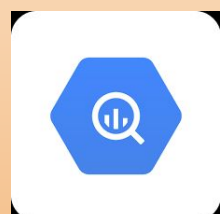




# Predictive Factors of Team Success in the WNBA: A Comparative Analysis with the NBA



Zubair Lakhia

Carsten Savage

Shusaku Matsuda

Margaret Reinhard



## MOTIVATION/ INTRODUCTION

Can player stats like assists or team metrics like turnovers predict success in the WNBA? This project aims to build predictive models to forecast team performance using data from individual players and overall team stats. By examining how these factors influence success, we can gain insights into team-building strategies and optimize player development. Additionally, comparing these predictors between the WNBA and NBA helps uncover unique patterns in women's professional basketball, potentially leading to new approaches in coaching, recruiting, and fan engagement. Understanding these differences can ultimately elevate the game and drive greater recognition for women's sports.

## APPROACHES

We employ a mix of predictive models and interactive visualizations to uncover patterns in WNBA and NBA team success.

- Predictive Modeling:** Using LightGBM, OLS linear regression, and random forest regressor models, we forecast team wins based on player stats (e.g., assists, field goal percentage) and team metrics (e.g., turnovers, defensive rebounds). This helps identify key predictors of success, highlighting differences between the WNBA and NBA.
  - Intuition:* These models capture complex patterns to reveal which metrics drive team performance.
  - What's new:* We explore not just traditional factors but also how these metrics vary between men's and women's leagues.
- Interactive Visualizations:**
  - What-If Scenario Tool:** Adjust team metrics to see their impact on projected wins, enabling data-driven strategy planning.
  - Radar Chart:** Compares high- and low-win teams across multiple metrics, highlighting key performance differences.
  - Shot Coordinate Chart:** Analyzes shot patterns to identify high-efficiency zones in both leagues.
  - Stacked Bar Chart:** Compares play type distributions (e.g., pick-and-roll, isolation), showcasing strategic differences between the leagues.

