

An aerial night photograph of a large, modern stadium with a complex, illuminated steel structure. The stadium is filled with spectators, and the field is brightly lit. Numerous fireworks are exploding in the dark sky above the stadium, creating a festive atmosphere. The surrounding city lights are visible in the background.

Historia de los Juegos Olímpicos

INTEGRANTES

ALEJANDROLARA MONDRAGON
ALEJANDRO SENABRE

PROYECTO FINAL DE CODERHOUSE



Introducción.....	2
Descripción de la temática de los datos.....	2
Alcance	2
Hipótesis	3
Tipo de Análisis.....	3
Dataset	3
Transformación del Dataset	3
Separación en tablas.....	4
Tabla [Países]	5
Tabla[Atletas].....	5
Tabla [Evento].....	5
Tabla [Genero]	6
Tabla [Ganadores]	6
Tabla [Disciplina].....	7
Diagrama Entidad-Relación	7
Listado final de las tablas.....	8
Listado columnas por tabla	9
Modelo relacional en Power Bi	10
Segmentaciones elegidas	11
Medidas Calculadas.....	11
Implementación en Power Bi	19
Visualización de los datos:	20
Pagina Portada.....	20
Pagina Medallero.....	21
Pagina Atletas, Edad y Peso.....	21
Página sedes	23
Pagina Invierno/Verano.....	25
Pagina atletas.....	26
Conclusión.....	27
Próximos pasos:	28
Bibliografía.....	28
Herramientas utilizadas.....	28





Introducción

En 776 AC en Olimpia Grecia, se realizaron los primeros juegos olímpicos en dónde el premio para los atletas consistía en una corona de olivos, los atletas participantes eran considerados los mejores en cada ciudad y competían en honor a Zeus. Tras unos cuantos siglos se tomó de inspiración para que, en 1894 se creara el comité olímpico en 1894 y se designara como sede Atenas para la edición, en ella participaron 14 países con 241 deportistas en 43 pruebas de 10 disciplinas.

Los Juegos Olímpicos han sido un evento que ha marcado la historia en cada una de sus ediciones, ya que es un evento deportivo multidisciplinario a nivel internacional con miles de atletas en búsqueda de la preciada medalla de oro. En muchos deportes, la medalla de oro es más ansiada que el propio mundial por lo difícil que es llegar a unos juegos olímpicos y por lo que representa a nivel mundial.

El logo de los juegos olímpicos son 5 anillos de colores “según la Regla 8 de la Carta Olímpica, “El símbolo olímpico expresa la actividad del Movimiento Olímpico y representa la unión de los cinco continentes y el encuentro de atletas de todo el mundo en los Juegos Olímpicos”. Los cinco continentes a los que se hace referencia son África, Asia, América, Europa y Oceanía”. (Chris McCrory, 2012)

En el presente trabajo se decidió abordar este tema, ya que consideramos que los Juegos Olímpicos es un tema muy importante a nivel deportivo que no podemos dejar atrás en los análisis que se realizan día a día, pues con ello podemos tener una mejor visión de lo que hace que los países se conviertan en los más ganadores debido a ciertas características de los atletas y/o podamos analizar más a detalle a cada uno de ellos.

Consultar la información de los ganadores y algunos detalles físicos fue complicado, sin embargo, tuvimos que hacer una gran depuración para simplificar la data base y evitar la repetición de datos con identificadores.

Este análisis se basó en los datos históricos desde la creación del comité olímpico de 1894 y su primera edición ya que consideramos que ese fue el inicio en la era moderna de los Juegos Olímpicos.

Descripción de la temática de los datos

En este trabajo se encontró información acerca del medallero en histórico de los juegos olímpicos, el cual contiene información como acerca de cada participante qué logró una medalla con sus características principales como la nacionalidad, la edición, el deporte, la disciplina, la medalla, el género y el evento. La base de datos incluye desde los primeros juegos olímpicos realizados en Atenas en 1896 hasta los juegos de Río de Janeiro en 2016.

