

The MinneMUDAC 2023 Challenge is centered on America’s National Pastime – Baseball! Participants of this year’s student data challenge will be required to gain an understanding of how various factors influence attendance at Major League Baseball (MLB) games.

MinneMUDAC 2023 has teamed up with the MN Twins for this year’s challenge. Student teams should consider the Minnesota Twins / MLB as their clients for this year’s challenge. Each team will be required to present their outcomes/findings in-person at Target Field on March 25. Student teams will give their 6-8 minute presentation to a group of judges (a judge group consists of 2-3 working professionals in analytics). Teams will give their presentation repeatedly – likely between 4 and 6 times. Outcomes will be evaluated using the published rubric for this year’s challenge.

**Novice Division**

* Understand and analyze the distribution of home-game attendance for the MN Twins and other MLB teams
* Identify factors that tend to influence home-game attendance for the MN Twins
* How do the important factors that tend to influence home-game attendance differ between the MN Twins and other MLB teams?
* Provide insights into how the important factors identified above could be used by the MN Twins (or other MLB organizations) to better inform marketing strategies, staffing strategies, supply chain strategies, fan experience, etc.

Bonus Element for Novice Division

* Predict the game-by-game attendance for the MN Twins home games

**Undergraduate Division**

* Complete all elements of the Novice Division
* How do the important factors that tend to influence home-game attendance differ across teams / leagues / geographic regions / etc?
* Build/construct predictive models for the game-by-game attendance for home games for all MLB teams for the 2023 season
* Discuss the strategies employed to build/construct your predictive models

Bonus Element for Undergraduate Division

* Build/construct a dynamic predictive algorithm that could be used at any time-point in the season to improve upon the predictions made for attendance at home games to be played, say, in the following week.

For example, the MN Twins have a six day home stand starting May 9 - May 14 against the Padres for a 3-game series and Cubs for a 3-game series. Your dynamic predictive algorithm should utilize the important factors identified above plus any additional factors (e.g., current record, player performance, weather forecast, etc) that are currently available through a specified time-point (e.g., May 5) to predict attendance for each of the six games in this home stand.

**Graduate Division**

* Complete all elements of the Undergraduate Division
* Build/construct a dynamic predictive algorithm that could be used at any time-point in the season to improve upon the predictions made for attendance at home games to be played, say, in the following week. (see description above for Bonus Element for Undergraduate Division).

Bonus Element for Graduate Division

* Twins 2023 Schedule Optimizer: Develop a model-driven approach for creating an optimal 2023 game schedule for the MN Twins that maximizes game-by-game attendance for home games. You must recommend a MN Twins home game schedule for 2023 that consists of 81 total home games. You may choose the dates, times, and opponents for each game with the objective of garnering the highest possible MN Twins home game attendance for the 2023 season. Explain why you believe your proposed schedule would allow the MN Twins to maximize attendance across the 81 home games. The following guidelines must be followed when constructing the optimal schedule for the MN Twins.

Guidelines:

* + The 81 home games must be scheduled on dates between April 1, 2023 and September 30, 2023.
  + A maximum of 41 weekend game dates (Fri, Sat, or Sun) are allowed.
  + The optimal schedule must include the same quantity of games against the opponents listed on the current 2023 MN Twins’ home game schedule. You may re-allocate these opponents across the season as you see fit.
  + To minimize travel for the teams, each MN Twins opponent must be scheduled for a minimum of two consecutive games before a new opponent can be played.

**Additional Considerations**

* The use of relevant data from other sources is encouraged. The data must be free to use and publicly available. Teams are not allowed to use proprietary/private data.
* Disruptions (e.g. COVID, 9/11 tragedy, etc) have occurred over time that have adversely affected attendance. Attendance for games affected by such tragedies can be removed from consideration for this challenge.
* Games may be postponed (or canceled) for various reasons. Postponed games are sometimes rescheduled as double-headers. Attendance outcomes for double-headers are not recorded in a consistent manner. Thus, all games that are identified by MLB as being postponed will be excluded from consideration in the quality of prediction metrics.
* MLB baseball has a small number of games being played at atypical locations (Estadio Alfredo Harp Helu, London Stadium, and Muncy Bank Ballpark). These games will be excluded from consideration in the quality of prediction metrics.