Summary

A town hosting a triathlon has asked us to organize the event in a way that will be fun for participants in the race but will not seriously disturb the residents of the town. To do this, we were given a data set of a prior triathlon to base our model off which included the bib number, age, gender, category, split times (the time for each segment of the race), and final times of 3218 participants. In our triathlon, we must account for two main things: we cannot have slower athletes blocking the fastest athletes, and we cannot allow the roads to close for more than five and a half hours.

To begin our model, we learned more about triathlons, and we found that a typical triathlon consists of swimming, biking, and running, in that order to promote safety due to swimming becoming increasingly dangerous as an athlete becomes exhausted. Knowing this, we realized that the roads did not have to close while contestants were still in the water because when the first athletes are in the water, there will be no athletes biking or swimming. This gave us extra time: however long it takes the first person to exit the swimming portion of the race. Utilizing the data set given to us, this would only realistically extend the race by at most 20 minutes. This is still significant though, as it would give those who are slower more time to finish the race. We figured it would be best if everybody could finish the race, as that would make participants happiest. However, not everybody is able to complete the entire race before the roads are required to open back up due to differences in physical abilities. For simplicity's sake, we decided that for those unable to finish the race in the allotted five-and-a-half-hour limit, they would be taken out of the race by sweepers. In the given data set, only two out of over 3000 participants would fall under this category, taking too long to finish the race, so with our planned triathlon consisting of just 2000 people, sweeping should only apply to very few people.

Although utilizing swim times to maximize time on the roads, we eventually decided to put the Professional athletes and Premier athletes in waves together and send them first. This was to draw attention to the advanced runners and to not mix them in with the rest of the crowd, to lessen congestion, and promote the competitive runners. After this, we decided to break up the remainder of our groups, the Female Open, Male Open, Clydesdale (men over 220lbs), and Athena (women over 165lbs) categories. We decided to send these groups in order from slowest to fastest, as this would allow most people to finish the race as the slower runners would have more time. Each wave would have a small buffer period between them of 3 minutes, which would be enough time to line up at the starting line, and considering most open waves have relatively similar times, faster contestants should not mix in with slower ones. With 20 waves in total, the first and second being for Pro and Premier athletes, the third for Athena and Clydesdale to give them sufficient time, and the rest being slowest to fastest open groups; each of our open waves will have around 180 people. With our model, we found that few people would be swept while maintaining the road closure time limit and low congestion rates.

Table of Contents

Introduction and Problem Statement

A town is hosting a triathlon, and we are trying to organize it to maximize the enjoyment of the participants without closing the roads for too long. The triathlon consists of 3 parts, in this order: a 1500m swim, a 40km bike ride, and a 10km run. To maximize the enjoyment of the participants, the faster racers should not be stuck behind slower racers, so that the racers are able to go to their top speed, so both they and the spectators may enjoy the race. The mayor also requires that the roads of the town must not be cut off for more than 5.5 hours. We need to find the waves and scheduled start times to minimize congestion and road closure time.

Assumptions and Justifications

1. The data set provided accurately represents the anticipated participants by distributions of athletes in terms of race categories, ages, and times.
2. Roads do not need to be closed until the first runner reaches T1 (the period to transition between swimming and biking). This is because the athletes are not yet occupying the roads.
3. Wave start times should be separated by between 2 and 10 minutes to give ample time for each wave to line up during the buffer time.
4. Waves should have a size of approximately 100 to 200 people to limit congestion at the start of the race and allow room for individuals to pass waves when needed while still finishing the race in a timely manner
5. In the later stages of the race, congestion is negligible because as the race progresses, participants in the Open groups will end up maintaining a separation in distance incurred via wave times due to similar speeds. The Open groups may end up colliding with the Athena and Clydesdale groups; however, this is insignificant due to the small number of these racers.
6. The final wave should enter the water within one hour of beginning the race to give athletes as much time as reasonably possible to complete the race.

Model

Our model utilizes the 3217 data points acquired from a prior triathlon, assuming the athletes to be close representations to the anticipated athletes of our triathlon.

