Presentación: Propuesta de proyecto

Análisis de datos con SQL – Juegos Olímpicos (120 años)

Roberto Ernesto Cárdenas Rodríguez

Selección: **Cliente 3: SportsStats**

Conjunto de datos de los Juegos Olímpicos – 120 años de datos.

¿Por qué seleccionarlo?

Me parece interesante analizar la relación que existente en los deportistas al analizar por grupos de países, disciplinas deportivas, años de los eventos, etc. Creo que hay un valor importante en aprender lo mayormente posible de personas que destacan en cualquier tipo de disciplina para aprovechar su posición como atleta para aprovechar la imagen mediática en un enfoque empresarial.

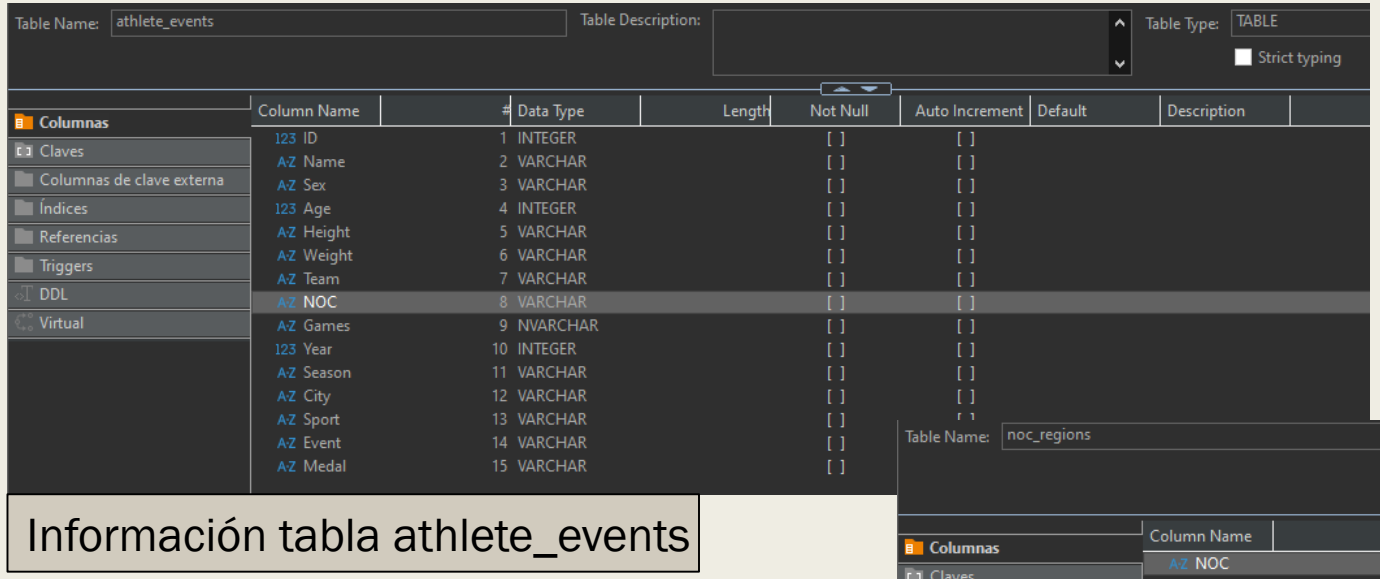
Conjunto de datos:

1. El conjunto de datos lo descargué en un archivo comprimido en el Dropbox del cliente.
2. En este hay 2 archivos CSV que contienen las siguientes tablas de información:
   - *athlete_events*
   - *noc_regions*

Importación y exploración inicial de los datos:

