

# Análisis de clasificación a un campeonato mundial de Ironman

Brian Britos, Centro de Matemática,  
*Facultad de Ciencias, Universidad de la República*  
Montevideo, Uruguay  
bbritos@cmat.edu.uy  
Juan Carlos Pellegrini  
*Facultad de Ingeniería, Universidad de la República*  
Montevideo, Uruguay  
jpellegrini11@gmail.com

## Resumen

El triatlón ha ganado popularidad como un desafiante deporte de tres etapas: natación, ciclismo y pedestrismo. Este estudio se centra en las modalidades de triatlón half Ironman, y tiene como objetivo analizar los datos para determinar la dificultad de clasificar en el Campeonato Mundial. También se busca identificar la categoría con más oportunidades de clasificación y si existen ubicaciones más propicias en el mundo. Se plantea la hipótesis de que la proximidad geográfica al lugar del evento podría conferir ventajas a los atletas. Además, se realiza un ranking de países con mayor práctica de triatlón y clasificación de competidores. En este estudio, se incorpora información de los half Ironman y se utiliza Python y R para el pre-procesamiento de datos y la creación de una base de datos de grafos en neo4j. Los resultados proporcionarán información valiosa para los atletas que buscan alcanzar sus objetivos en el triatlón.

## I. INTRODUCCIÓN

El triatlón es un deporte cada vez más popular y desafiante que combina natación en aguas abiertas, ciclismo y pedestrismo. Existen diferentes modalidades, como sprint, standard, half Ironman y full Ironman, que se diferencian por la distancia de cada disciplina. Los triatlones half Ironman consisten en 1,9 km de natación, 90 km de ciclismo y 21 km de pedestrismo, mientras que los full Ironman tienen el doble de distancia en cada etapa. Estas últimas dos modalidades son organizadas por la World Triathlon Corporation (WTC) y se llevan a cabo en todo el mundo.

Cada carrera ofrece cupos para el Campeonato Mundial, que se celebra anualmente en una fecha específica para cada versión del triatlón. Estos cupos se asignan a los primeros puestos de cada categoría por edad y género, teniendo en cuenta la cantidad de participantes en cada categoría. A partir del estudio de los datos de cada carrera es de interés comprender la dificultad de clasificar para el campeonato mundial y responder preguntas clave, como en qué categoría hay más oportunidades de clasificación y si hay lugares más propicios en el mundo para lograrlo.

Además, se busca determinar el tiempo máximo necesario para asegurar un cupo y si hay fechas más convenientes durante el año. También se plantea la hipótesis de que tener como procedencia el mismo país o continente donde se realiza el evento puede ser beneficioso para la clasificación, debido a la familiaridad con los circuitos y a la menor fatiga por los desplazamientos.

Finalmente, se pretende realizar un ranking de los países con mayor práctica de triatlón y de aquellos con más competidores clasificados, con el objetivo de observar la maduración de este deporte en el mundo. El equipo se propone responder estas preguntas para comprender los requisitos necesarios para que los deportistas que entrenan regularmente puedan alcanzar sus objetivos en el triatlón.

## II. SECCIÓN PRINCIPAL

### II-A. *Preprocesamiento de los datos*

En esta sección se presenta el preprocesamiento realizado a los datos para la correcta creación de la base de datos, la misma contiene la siguiente información sobre los competidores en triatlones Half Ironman.

- Gender: Género del competidor.
- AgeGroup: Es un intervalo de edades, al cuál pertenece el competidor.

- AgeBand: Es un número que representa el AgeGroup (el primero número del intervalo).
- Country: País de procedencia del competidor.
- CountryISO2: Código ISO del país anterior.
- EventYear: Año del evento.
- EventLocation: Contiene información del lugar del evento, y si es una competencia nacional, continental o un campeonato mundial.
- SwimTime: Tiempo empleado en la sección de natación.
- Transition1Time: Tiempo empleado en la primera transición.
- BikeTime: Tiempo empleado en la sección de ciclismo.
- Transition2Time: Tiempo empleado en la segunda transición.
- RunTime: Tiempo empleado en la sección de pedestrismo.
- FinishTime: Tiempo total empleado para completar la prueba.