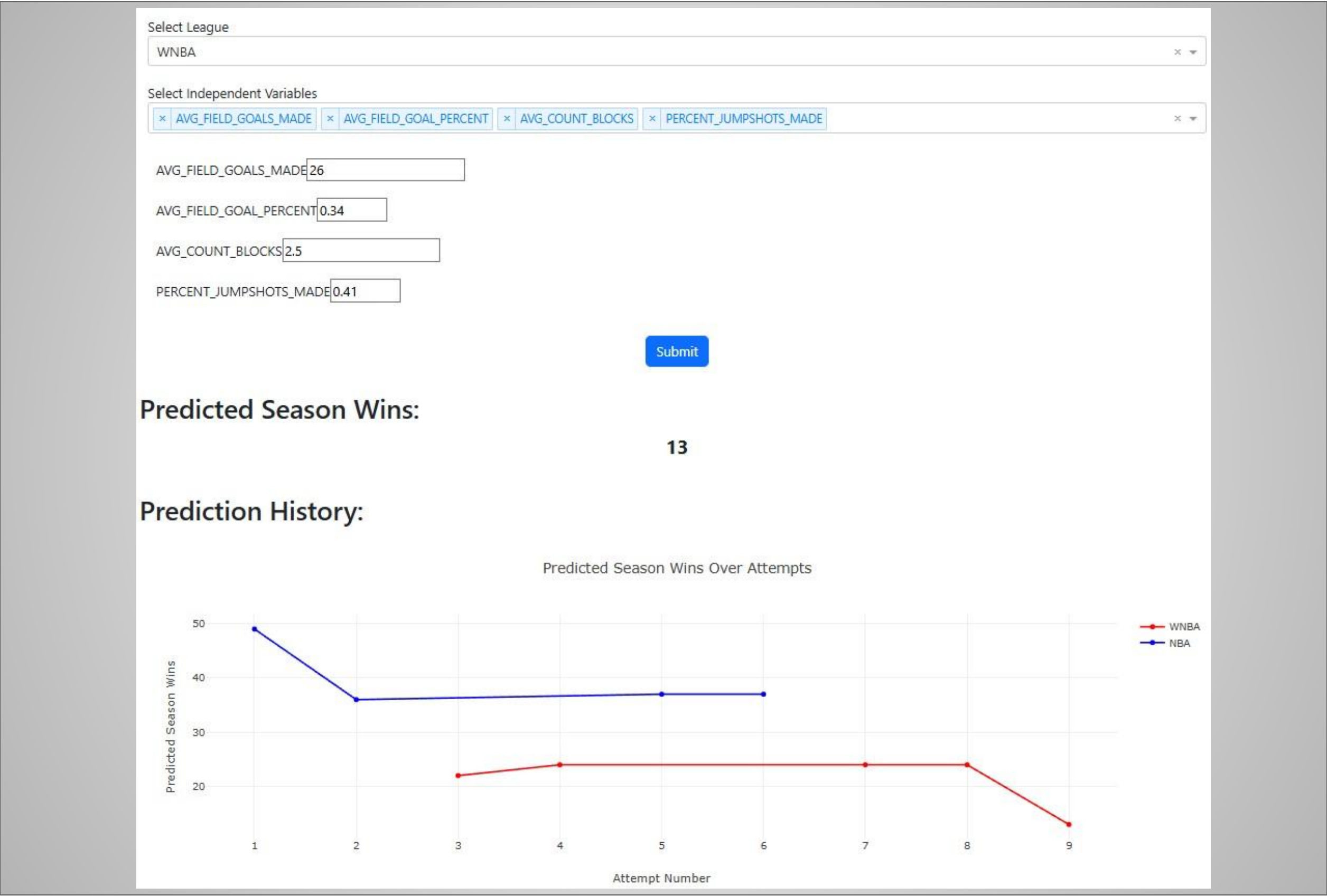
## DATA

We sourced data from the most recent 6 NBA seasons and most recent 12 WNBA seasons from nba\_api and wehoop APIs, resulting in a dataset with the following characteristics:

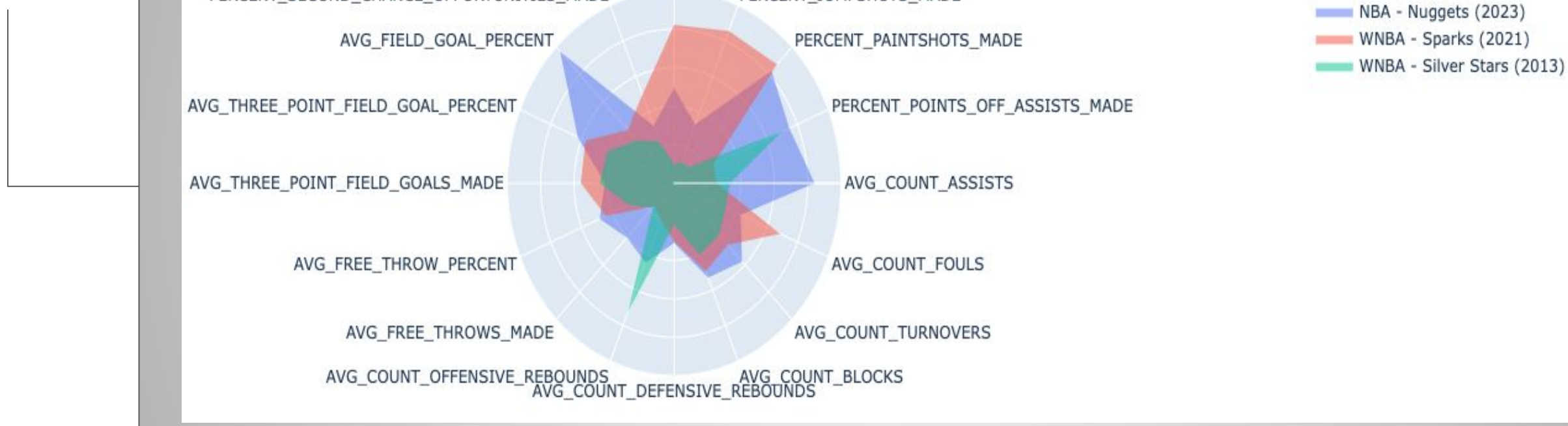
- 1.3GB disk size
- 20K records representing individual game statistics
- 4.3M records representing individual plays and location on the court

To manage data transformation efficiently, we built a data warehouse in Google BigQuery using the data build tool (dbt). The pipeline produces clean, structured tables with the following key datasets:

- Team averages:** Game stats (e.g., assists, turnovers) aggregated per season.
- Season outcomes:** Total wins and identification of the champion each season.
- Play-by-play stats:** Advanced stats highlighting how teams score points, e.g., assisted points, % points in paint

