Step-by-Step Methodology

Data Acquisition

- Pulled play-by-play (PBP) data from nflfastR for the 2021–2024 NFL seasons.
- Retrieved Next Gen Stats participation data from nflreadr, which contains player-level tracking details such as routes, pressures, and air yards.

Initial Filtering of Plays

- Restricted dataset to pass plays only, excluding two-point conversions (as these represent atypical, high-leverage situations that distort efficiency metrics).
- Applied a neutral game script filter: only included plays where the score differential was within 16 points. This ensures analysis focuses on standard game flow rather than garbage time or desperation scenarios.

Column Selection for Core Dataset

- Trimmed the PBP dataset down to variables essential for modeling decision-making and evaluating quarterback efficiency.
- Key retained variables include:
    - Identifiers: game_id, play_id, season
    - Play context: down, ydstogo, yardline_100, qtr, season_type
    - Outcome metrics: yards_gained, epa, air_epa, yac_epa, wpa, qb_epa, etc.
    - Passing details: pass_length, pass_location, air_yards, yards_after_catch, complete_pass, incomplete_pass, interception, etc.
    - Advanced efficiency stats: cp, cpoe, xpass, pass_oe
    - Player identifiers: passer_player_id, passer_player_name, receiver_player_id, receiver_player_name
    - Game state metrics: posteam_score, defteam_score, score_differential, wp, def_wp
- This selection preserves all contextual, efficiency, and outcome-related measures necessary for computing weighted EPA (wEPA).

Integration with Participation Data

- Merged PBP data with participation data on game_id and play_id.
- Restricted participation dataset to only the fields relevant to quarterback decision-making and target evaluation:
    - ngs_air_yards (Next Gen Stats version of air yards)
    - time_to_throw
    - was_pressure (binary indicator of QB pressure)
    - route (route run by the targeted receiver)
    - defense_man_zone_type
    - defense_coverage_type

- This step enriches each pass play with tracking-based context, allowing route-level adjustments in the wEPA framework.

Data Manipulation

- Adjusted 'na' values in coverage type column to be 'unknown' coverages
- Merged multiple player name values of the same playerid to reflect only one player name
  - Aaron Rodgers ← A.Rodgers, Aa. Rodgers
  - Josh Allen ← J.Allen, Jos.Allen
  - Tyrod Taylor ← T.Taylor, Ty.Taylor
- Generated CPOE (completion percentage over expected) summary table by pass location, pass depth, and defensive coverage type
  - Minimum 10 attempts for a given combination, otherwise imputed 0s
- Merged QB-specific CPOE onto baseline completion probability (cp) to create **qbxcp**
  - Thought Process / Rationale:
    - Goal: Adjust each pass attempt's baseline completion probability (cp) by the quarterback's observed over- or under-performance in similar situations (CPOE) to get a QB-specific expected completion probability (qbxcp).
      - Why not just add in raw probabilities?
      - Simply adding cp + qb_cpoe can produce invalid probabilities (<0 or >1), especially when both values are near the extremes.
      - Probabilities are bounded [0,1], so naive addition could distort the metric and reduce interpretability.
    - Solution: Use logit transformation (log-odds space)
      - Transform cp into log-odds using logit(cp) = log(cp / (1-cp)).
      - Apply the QB-specific adjustment (qb_cpoe) in log-odds space. Additive shifts in log-odds are mathematically sound and naturally respect probability limits.
      - Convert back to probability space with the logistic function: qbxcp = 1 / (1 + exp(-(logit(cp) + qb_cpoe))).
    - Result:
      - Each play now has a QB-adjusted completion probability that reflects both the difficulty of the pass and the QB's demonstrated ability in that type of throw.
      - This provides a foundation for weighted EPA (wEPA) calculations where each pass's value is adjusted for decision-making skill, not just raw league-average completion probability.
- Modeled for xEPA using air yards, end of play field location, xYAC, binaries for first downs and TDs expected to occur based on the expected resulting yards gained.
  - Calculated total expected yards (air_yards + xyac_mean_yardage) for each play.
  - Generated xendyardline_100 as projected field position after the play.
  - Created binary variable first_down_flag to indicate whether the play would result in a first down.