Todos los tiempos están en segundos. En el cuadro I puede encontrar un ejemplo de estos datos.

Además de los datos mencionados, fue necesario agregar datos adicionales sobre los eventos, como el país donde ocurrieron y sus coordenadas geográficas. Estos datos geográficos son relevantes para analizar la proximidad geográfica de los competidores a los lugares de los eventos y su posible influencia en la clasificación. Para esto se utilizaron librerías de Python especializadas en estas tareas: *locationtagger*, *geopy*, *spacy* y *geonamescache*. Para algunos de los eventos, dichas librerías no lograron extraer la ciudad, por lo que se procedió a agregarlas manualmente.

En este estudio, se prestó especial atención a los competidores que lograron clasificar en los eventos. Esto implica identificar y analizar las características y tiempos de aquellos competidores que alcanzaron los cupos para el Campeonato Mundial. Esta información será clave para comprender los factores que contribuyen a la clasificación y determinar qué atributos y rendimientos son necesarios para lograr este logro en el triatlón. Para saber que competidores lograron clasificar se realizó un script en R siguiendo la siguiente lógica:

1. Se asigna un cupo para cada categoría Gender-AgeBand que tenga al menos un participante. Esto garantiza que cada categoría tenga al menos un representante en el Campeonato Mundial.
2. Los cupos restantes se distribuyen según la proporción de cada género de los competidores.
3. Dentro de cada género, los cupos se distribuyen en función del peso o importancia de cada AgeBand. Esto implica que se asignan más cupos a las categorías de edad que tienen una mayor participación o relevancia en el evento.

Es importante destacar que la cantidad de cupos puede variar de un evento a otro y de un año a otro. Para simplificar el análisis, se ha supuesto una asignación de 40 cupos para eventos nacionales y 80 cupos para campeonatos continentales. Estas cifras son utilizadas como referencia, pero es importante tener en cuenta que en la práctica pueden existir variaciones en la asignación real de cupos en cada evento.

Finalizados ambos procesos, se agregan las siguientes variables a los datos:

- Lugar: Es el lugar donde ocurre el evento.
- Latitud y Longitud: Son las coordenadas geográficas del Lugar.
- Clasifica: Es una variable binaria que identifica a los competidores que clasifican al mundial.
- EsMundial: Es un booleano que identifica si un evento es un mundial.

Todos los scripts utilizados para este preprocesamiento pueden encontrarse en el siguiente repositorio de Github: Ironman analysis and graph database.

Cuadro I: Ejemplo de los datos originales

Gender	AgeGroup	AgeBand	Country	CountryISO2	EventYear	
M	45-49	45	Andorra	AD	2020	
EventLocation	SwimTime	Transition1Time	BikeTime	Transition2Time	RunTime	FinishTime
IRONMAN 70.3 Bariloche	1679	119	9107	95	5515	16514

Cuadro II: Estructura de la base de datos de grafo.

Participante	Evento	Lugar	País
Gender	EventYear	Latitud	Latitud
AgeGroup	EventLocation	Longitud	Longitud
AgeBande	EsMundial	NombreLugar	NombrePais
Country			
Contry ISO			
Clasifica			
SwimTime			
Transition1Time			
BikeTime			
Transition2Time			
RunTime			
FinishTime			

### II-B. Creación de la base de datos

Se construyó una base de datos de grafos utilizando la tecnología Neo4j para representar la información de manera eficiente y establecer las relaciones entre los distintos elementos. La base de datos está compuesta por cuatro tipos de nodos: Evento, Competidor, País y Lugar, y relaciones entre los mismos. El cuadro II muestra que atributos pertenecen a cada nodo. Dado que no se cuenta con información para identificar a los competidores, se decidió guardar la información sobre los tiempos de competencia directamente en el nodo competidor.

En la figura 1 se muestra un ejemplo ilustrativo de la apariencia de la base de datos. En verde los nodos *Participante*, en rojo los nodos *Lugar*, en beige los nodos *Evento* y en azul el nodo *País*.

