

SYS-611 Term Project

Monte Carlo Simulation of NFL Fourth-Down Decision Strategies

Changye Chen Daniel Caracitas

December 2025

1 Introduction

1.1 Problem statement and study objectives (What decision should be made?)

In the National Football League (NFL), coaches repeatedly face fourth-down decisions: punt, attempt a field goal, or “go for it” to try to convert for a new set of downs. These choices affect field position, expected points, and ultimately the probability of winning.

Decision problem: Given a fourth-down state (field position, yards to go, score situation, and time remaining), what level of fourth-down aggressiveness produces the best outcomes?

Decision objective: We frame the coaching decision as a policy-selection problem. For a given state s (field position, $ydstogo$, score differential, and time remaining), the coach selects an action $a \in \{\text{punt, FG, go}\}$ (or equivalently a policy class that maps states to actions) to maximize performance. In this project, we focus on two practical criteria: (i) maximizing expected scoring output and (ii) maintaining realistic behavioral rates (e.g., fourth-down attempt frequency) relative to historical data.

Study objective: Using historical NFL play-by-play data (2009–2018), we build a Monte Carlo simulation of simplified game outcomes under three fourth-down strategies: *conservative*, *balanced*, and *aggressive*. We estimate how each strategy affects scoring, fourthdown behavior, and drive-level outcomes, and we quantify uncertainty using confidence intervals across repeated simulation runs.

1.2 System boundary (What is fixed and what is variable?)

The system under study is one NFL game played between two teams. The simulation represents the game as alternating possessions (drives). Each drive is modeled as a sequence of plays that advances the ball until one of the following terminal events occurs: touchdown, successful field goal, punt, turnover, turnover on downs, missed field goal, or the game clock expiring.

Fixed elements:

- NFL scoring rules (TD=7, FG=3) and possession-change mechanics.
- The action set on 4th down: {punt, field goal, go for it}.
- Data-driven rate models estimated from 2009–2018 play-by-play data (yards gained distributions, turnover probability, fourth-down conversion probability, field goal make rates by kick distance, and punt/return outcomes).

Variable elements:

- Fourth-down decision strategy (Conservative vs. Balanced vs. Aggressive).
- Stochastic play outcomes drawn from the fitted models.
- Game evolution over time (clock decreases with each play; late-game hurry-up behavior is triggered when trailing).

1.3 Key performance measures (How is a good solution identified?)

We evaluate each policy using the following key performance measures (KPMs):

- **Average points per team per game:** mean points scored per team.
- **Home win rate:** proportion of games the home team wins.
- **Tie rate:** proportion of games ending tied.
- **Fourth-down attempts per team per game:** average number of “go for it” attempts per team.
- **Fourth-down success rate:** fraction of fourth-down attempts that convert.

- **Drive outcome rates:** fraction of drives ending in TD, FG, punt, missed FG, turnover, or turnover on downs.

2 Modeling Approach

2.1 Collecting and processing real data (Observe new data or use existing data)

We use a public NFL regular-season play-by-play dataset covering seasons 2009–2018 (NFL Play by Play 2009-2018 (v5).csv). We load only the columns required for modeling: down, yards-to-go, yardline (distance to opponent end zone), yards gained, play type, turnover indicators (interception, fumble lost), fourth-down conversion flags, field goal attempts/results (with kick distance), and punt attempts.

The data is processed to estimate the following components:

- **Normal plays (downs 1–3):** Restrict to run/pass plays. Bucket by down $\in \{1, 2, 3\}$ and yards-to-go bucket (1–2, 3–5, 6–9, 10+). For each bin we compute mean/std of yards gained and turnover probability. During simulation, yards gained are sampled from a Gaussian approximation (with truncation on extreme negative outcomes).
- **Fourth downs:** Bucket by yards-to-go bucket (1–2, 3–5, 6–9, 10+). For each bucket we compute conversion probability and mean/std of yards gained. A fourth-down attempt is simulated using a Bernoulli conversion trial and a sampled gain.
- **Field goals:** Bucket kick distance into short (≤ 35), mid (≤ 45), long (≤ 55), and bomb (> 55). For each bucket we estimate FG make probability.
- **Punts:** Estimate punt net distance mean/std from punt attempts; during simulation, net punt distance is sampled (truncated to a reasonable range) with a simple return model.