    - Created binary variable xTD to flag plays expected to end in a touchdown (xendyardline_100 == 0).
    - Imputed missing values in first_down_flag with 1 to preserve positive outcome signal.
    - Fitted multiple models (linear regression, random forest, XGBoost) using the above predictors to estimate expected play value (xEPA).
    - Cross-validated each model and retained predicted xEPA for each play, aligned with game_id and play_id for downstream merging.
- Used qbxcp and regular epa to generate WEPA (or raw WEPA)
- Used xEPA modeled values paired with qbxcp to generate xWEPA

Stress Testing / Game-Level Analysis

1. **Simplifying the play-by-play dataset**
   To begin, I distilled the full play-by-play data into a smaller subset that included only essential information: game IDs, team identities, passer details, and scoring results. This was done to streamline subsequent analysis and reduce computational complexity while retaining the necessary context to examine quarterback performance relative to outcomes.

2. **Identifying primary quarterbacks per game**
   For each team in every game, I determined which quarterback was the primary passer by counting passing attempts. This allowed me to focus the analysis on the main contributor at the position, as backup QBs with limited attempts would not meaningfully influence team-level outcomes.

3. **Combining game outcomes with QB data**
   I merged the identified primary quarterbacks with final game results to create a complete dataset that linked each QB to their team's performance and whether their team won or lost. This created a game-level view where each matchup could be evaluated in terms of quarterback statistics and actual outcomes.

Comparing Per-Attempt Stats to Winning Outcomes

4. **Assigning hypothetical "stat winners"**
   Using per-attempt performance metrics (EPA, Raw WEPA, and xWEPA), I identified which quarterback had the superior value in each game for each metric. The purpose of this step was to test whether higher efficiency per attempt predicted actual game outcomes.

5. **Evaluating metric predictive accuracy**
   I compared the identified "stat winners" against the real game winner to calculate whether the QB with the higher per-attempt metric actually won the game. This allowed me to quantify the predictive strength of each metric at the game level.

6. **Aggregating results across seasons and metrics**
   To understand broader trends, I computed overall win rates by metric across all games and also examined season-by-season trends. This highlighted whether certain metrics consistently outperformed others in predicting wins and whether their reliability changed over time.

Relationship Between QB Stats and Team Performance

7. **Calculating team scoring averages**
   For each team in each season, I calculated average points per game. This provided a measure of team success that could be compared against QB efficiency metrics.

8. **Linking QB efficiency to team scoring**
   I merged per-QB efficiency metrics with their corresponding team scoring averages. This step allowed for a direct comparison between a quarterback's efficiency on a per-attempt basis and their team's overall offensive success.

9. **Visualizing the relationships**
   Scatter plots were used to display the relationship between QB efficiency metrics and team points per game, including logos to identify players visually. Linear trend lines were added to quantify the strength and direction of these relationships.

Assessing Game-to-Game Consistency (Stat Stickiness)

10. **Measuring variation in per-attempt metrics**
    For each quarterback in each season, I calculated the standard deviation of their per-attempt metrics across games. This quantified how much a QB's performance fluctuated from game to game.

11. **Visualizing distribution of variability**
    Violin and boxplots were employed to display the overall distribution of game-to-game variability, with individual QB-season points overlaid for context. This helped identify which metrics were more consistent and which were prone to fluctuation.

High-Leverage Play Analysis

12. **Selecting key plays**
    I filtered for the top 10% of plays in terms of win probability added (WPA) magnitude. These

"high-leverage" plays were considered critical moments in games, where individual QB decisions had outsized impact on outcomes.

13. **Aggregating performance on key plays**
For each quarterback, I summed both the actual WPA and the expected WPA based on xWEPA for these high-leverage plays. This generated a measure of how well expected efficiency aligned with game-impacting results.

14. **Quantifying predictive power of xWEPA**
A linear relationship between total xWEPA and total WPA on key plays was evaluated, with R^2 calculated to quantify how much of the variance in real-world impact could be explained by the expected metric. This step tested the practical validity of xWEPA in predicting meaningful outcomes.

Purpose and Rationale Summary

- The overarching goal of this methodology was to examine the predictive value and reliability of different per-attempt quarterback metrics (EPA, Raw WEPA, xWEPA) at multiple levels: game outcomes, team scoring, game-to-game consistency, and impact on critical plays.

- Each step moved from broad (season-level summaries) to granular (high-leverage plays), ensuring that insights were both statistically robust and contextually meaningful.

- By combining efficiency metrics with actual results and team performance, the approach provides a comprehensive evaluation of QB effectiveness, highlighting both predictive power and practical implications for coaching and performance analysis.