Alcance

El dashboard que planteamos pretende ser orientado a nivel operativo, con el fin de detallar cada uno de los eventos que han marcado la historia de los deportes para que el público pueda conocer todos los detalles con un simple clic.

El mapa interactivo tiene la función de encontrar detalles que muchos quisieran saber sobre las características que hacen ganadores a los atletas, dependiendo el deporte, la nacionalidad y su aspecto corporal



Hipótesis

Nuestro objetivo es mejorar la forma en la cual se puede entender la historia de los resultados de las ediciones de los juegos olímpicos. Por ello verán cada aspecto, desde las medallas por nación, hasta las diferencias de edad, estatura y peso sobre los ganadores de las medallas. Para poder comprobar si tiene algo que ver, la estatura, edad o peso de cada atleta para la obtención de las medallas.

Tipo de Análisis

Considerando la base de datos existente, realizaremos 3 tipos de análisis; El descriptivo en el cual nos enfocaremos en observar y describir cierto tipo de comportamientos percatados en la base de datos; El diagnóstico en donde realizaremos cierto tipo de conclusiones una vez identificando las cuestiones más importantes de nuestros datos históricos; así mismo identificaremos comportamientos futuros en base a nuestra información con un tipo de análisis predictivo en el cual podremos predecir qué países podrán ganar más medallas en las próximas ediciones y en qué deportes son mejores cada país para entrar al medallero, entre otras predicciones.

Dataset

La base de datos, extraída por medio de Kaggle¹, recopila la información individual de los atletas empezando por las olimpiadas de Atenas 1896 hasta las de Río de Janeiro de 2016.



Los datos en crudo abierto mediante la aplicación Bloc de notas

Los datos se encontraban inicialmente almacenados en un archivo tipo CSV², lo que quiere decir que las columnas están separadas mediante comas y toda la información junta en una sola tabla. Este tipo de archivo es beneficioso por su poca exigencia de cómputo y bajo espacio de almacenamiento, pero era necesario llevar a cabo una transformación de los datos con el fin de manejarlos a nuestro gusto.

Transformación del Dataset

- Para esto se empleó Microsoft Excel® con su función “Texto en columnas” pudiendo indicar el limitador usado en el CSV, en este caso el uso de las comas, para poder separar los datos en columnas individuales. Esta función también nos permite indicar el formato de las columnas, ya sea texto, fecha, número, etc.

¹ Kaggle es una página web que dispone de miles de bases de datos para poder descargar y usar de manera gratuita.

² Las siglas CSV vienen del inglés "Comma Separated Values" y significan valores separados por comas.



además de poder especificar el uso de decimales usados en el dataset, en este caso, el separador de decimales usado era mediante el punto, algo que en esta parte del mundo es reemplazado por la coma. Si no se hubiera realizado este último cambio, en los datos del peso de los atletas, hubieran figurado gran cantidad de datos nulos, ya que a los números que contengan un punto, Excel automáticamente los trataría con el formato texto, y no con el numérico como necesitamos.

