DSCI 510 Final Project Report

Project Name: The Unrivaled Basketball League 2026 Season Ranking Predictions

Team Members:
Mengxiang Xiao (USC Email: mxiao712@usc.edu; USC ID: 2051232216)
Zike Zeng (USC Email: zikezeng@usc.edu; USC ID: 5035384106)

Project Description:

Our project is designed to predict the ranking of the eight teams in the Unrivaled Basketball League for the upcoming 2026 season. For the 2026 season, the Unrivaled League will have eight teams and forty-eight players.

To determine the key metrics used to calculate the player's individual score and the overall team score, we collected the team statistics from the Unrivaled League 2025 Season on the Unrivaled official website. We used the statistics from the top 3 teams as a benchmark and compared them with the overall league averages. The key metrics were standardized and weighted according to their effect sizes. Then, we generated a formula to generate each player's personal score. We collected the 2025 season individual statistics from the WNBA website and calculated a final score for each player in the 2026 Unrivaled League using the scoring formula. Then, for each team, we aggregated the scores of all its players to get the final team score by calculating an average. Our final result was a ranking of the eight teams in the Unrivaled Basketball League for the 2026 season.

Data Sources:

1) The first source of our data was the Unrivaled Basketball League official website (https://www.unrivaled.basketball/). Specifically, we retrieved the Unrivaled League 2025 Season team statistics (https://www.unrivaled.basketball/stats/team) and team information (https://www.unrivaled.basketball/clubs), including the club names and their corresponding members, from the official website. We scraped this website page using BeautifulSoup.

2) The official WNBA statistics API is the second data source of our project. By sending an HTTP GET request with a custom HTTP header including a user-agent string simulating a standard web browser and a referrer header matching the official WNBA statistics webpage, the performance data of all the players appearing in the 2025 WNBA Regular Season is retrieved. This is because the individual data of each player contributes significantly to the performance data of their team as a whole, which is necessary for predicting the team rankings.

Number of Data Samples: 158 data samples were collected in total

1) **Unrivaled 2025 Season Teams Dataset:** The Unrivaled 2025 season teams dataset contains 6 data samples, including the statistics for all six teams in the Unrivaled League for the 2025 season.
2) **WNBA Dataset:** The WNBA dataset consists of player-level statistics for all players who appeared in the 2025 WNBA Regular Season, 104 in total. Each observation represents an individual player, resulting in a dataset containing statistics for the full population of players recorded by the league in 2025.

3) **Unrivaled Club Dataset:** The Unrivaled Club dataset includes 48 publicly listed members across eight league clubs. Each row corresponds to a single club member, providing a complete overview of club composition for team-level analysis.

Data Cleaning, Analysis, and Visualization:

- **Process:**
  1) **Data Cleaning:**

     Step 1: Key Metric Filtering

     After retrieving the team statistics from the Unrivaled League website, we performed data cleaning and preparation to identify the key metrics most relevant to our data analysis and ranking prediction. The following steps describe this data cleaning process. First, we imported the pandas library and read the CSV file that contained the Unrivaled team statistics from the 2025 season. Second, we identified the top three teams from the last season and filtered their corresponding statistics. Next, we calculated the mean value of each metric across the top three teams. Meanwhile, we calculated the mean of each metric across the entire league. Then, we computed the performance difference between the top three teams' averages and the entire league's averages for each metric. Those performance differences were organized using a pandas series. By analyzing the performance differences, we selected four key metrics: PTS (points per game), STL (steals per game), AST (assists per game), and TOV(turnovers per game). Therefore, after collecting the full WNBA player statistics dataset, we performed an initial data cleaning step by retaining only these four key metrics as the starting point of our scoring model.

     Step 2:Cross-Table Linking and Intersection Filtering

     In the second cleaning step, we integrated the Unrivaled club data with the WNBA player statistics by performing a cross-table merge using a standardized player identifier (NAME_KEY). To enable accurate matching across datasets, player names from both sources were first standardized by converting all characters to uppercase and removing non-alphabetic characters, thereby resolving inconsistencies in name formatting. After merging the two tables, we removed all non-overlapping records to ensure that the final dataset included only players who are both listed as Unrivaled club members and present in the WNBA 2025 statistics dataset.