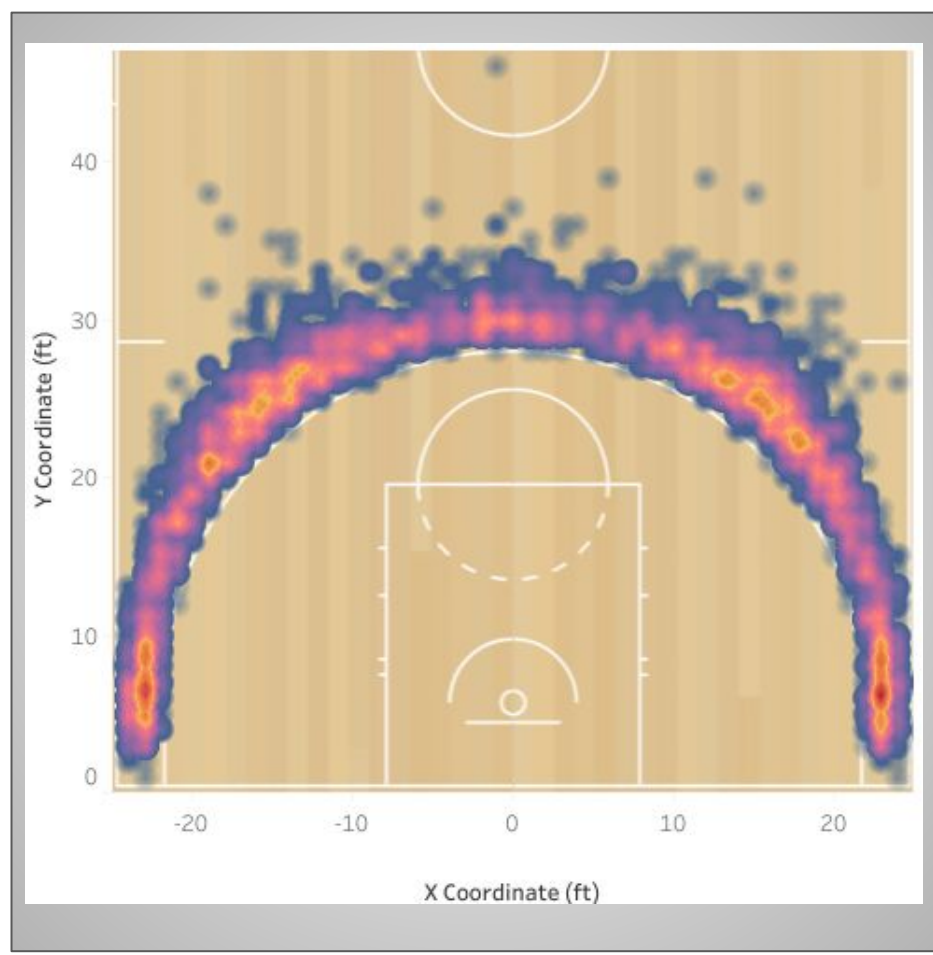


## Radar Chart

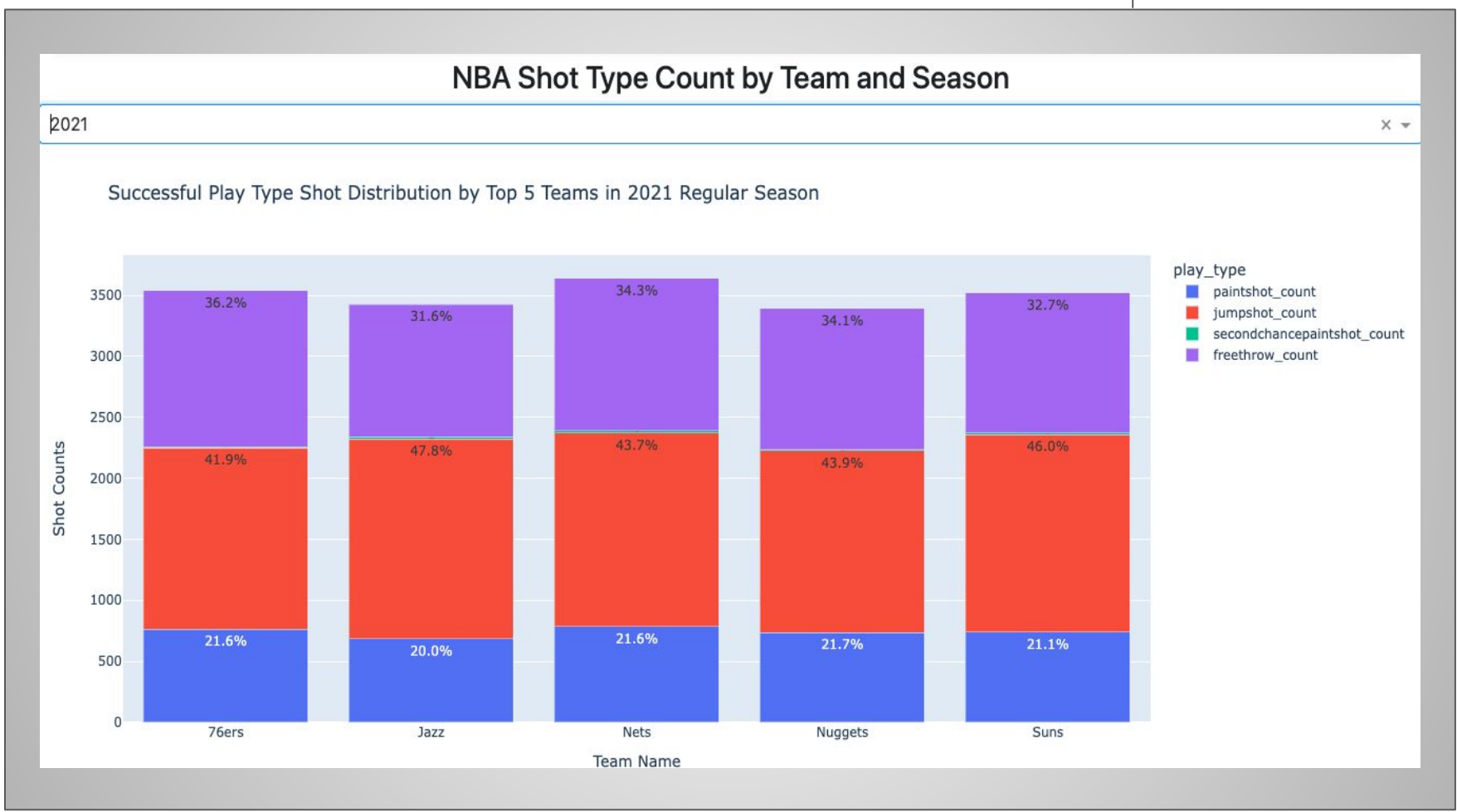


## Season Win Predictor

## Stacked Bar Chart



## Shot Coordinate Chart



## EXPERIMENTS & RESULTS

- Scalability Evaluation:**
  - We assessed our solution's efficiency by measuring build time of our data pipeline and response times for our What-If Scenario tool for 6 season and 12 season WNBA datasets. Results showed a less-than-linear increase of 58% in total computation time from 5.8 minutes to 9.2 minutes, confirming solution scalability.
- Prediction Accuracy:**
  - We compared LightGBM, Linear Regression, and Random Forest models. The models were evaluated on two criteria:
    - Predicting regular season wins with a margin of  $\pm 30\%$ .
  - NBA OLS regression outperformed the other models, with the lowest mean absolute error (MAE), lowest mean squared error and highest R squared
  - WNBA LightGBM MAE and MSE were substantially lower than the NBA's models
- Model Insights:**
  - Feature Importance:** For the NBA, the most significant predictors were **field goal percentage**, and **defensive rebounds**. In contrast, the WNBA model emphasized a broader split across features and relied much more on **per\_second\_chance**—indicating different strategies between the leagues.
  - The WNBA model showed lower prediction errors, possibly due to a smaller average season length (18.52 games) compared to the NBA (38.72 games). However, relative errors were comparable at 19.49% for the NBA and 23.43% for the WNBA.
- Key Findings from Visualizations:**
  - Radar Chart:** The 2023 champions from both leagues showcased contrasting strategies—while the WNBA Sparks had minimal turnovers and fouls, the NBA Nuggets excelled in many field goals
  - Shot Type Analysis:**
    - Three-Point Shots:** The NBA shows a higher concentration of corner threes due to a shorter distance from the basket compared to the WNBA, highlighting differences in court layouts.
    - Two-Point Shots:** The WNBA favors mid-range shots, while the NBA focuses more on high-efficiency shots near the rim or beyond the arc.

These results demonstrate how team metrics and player stats can be leveraged to predict performance and highlight strategic differences between the NBA and WNBA.