Una vez realizado este cambio, el archivo de texto pasa a ser un XLSX, un archivo de Excel cuyas tablas ya están habilitadas para una posterior operación.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
2	1	A Diliang	M	24	180	80	China	CHN	1992 Summer	1992 Summer	Barcelona	Basketball	Basketball M	NA	
3	2	A Lamusi	M	23	170	60	China	CHN	2012 Summer	2012 Summer	London	Judo	Judo Men's E	NA	
4	3	Gunnar Niel	M	24	NA	NA	Denmark	DEN	1920 Summer	1920 Summer	Antwerpen	Football	Football Mei	NA	
5	4	Edgar Linder	M	34	NA	NA	Denmark/Sw	DEN	1900 Summer	1900 Summer	Paris	Tug-Of-War	Tug-Of-War	Gold	
6	5	Christine Jac	F	21	185	82	Netherlands	NED	1988 Winter	1988 Winter	Calgary	Speed Skatir	Speed Skatir	NA	
7	5	Christine Jac	F	21	185	82	Netherlands	NED	1988 Winter	1988 Winter	Calgary	Speed Skatir	Speed Skatir	NA	
8	5	Christine Jac	F	25	185	82	Netherlands	NED	1992 Winter	1992 Winter	Albertville	Speed Skatir	Speed Skatir	NA	
9	5	Christine Jac	F	25	185	82	Netherlands	NED	1992 Winter	1992 Winter	Albertville	Speed Skatir	Speed Skatir	NA	
10	5	Christine Jac	F	27	185	82	Netherlands	NED	1994 Winter	1994 Winter	Lillehammer	Speed Skatir	Speed Skatir	NA	
11	5	Christine Jac	F	27	185	82	Netherlands	NED	1994 Winter	1994 Winter	Lillehammer	Speed Skatir	Speed Skatir	NA	
12	6	Per Knut Aal	M	31	188	75	United State	USA	1992 Winter	1992 Winter	Albertville	Cross Countr	Cross Countr	NA	
13	6	Per Knut Aal	M	31	188	75	United State	USA	1992 Winter	1992 Winter	Albertville	Cross Countr	Cross Countr	NA	

La visualización correspondiente en Microsoft Excel® una vez separado los datos en columnas

Separación en tablas

Con el fin de simplificar la vista de la información y a su vez de eliminar mucho dato redundante, el siguiente paso realizado fue separar la información en diferentes tablas que recopile determinada información. Esto nos dio 6 diferentes tablas:

- Ganadores
- Atletas
- Disciplina
- Evento
- Países
- Genero

El objetivo de estas tablas era crear un identificador propio para cada información con el fin de evitar la repetición de los datos. Ejemplo: En vez de tener repetido miles de veces el nombre del país “Estados Unidos” fue reemplazado por un identificador más acotado, en este caso por “USA”. Así para cada uno de los datos. Este cambio, además de beneficiar a la lectura y al manejo de los datos, también tuvo un impacto positivo en el tamaño del archivo, logrando obtener una reducción del 58,5% en el tamaño del mismo.



Tabla [Países]

Para realizar esta tabla utilizamos dos de las columnas originales de la base de datos. La columna NOC³ y TEAM. Para agarrar la totalidad de los países se utilizó la función =UNICO para crear una tabla evitando la repetición de los nombres. Finalmente para la columna id_pais se realizó con =BUSCARV, para que encuentre la coincidencia entre el nombre del país de la columna original TEAM con las primeras siglas de la columna NOC. El resultado fue el siguiente:

	A	B
1	id_pais	Pais
2	CHN	China
3	DEN	Denmark
4	NED	Netherlands
5	USA	United States
6	FIN	Finland
7	NOR	Norway
8	ROU	Romania
9	EST	Estonia
10	FRA	France
11	MAR	Morocco
12	ESP	Spain
13	EGY	Egypt

Tabla[Atletas]

Esta tabla prácticamente se conservaron todos los datos originales, ya que afortunadamente, en la base de datos original se encontraba una columna ID único para cada participante, por lo que facilito mucho el manejo de los datos.

	A	B	C	D	E	F	G
1	id_atleta	id_genero	Nombre	id_pais	Peso	Altura	Edad
2	1	M	A Dijiang	CHN	80	180	24
3	2	M	A Lamusi	CHN	60	170	23
4	3	M	Gunnar Nielsen Aaby	DEN	NA	NA	24
5	4	M	Edgar Lindenau Aabye	DEN	NA	NA	34
6	5	F	Christine Jacoba Aaftink	NED	82	185	21
7	6	M	Per Knut Aaland	USA	75	188	31
8	7	M	John Aalberg	USA	72	183	31
9	8	F	Cornelia "Cor" Aalten (-Strannoc	NED	NA	168	18
10	9	M	Antti Sami Aalto	FIN	96	186	26
11	10	M	Einar Ferdinand "Einari" Aalto	FIN	NA	NA	26
12	11	M	Jorma Ilmari Aalto	FIN	76,5	182	22
13	12	M	Jyri Tapani Aalto	FIN	70	172	31
14	13	F	Minna Maarit Aalto	FIN	55,5	159	30
15	14	F	Pirjo Hannele Aalto (Mattila-)	FIN	65	171	32
16	15	M	Arvo Ossian Aaltonen	FIN	NA	NA	22

