

Humana-Mays

2023 Healthcare Analytics Case Competition

Seventh Annual

Humana



TEXAS A&M UNIVERSITY

Mays Business School



Competition Overview

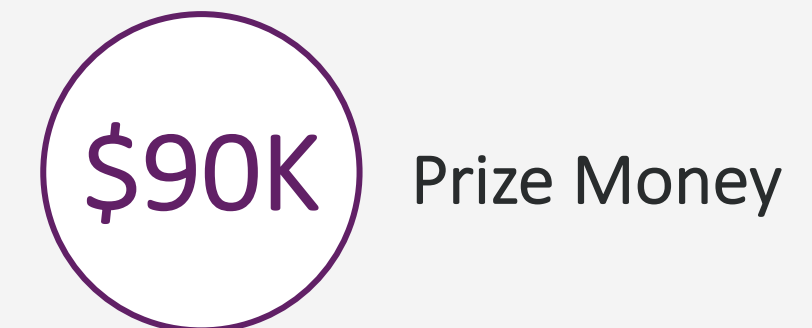
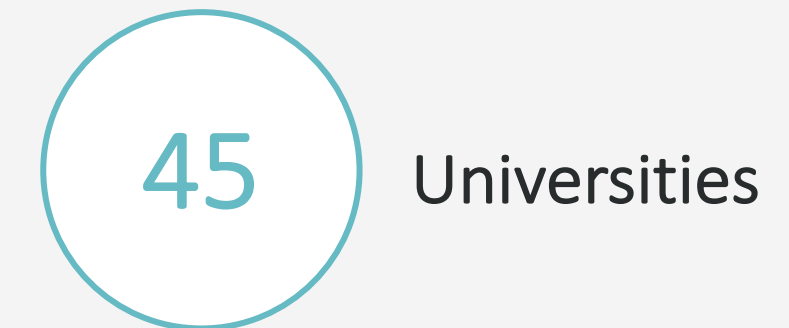
The Challenge

This is an opportunity for students to showcase their analytics skills to solve real-world business problems using Humana's data

Eligibility Requirements

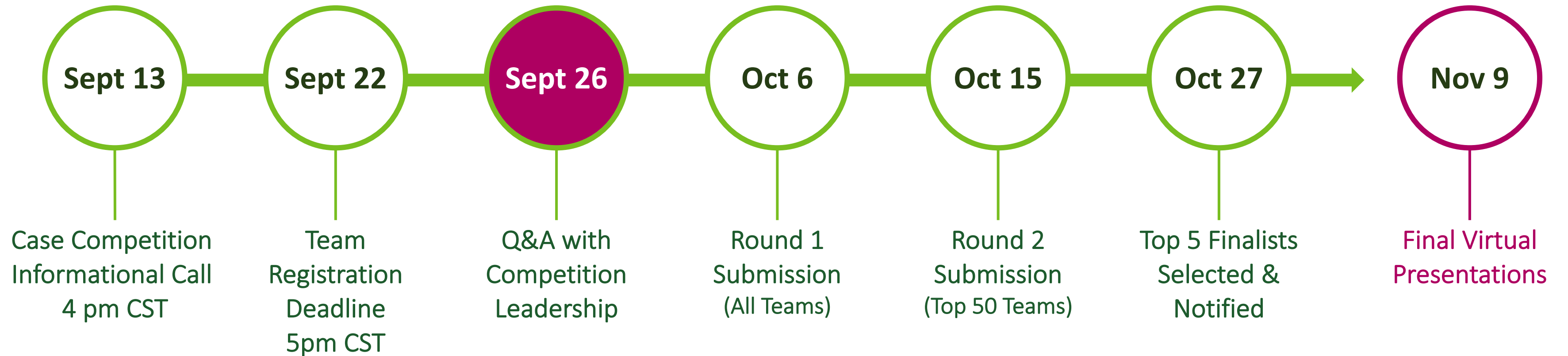
- Student must be enrolled part- or full-time in an accredited Master of Science, Master of Arts, Master of Information Systems, Master of Public Health, Master of Business Administration, or similar master's programs that educate in business, healthcare, or analytics located within the US
- One entry per team; teams must have 2-4 members from the same school
- If a participant drops out of the competition, no substitution is permitted
- If the team falls below the 2-person minimum due to a member leaving, the team is no longer eligible to compete
- Students may only participate on one team
- Humana Employees are not eligible

2023 Participation





Competition



Submitted Questions



Competition Related

Can we use any program or language that we would like in organizing the data? In other words, are there any restrictions as to how we do our analysis?