2.2 Formulate and develop model (Monte Carlo simulation)

2.2.1 State representation and drive evolution

We represent field position using an absolute coordinate $abs\ pos \in [0, 100]$ measured as yards from the offense's own goal line. Thus $abs\ pos = 25$ corresponds to the offense on its own 25-yard line. The distance to the opponent end zone is $yards\ to\ goal = 100 - abs\ pos$.

A drive begins at a kickoff-derived starting position (mean around the 25 with variance) and proceeds play-by-play. On downs 1–3, the model samples yards gained from the fitted normal-play distribution for the current (down, to-go bucket). A turnover may occur with probability estimated from historical data; if a turnover occurs, possession switches at the current spot (field position flips).

When a first down is achieved, the down resets to 1 and the to-go resets to 10 (or goal-to-go inside the 10). A touchdown is scored if $\text{abs pos} \geq 100$.

2.2.2 Fourth-down decision policies (explicit rules implemented in code)

On 4th down, the strategy chooses among punt, field goal, or go for it based on field position and yards-to-go. Kick distance is approximated as $\text{kick distance} = (100 - \text{abs pos}) + 17$. Table 1: Fourth-down decision rules implemented in the code

Policy	Fourth-down action rule (from choose fourth action)
Conservative	If goal-to-go ($\text{yards to goal} \leq 1$), always go . Otherwise, go only if in plus territory ($\text{abs pos} \geq 50$) and $\text{ydstogo} \leq 1$. Else FG if $\text{kick distance} \leq 55$; otherwise punt .
Balanced	If goal-to-go ($\text{yards to goal} \leq 1$), always go . Otherwise, go if ($\text{abs pos} \geq 50$ and $\text{ydstogo} \leq 3$) or ($\text{abs pos} \geq 65$ and $\text{ydstogo} \leq 5$). Else FG if $\text{kick distance} \leq 55$; otherwise punt .
Aggressive	If goal-to-go ($\text{yards to goal} \leq 1$), always go . Otherwise, go if ($\text{abs pos} \geq 35$ and $\text{ydstogo} \leq 4$) or ($\text{abs pos} \geq 60$ and $\text{ydstogo} \leq 6$). Else FG if $\text{kick distance} \leq 57$; otherwise punt .

2.2.3 Special teams, penalties, and clock (simplified mechanisms)

To increase realism while keeping the model lightweight, the simulation includes:

- **Penalties:** With probability 0.08 per play, a penalty occurs. Offensive penalties back the offense up (typically 5 or 10 yards). Defensive penalties advance the offense (typically 5 or 10 yards) and may award an automatic first down with probability 0.5.

- **Field goal and punt blocks:** A 1% block probability is applied for punts and field goals.
- **Clock model:** The game starts with 3600 seconds. Each play consumes a random amount of time (uniformly sampled between 20 and 32 seconds). If the offense is trailing in the final 300 seconds, a hurry-up adjustment reduces time consumption. The simulation ends when the clock reaches zero.

2.3 Simulation experiments (Study design and conditions)

We evaluate each strategy under **mirror matchups** where both teams follow the same policy (e.g., Conservative vs. Conservative) to control for opponent behavior. We run repeated Monte Carlo experiments and report uncertainty using 95% confidence intervals (CI) computed across repeated runs. Each run simulates 300 games; summary metrics are then aggregated across runs to obtain mean and CI.

For each game we record: home points, away points, home win indicator, tie indicator, total fourth-down attempts and conversions, and the terminal outcome of each drive (TD, FG, punt, missed FG, turnover, or turnover on downs).

2.4 Validate and document model (Compare with collected data)

We validate the simulation by comparing historical and simulated aggregate statistics computed from the same underlying dataset (2009–2018). We compare: (i) mean points per team per game, (ii) fourth-down attempts *per team per game*, and (iii) fourth-down success rate. Historical values are computed directly from the play-by-play dataset, while simulated values are computed from the Monte Carlo outputs.

Because the simulation uses mirror matchups, simplified clock logic, and does not explicitly model overtime, perfect agreement is not expected; the goal is to verify that simulated magnitudes and behavioral patterns are reasonable relative to the data. In particular, home win rate in historical data reflects real-world home-field advantage, while mirror matchups in our symmetric simulator are expected to yield win rates near 50%.