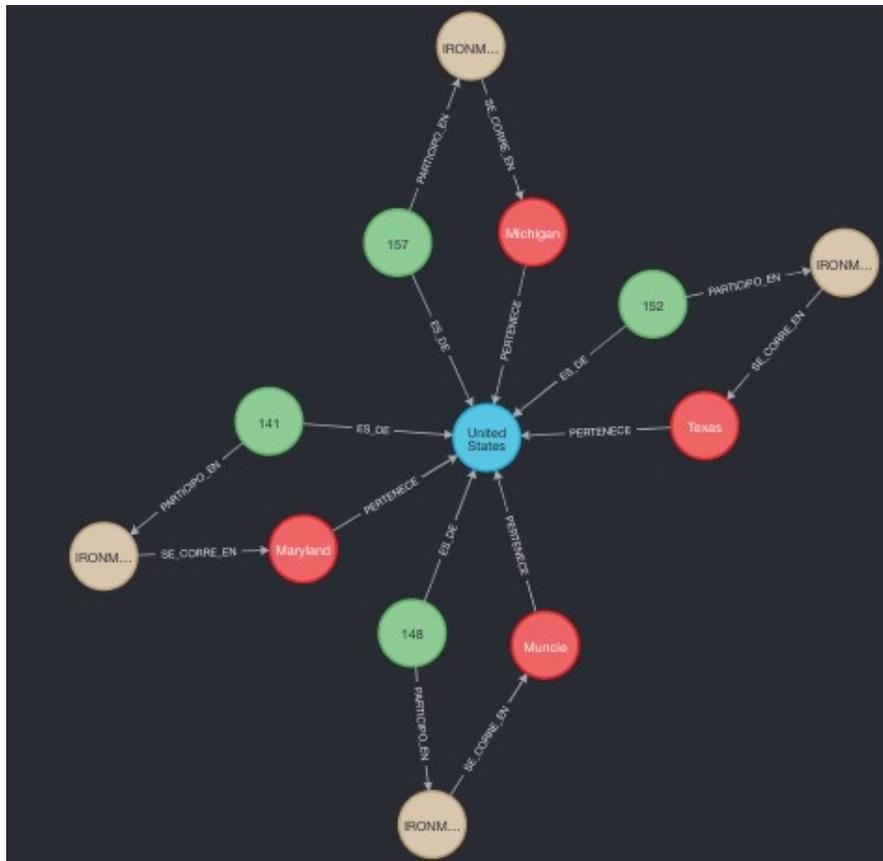


Figura 1: Ejemplo de los nodos y relaciones en la base de datos.

### II-C. Análisis de los datos

Para un mejor entendimiento de los datos se procede a estudiar la información obtenida mediante diferentes consultas, se investiga cual es el comportamiento de personas clasificadas según su país de origen, el país donde compiten teniendo en cuenta si es residente o extranjero, los tiempos máximos necesarios para clasificar a un mundial según lugar, género y categoría. Esta información se detalla a continuación.

*II-C1. Cantidad de clasificados por cantidad de participantes y país:* Se realizó esta consulta para ver si existe una relación en la que, cada país pueda tener o no mayor cantidad de clasificados en relación al total de participantes del mismo. Sin tener en cuenta los lugares donde compiten. Como se puede ver en el cuadro III los valores son bajos ya que es de esperarse que se inscriban muchas personas y clasifiquen pocas. De todas formas se visualizan algunos países con mayor porcentaje lo cual puede ser un indicador de qué tan buenos son sus deportistas.