Tabla [Evento]

Esta tabla contiene los 51 eventos comprendidos entre los años trabajados, se extrajeron usando otra vez =UNICOS para conseguir los eventos sin repetición de los mismo, posteriormente para cada evento se le creó un id único utilizando la herramienta =CONCAT para unir los valores de distintas celdas en una, y también las funciones =IZQUIERDA y =DERECHA que devuelven un numero específico de caracteres comenzando de la izquierda o derecha respectivamente. Con esto se agarró los últimos dos dígitos el año, la inicial de la temporada, ya sea Summer (S) o Winter (W), y las

³ Estas siglas hacen referencia al Comité Olímpico Internacional por sus siglas en ingles "National Olympic Committee"



primeras 3 letras del nombre de la ciudad. Específicamente la función completa utilizada fue la siguiente:

```
=CONCAT(IZQUIERDA(F3;3);DERECHA(D3;2);IZQUIE
```

Luego se creó manualmente una columna para id_pais, donde se buscó uno por uno para los 51 eventos la ubicación de la ciudad y escribir el respectivo ID correspondiente al país.

	A	B	C	D	E
1	Id_evento	Año	Temporada	Ciudad	id_pais
2	Ath96S	1896	Summer	Athina	GRE
3	Par00S	1900	Summer	Paris	FRA
4	St.04S	1904	Summer	St. Louis	USA
5	Ath06S	1906	Summer	Athina	GRE
6	Lon08S	1908	Summer	London	GBR
7	Sto12S	1912	Summer	Stockholm	SWE
8	Ant20S	1920	Summer	Antwerpen	BEL
9	Cha24W	1924	Winter	Chamonix	FRA
10	Par24S	1924	Summer	Paris	FRA
11	Ams28S	1928	Summer	Amsterdam	NED
12	San28W	1928	Winter	Sankt Moritz	SUI
13	Lak32W	1932	Winter	Lake Placid	USA
14	Los32S	1932	Summer	Los Angeles	USA
15	Ber36S	1936	Summer	Berlin	GER
16	Gar36W	1936	Winter	Garmisch-Partenkirchen	GER

Los demás valores se consiguieron con la función =BUSCARV

Tabla [Genero]

Esta es la tabla más llana a comparación de las otras. Se hizo completamente manual por la poca cantidad de datos usados. Se incluyó un tercer genero con el identificador “X” para hacer referencias a los deportes donde permite la participación de ambos géneros.

	A	B
1	id_genero	Genero
2	M	Masculino
3	F	Femenino
4	X	Mixto

Tabla [Ganadores]

Para integrar todos los identificadores anteriormente creados y resumir bastante la información, se creó la tabla “ganadores” siendo esta la tabla principal que tiene relación con la mayoría de las demás tablas.





	A	B	C	D	E
1	id_atleta	id_evento	id_disciplina	Medal	Id_ganador
2	1	Bar92S	1BaskM	NA	A1ID1
3	2	Lon12S	2JudoM	NA	B2ID2
4	3	Ant20S	3FootM	NA	C3ID3
5	4	Par00S	4Tug-F	Gold	D4ID4
6	5	Cal88W	5SpeeF	NA	E5ID5
7	5	Cal88W	6SpeeM	NA	F6ID5
8	5	Alb92W	5SpeeF	NA	G7ID5
9	5	Alb92W	6SpeeM	NA	H8ID5
10	5	Lil94W	5SpeeF	NA	I9ID5
11	5	Lil94W	6SpeeM	NA	J10ID5
12	6	Alb92W	7CrosM	NA	K11ID6
13	6	Alb92W	8CrosM	NA	L12ID6
14	6	Alb92W	9CrosM	NA	M13ID6
15	6	Alb92W	10CrosM	NA	N14ID6

Aparte de esto se creó un identificador único para esta tabla, que fue creado uniendo id_atleta junto con otras columnas que contenían la combinación de una letra del abecedario junto con un número que subía progresivamente para asegurarse que fuese un ID único.