A: There is no restriction on how you do your analysis.

What, exactly, needs to be submitted on October 6, for Round 1?

A: Only one thing is required on 10/6 to be considered a complete submission for Round 1. Please be sure to follow requirements listed below:

Scored Holdout File: Your final model applied to the holdout file

*File Name: **2023CaseCompetition_TeamCaptainFirstName_TeamCaptainLastName_2023mmdd.csv***

Fields to include: ID, SCORE, RANK

Accepted Format: CSV

(Failure to comply with submission requirements may result in disqualification)

Assuming we make it in the Top 50 after Round 1, what, exactly, needs to be submitted on October 15th, for Round 2?

A: Only the Top 50 from Round 1 need to submit a write-up for Round 2. Please be sure to follow requirements listed below to be considered a complete submission:

Written submission: Summary of analysis, insights, and recommendations.

*File Name: **2023CaseCompetition_TeamCaptainFirstName_TeamCaptainLastName.doc***

Accepted format: Word, PDF

(Failure to comply with submission requirements may result in disqualification)

Is the PowerPoint for the Finals due along write-up at the time of submission on October 15th?

A: No, the presentation for the finals is not due on Oct 15th. The PowerPoint required for the final presentations are due on 11/8 and are only expected from the Top 5 finalists.

Is there a limit to the number of times a team can submit a holdout file to the Leaderboard for mid-cycle evaluation?

*A: A team can submit a holdout file **once per day** for the mid-cycle evaluation between Sept 25 and Oct 5. However, only the highest performance value for your team will be used in positioning you on the Leaderboard.*

If our team submits a scored holdout file to the leaderboard between Sept 25 and Oct 5 for mid-cycle evaluation, do we still need to submit a 'final' scored file at the Round 1 deadline on Oct 6th?

A: No, it is not required. If you have participated in the leaderboard and are happy with your highest performing scored file - submitted prior to the Round 1 deadline of Oct 6th - that highest performing file will be recorded as your qualifying entry for Round 1.

Competition Related

Can teams have a faculty advisor, or receive support from outside sources?

*A: As stated in the [Official Rules](#), coaching and mentoring from outside sources, other than your registered teammates, is **not** allowed. These outside sources include but are not limited to, university faculty, university teaching assistants, university staff, or other professional consultants in related fields.*

Once the Q&A session is over, can we still submit questions pertaining to the case? If so, when can we expect to receive answers?

A: We will accept questions through Thursday, October 5th. ALL questions will be answered within two business days and posted in the FAQ section of the website.

What are the judging criteria for accuracy and fairness?

A: Round 1 is evaluating modeling accuracy & fairness using objective metrics based on the HOLDOUT file returned by the participants.

Accuracy: ROC/AUC measure will be calculated

Fairness: Disparity score calculated using RACE & SEX

Additional details are contained in the "Fairness in AI Guide 2023" available on the website under FAQs.

Where can I find additional information about the Leaderboard?

You can find the all the details related to the 2023 mid-cycle leaderboard on the website under the FAQs.

How are Round 1 and Round 2 different in terms of the analysis for the participants?

A: Round 1 submissions are due no later than October 6th. Round 2 submissions, for the Top 50 from Round 1, are due no later than Oct. 15th.

Round 1: *This is all about prediction accuracy and 'fairness' (i.e equal opportunity) in your solution. Using the HOLDOUT file, a ROC/AUC metric will be calculated for each team to measure the accuracy of the prediction. Similarly, a DISPARITY SCORE will be calculated to measure 'fairness' in the modeling solution. These two measures will be combined for each submission and compared across all participants. The top 50 submissions from Round 1 will move on to Round 2.*

Round 2: *Only the top 50 teams from Round 1 will need to provide the written submission where you will need to summarize the entirety of the solution: approach, analytics, insights, recommendations, and actionability. Please refer to the kick-off deck and/or the official rules for more details on the judging criteria. The website also contains examples of Top 5 written submissions from previous years to provide an example of what previous success has looked like.*

The instructions want us to submit a CSV file with 3 columns. But it does not state which version of the 3rd column should be used: Rank-Low-2-High, or Rank-High-2-Low .