To begin our model, we created a pivot table to split all our data points into separations by categories and ages, as follows:

t	AGE	GENDER	MAX of SWIM	MAX of T1	MAX of BIKE	MAX of T2	MAX of RUN	MAX of FINALTI	COUNT of AGE	MEDIAN of FINA
ATH Total			00:31:49	00:22:02	02:47:56	00:09:11	01:44:00	05:19:48	32	03:37:04
CLY Total			00:35:10	00:16:54	01:59:18	00:13:48	01:41:17	04:15:47	60	03:11:21
F OPEN Total			00:38:44	00:26:01	03:12:42	00:16:02	02:07:17	05:57:48	898	03:13:01
F PREMIER Total			00:19:50	00:06:27	01:25:33	00:02:41	00:49:42	02:43:33	18	02:23:06
F PRO Total			00:13:29	00:03:56	01:10:10	00:01:08	00:41:48	02:09:26	6	02:04:39
M OPEN Total			00:54:05	00:31:19	02:59:46	00:13:44	02:09:09	05:02:09	2147	02:53:56
M PREMIER Total			00:20:28	00:06:16	01:22:13	00:02:15	00:48:32	02:30:36	49	02:09:07
M PRO Total			00:15:22	00:03:52	01:05:24	00:01:00	00:38:45	02:03:58	7	01:53:52
Grand Total			**00:54:05**	**00:31:19**	**03:12:42**	**00:16:02**	**02:09:09**	**05:57:48**	**3217**	**02:58:46**

In the figure, maximum split times are shown between each category (also split by gender) of racer as well as the count of each category and the medium final time in the categories. On the pivot table itself, if the plus button beside a category is hit, it expands to show all of the athletes ages as follows:

t	AGE	GENDER	MAX of SWIM	MAX of T1	MAX of BIKE	MAX of T2	MAX of RUN	MAX of FINALTI	COUNT of AGE	MEDIAN of FINA
ATH	28	F	00:29:46	00:12:25	02:04:56	00:04:54	01:35:36	04:27:37	2	03:42:33
	28 Total		00:29:46	00:12:25	02:04:56	00:04:54	01:35:36	04:27:37	2	03:42:33
	29	F	00:21:53	00:09:34	01:50:59	00:02:24	01:09:07	03:33:57	1	03:33:57
	29 Total		00:21:53	00:09:34	01:50:59	00:02:24	01:09:07	03:33:57	1	03:33:57
	30	F	00:22:27	00:10:05	01:44:01	00:04:28	01:23:24	03:43:13	2	03:34:45
	30 Total		00:22:27	00:10:05	01:44:01	00:04:28	01:23:24	03:43:13	2	03:34:45
	31	F	00:21:38	00:12:08	01:36:22	00:06:11	01:15:05	03:31:24	1	03:31:24
	31 Total		00:21:38	00:12:08	01:36:22	00:06:11	01:15:05	03:31:24	1	03:31:24
	32	F	00:31:05	00:20:23	02:47:56	00:08:18	01:32:06	05:19:48	1	05:19:48
	32 Total		00:31:05	00:20:23	02:47:56	00:08:18	01:32:06	05:19:48	1	05:19:48
	33	F	00:23:07	00:11:27	01:50:00	00:04:29	01:26:43	03:55:46	1	03:55:46
	33 Total		00:23:07	00:11:27	01:50:00	00:04:29	01:26:43	03:55:46	1	03:55:46
	34	F	00:22:36	00:13:04	01:41:25	00:07:34	01:14:19	03:38:58	2	03:07:57

From the data acquired by our table, we broke down which categories would go first to go from slowest to fastest (excluding Pro and Premier athlete). We found that the order would be Athena, Female Open, Clydesdale, then Male Open. However, due to the low number of athletes participating in the weight-based categories, we decided to combine the two groups.

Using this, we then broke up this data into the desired number of waves we wished to have in our own model. For our triathlon, we decided to have 20 waves of a similar number of people in each wave because this gave us a buffer time of 3 minutes between waves (60 minutes to get all athletes into the water / 3) and gave us a good number of average participants in each wave (100). However, due to the restriction of advanced athletes needing limited congestion we chose to make our first two waves of athletes our Pro and Premier runners.

To create our waves, we found the total number of people remaining in the Female Open and Male Open categories, as we removed the other three categories from our pool of categories to accommodate different expected racing times.

Then, with the remaining participants, we divided it by our remaining number of waves, so, in our model's case, it was 3045/17 which gave us approximately 179, so that was the approximate number of people we want in each of our waves to even them out.

