

7.2 System Components

The deployment system consists of three main components. The data layer ingests weekly CSV files, using the same inputs as in the analysis. The modeling service is implemented in an R script (nyg availability visual pack weekly cap.R) that generates predictions and computes expected cap risk. Finally, the outputs including tables, figures, and other visual artifacts are distributed to the front office.

7.3 Governance

Governance includes monitoring weekly model discrimination and calibration, with retraining scheduled annually or whenever performance declines significantly, ensuring predictions remain accurate and actionable.

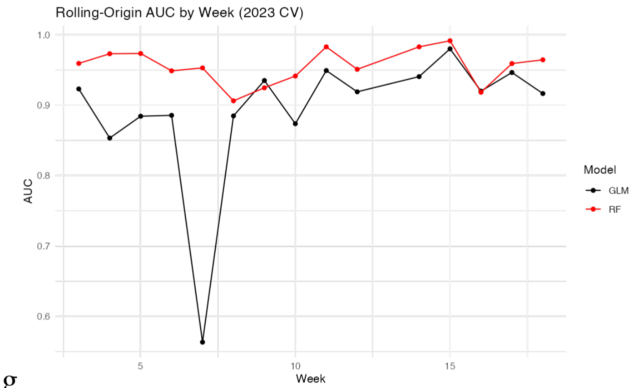
7.4 Risks and Mitigation

Key risks include overreliance on model predictions, data gaps, and ethical concerns. These risks can be reduced by requiring human review, cross-checking with medical reports, and keeping player data private for internal use only.

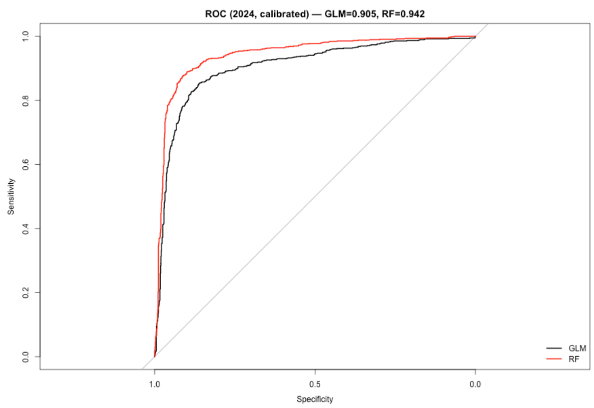
8 Conclusion and Next Steps

The calibrated RF model provides strong weekly predictions (AUC 0.942), reliable calibration, and direct translation to weekly cap risk. Our next steps would be to extend to league-wide data, incorporate practice participation, test gradient boosting, and deliver insights via a Shiny dashboard.

