**Project\_0507\_03\_Readme File**

**Brand Name: StatsPlay Insights**

**Project Description:**

Our client, *National Endowment for the Arts* consulted StatsPlay Insights to analyze the previous years University of Maryland Women's Softball Team’s performance and help their upcoming artists to identify few matches where they can perform to get a good recognition in the start for their career in the world of arts.

**Data Sources:**

UMTerps Women’s Softball Match Schedule:

<https://umterps.com/sports/softball/schedule>

UMTerps Women’s Softball Coach Data:

<https://umterps.com/sports/softball/coaches>

**Data Highlight:**

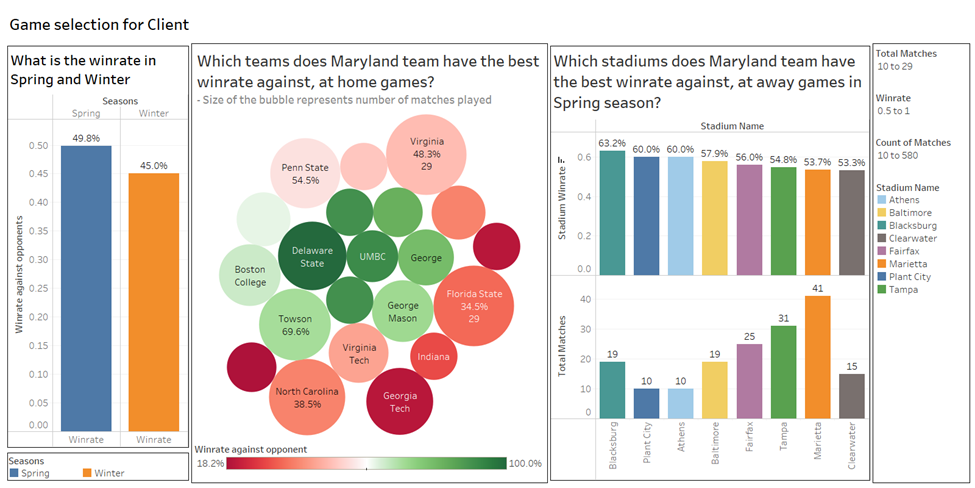
The database consists:

* A total of #1597 games spanning over years from 1995 to 2003
* Played against #242 Opponents
* Located in #101 stadiums in USA & Mexico

**Data Cleaning**

1. Removed unnecessary columns: Checked what all columns are necessary for our analysis and deleted the ones which don’t contribute. (Ex. time and day of matches played, tournament names, etc.)
2. Identified missing data, put NULL onto the missing data entries. Checked for duplicates, deleted all duplicate entries.
3. Fixed data inconsistency: this included converting all state names to 2 letter State acronyms in Capital, renaming opponent team names to make data consistent, etc.
4. Validate and Correct Data Types: Ensure that each variable has the correct data type (numerical, categorical, date, etc.). Convert variables to the appropriate data type if needed.
5. Document Changes: Keep a record of all the changes made during the data cleaning process. We did this to maintain data transparency and reproducibility.
6. Perform Exploratory Data Analysis: After cleaning the data, we conducted exploratory data analysis to gain insights and to identify any additional issues.

**Data Visualization:**



*Q1. What is the winrate in Spring and Winter?*

This bar chart represents the cumulative win rate percentages played in spring and winter seasons. We analyzed this to understand seasonal win rate patterns, since our client had informed about their inclination to have their performances planned during comfortable outdoor weather conditions.

This visualization can help us to determine that we have an overall higher winrate in the Spring season, when compared to Winter season.

*Q2. Which team does the Maryland team have the best win rate against, at home games?*

In this visualization, the size of the bubble represents the number of matches played, which means bigger the bubble size, more the number of matches played against that particular team.

Secondly, the color of the bubble signifies the percentage of wins against that particular team. Deep red represents the lowest percentage of win rate and dark green represents the highest percentage of win rate.

This visualization helped us to assess games where we have a high chance of winning at home games, and we can recommend our client accordingly where they can plan to organize the performances.

*Q3. Which stadiums does Maryland team have the best win rate against, at away games in spring season?*

This visualization consists of 2 bar charts (one above another).

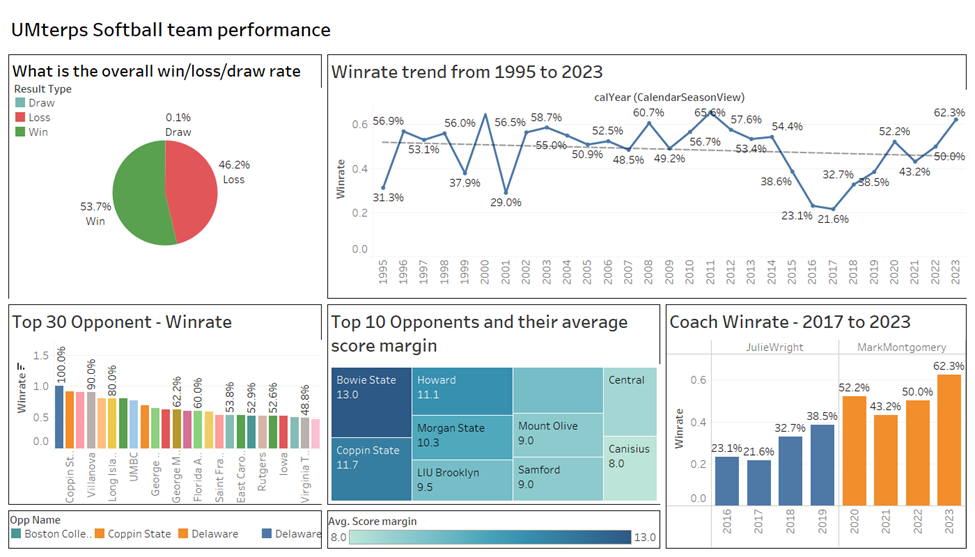
The upper chart shows the percentage of win rates of Maryland teams against opponents, sorted in descending order of percentage.