*II-C2. Cantidad de clasificados cuando compiten en su país de origen:* Se encuentra que, para el total de personas clasificadas a un mundial, aproximadamente el 61 % de los participantes clasificó en su país de origen. No obstante, esto puede deberse a que en cada país se inscriben mayor cantidad de personas oriundas del mismo. Para analizar esto se utilizan los siguientes ratios:

$$\text{Ratio Residentes Clasificados} = \frac{\text{Residentes Clasificados}}{\text{Residentes Competidores}} \times \frac{\text{Total Competidores}}{\text{Total Clasificados}}$$

A mayor cantidad de residentes clasificados mayor es el ratio, también si aumenta la cantidad de competidores; ya que es más meritorio que se mantenga un porcentaje de clasificados residentes cuando aumente la cantidad de participantes totales. De manera inversa actúan el total de clasificados y la cantidad de participantes residentes. En el cuadro IV se puede ver el Top 10 de países con mayor ratio de residentes clasificados. Se aprecia que algunos países tienen un alto porcentaje de clasificados en su propio país, lo que hace que el ratio sea alto. Además la mayoría de estos países cuentan con ratio mayor a 1 lo cual significa que la relación entre clasificados residentes es mayor en proporción a la relación entre la cantidad de personas pertenecientes al país y la totalidad de las personas que se anotaron.

Cuadro III: Top10 de Total de clasificados / total de inscriptos para cada país

Posición	País	Participantes del país	%Clasificados
1	Bermuda	269	7,8
2	Letonia	465	7,5
3	Nueva Zelanda	7015	7,4
4	Panamá	1472	6,9
5	Emiratos Árabes Unidos	441	6,8
6	Hong Kong	2212	6,4
7	República Checa	1464	6,2
8	Japón	8806	6,1
9	Suiza	12335	6,0
10	Korea del Sur	2279	5,9
43	Uruguay	926	3,2

*II-C3. Cantidad de clasificados cuando compiten en un lugar fuera de su país de origen:* En esta sección se estudia el comportamiento de clasificados de un país en un evento que no corresponde a su lugar de origen. Para esto se utiliza el siguiente ratio, muy similar al utilizado en el caso de los competidores residentes, calculado para participantes de distintos países en un país de competencia determinado:

$$\text{Ratio Extranjeros Clasificados} = \frac{\text{Extranjeros Clasificados}}{\text{Extranjeros Competidores}} \times \frac{\text{Total Competidores}}{\text{Total Clasificados}}$$

A mayor cantidad de clasificados de un país o mayor cantidad de competidores, mayor es este valor y a menor cantidad de competidores extranjeros o clasificados totales también aumenta el ratio. En el cuadro V se puede observar que, para competidores de países como Australia, Reino Unido o Argentina tienen mayores oportunidades en India, Irlanda, Chile y Uruguay respectivamente. Al final de la tabla se muestra el resultado de competidores de Uruguay donde, su mejor resultado siendo extranjeros es en Argentina.

*II-C4. Cantidad de clasificados cuando compiten en un lugar fuera de su país de origen:*

Cuadro IV: Top10 Ratio de residentes clasificados

Posición	País	%Clasificados residentes	Ratio residentes clasificados
1	Suiza	71,1	1,35
2	Croacia	2,5	1,31
3	Emiratos Árabes Unidos	3,1	1,24
4	Sudáfrica	93,1	1,16
5	Francia	57,9	1,10
6	Australia	90,3	1,05
7	Panamá	24,9	1,04
8	Canadá	72,6	1,03
9	Brasil	91,6	1,00
10	Nueva Zelanda	66,6	0,98
41	Uruguay	8,1	0,43

### III. CONCLUSIONES Y TRABAJO FUTURO

En primer lugar, al observar los resultados en términos de los países con más clasificados en el top 3, se destaca la notable actuación de Bermuda, Letonia y Nueva Zelanda. Estos países han logrado una destacada participación en los eventos de Half Ironman, lo que demuestra un alto nivel de rendimiento y dedicación por parte de sus atletas. Por otro lado, Uruguay se encuentra en el puesto 43 en términos de clasificados, lo que indica que hay margen de mejora en cuanto al desempeño de los competidores uruguayos en este tipo de eventos.

Al considerar el ratio de residentes clasificados, Suiza, Croacia y los Emiratos Árabes Unidos ocupan los primeros puestos. Estos países han logrado un alto porcentaje de clasificados en relación con su población residente, lo que sugiere un enfoque destacado en el desarrollo de atletas locales y en la promoción del deporte de resistencia. Uruguay, por su parte, se ubica en el puesto 41, lo que indica que hay espacio para mejorar en términos de clasificación de sus residentes en eventos de Half Ironman.