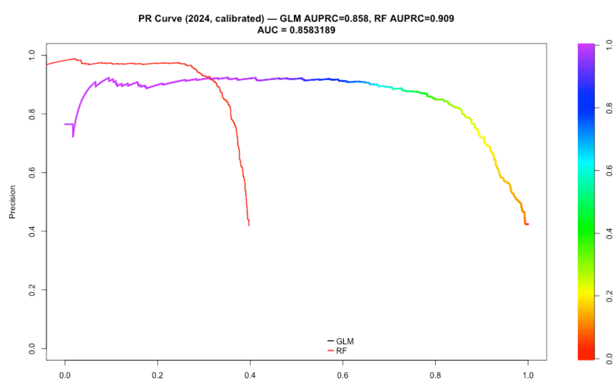
Appendix A: Figures



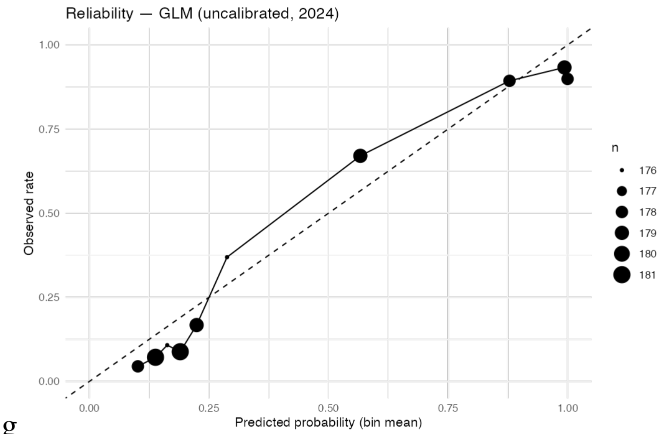
• cv rolling auc 2023.png



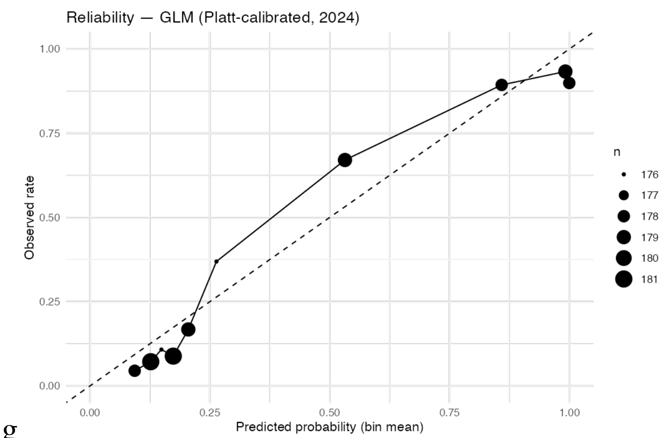
• roc calibrated 2024.png



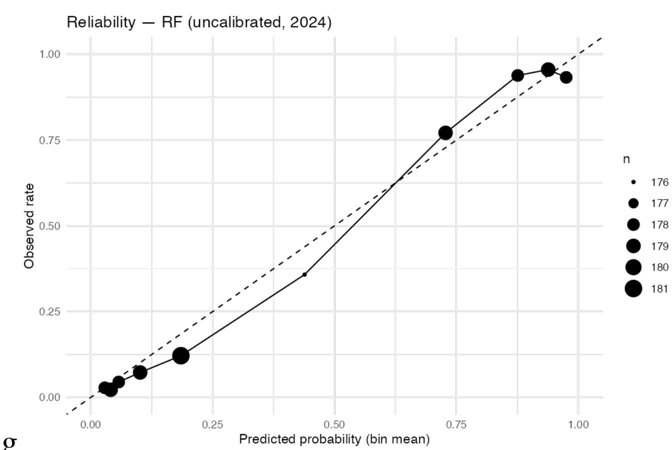
• pr calibrated 2024.png



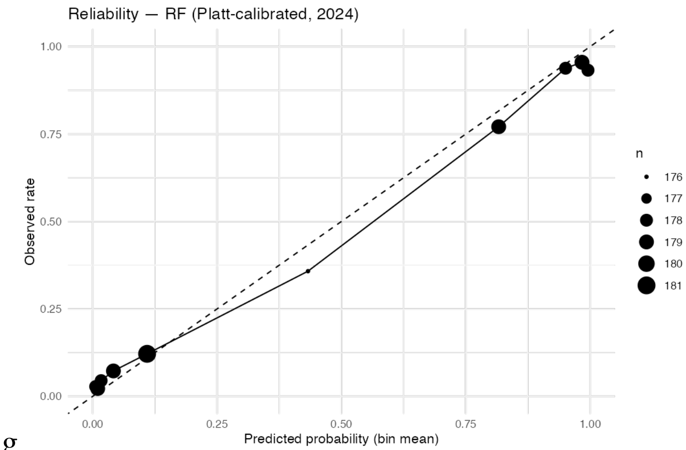
• reliability glm 2024.png



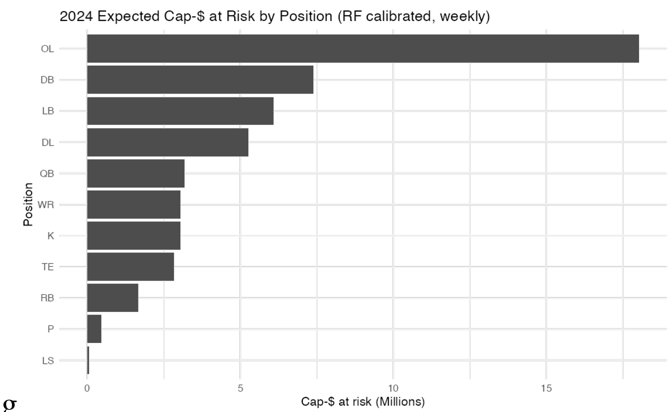
• reliability glm cal 2024.png



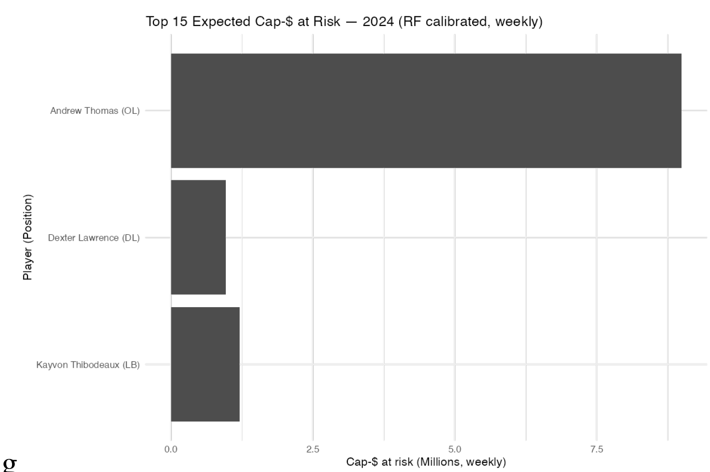
• reliability rf 2024.png



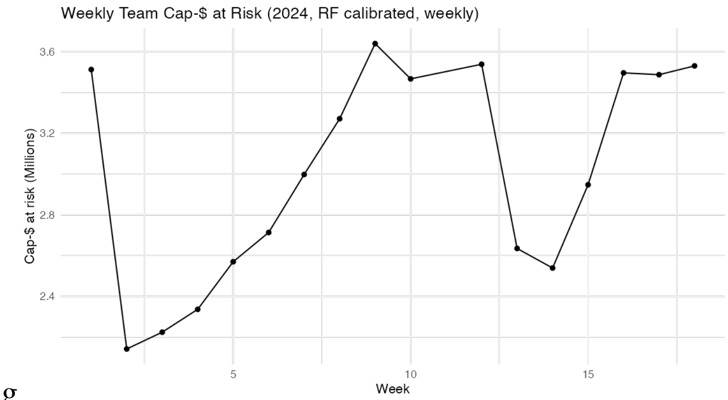
• reliability rf cal 2024.png



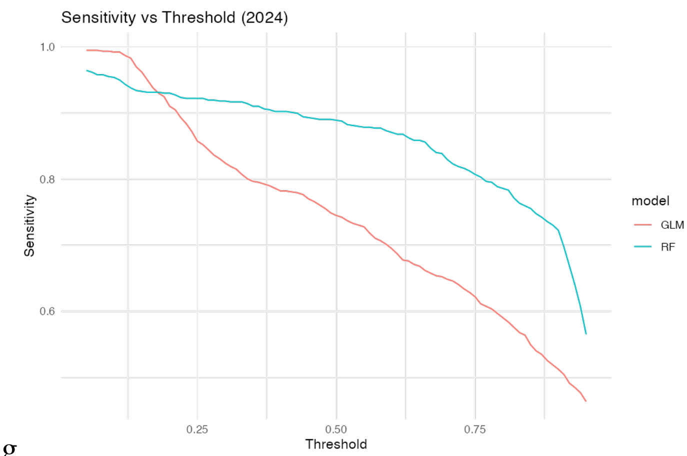
• cap at risk by position 2024.png



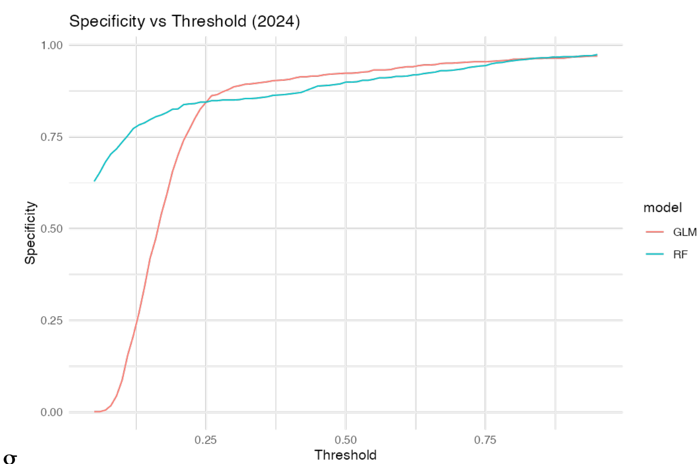
• top15 cap risk players 2024.png



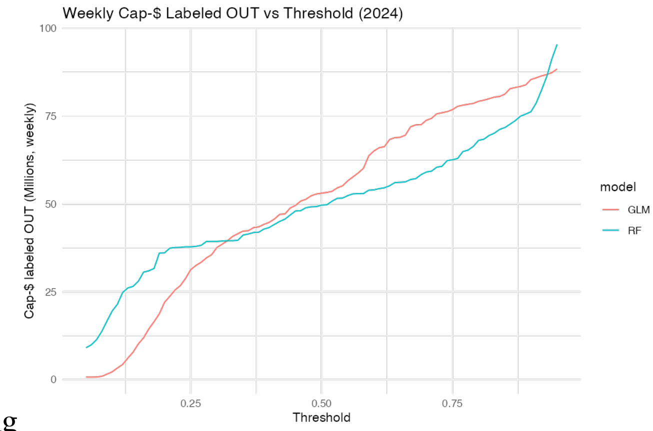
• weekly cap risk 2024.png



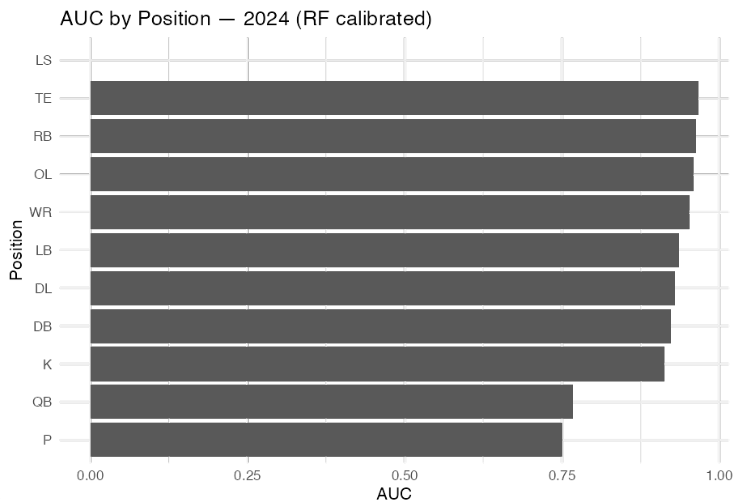
• threshold sensitivity 2024.png



• threshold specificity 2024.png



• threshold cap flagged out 2024.png



• auc by position 2024 bar.png

B Appendix B: Tables

• cv rolling 2023 weekly auc.csv

• availability predictions 2024 with calibration.csv

• auc by position 2024.csv

• cap at risk by position 2024.csv

• top15 cap risk players 2024.csv

• weekly cap risk 2024.csv

• threshold sweep metrics 2024.csv

Appendix C: Replication Instructions

• Files: availability model frame clean.csv, R script.

• Run in R: source("nyg availability visual pack weekly cap.R")

• Outputs: artifacts/ (CSVs), visuals/ (PNGs).

Appendix D: Author Contributions