The lower chart represents the number of matches played against that opponent.

Note that:

1. These charts only show wins and number of matches played during the spring season, following the client’s request that they wanted to perform during comfortable outdoor weather conditions.
2. The minimum number of matches played have been set to 10 to make substantial analysis.
3. The minimum win rate percentage is set to 50%.

This visualization helped us to assess games where we have a high chance of winning at away games during spring season and we can recommend our client accordingly where they can plan to organize the performances.



*Q4. What is the overall win/ loss/ draw rate?*

Here, we have visualized the overall performance of the Maryland team using a pie chart. We can see that overall, the team has a majority percentage of wins.

This helped our client to gain an overall insight to the Women’s Maryland Softball team’s performance.

*Q4. How has the Winrate trend been for the team from 1995 to 2003?*

We chose to visualize Winrate trend of the team from the year 1995 to 2003. We notice that though, the performance over the years has been quite volatile, there has been an upward trend that can be seen from the years 2016-2017 till 2023.

*Q5. What are the top 30 Winrate percentages against opponents?*

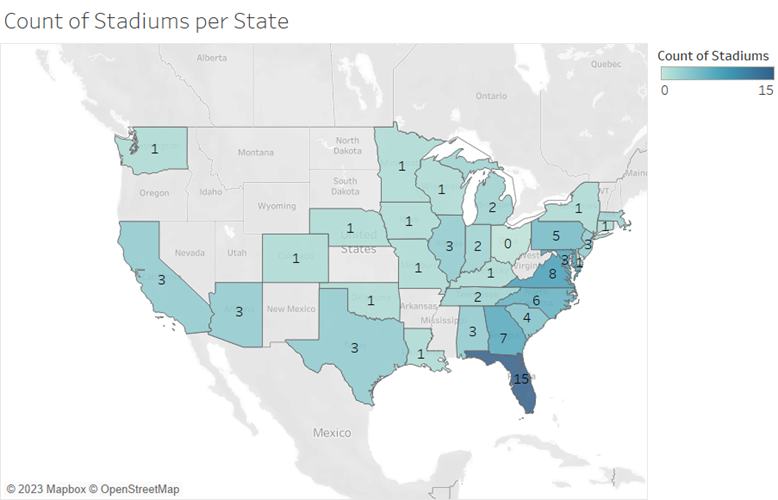
This plot helps us list the top 30 opponents against which the Maryland team has the top Winrate against. We came to know that out of these 30, Maryland has maintained a win rate of over 50% against 28 teams in the last 28 years.

*Q6. What are the top 10 opponents and their average score margin?*

This treemap shows the top 10 opponents of the Maryland Softball team and the average score margins of games played between the years 1995 and 2003. The plot represents the teams against which the Maryland team stand the strongest against and can be classified as safe teams to play with.

*Q7. How has the Maryland team’s Winrate varied under different coaches over the years?*

Though our dataset only includes coach data from the year 2016, we plotted Maryland team’s Winrate percentage under the respective coaches and observed an upward trend. This hints that the team has been under effective leaders, and skilled trainers, which means we can expect consistency in the team’s performance.



*Q8. What is the count of Stadiums per state in which University Softball matches are held?*

The above geographical plot shows the number of Softball stadiums present in each state of the USA, where Maryland teams have played. We can see that the stadium density is high towards the East, specifically the South-East of the USA. This analysis can be used to assess if the performances are planned towards the dense side, it can mean that the performers can travel less, and cover the target audience more effectively.

**Project Deliverables**:

1. Lucid Chart (ER Diagram) : <https://lucid.app/lucidchart/d8b36fee-6525-446d-ac14-7ec329746453/edit?viewport_loc=-2227%2C-433%2C3387%2C1743%2C0_0&invitationId=inv_16efadec-dc0c-4aa0-9dd3-b628da4df214>
2. Project Proposal (Project\_0507\_03\_Proposal.docx)
3. Project SQL Create Table File (Project\_0507\_03\_Create.sql)
4. Project SQL Insert Table File (Project\_0507\_03\_Insert.sql)
5. Project SQL Queries (Project\_0507\_03\_DML.sql)
6. Project Visualization in Tableau (Project\_0507\_03\_Business Transactions.twb)
7. Project Presentation (Project\_0507\_03\_Presentation.pptx)
8. Project Readme File (Project\_0507\_03\_Readme.docx)

**Project Contributors**:

1. Athul Ramachandran *(*[*athulr@umd.edu*](mailto:athulr@umd.edu)*)*
2. David Dapaah *(*[*ddapaah@umd.edu*](mailto:ddapaah@umd.edu)*)*
3. Linta Joly *(*[*ljoly@umd.edu*](mailto:ljoly@umd.edu)*)*
4. Sahil Gupta *(*[*sahil993@umd.edu*](mailto:sahil993@umd.edu)*)*

**Document Version**:

1. 1.01, last edited 12/06/2023