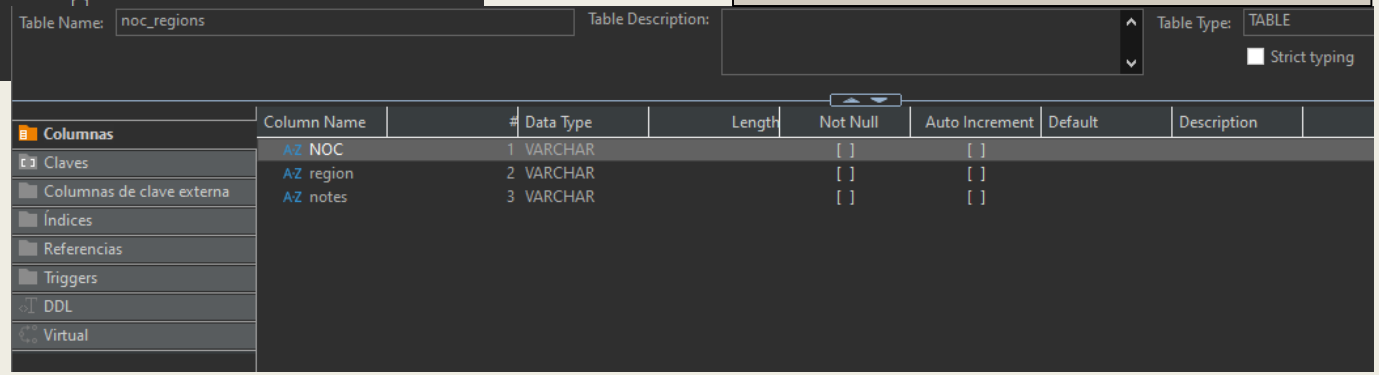
Importé los archivos CSV a una base de datos relacional en dbeaver conectado a SQLite, ahí encontré:

1. La tabla de athlete_events cuenta con las columnas: ID, Name, Sex, Age, Height, Weight, Team, NOC, Games, Year, Season, City, Sport, Event, Medal.
2. La tabla noc_regions contiene las columnas: NOC, región, notes.



Información tabla athlete_events

Información tabla noc_regions



Relación:

Encontré que ambas tablas tienen una columna en común de nombre “NOC”, hice la relación de ambas tablas a través de ellas con una clave foránea virtual.

athlete_events 1

Enter a SQL expression to filter results (use Ctrl+Space)

	ID	Name	Sex	Age	Height	Weight	Team	NOC
1	1	A Dijiang	M	24	180	80	China	CHN
2	2	A Lamusi	M	23	170	60	China	CHN
3	3	Gunnar Nielsen Aaby	M	24	NA	NA	Denmark	DEN
4	4	Edgar Lindenau Aabye	M	34	NA	NA	Denmark/Sweden	DEN
5	5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED
6	5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED
7	5	Christine Jacoba Aaftink	F	25	185	82	Netherlands	NED
8	5	Christine Jacoba Aaftink	F	25	185	82	Netherlands	NED
9	5	Christine Jacoba Aaftink	F	27	185	82	Netherlands	NED
10	5	Christine Jacoba Aaftink	F	27	185	82	Netherlands	NED

noc_regions 1

Enter a SQL expression to filter

	NOC	region	notes
1	AFG	Afghanistan	
2	AHO	Curacao	Netherlands Antilles
3	ALB	Albania	
4	ALG	Algeria	
5	AND	Andorra	
6	ANG	Angola	
7	ANT	Antigua	Antigua and Barbuda
8	ANZ	Australia	Australasia
9	ARG	Argentina	
10	ARM	Armenia	

DESCRIPCIÓN DEL PROYECTO

Este proyecto contiene información que puede ser útil para **medios de información deportiva, empresas o marcas con giro empresarial dirigido al deporte y analistas de deportistas.**

Al tener información de los deportistas que han participado en distintas disciplinas deportivas en los juegos olímpicos (JO) durante un periodo de 120 años puede utilizarse como **impulso a nuevas carreras deportivas de futuras generaciones de cada país.**

Identificar a los deportistas por su carrera profesional en relación al país que representan en los JO permite que las empresas que manejan productos o servicios del giro deportivo **generar campañas de marketing** enfocando sus esfuerzos en el país donde el deporte tiene historial de triunfos o campeonatos.

PREGUNTAS

1. ¿Los atletas medallistas tienden a ser mayores que el promedio de participantes en su disciplina?

2. ¿Cuáles deportes tienen mayor participación de los atletas en los Juegos Olímpicos?

3. ¿Qué país tiene la mayor cantidad de medallas olímpicas en todas las disciplinas registradas?

HIPÓTESIS INICIAL

1. El promedio de los deportistas que más han ganado medallas olímpicas considero que será mayormente en edades longevas que los más jóvenes, debido a la experiencia que pueden acumular durante toda su carrera deportiva, mientras que los más jóvenes apenas están comenzando su carrera deportiva y no cuentan con la confianza que puede tener alguien veterano de cualquier disciplina.

2. Existe una relación proporcional entre la cantidad de deportistas que ganan medallas con la cantidad de deportistas que representan cada país. Mientras más deportistas tenga un país en los JO, más medallas acumularán.

