

Challenge Questions

| MinneMUDAC 2025 has teamed up with Big Brother Big Sisters Twin Cities for this year’s challenge. Big Brother Big Sisters Twin Cities the largest and oldest youth award-winning mentoring organization in the greater Twin Cities. Each year, we match up youth (Littles age 8-13) and their families with caring adults (Bigs) who volunteer as mentors. Through a variety of community-based, school-based, and workplace-based mentoring programs, and together with our community, we want every youth to have a mentor, be affirmed in who they are, and explore who they want to be. |  |
| --- | --- |
|  | |

**Novice Division**

* Data: Novice.xlsx [ [Link](https://docs.google.com/spreadsheets/d/1pk57TQPlUxO6ZDX71s-wvq868GuG5WZF/edit?usp=sharing&ouid=107480654748844323422&rtpof=true&sd=true) ]
* Response variables of interest
  + Match Length
  + Closure Reason
* Understand and analyze the *response variable* distributions
* How do the *response variable* distributions vary across Program Type?
* Have the *response variable* distributions changed over time?
* What influence do the various Big and/or Little demographic variables have on the *response variable* distributions?
* What influence does the alignment of the demographic and/or interest variables (e.g. same gender, similar Hobbies, etc) have on the *response variable* distributions?
* A successful match is 1) a match that remains active, 2) a match that lasts a long time, 3) a match that closes due to the age of Little, or 4) Successful Match Closure for the Closure Reason. Provide insights for Big Brothers Big Sisters Twin Cities into what factors tend to lead to a successful (or unsuccessful) match

**Undergraduate Division**

* Dataset #1: Training.xlsx [ [Link](https://docs.google.com/spreadsheets/d/1qONX0t_4QiOz6H5wKWsIR2XExobXh8tR/edit?usp=sharing&ouid=107480654748844323422&rtpof=true&sd=true) ]
* Dataset #2: Test-Truncated.xlsx [ [Link](https://docs.google.com/spreadsheets/d/1beGPQyBlAFJPSuMR-byHfMUdw9oyX0Cr/edit?usp=sharing&ouid=107480654748844323422&rtpof=true&sd=true) ]
* Complete all elements of the Novice Division
* Does the cadence (i.e. days between calls) of the Match Support Calls have an impact on the distribution of Match Length?
* Consider the Match Support Contact Notes open text field.
  + Are there key words, phrases, or sentiments within this field that are related to Match Length?
  + Do the key words, phrases, or sentiments differ between the early stages and late stages of a match?
* Big Brother Big Sisters Twin Cities would like to improve upon their ability to identify whether or not a match is likely to remain active or close in the near future. Provide a small set of questions that a call coordinator should ask when making Match Support Calls that would be useful in understanding whether or not a match is at-risk for closure in the next 3-6 months. These questions should leverage information from all fields including insights gained from the Match Support Contact Notes and should be structured as drop-down options with a limited set of responses.
* Predictive Component
  + Create a predictive model for Match Length
  + Build the model using the Training set. You will be required to submit predictions for Match Length for cases in the Test-Truncated set. Additional information will be forthcoming regarding the submission of your predictions.
  + Root Mean Square Error (RMSE) will be the metric used to evaluate predictions for the Test cases.

Comments:

1. The difference between Match Activation Date and most recent Completion Date is strongly correlated with Match Length. Thus, the Match Support Contact Notes for the cases in the Test set have been truncated. That is, the last few instances of Match Support Contact Notes have been removed from the Test-Truncated. Additional explanation of this process is provided in the Data Sources document. [ [Link](https://docs.google.com/document/d/1bOtrG8VYk-L38WWTrmre1P3_Igj4lpZiNErEe6Xr_mA/edit?usp=sharing) ]
2. The quality of your prediction will be evaluated objectively using RMSE. In addition, the rubrics used by the judges will contain items related to your predictive modeling strategy and your ability to leverage additional information beyond the cadence or number of Match Support Contact Notes related to each match.

**Graduate Division**

* Dataset #1: Training.xlsx [ [Link](https://docs.google.com/spreadsheets/d/1qONX0t_4QiOz6H5wKWsIR2XExobXh8tR/edit?usp=sharing&ouid=107480654748844323422&rtpof=true&sd=true) ]
* Dataset #2: Test-Truncated.xlsx [ [Link](https://docs.google.com/spreadsheets/d/1beGPQyBlAFJPSuMR-byHfMUdw9oyX0Cr/edit?usp=sharing&ouid=107480654748844323422&rtpof=true&sd=true) ]
* Complete all elements of the Novice Division and all elements of the Undergraduate Division
* Consider the Match Support Contact Notes open text field. Are there key words, phrases, or sentiments within this field that are related to Closure Reason?
* In addition to identifying key words, phrases, and sentiments in the Match Support Calls, are there emotional tones present in this field that are useful in understanding Match Length or Closure Reason.
* Consider the Rationale for Match open text field. Are there key words, phrases, or other information contained in this field that are related to Match Length or Closure Reason?
* Consider the Closure Details open text field. Are there key words, phrases, or other information contained in this field that could be used in the development of the small set of questions that a call coordinator should ask when making Match Support Calls.
* Provide an intervention strategy that could be used by Big Brothers Big Sisters Twin Cities to reduce the risk of a match closing for the common Closure Reasons. What factors are most important in developing an early intervention model to reduce the risk of a match closing?
* The data provided contains information regarding the Census Block Group ID for the Big and Little for the Community Program Type. Obtain any relevant Census data that would prove useful to Big Brothers Big Sisters Twin Cities in their goal of understand the success of a match.