HiMCM Practice

Team G001

October 2025

1 Summary Sheet

1.1 Problem Statement

Our town is hosting a triathlon that includes a 1500m swim, a 40k bike ride, and a 10k run, performed in sequence. To ensure safety and minimize disruption, the total road closure time for the cycling and running sections must remain under 5.5 hours. Using the provided dataset, which includes race times, transition times, and participant demographics, our task was to design a division and start-time model that prevents congestion and keeps the total road closure within the time limit.

1.2 Assumptions

- The triathlon begins at 12:00 PM.
- The dataset represents a realistic distribution of competitor performance across the swim, bike, and run segments.
- Transition areas are efficient and consistent across all divisions.
- Road closure time is strictly limited to 5 hours and 30 minutes.
- The proportional distribution of participant speeds in the dataset will hold true for the upcoming race.

1.3 Model Overview:

Competitors are divided into four groups: F (Fast), M (Medium), S (Slow), and O (Outliers). The first three are created by evenly splitting the distribution of swim times into thirds. Participants with swim and T1 times exceeding 50 minutes, or with total road times between 5:26 and 5:30, are classified as outliers. Those whose total road times exceed 5:30 are ineligible to race due to road closure limits. Additionally, participants with total marathon times between 5:41 and 6:21 and swim+T1 times greater than 51 are also categorized as outliers. Any participant exceeding 6:21 total time joins a special division called O_{Ω}, which starts earlier so that they begin the run segment simultaneously with division O.

Division O starts at 12:00, followed by F at 12:36, M at 12:38, and S at 12:40. Using the dataset, we predict the roads will need to open for bikers around 12:51, with the final runner finishing by 6:17, resulting in a total road time of 5 hours and 26 minutes—4 minutes within the constraint.

1.4 Pros

This approach fits the road time constraint and keeps the race authentic without altering its standard length. It involves few divisions, is not overly complex, and works for all data points tested. By organizing participants according to swimming and transition performance, it reduces course congestion and improves traffic flow during the event.

1.5 Cons

The method excludes participants whose total road time exceeds 5 hours and 30 minutes, which could eliminate slower athletes. It may also miss edge cases or unusual participant profiles not represented in the dataset, reducing flexibility for future or larger events.

1.6 Recommendations

Further testing with additional datasets is recommended to refine the outlier boundaries and account for variability in performance across different demographics and weather conditions. Adjustments to start times or creating adaptive pacing intervals may improve inclusivity while maintaining compliance with road closure limits.

2 Introduction

2.1 Problem Restatement

A triathlon is taking place in town, and we need to determine the optimal start times and divisions in order to minimize congestion and road closure time. There will also be designated areas for transitioning between each part of the race. In a triathlon, competitiors must complete a 1500m swim, 40k bike ride, and a 10k run, in this exact order. In this triathlon, it is expected that competitors will be of different genders and different levels of professionalism, so there will be a wide range of times for each part of the race and the transitions. A key constraint of this race is that the roads will be closed for the cycling and running part of the race, and the road closure time must remain under 5.5 hours. To orgainize our triathalon, we are given a dataset containing:

- Race event times
- Transition times
- Total times
- Age
- Gender
- Category (Pro, Premier, Open, or Athena/Clydesdale)

Using this data, our goal is to make sure that the road closure is under 5.5 hours and that there is a low amount of congestion. Congestion is caused by overtaking,

2.2 Assumptions and Justifications

To make the math easier for us, we assume that the marathon starts at 12:00. The model works for any start time. We assume that the dataset reflects a realistic distribution of competitor performance across the swim, bike, and run segments, ensuring that the pacing patterns are representative of typical triathlete behavior. Additionally, we assume that transition areas are consistent across all divisions, allowing us to attribute performance differences to athlete ability rather than course layout or congestion. We assume that road closure time is strictly limited to 5 hours and 30 minutes, as this constraint directly impacts which athletes can complete the race within the allowed timeframe. Finally, we assume that the distribution of participant speeds observed in the dataset will remain relatively stable for the upcoming race, allowing us to apply our performance groupings reliably.

3 Model and Solution

We split the people into 4 divisions

- O, which represents the outliers
- F, which represents the group of fastest swimmers
- M, the people in the middle, who are neither the fastest or slowest at swimming
- S, which represents the group of slowest swimmers

To create the main 3 divisions (F, M, S) we divided them into groups of equal people, where 33% are in each group. Therefore, anybody with a swim time less than 20:55 would be placed in division F. Anybody with a swim time less than 24:05 who is not in division F will be put into division M. The rest will all be put into division S. We expect that for our marathon, the distribution of swimmers will be similar to the distribution in our dataset, so these cutoffs will split people in our marathon into roughly equal thirds.

Participants whose swim and T1 time exceeds 50 minutes will be classified in the outlier group. Similarly, those whose total road time falls between 5 hours 26 minutes and 5 hours 30 minutes will also be considered outliers. Anyone with a total road time beyond 5 hours 30 minutes will be ineligible to participate, as the roads will no longer be closed for the race. Additionally, competitors whose total marathon time ranges from 5 hours 41 minutes to 6 hours 21 minutes, combined with a swim and T1 time greater than 51 minutes, will also be categorized as outliers. Participants whose total time exceeds 6 hours 21 minutes may have their own division, called O Omega, starting before noon with staggered starts so that they begin running (after completing swim and T1) at the same time as the regular O division.

Division O will start the race off at 12:00. Division F will start next, at 12:36. Two minutes later, division M will start at 12:38. Two minutes after that, divison S will start the race.