ENFOQUE

Trabajar los datos al realiza la relación faltante de los atletas con el país que representan al hacer un JOIN, seguido por arreglar la columna del país “Singapore” para que conserve las siglas del NOC que tiene la tabla de athlete_events.

Las columnas que considero tienen mayor importancia están en la tabla athlete_events con los nombres de **Sex, Age, Year, NOC, Sport, Medal**. Mientras que en la tabla de noc_regions son las columnas de **NOC y region**.

La relación que más interesa abordar es la de las edades de los participantes con el número de medallas obtenidas y la disciplina deportiva en la que participan. Creo que puede ser de interés comercial para las marcas e individuos del rubro deportivo por el mercadeo que pueden lograr.

Las métricas que considero son útiles son gráficas de barras, gráficas de pastel y graficas con líneas.

RESUMEN METODOLÓGICO

Comencé ejecutando consultas para conocer el contenido de las tablas para saber el nombre de sus columnas y la información que podía encontrar en ellas. La tabla athlete_events contenía mucha información de mucha relevancia para conocer a los atletas que compitieron en Juegos Olímpicos. Por otra parte, la tabla noc_regions ayuda a conseguir la relación entre los atletas con la región que representan en situaciones donde su nomenclatura ha variado. Por ejemplo, a la región de Rusia la han representado atletas cuyos equipos representaban a los equipos de Russia, Unión Soviética o Equipo Unificado.

Enfoqué el análisis en la tabla athlete_events comenzando a contar la cantidad de medallas ganadas por los atletas agrupándolos por el país que representan y saber que países eran los que tenían más medallas olímpicas, además de conocer las edades en que estos atletas ganaban medallas según la disciplina en la que estaban compitiendo.

También hice consultas donde comparé promedios de edades entre los atletas ganadores en sus disciplinas con las edades promedio de otros atletas de la misma disciplina y con todos los atletas que participan en juegos olímpicos.

Durante el análisis encontré información que me parece interesante para responder a la propuesta inicial del proyecto, que era encontrar patrones entre los deportistas olímpicos para el uso de ellos en la mercadotecnia de marcas deportivas.

PUNTOS CLAVE

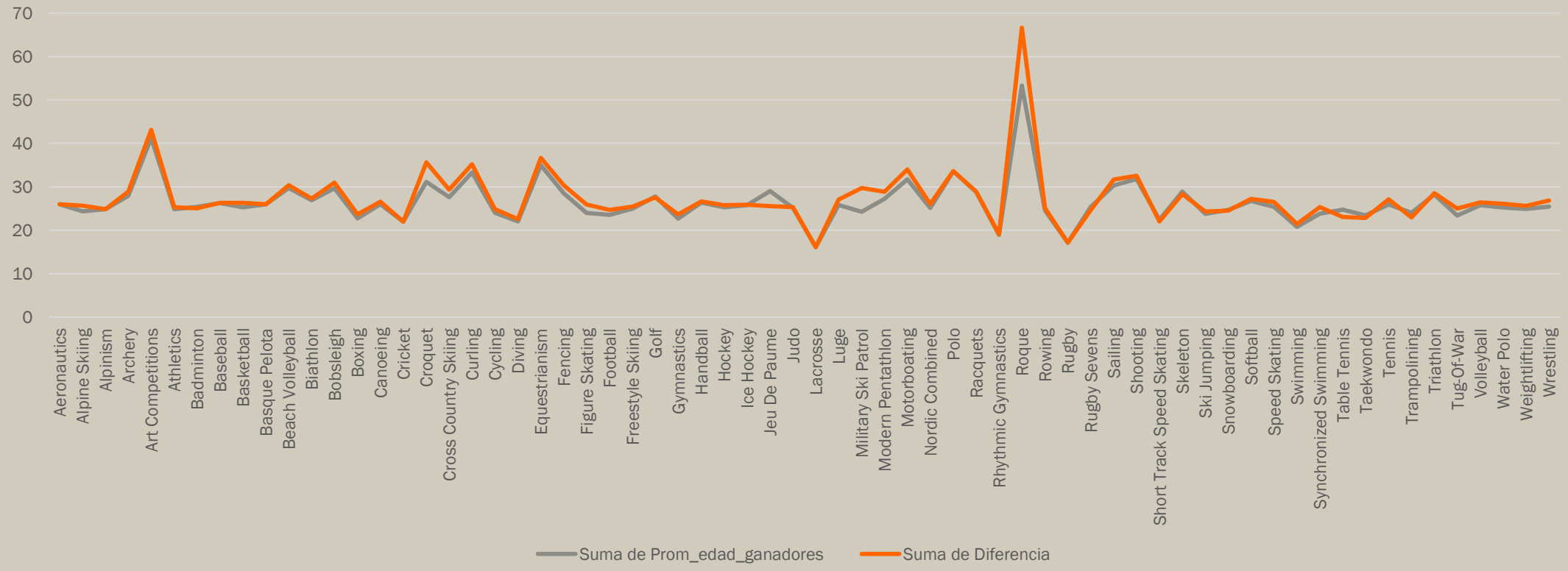
1. La región de USA (Estados Unidos de América) es la que tiene el mayor número de medallas olímpicas ganadas en los 120 años. **Tiene el 29.89% de ocasiones donde ha conseguido medallas divido en la cantidad de participaciones**, siendo así, **el más alto entre todas las regiones que compiten en los JO**. Es una región importante para dar seguimiento a las carreras de los atletas.