Starting from the oldest athletes from Female Open and ending with youngest from Male Open we created our waves for the data provided:

Bucket	Number of people in Data	Anticipated Athletes	Age
Male Premier and Pro	56	35	N/A
Female Premier and Pro	24	15	N/A
Clydesdale and Athena	92	57	N/A
Female Open 1	201	125	48+
Female Open 2	165	103	41-47
Female Open 3	176	109	35-40
Female Open 4	169	105	31-34
Female Open 5	187	116	<30
Male Open 1	185	115	57+
Male Open 2	181	113	52-56
Male Open 3	169	105	49-51
Male Open 4	159	99	46-48

Male Open 5	206	128	43-45
Male Open 6	146	91	41-42
Male Open 7	225	140	38-40
Male Open 8	155	96	36-37
Male Open 9	157	98	34-35
Male Open 10	224	139	31-33
Male Open 11	183	114	28-30
Male Open 12	157	98	<27

Next, using this data, we found the number of anticipated athletes in each of our waves. To do this, we took the proportion of athletes in each wave from our given data and scaled it by our total expected athletes to give us our new wave sizes. (Column 3)

Finally, we decided our start times. Because every athlete should be in the water within an hour of beginning the race to maximize the time to complete the triathlon, we took 60, divided by the number of waves, and used the result as our interval in minutes between sending waves.

Event	Time
Male Premier and Pro	0:00
Female Premier and Pro	0:03
Clydesdale and Athena	0:06
Female Open 1	0:09
Female Open 2	0:12 <- Roads close around here
Female Open 3	0:15
Female Open 4	0:18
Female Open 5	0:21
Male Open 1	0:24
Male Open 2	0:27
Male Open 3	0:30
Male Open 4	0:33

Male Open 5	0:36
Male Open 6	0:39
Male Open 7	0:42
Male Open 8	0:45
Male Open 9	0:48
Male Open 10	0:51
Male Open 11	0:54
Male Open 12	0:57
End of Race	About 5:42

Solution

Use these variables:

T = Total Samples from a past Dataset

T2 = Number of Pro/Premier/Clydesdale/Athena Athletes from a past triathlon

W = Desired Number of Waves

W2 = Number of waves desired for Pro/Premier/Clydesdale/Athena athletes

In this formula:

(T – T2) / (W – W2)

- To get this formula, we only accounted for the people who aren't part of the Pro/Premier athletes or the Clysdale/Athena since they are going to be released as part of the first 3 waves. Thus, their waves sizes will definitely fall under the limit of 180 people per wave, allowing us to ignore them from the formula for calculating wave size.
- The result of the formula gives the approximate number of people needed per wave. With this number, waves can be created, starting from the oldest Female Open athletes to the youngest Female Open athletes, then the oldest Male Open athletes to the youngest Male Open athletes.
- And finally, create your wave order, starting with the Pro and Premier wave(s), followed by the Clydesdale and Athena wave(s), then the Female Opens in old to young order, and finished up by the Male Opens in old to young order.
- To decide start times, simply do 60/W and use the result as the interval in minutes between every wave.
- The diagram above displays when each wave listed above starts the race. These times are relative to wave 1, which is why wave 1 starts at 0:00.

Analysis

Our model successfully develops expected waves and wave start times in a successful order to maximize efficiency and meet limitations. It demonstrates key strengths while acknowledging important limitations. Our primary strength lies in our model's ability to balance multiple objectives: maintaining road closure within the 5.5-hour limit, minimizing congestion through appropriately sized waves of athletes, and prioritizing athletes finishing the race in time. We strategically placed the Pro and Premier runners at the front for sponsors and showcased the athletes' competitive performances. On top of this, placing the advanced runners at the beginning of the race reduces congestion in our model as the runners will not need to go around any of the waves that follow. After the fastest runners, we decided to send the Clydesdale and Athena, because on average those two groups have the slowest times. Sending the slowest group directly after the fastest group gives the slower racers as much time as possible to finish the race while still giving the fast runners a chance to shine in the front. Following the Clydesdale and Athena wave, we send Female Open athletes and Male Open athletes from oldest waves to youngest waves because this setup typically accounts for the completion times by each group, going from the slowest to the fastest. Although this sounds like the faster runners could get congested throughout the race in the back of the group, the times among different ages are not significantly different enough to cause notable congestion in the race. In between every wave there is a buffer that allows every athlete to enter the water within one hour of beginning the race.

Our model, however, only accounts for up to 30 waves of athletes, as 2 minutes is the minimum buffer time between waves to be sufficient given the natural spread of races as athletes progress. In addition, our model doesn't account for water facility closure times, which could be addressed by incorporating logistics for swim venues and transition time analysis into the model.

In a sensitivity analysis, if our data from a previous triathlon was not similarly spread out among categories, genders, and ages, our waves would be off balanced. Our model bases the age ranges for each wave, as well as the number of waves in a category and gender are allocated based off a previous model. If the data is different and disproportionately over accounts for male athletes, the female waves would be overcrowded and congested throughout the race, as would over accounting for different ages.