Table 2: Historical vs. simulated aggregate statistics (2009–2018 data vs. Monte Carlo outputs)

Model	Avg. Points	Tie Rate (%)	4th Attempts	4th Success Rate (%)
Historical	22.34	0.55	0.95	48.94
Conservative	23.13	0.33	0.63	48.28
Balanced	23.50	0.00	1.52	48.63
Aggressive	23.60	0.33	2.27	47.57

3 Results and Analysis

3.1 Summary of key performance measures (with uncertainty)

Table 3 summarizes the simulation outcomes for each mirror-matchup strategy. We report mean outcomes with 95% confidence intervals across repeated simulation runs.

Table 3: Simulation results by strategy (mirror matchups; mean with 95% CI)

Strategy	Avg. points per team (CI)	Home win rate (CI)	Tie rate	4th attempts per team
Conservative	23.13 [22.68, 23.58]	0.502 [0.462, 0.542]	0.0033	0.63
Balanced	23.50 [23.01, 23.99]	0.472 [0.432, 0.512]	0.0000	1.52
Aggressive	23.64 [23.17, 24.11]	0.517 [0.477, 0.557]	0.0033	2.27

Across these simulations, average scoring increases slightly with aggressiveness. However, the confidence intervals overlap across strategies, indicating limited statistical separation in scoring under the current design. By construction, mirror matchups yield win rates near 50% in expectation.

3.2 Drive-level mechanism (risk-reward trade-off)

To explain *why* strategies differ, we decompose each possession into terminal outcomes.

Table 4 shows the fraction of drives ending in each terminal event.

Table 4: Drive outcome rates by strategy (mirror matchups)

Strategy	TD	FG	Punt	Miss FG	Turnover	Downs
Conservative	0.215	0.166	0.429	0.036	0.126	0.028
Balanced	0.245	0.120	0.412	0.028	0.125	0.069
Aggressive	0.253	0.118	0.365	0.034	0.123	0.107

Aggressive policies reduce punts and field goals while increasing both touchdown rate and turnover-on-downs rate, revealing a clear risk-reward trade-off: more attempts create more high-value touchdown outcomes but also increase the probability of losing possession via downs.

3.3 Figures

Figures 1–4 visualize scoring, aggressiveness, conversion efficiency, and win/tie outcomes.