A: Rank things High-to-Low. Therefore, the member with the highest probability of having an ADE and discontinuing therapy will have the lowest rank. (i.e. 0.93 Score/Rank 0, 0.75 Score/Rank 1, 0.53 Score/Rank 2, etc.)

Attribute / Data Specific

Does the holdout data contain the same information & timing as the training data?

A: Yes, the holdout data contains the same attributes and timing as the training data. (With one exception: the target indicator (ie. a member experienced an ADE and discontinued therapy is not included in the holdout file)

Can you please explain why there are instances in the claims data where the process date is less than the service date? How is that possible?

A: You are correct in that process date should not occur prior to the service/visit date. 2 things may be happening: (a) there is a problem with your data (i.e. read-in or join error) (b) the data is erroneous. Data is messy...you must decide which it is and how to handle it.

It has been mentioned that "medclm_key" is an indicator of the medical claim, while "clm_unique_key" represents a unique medical claim. We are curious to understand the difference between these two variables. Upon inspecting the "medclms_trian" dataset, we noticed that "medclm_key" actually appears to have a unique value for every row. However, we observed duplicate values for "clm_unique_key." Could you please explain why this is the case?

A: You should think of medclm_key as the primary key for the medical claims table and it is unique for every claim line. The clm_unique_key groups together a single "claim" which can consist of multiple "claim lines". These unique claims can be combined together to form a logical claim that group together claims from the same provider/member combo with overlapping service dates. We typically use the logical claim to count utilization/visits rather than clm_unique_key, but the explanation above should provide the reasoning why there are duplicate values for clm_unique_key.

Could you clarify whether the two claim datasets are claims filed by healthcare providers or by patients?

A: The way this usually works for medical claims is this: Someone goes to the doctor, they show their insurance card to the doctor, and after the appointment a billing person submits a claim to the insurance company. For prescription claims, it's a lot faster but generally the same process. When a patient fills a claim at a pharmacy, the pharmacy submits the claim to the insurance company. Since we're the insurance company, this is the data we have. Generally, it comes from the provider or pharmacy directly to us.

Attribute / Data Specific

Does the holdout data contain the same information & timing as the training data?

A: Yes, the holdout data contains the same attributes and timing as the training data. (With one exception: the target indicator (ie. a member experienced an ADE and discontinued therapy is not included in the holdout file)

How much on average does the insurance covers for patients in both medical claims and prescription drugs. Do we have that data?

A: Coverage amounts vary on myriad of factors including, but not limited to, the plan an individual member has chosen. However, for the purpose of this case competition, those details are not included in the data that was provided to the participants.

Why do you think AUC is more important than recall for this problem? To explain: If we maximize recall, AUC would drop a little but we would be capturing more of the positive class. Even if the model's precision is low, the false positives we follow up with are going to benefit anyway, even if we know they are not going to stop the treatment.

A: It's not that we believe AUC is more important, it is simply the measure of accuracy chosen for Round 1 evaluation where every team's model will be evaluated using the same measuring stick. However, feel free, in subsequent rounds, to make a case for a model that maximizes recall, why that makes sense, and what (if any) implications it carries.

Live Q&A





Final Submissions - <https://mays.tamu.edu/humana-tamu-analytics>

Late submissions will not be accepted*

- **Round 1 Submission:** Due on Friday, October 6th @ 5:00 PM CST
- **Round 2 Submission:** Due Sunday, October 15th @ 11:59 PM CST
 - Only Top 50 from Round 1 need to submit for Round 2

Must follow prescribed file formats to be accepted*

- **Round 1 Holdout File Format:** CSV (Fields include: ID, SCORE, RANK) | Example:
- **Round 2 Written Submission Format:** MS Word, PDF

ID,SCORE,RANK
1545,0.8954,1
32,0.8532,2
368,0.7976,3

Must follow the following naming conventions* (using the first and last names of your team captain)

- **Leaderboard/Round 1:** 2023CaseCompetition_*FirstName_LastName*_2023mmdd.csv
- **Round 2:** 2023CaseCompetition_*FirstName_LastName*.doc

Judging is Blinded | Do not include names or school in content of submission

If you have any issues with your submission, please email: humanacasecomp@tamu.edu

*Failure to comply with submission requirements may result in disqualification

Ongoing Questions?

Website: <https://mays.tamu.edu/humana-tamu-analytics>

Email: humanacasecomp@tamu.edu