2. Existen 66 deportes registrados en el conjunto de datos, en **el 75.75% de esos deportes el promedio de edad de los atletas que han ganado la medalla olímpica superan al promedio de edad de otros competidores en la misma disciplina.**

3. El deporte de la **natación aparece 3 veces en el top 7** en las regiones de Estadios Unidos de América, Alemania y Australia. El deporte de **atletismo también aparece 3 veces en este top 7** con ganadores de las regiones de Estados Unidos de América, Russia y Alemania.

CONCLUSIÓN DE HIPÓTESIS

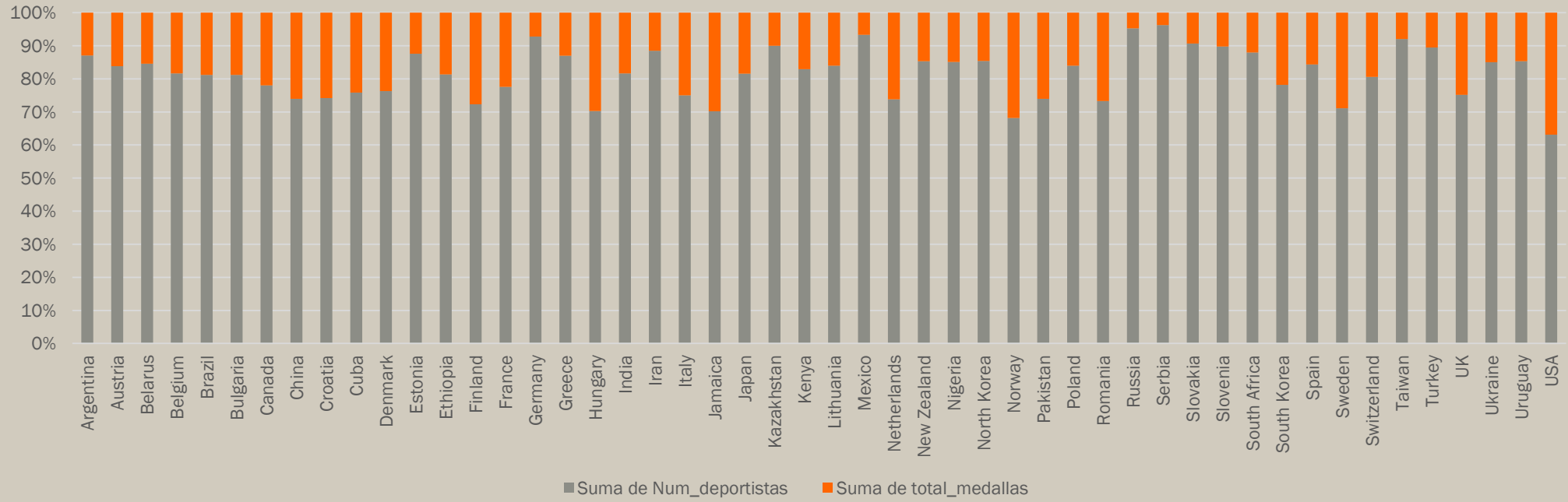
En 50 de los 66 deportes donde hay competencias olímpicas **el promedio de la edad de los atletas medallistas superan el promedio de edad registrada para cada uno de sus deportes.** <u>En promedio los atletas más longevos han dominado su competición.</u>