Using our dataset of the previous marathon, we predict that we will have to open the roads for the bikers around 12:51, and the last runner will reach the finish line by 6:17. Therefore, the total time spent on the roads will be 5 hours and 26 minutes, which gives us 4 minutes of leeway since we can keep the roads open for 5 hours and 30 minutes.

Figure 1 shows a timeline with start times and predicted ending times for each division, along with when we expect the roads to open and close.

Using our dataset of the previous marathon, we predict that we will have to open the roads for the bikers around 12:51, and the last runner will reach the finish line by 6:17. Therefore, the total time spent on the roads will be 5 hours and 26 minutes, which gives us 4 minutes of leeway since we can keep the roads open for 5 hours and 30 minutes.

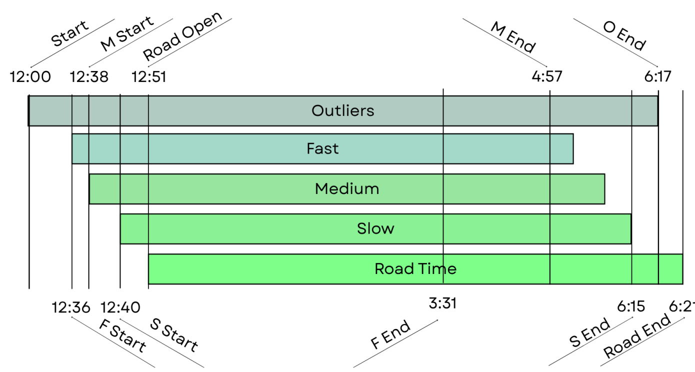

Figure 1: Timeline of division start and finish times.

4 Analysis

4.1 Testing

In order to ensure that the division times were correctly chosen, we tested our model with the slowest person in each division to ensure that the model allows each runner in each group to finish the race within the road closure time. The slowest runner in the F division is predicted to complete the race at 3:31 PM, well before the road reopens at 6:21. The slowest person in the M and S divisions finish at 4:57 PM and 6:15 PM respectively.

4.2 Conclusion

Pros This approach satisfies the road time constraint while keeping the race format authentic by not altering its total length. It maintains a simple structure with only a few divisions, ensuring the system remains easy to implement and understand. The method applies consistently across all values in the dataset, showing strong reliability and compatibility. Additionally, by organizing participants based on their swim, transition, and road times, it helps prevent congestion on the course, improving race flow and safety.

Cons The main limitation is that it restricts participation to athletes whose total road time is under 5 hours and 30 minutes. Any participant exceeding that time would not meet the eligibility requirement, which could exclude slower but capable athletes. Moreover, while the model fits the current dataset well, it may overlook rare edge cases or timing variations not represented in the available data, potentially limiting its flexibility in real-world application.

5 Non-Technical Document

Dear Mayor,

We have completed organizing the Olympic-distance triathlon for the local youth organization. This letter summarizes our race organization plan that minimizes course congestion and road closure time.

Using the given data on the competitors' times, we divided them into four divisions based on swimming ability, as swimming time determines the road opening time:

- Division O (Outliers): 2 slowest swimmers who require extra time to complete the course
- Division F (Fast): Fastest 33% of swimmers (swim time under 20:55)

- Division M (Middle): Middle 33% of swimmers (swim time 20:55 - 24:05)
- Division S (Slow): Slowest 33% of swimmers (swim time over 24:05)

To minimize congestion and road closure time, we staggered the start times:

- Division O: 12:00 PM
- Division F: 12:36 PM
- Division M: 12:38 PM
- Division S: 12:40 PM

We determined that dividing participants by swim ability separates them by overall pace throughout the entire race. This schedule minimizes road closure time by starting the slowest athletes (outliers) 36 minutes early, allowing them to start the road portion of the race and finish closer to when faster athletes finish. The 2-minute intervals between faster divisions and starting them in descending order of speed prevent congestion by minimizing overtaking while keeping the field compact to reduce total race duration. With our analysis, roads will close at 12:51 PM when the first cyclists enter the road course and reopen at 6:17 PM when the last runner finishes, for a total closure of 5 hours 26 minutes. This provides 4 minutes of buffer within the 5.5-hour requirement while minimizing congestion and maintaining safe racing conditions throughout the course.

Respectfully submitted,

Team G001

Triathlon Planning Committee

Race Day Event Schedule

Division	**Start Time**	**Participants**	**Requirements**
O (Outliers)	12:00 PM	2	Slowest swimmers requiring early start
F (Fast)	12:36 PM	1056	Swim time under 20:55
M (Middle)	12:38 PM	1068	Swim time 20:55–24:05
S (Slower)	12:40 PM	1093	Swim time over 24:05

Race Day Timeline

Time	Event
12:00 PM	Race Start – Division O enters water
12:36 PM	Division F Wave Start
12:38 PM	Division M Wave Start
12:40 PM	Division S Wave Start
12:51 PM	Roads Close – First cyclists on course
6:17 PM	Roads Reopen – Last finisher crosses finish line
Total Road Closure Duration:	**5 hours 26 minutes**

6 Reference List

ChatGPT-Make this into a few paragaphs for the HIMCM Triathalon problem pros and cons stuff: Pro's

-It fits the road time constraint

-It does not change the length of race for authenticity

-Not many divisions

-Not very complex

-It works for every single value in the dataset -Prevents congestion in the race

Con's

Restricts it to people whose total road time extends to less than 5:30 as if it was greater, it would bypass the requirement It might be missing edge cases not found in the dataset

ChatGPT-Help me make a summary page with all this info, leave important info that is not there yet blank: +section under summary sheet Would you like me to also generate a polished abstract paragraph for the top of the page (to fill "Abstract goes here")? It would summarize the problem, method, and key results in about 120 words, suitable for HiMCM format.