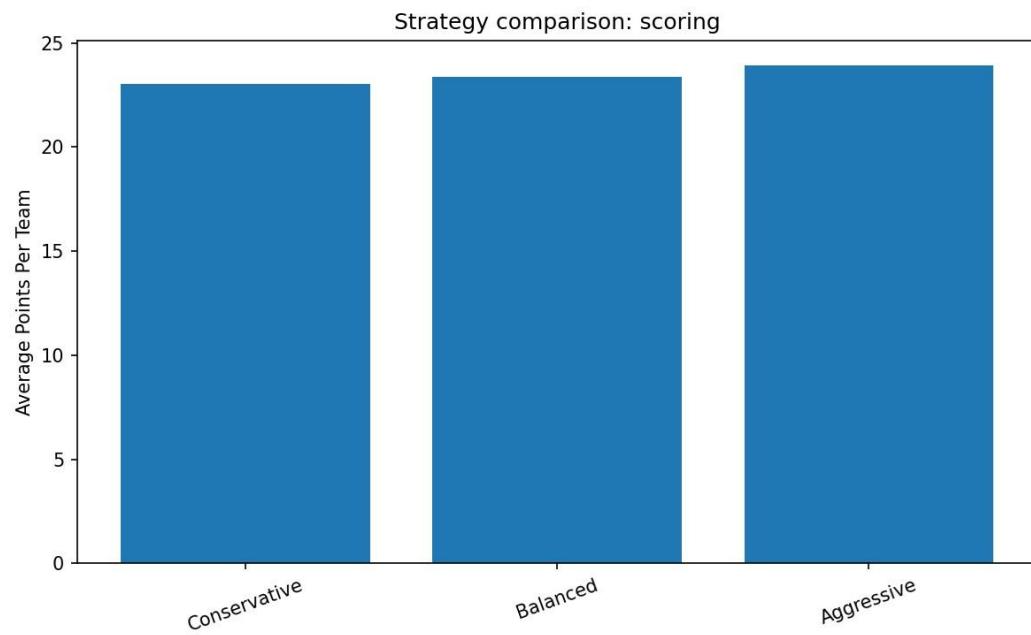


Figure 1: Strategy comparison: average points per team per game

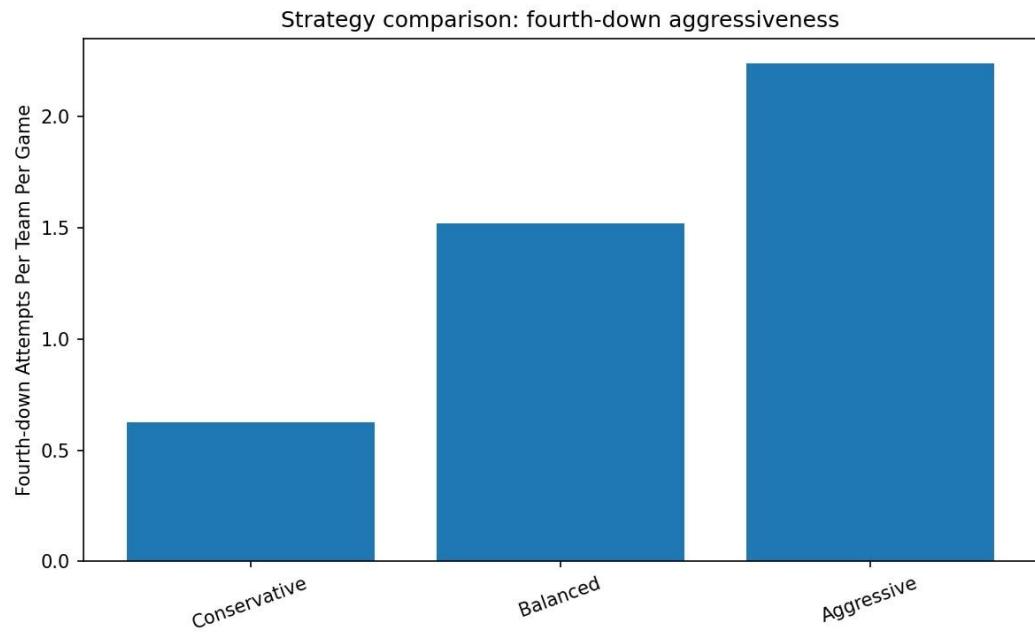


Figure 2: Strategy comparison: fourth-down attempts per team per game

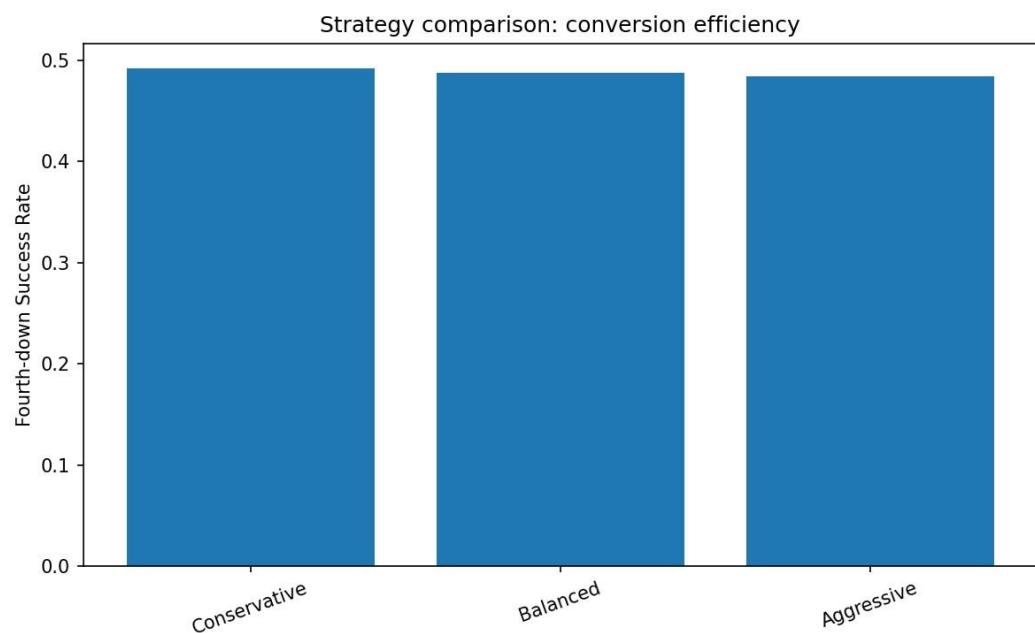


Figure 3: Strategy comparison: fourth-down conversion success rate

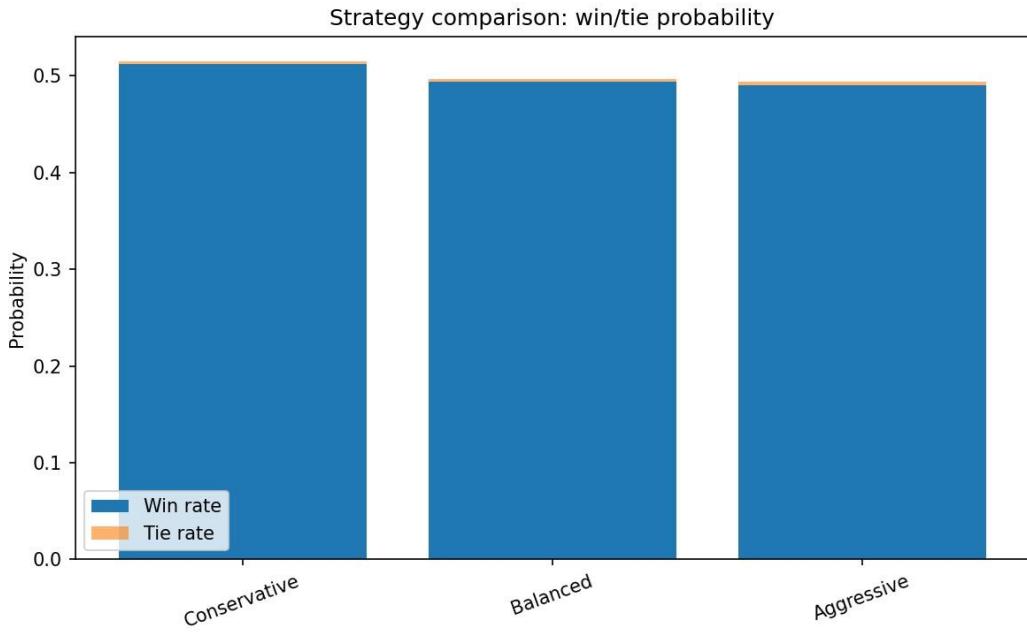


Figure 4: Strategy comparison: win/tie probability

4 Discussion and Conclusion

This project used a Monte Carlo simulation driven by rates estimated from 2009–2018 NFL play-by-play data to compare three fourth-down strategies.

Across mirror-matchup experiments, average points per team increased modestly with aggressiveness (23.13 for Conservative, 23.50 for Balanced, and 23.64 for Aggressive), but confidence intervals overlap, suggesting limited statistical separation in scoring under the current design. Behavioral metrics, however, show large differences: fourth-down attempts per team per game increase sharply (0.63, 1.52, 2.27), while fourth-down success rates remain close to the historical baseline (approximately 0.49), indicating that the model produces realistic fourth-down efficiency. Seen in Figure 4, we can see that aggressive plays can lead to higher win rate compared to balanced and conservative, while also introducing more risk at the same time, giving us realism within our model.

Drive-level decomposition explains the mechanism: aggressive strategies convert a portion of punt/field-goal drives into touchdowns while also increasing turnover-on-downs risk. Thus, the strategic effect is best understood as a trade-off between higher upside (more TD drives) and increased failure risk (more downs).

Recommendation: If the objective is to increase aggressiveness while retaining realistic fourth-down efficiency and a balanced risk profile, the balanced policy is a reasonable baseline. If the objective is to maximize expected points under this simplified model, the aggressive policy performs best on average, but with substantially more fourth-down attempts and higher turnover-on-downs frequency.

Limitations: The model uses coarse bins and Gaussian approximations for play gains, and fourth-down conversion is conditioned only on yards-to-go (not yardline, formation, or opponent). The simulator also simplifies special teams and clock dynamics and does not explicitly model overtime. Mirror matchups provide a controlled baseline but do not directly measure competitive advantage against a different opponent policy.

Future work: Extend the model with finer state conditioning (including yardline bins for fourth-down), explicit overtime logic, asymmetric (cross-policy) matchups to quantify competitive advantage, and calibration of additional drive-level statistics (FG attempts by distance, punt net distribution) to further improve realism.

5 References

- NFL Play by Play dataset (2009–2018), file: NFL Play by Play 2009-2018 (v5).csv.
- Supporting code and outputs included with submission (e.g., simulation.py, strategy summary.csv-historical vs simulated.csv, and figures).