CONCLUSIÓN DE HIPÓTESIS

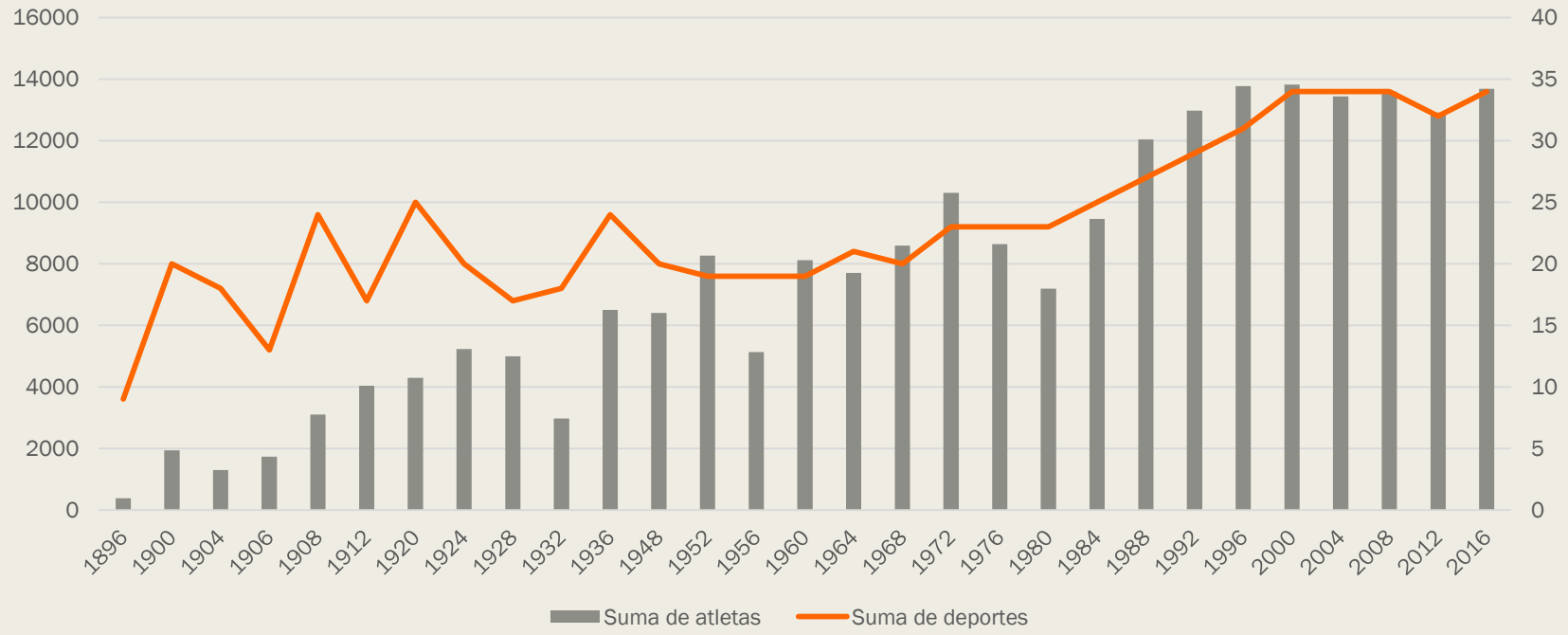
No existe una relación directa en todas las regiones de la cantidad de medallas ganadas con la cantidad de participaciones de atletas para cada región en los JO. Por ejemplo, hay regiones que tienen muchas participaciones pero pocas medallas. **Destaca el porcentaje de medallas obtenidas en las participaciones de USA**. Las participaciones no equivalen a atletas únicos, lo cual puede afectar la proporción.