En cuanto al ratio de extranjeros clasificados, Australia, Reino Unido y Argentina ocupan el top 3. Estos países han demostrado una destacada capacidad para que sus atletas logren clasificarse en eventos de Half Ironman fuera de sus respectivos países de origen. Por otro lado, Uruguay ocupa el puesto 359 en esta categoría. Es inevitable la comparación con la conclusión anterior, donde Uruguay quedó en el puesto 41.

Como posibles mejoras, se identificó una limitación en los datos utilizados en este estudio, no se cuenta con la identificación de los competidores en la base de datos, lo que impide establecer relaciones entre un competidor y múltiples competencias en el grafo.

Para futuras investigaciones, sería de gran utilidad contar con información adicional que permita identificar de manera única a cada competidor en la base de datos. Al tener la capacidad de relacionar a los competidores con varias competencias en la base de datos de grafos, se podrían obtener beneficios significativos. Por ejemplo, sería posible analizar el desempeño y los tiempos de un competidor a lo largo del tiempo, identificar tendencias y patrones en su rendimiento, y comprender mejor su trayectoria en el triatlón.

La inclusión de la identificación de los competidores en la base de datos de grafos permitiría aprovechar al máximo la riqueza de la estructura del grafo y brindaría nuevas oportunidades para el análisis y la comprensión del mundo del triatlón.

Cuadro V: Top10 Ratio de extranjeros clasificados

Posición	País de origen	%Clasificados	%Clasificados/%Participantes	País del evento
1	Australia	12,5	1,53	India
2	Reino Unido	29,0	1,08	Irlanda
3	Argentina	19,2	0,93	Chile
4	Rusia	41,2	0,88	Kazakhstan
5	Estados Unidos	52,5	0,77	Puerto Rico
6	Estados Unidos	22,0	0,75	Canadá
7	Argentina	81,2	0,70	Uruguay
8	Chile	11,5	0,66	Ecuador
9	Estados Unidos	15,4	0,62	México
10	Japón	5,0	0,61	India
359	Uruguay	2,22	0,087	Argentina

Cuadro VI: Mejores lugares para clasificar según categoría para mujeres

Categoría	Lugar de clasificación	País	Tiempo de clasificación necesario	Tiempo promedio de la categoría
18-24	Panama City	Estados Unidos	5,23	5,26
25-29	Panama City	Estados Unidos	5,77	6,12
30-34	San Carlos de Bariloche	Argentina	5,34	5,90
35-39	San Carlos de Bariloche	Argentina	5,40	5,90
40-44	Panama City	Estados Unidos	5,63	5,70
45-49	Taitung	Taiwan	6,80	6,81
50-54	Gdynia	Polonia	5,39	5,82
55-59	Mooloolaba	Australia	5,44	5,65
60-64	Dubai	Emiratos Árabes Unidos	5,64	5,69
65-69	Cozumel	México	7,42	7,70
70-74	Canadá	Canadá	7,99	8,10

Cuadro VII: Mejores lugares para clasificar según categoría para hombres

Categoría	Lugar de clasificación	País	Tiempo de clasificación necesario	Tiempo promedio de la categoría
18-24	Panama City	Estados Unidos	5,65	5,96
25-29	Panama City	Estados Unidos	5,40	5,93
30-34	Panama City	Estados Unidos	5,15	5,67
35-39	Taupo	Nueva Zelanda	4,36	5,09
40-44	Panama City	Estados Unidos	5,18	5,68
45-49	Panama City	Estados Unidos	5,45	6,01
50-54	Cairns	Australia	5,37	5,96
55-59	Panama City	Estados Unidos	5,52	6,14
60-64	Cairns	Australia	5,86	6,12
65-69	Turquía	Turquía	5,33	5,49
70-74	Bangsaen	Tailandia	5,98	5,98
75-79	Canadá	Canadá	7,23	7,32
80-84	Maryland	Estados Unidos	6,93	7,83