Tabla [Disciplina]

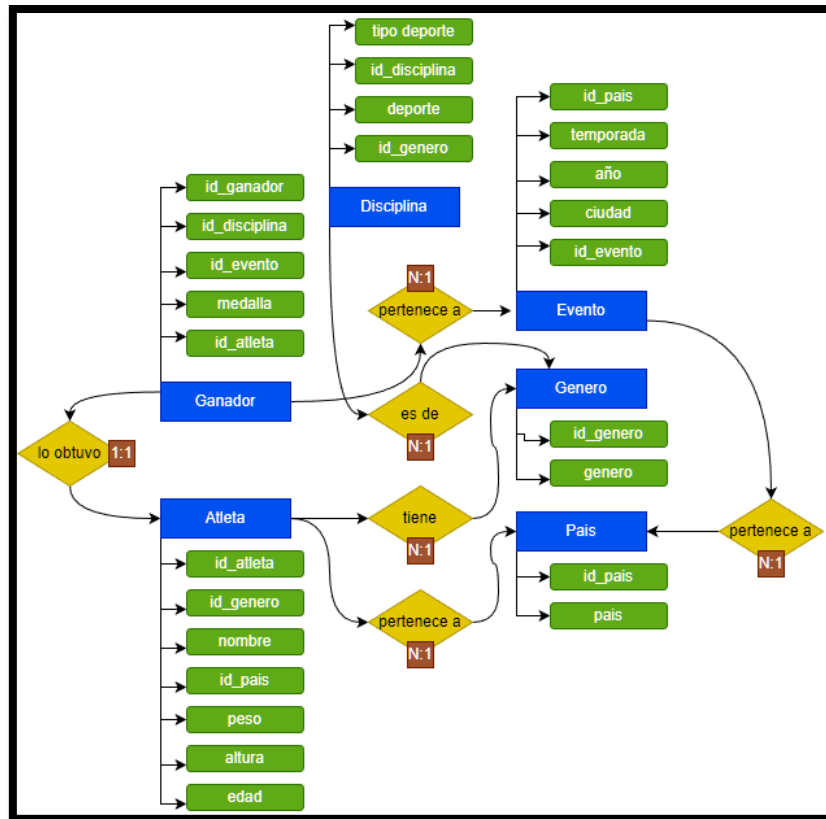
Por último, tenemos la tabla disciplina que más complejidad tomó a la hora de crearla.

	A	B	C	D
1	id_disciplina	deporte	Tipo	id_gener
2	1BaskM	Basketball	Basketball	M
3	2JudoM	Judo	Extra-Lightweight	M
4	3FootM	Football	Football	M
5	4Tug-M	Tug-Of-War	Tug-Of-War	M
6	5SpeeF	Speed Skating	500 metres	F
7	6SpeeF	Speed Skating	1,000 metres	F
8	7CrosM	Cross Country Skiing	10 kilometres	M
9	8CrosM	Cross Country Skiing	50 kilometres	M
10	9CrosM	Cross Country Skiing	10/15 kilometres Pursuit	M
11	10CrosM	Cross Country Skiing	4 x 10 kilometres Relay	M
12	11CrosM	Cross Country Skiing	30 kilometres	M
13	12AthIF	Athletics	100 metres	F
14	13AthIF	Athletics	4 x 100 metres Relay	F
15	14Ice M	Ice Hockey	Ice Hockey	M
16	15SwimM	Swimming	400 metres Freestyle	M
17	16BadmM	Badminton	Singles	M
18	17SailF	Sailing	Windsurfer	F
19	18BiatF	Biathlon	7.5 kilometres Sprint	F