Más atletas ≠ Mayor probabilidad de medalla



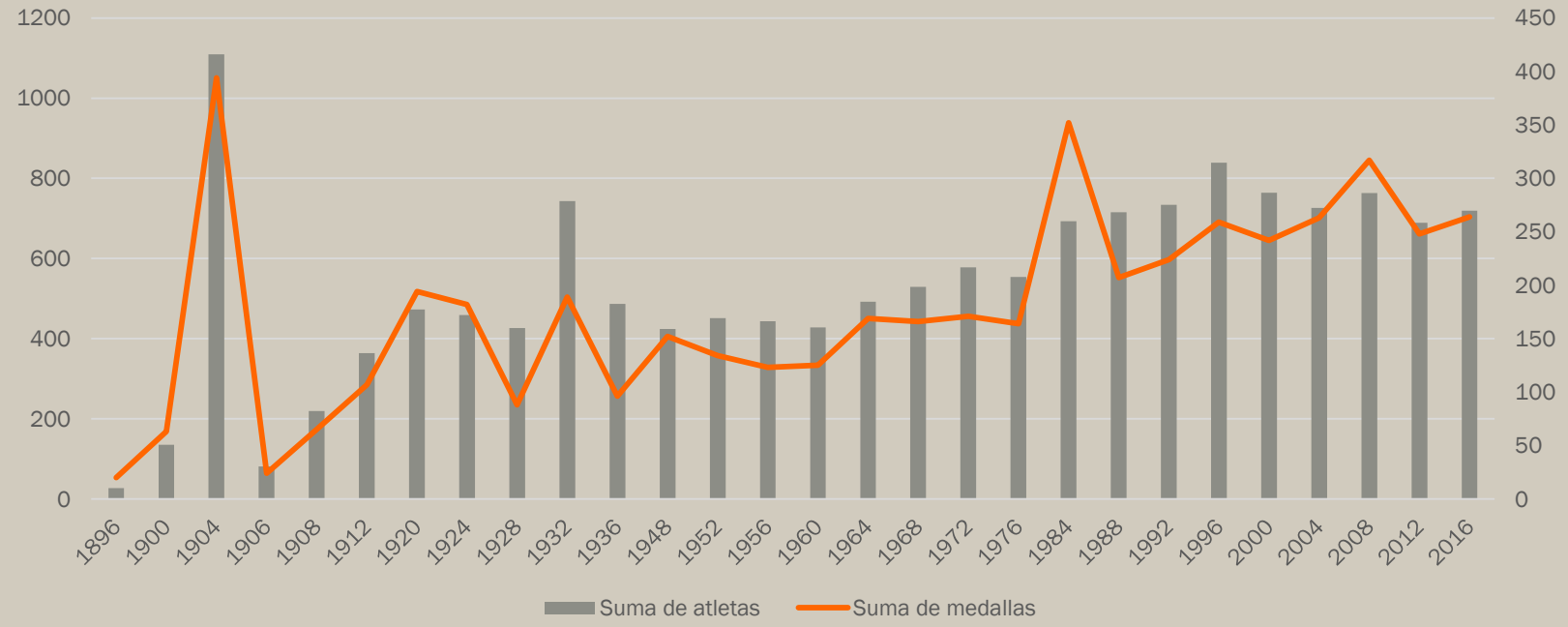
CORRELACIONES DESCUBIERTAS

Al hacer un conteo de los distintos deportes practicados en cada edición de los Juegos Olímpicos agrupándolos por el año en que ocurrieron pude encontrar una pequeña correlación que mientras más disciplinas deportivas se están realizando en esa edición de JO, más atletas están compitiendo en esa edición. Aunque no ocurre en todas las ediciones generalmente está ocurriendo la relación. Las pequeñas variaciones pueden deberse a incidentes geopolíticos en ese periodo de tiempo, como la poca participación de atletas en 1980 por el boicot a los JO.



CORRELACIONES DESCUBIERTAS

En el caso específico de USA como la región que consideramos que tiene un mercado de deporte de gran importancia mediática también encontramos una correlación entre la cantidad de atletas que participan en JO con la cantidad de medallas que consiguen en su participación en la competencia.



DESCUBRIMIENTOS RESPECTO HIPÓTESIS

Los ganadores de medallas para cada disciplina suelen ser mayores al promedio de la edad en cada deporte, por ello **se debe aprovechar cuando los atletas están empezando sus carreras deportivas desde jóvenes**, realizar contratos que aprovechen su imagen deportiva para destacar en el deporte alcanzar en las edades en que suelen destacar. Con esto se conseguiría **trabajar la carrera del mismo desde joven y permitir dar seguimiento de la misma para los fines empresariales o deportivos deseados.**

Se descubrió que si una región envía a muchos atletas a competir no significa que muchos de esos atletas regresarán a casa con medalla de ganador. Con este análisis se logra detectar **cuales son los países que tienen mayor porcentaje de victoria** con sus respectivos representantes, identificando **en que regiones se puede trabajar con mayor seguridad y probabilidad de logro** para encontrar un posible campeón olímpico.