Conor Patten: Led technical work including data preparation, feature engineering, model design, validation, calibration, and production of all artifacts and visuals. Drafted the detailed outline, provided deployment recommendations, and coordinated integration of the technical results into the business narrative.

Peggy Su: Contributed substantially to the written report, especially in sections on Business Understanding and Business Evaluation. Assisted with editing and refining prose throughout the document.

Stacy Chen: Contributed substantially to the written report, particularly the Executive Summary, Deployment Plan, and Conclusion sections. Assisted with organization and editing.

Mercy Zhao: Contributed to creation of the slide deck, focusing on the business narrative and integration of visuals. Assisted with polishing report sections.

Emma Zhu: Contributed to creation of the slide deck, focusing on communicating evaluation results and business impact. Assisted with editing report text.

Hanson Chandra: Contributed to creation of the slide deck, with emphasis on deployment plan and scenario analysis. Also supported the write-up of the business deployment section and final proofreading.

Appendix E. AI Use Disclosure

In line with the course policy on the use of Large Language Models (LLMs), the following prompts represent the ways in which ChatGPT was used to support coding and conceptual understanding during the development of this project. At no point were modeling choices or assignment text pasted directly into the system; all modeling design and business framing were independently developed by the project team.

Coding Support

- Write R code to split data into training and test sets by season.
- How do I implement rolling-origin cross-validation in R without target leakage?
- Show me an R function to compute AUC safely when the target has only one class.
- Can you give me an R snippet to generate a calibration/reliability plot?
- What's the best way to pro-rate annual cap hit into weekly values in R?

Debugging

- Why am I getting an error: object 'cap hit' not found in dplyr::mutate()? How can I fix this?
- Explain why glm() is throwing the warning "non-integer #successes in a binomial glm".
- Help me join two data frames by player and week while avoiding duplication.

Concept Explanations (Class Material)

- Explain in simple terms what AUC measures in binary classification.
- What's the difference between discrimination and calibration in a predictive model?
- How do class weights help address imbalance in logistic regression and random forests?
- What is rolling-origin cross-validation and why is it better than random splits for time-series?
- How can threshold sweeps be used to analyze business trade-offs in classification?

Visualization Help

- Give me an R ggplot template for a bar chart ordered by value.
- How do I make a reliability plot with predicted probability bins on the x-axis?
- How can I generate ROC and PR curves in R and save them as PNGs?