#### MLS Penalty Outcome Analysis (2008-2022)

#### **Abstract**

#### This research project investigates penalty outcomes in the Major League Soccer (MLS) from 2008 to 2022, providing a data-driven analysis of one of the most pivotal moments in soccer. The study aims to provide insights into various aspects of penalty events, including home vs. away performances, player and goalkeeper statistics, and the influence of match attendance. By combining descriptive analytics with predictive modeling, this project offers a deeper understanding of the dynamics behind penalty events in the MLS.

#### **Introduction**

#### Penalty events are critical moments in soccer matches, often determining the outcome of games. This capstone project combines advanced data analysis and machine learning techniques to analyze penalty events in the MLS over 15 seasons, uncovering patterns and trends that influence success rates. Through a blend of exploratory and predictive analytics, this project aspires to provide valuable insights into the mechanics of penalty events.

#### **Data Collection**

#### The primary dataset comprises 1328 entries corresponding to penalty events in MLS matches from 2008 to 2022. These entries were derived from 5 major datasets each containing detailed information about matches, events, players, goalkeepers, and seasonal statistics. Sourced from official MLS records and publicly available databases, ensuring comprehensive coverage and accuracy for the analysis; [kaggle datasets](https://www.kaggle.com/datasets/josephvm/major-league-soccer-dataset).

Applied data processing techniques to simplify the datasets and gather the data I was interested in, through the next steps:

* Identifying Penalty Events:
  + From the events dataset, all penalty-related events were extracted and classified into three categories:
    - Scored; Successful penalties resulting in a goal.
    - Missed; Penalties that were off-target or hit the post.
    - Saved; Penalties successfully blocked by the goalkeeper.
* Relating matches to events:
  + A relational mapping was established between matches and events using the match ID column as the key. This step ensured that penalty events were contextualized within their respective matches.
* Player Analysis:
  + The players’ dataset was analyzed to identify penalty takers and collect their performance metrics to evaluate the effectiveness of individual players over time.
* Goalkeeper Analysis:
  + The goalkeepers’ dataset was explored to obtain a comprehensive list of goalkeepers who participated in penalty events and extracted key statistics to assess individual performance trends.
* Seasonal Context and Derived Features:
  + Contextual features like competition and team standings were extracted to contextualize games.
  + Additional statistics such as home vs away venue, match timing, and scoreline, were engineered to enhance the analysis.

#### **Methodology**

#### The research is divided into six key topics:

#### **Dataset Description**

#### Detailed exploration of the dataset, including data preprocessing and descriptive statistics.

#### Techniques used: data wrangling, exploratory data analysis (EDA), and statistical summaries.

#### **Home vs. Away**

#### Analysis of penalty success rates for home and away teams.

#### Techniques used: hypothesis testing, and visualization.

#### **Player Performances**

#### Evaluation of individual player performance in penalty situations.

#### Techniques used: success rate analysis, and performance ranking.

#### **Goalkeeper Performances**

#### Assessment of goalkeeper performance during penalty events.

#### Techniques used: save rate analysis and performance ranking.

#### **Attendance Influence**

#### Investigation of the impact of match attendance on penalty outcomes.

#### Techniques used: correlation analysis and visualizations.

#### **Technologies Used**

* **Programming Languages**: Python for data analysis and modeling, SQL and Excel for data management.
* **Libraries and Tools**:
  + Pandas and NumPy for data manipulation.
  + Matplotlib and Seaborn for data visualization.
  + Scikit-learn for machine learning models.
  + Jupyter Notebook for interactive data analysis.

#### **Results**

The analysis revealed significant insights into penalty event patterns in MLS:

* Home teams have a statistically higher success rate in penalty events.
* Experienced players consistently outperform in penalty situations.
* Goalkeepers' performance varies widely, with some exhibiting exceptional save rates.
* Higher attendance appears to positively influence home team performance during penalties.

#### **Conclusion**

This capstone project provides a comprehensive analysis of penalty events in MLS, offering valuable insights for games. The findings highlight the importance of considering various aspects such as home vs away venue, player and goalkeeper performances, and other external factors influence when preparing for penalty situations. Overall, the consistent positive trends expose throughout the research, underline the development of soccer in recent years

**References**

‘Learn to code with soccer’ by Nathan Braun