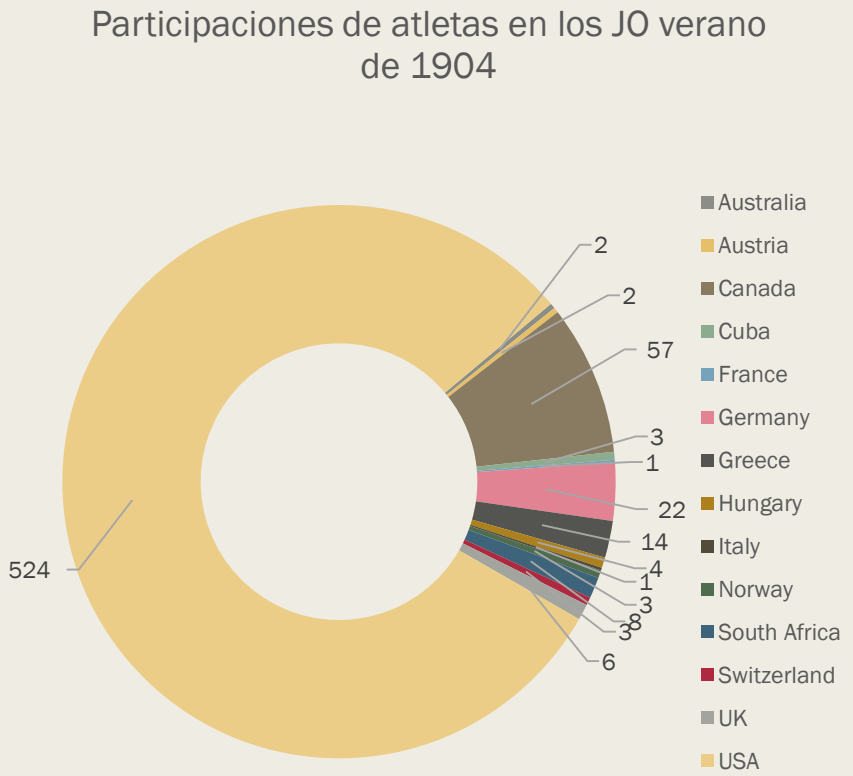
DESCUBRIMIENTOS

Existen deportes que han estado presentes en los Juegos Olímpicos a lo largo de su historia, dejando a su paso un historial de atletas participantes, medallistas e incluso records, sin embargo en algunas ediciones de este evento deportivo **se han celebrado competencias únicas en algún deporte**, el cual tiene su valor como deporte pero que no volvió a considerarse un deporte olímpico, dejando de aparecer en el historial de la competición en otras ediciones, tal es el caso de deportes como el ROQUE, AERONAUTICA, BASQUE PELOTA, entre otros.

DESCUBRIMIENTOS

A la vez en el recorrido histórico de estos juegos han ocurrido situaciones geopolíticas que han tenido importancia en las estadísticas de los registros de este evento, como en el año 1904 que la competencia tuvo una duración de casi 5 meses teniendo una cantidad abrumadora de atletas (en especial de la región de Estados Unidos) compitiendo en los JO.

También en el año de 1980 donde el país de USA decidió manifestarse y no participar en los JO de verano, invitando a otras regiones a hacer lo mismo en exigencia de sus demandas como atletas, haciendo que ese evento tuviera poca participación de atletas; situaciones como ésta pueden afectar muchísimos a los intereses económicos y tienen un peso importante en el mundo deportivo.



PRÓXIMOS PASOS / RECOMENDACIONES

- Realizar un análisis de los atletas que compiten en JO en categorías ganadoras filtrando los países que no causan interés a la empresa, destacando el país o países que si se adecuan al mercado deseado por la misma para enfocar esfuerzos en esos atletas.

- Diseñar un plan de acción para trabajar con los atletas jóvenes que están teniendo su primera o primeras participaciones en competencias de JO, haciendo que consigan más experiencia y continuidad en su disciplina deportiva para llevarlos al camino de ser medallistas olímpicos. Es un ganar-ganar, si el deportista que apoyas consigue medalla, su imagen y valor en el mercado comercial aumenta y la empresa que lo representa haciendo uso de su imagen también adquiere más valor.

- Después de seleccionar el país o países que sean del interés de la empresa hay que relacionar la tabla de atletas olímpicos con los atletas en otras competiciones (que no sean JO) para encontrar atletas jóvenes que sean ganadores o que han mostrando buen desempeño.