  2) **Data Analysis and Visualization:**

     After identifying the key metrics most relevant to the player and team performance, we calculated the weight for each key metric by analyzing the performance difference. Each weight was obtained by dividing the performance difference of each metric by the sum of the performance differences of all metrics. We got the weighted score formula: PTS * 0.61 + STL* 0.12 + AST * 0.07 - TOV * 0.21. We decided to manually adjust this weight formula to PTS * 0.61 + STL* 0.11 + AST * 0.07 - TOV * 0.21. We made this change because steals were less important in 3 vs. 3 games, and the adjusted formula would make future calculations more fluent.

After data cleaning, we analyzed the data at both the player and club levels. At the player level, we applied a custom weighted scoring model (PTS * 0.61 + STL* 0.11 + AST * 0.07 - TOV * 0.21) to calculate each player's individual score. At the club level, we computed the average player score for each Unrivaled club to produce a team-level performance indicator.

For visualization, first, we used Matplotlib to generate a bar chart to show the performance difference by each metric. This bar chart provided a clear visualization of the relative importance of each metric to the team's performance. Second, we used Matplotlib to generate a pie chart to demonstrate the weight of each key metric. Finally, we used Matplotlib to create a line chart showing average scores across clubs. Because the values were closely clustered, the y-axis scale was adjusted to make differences more visible.

- **Hypothesis/Premise and Conclusions:**
  1) **Hypothesis:** Before conducting the prediction, we hypothesized that Mist, Lunar Owls, and Phantom were going to be the three most competitive teams in the Unrivaled League for the 2026 Season. We believed Mist was very competitive for this upcoming season because Allisha Gary joined the team while Breanna Stewart was still on the team. Last season, Allisha Gary was a member of Lunar Owls, the championship-winning team. Breanna Stewart is the founder of the Unrivaled League, who has won three WNBA championships. We believed Lunar Owls had great potential because they had both experienced players like Napheesa Collier, 2025 Unrivaled League MVP, and young players like Aaliyah Edwards. We also believed that Phantom was competitive because they had two first overall draft picks: Aliyah Boston, the first overall pick in the 2023 WNBA draft, and Kelsey Plum, the first overall pick in the 2017 draft.

  2) **Conclusion:** As shown in the figure, Vinyl and Laces achieve the highest average club scores, while Mist, Phantom, Lunar Owls, and Breeze form a middle cluster with relatively close values, and Rose ranks lowest. This contradicts our former hypothesis. The results suggest that average-based scoring favors clubs with more consistently high individual scores across players. Teams with greater variance in player performance may therefore rank lower at the club level, even if they include high-performing individuals. Overall, the findings indicate that data-driven performance evaluation can challenge intuitive or expectation-based rankings, reinforcing the value of quantitative analysis in reassessing team strength.

Changes from Original Proposal:

In the original proposal, we mentioned that we were going to ESPN as our source because we thought that the official website of the Unrivaled Basketball League and the WNBA could not provide enough data that we needed. However, as we started our project, we found that the data provided by the Unrivaled Basketball League and the WNBA website was very comprehensive, and we were able to find all the metrics we sought. Therefore, in the final project, we only used the official website of the Unrivaled Basketball League and the WNBA as our two main sources of data. Also, previously, we planned to normalize player scores to a 0–100 scale to facilitate interpretation. However, during implementation, we decided not to apply explicit score normalization. This adjustment was made because the weighted scoring model already produces comparable relative scores across players, and additional normalization was not necessary for ranking or club-level comparison. Instead, we focused on preserving the original score distribution to maintain transparency in how individual statistics contributed to the final score. This change simplified the analysis while still supporting meaningful comparison across players and clubs.

Mention of Future Work:

After completing the project, we reached an agreement that future work on the Unrivaled League predictions could expand the dataset by collecting player statistics from multiple seasons, such as the past five years, to increase the overall sample size. Incorporating multi-season data would allow the model to capture longer-term performance trends and potentially improve the accuracy of the predictions.