Conclusion

Overall, our model was performed to meet all of the given conditions. According to the sensitivity analysis, we came to the conclusion that our model has its limits. While it succeeded in determining the best wave sizes and number of waves, it heavily relied on the expected data to be like prior data. Some interesting extensions include building a model that satisfied pool/ocean closure times. This would require us to not only keep in mind the 5.5-hour limit but also consider the slowest swimmers and how they will affect our overall timings suggested by our model. In this case, we could also need to consider both transition times as the pools and roads will be closed. Another interesting extension could include accounting for unexpected factors like weather. For example, the weather man says that the day before the triathlon, there is supposed to be rain during a certain period, and the goal would be to keep the triathlon running even with that unexpected factor.

Part 2

As there are still a couple of racers who are unable to finish the race within the 5.5-hour time limit, changing the race distances increases the time for these athletes to finish the race. Adjusting the race distances for the biking and running portions of the race specifically would have the greatest effect in maintaining a strict time limit because those two portions of the race affect the road-closure times. Changing the running distance to a 5k rather than a 10k would reduce the running time enough that, according to the data from the previous triathlon, every athlete would finish the race in under 5 hours and 15 minutes, assuming their pace remains relatively similar. Likewise, the biking distance could also be brought down to a shorter distance such as 30k, which would similarly bring every racer's time to under 5 hours 15 minutes.

However, changing the race distances does have its faults. Using standardized distances makes the event appeal to a much broader audience as many people look specifically for certain races and distances, so changing the race to be shorter could diminish the popularity of the race itself. To combat this, the triathlon could switch to a different standardized distance, such as a sprint triathlon, which lowers all of the distances of the race. This would keep the standardization of the race, allow more people to finish the race before the roads close, and could appeal to more people that are interested in a triathlon but are weary about competing over longer distances.

Non-Technical Document

Dear Mr. Mayor:

Over the past week, we have been working hard to find the best way to host the upcoming triathlon. We have been taking many of the parameters that you gave us into consideration in order to make this event enjoyable for participants while raising some funds for the local youth organization. We hope with our outline, the race will be a huge success, so that it can be in years to come as well.

As you may be aware, other triathlons in the area have attracted a few thousand participants. Expecting relatively similar numbers of participants coming to our race, around 2000 people, we decided how we would split up divisions. Congestion is one of the most important things that we have to minimize. With high congestion, it could lead to a less enjoyable race. We decided to split the participants into 20 waves of racers, with the average wave containing approximately 100 people. Knowing that the largest pool of people is within the Male and Female Open groups, we allocated the most waves to them, with 6 waves going to Female Open and 11 going to Male Open; wave numbers will correspond to the proportion of Male to Female Open participants. The other three waves will be given to the groups as following: Male Pros and Premiers, Female Pro and Premiers, and Clydesdale and Athena.

To ensure that the fast participants don't get stuck behind the slow runners and attention isn't distracted from them, which could be catastrophic to the race's success, we chose to put them first. These fast athletes really only include the premier and Pro groups, which in total is a small group of less than 100 people. From there, we chose to start groups based on how much time it would take them to finish the race. With each group having a buffer in between, even with it being short, the time adds up, and that may mean groups don't have enough time to finish the race. We determined that it would be better for more people to finish, otherwise it may harm participant satisfaction. So, to give the slower people more time, we chose to start them off first, and keep letting the faster runners go, until the very last group that goes is the fastest Male Open group. With the relatively small wave sizes and the good buffer times, we don't think congestion will be an issue, but most importantly, most, if not all people should be able to finish the triathlon before the end of the five and a half hours that the roads can be closed. We believe that using this strategy, the race will be a massive success, participants will be happiest, and the town's residents will still be able to operate their daily lives as normal. Below, we have listed the day schedule if you are curious.

Sincerely,

Your Race Organizers

Wave	Time
Male Premier and Pro	0:00
Female Premier and Pro	0:03
Clydesdale and Athena	0:06
Female Open 1	0:09
Female Open 2	0:12 <- Roads close around here
Female Open 3	0:15
Female Open 4	0:18
Female Open 5	0:21
Male Open 1	0:24
Male Open 2	0:27
Male Open 3	0:30
Male Open 4	0:33
Male Open 5	0:36
Male Open 6	0:39
Male Open 7	0:42
Male Open 8	0:45
Male Open 9	0:48
Male Open 10	0:51
Male Open 11	0:54
Male Open 12	0:5
End of Race	About 5:42

References

1. Google Sheets: Data analysis
2. Word: Report
3. Canva: Graphics