

Computer Science Department
Fall 24
Data Science

Project Proposal

Students' Names	IDs	Section
Maram Alhusami	S21207443	2
Bushra Alshehri	S21107272	2
Sara Imran	S22107878	2
Amani Albarazi	S19206062	2

I. Project Name:

Analysis of Women's Tennis Association (WTA) Match Data for the year 2023 and 2024

II. Description:

This project analyzes match data from the Women's Tennis Association (WTA) to identify key factors influencing player performance and develop predictive models for match outcomes and player rankings. The datasets include:

- **Match Statistics Dataset:** Contains detailed information about match results, player statistics, and performance metrics.
- **Player Ranking Dataset:** Provides historical player rankings and performance over time.
- **Tournament Dataset:** Details about different tournaments, including surface types and locations.
- **Player Attributes Dataset:** Includes demographic data such as age, handedness, and activity levels.

III. The Goal:

The primary goal of this project is to uncover patterns in player performance and predict match outcomes by leveraging the provided datasets. Specifically, the project aims to:

1. **Performance Analysis:** Identify significant factors influencing player performance in WTA matches.

Computer Science Department
Fall 24
Data Science

2. **Predictive Modeling:** Develop models to predict match outcomes based on player and match characteristics.
3. **Surface Type Analysis:** Investigate how different surfaces affect player performance.
4. **Ranking Insights:** Analyze the relationship between player rankings and match outcomes.
5. **Actionable Insights:** Provide recommendations for players and coaches based on data-driven insights.

IV. Dataset Reference:

https://github.com/JeffSackmann/tennis_wta

- Match Statistics
- Player Ranking
- Tournament Information
- Player Attributes

Proposal:

1. High-Level Statement of the Problem:

While extensive research has been conducted on men's tennis, studies focusing on the WTA remain limited. This project seeks to fill the gap in understanding how various factors contribute to player performance and match outcomes, particularly in the context of women's tennis.

Research Question:

How can the Factors of the Player's Performance in WTA Matches can be Utilized to Predict Match Outcomes and Player Rankings?

Existing Research:

Prior studies have emphasized player rankings and surface types as critical predictors. However, gender-specific dynamics and the interaction of various predictors in women's tennis have not been thoroughly explored.

2. Outcome Variable:

Winner Rank: representing the ranking of the winning player in a match.

- **Conceptual Relation to Research Question:** Winner rank is directly correlated with player performance, serving as an indicator of match competitiveness.

Computer Science Department
Fall 24
Data Science

- **Summary Statistics:**
 - Mean Rank: 586.35
 - Standard Deviation: 348.58
 - Range: 22 to 1373

Match Outcome: Indicates whether a player won or lost the match.

- **Conceptual Relation to Research Question:** The match outcome is crucial for understanding player performance in WTA matches and provides a binary classification for predictive modeling.

3. Predictor Variables:

To model the outcome variable, the following predictors will be utilized:

1. Player Age
2. Tournament Surface Type
3. Handedness
4. Serve Effectiveness
5. Match Duration
6. Opponent Ranking
7. Win-Loss Ratio
8. Break Point Conversion Rate
9. Player Activity
10. Country Representation

Data Sources: The Jeff Sackmann WTA dataset and other relevant sources.

4. Definition of “Success”:

Success will be defined by the ability to accurately predict match outcomes and player rankings using the selected variables. Key indicators of success include:

1. **Predictive Accuracy:** Achieving over 75% accuracy in match outcome predictions.
2. **Insights Generated:** Identifying key performance factors and their implications.
3. **Application of Findings:** Relevance of insights for real-world scenarios, such as coaching and player development.

5. Practical Applications:

Computer Science Department
Fall 24
Data Science

1. Recommendations for optimizing player training and match strategies.
2. Insights for enhancing fan engagement through data-driven narratives.
3. Contributions to the broader field of sports analytics by providing a comprehensive understanding of women's tennis dynamics.

By achieving these goals, this project will demonstrate a high proficiency in data analysis, provide meaningful contributions to the understanding of performance dynamics in women's tennis, and offer actionable insights for coaches, players, and analysts to enhance competitive strategies and player development.