Diagrama Entidad-Relación

A continuación, se puede observar el diagrama E-R que fue creado mediante la aplicación Diagrams:





Listado final de las tablas

En este apartado, se mencionará cada una de las tablas junto a una breve descripción de éstas y la definición del tipo de clave primaria y/o foránea:

Atletas: Contiene los datos principales de los atletas como el nombre, edad, peso, alturas agrupadas por atleta, país, género.

- PK id_atleta
- FK id_pais
- FK id_genero
- FK Nombre
- FK Edad
- FK Peso
- FK Altura

Género: Contiene el id agrupado del género de cada atleta.

- PK Id_genero
- FK género

Nacionalidad: Contiene la nacionalidad de cada atleta agrupada por números.

- PK id_pais
- FK país

Ganador: Contiene los id de los atletas por medalla conseguida, así como el evento en el que se consiguió y la disciplina.

- PK id_ganador
- FK id_disciplina
- FK id_evento
- FK medalla



- FK id_atleta

DisciplinaFinal: Contiene agrupada la disciplina en la que participa el atleta, así como el género.

- PK id_disciplina
- FK Deporte
- FK id_genero
- FK Tipo

Evento: Contiene la ciudad en la que se realizó el evento, el año del evento, la temporada en del año en la que fue, el país dónde se realizó el evento que hubo.

- PK id_evento
- FK ciudad
- FK año
- FK temporada
- FK id_pais

Año: Contiene los años en los cuales se realizó cada evento.

- PK Año

Medidas: Contiene todas las medidas calculadas creadas.

Listado columnas por tabla

Atletas		
Campo	Tipo de Campo	Tipo de Clave
Id_atleta	int	PK-INDEX
Id_pais	int	FK
Id_genero	int	FK
Nombre	Varchar/text	FK
Edad	Int	FK
Peso	Decimal	FK
Altura	Decimal	FK

Género		
Campo	Tipo de Campo	Tipo de Clave
Id_genero	int	PK
Género	Varchar/text	FK

País		
Campo	Tipo de Campo	Tipo de Clave
Id_pais	var	PK
pais	Varchar/text	FK

Ganadores(mod)		
Campo	Tipo de Campo	Tipo de Clave
Id_ganador	varchar	PK
Id_evento	int	FK
Id_atleta	int	FK
Medal	Varchar/text	FK
Id_disciplina	Int	FK



Evento		
Campo	Tipo de Campo	Tipo de Clave
Id_evento	int	PK-INDEX
temporada	Varchar/text	FK
año	Int o data/time	FK
Ciudad	Varchar/text	FK
Id_pais	Int	FK

Disciplina Final		
Campo	Tipo de Campo	Tipo de Clave
Id_disciplina	int	PK
Deporte	Varchar/text	FK
Id_genero	Int	FK
Tipo	Varchar/text	FK

Año		
Campo	Tipo de Campo	Tipo de Clave
año	Int o data/time	PK

Medidas		
Campo	Tipo de Campo	Tipo de Clave

Modelo relacional en Power Bi

Después de subir los archivos a Power Bi, y comenzar a trabajar en la visualización, se comenzaron a hacer ciertos ajustes que permitieran la correcta filtración de datos.

Evento:

- Se agrega la columna de Año Calculado para usar en modo calendario las fechas y mejorar el análisis.

Fecha de Actualización:

- Se añadió una tabla para modificar la fecha de actualización del documento para que se pueda ver cuándo fue la última fecha en la que se modificó el archivo.

Ganadores(mod):

- Se mejoró la relación de los ganadores puesto que a primera instancia sólo tomaba en cuenta a los atletas en sólo una edición, al final se transformó la base para que salieran todas las veces que un atleta había ganado alguna medalla

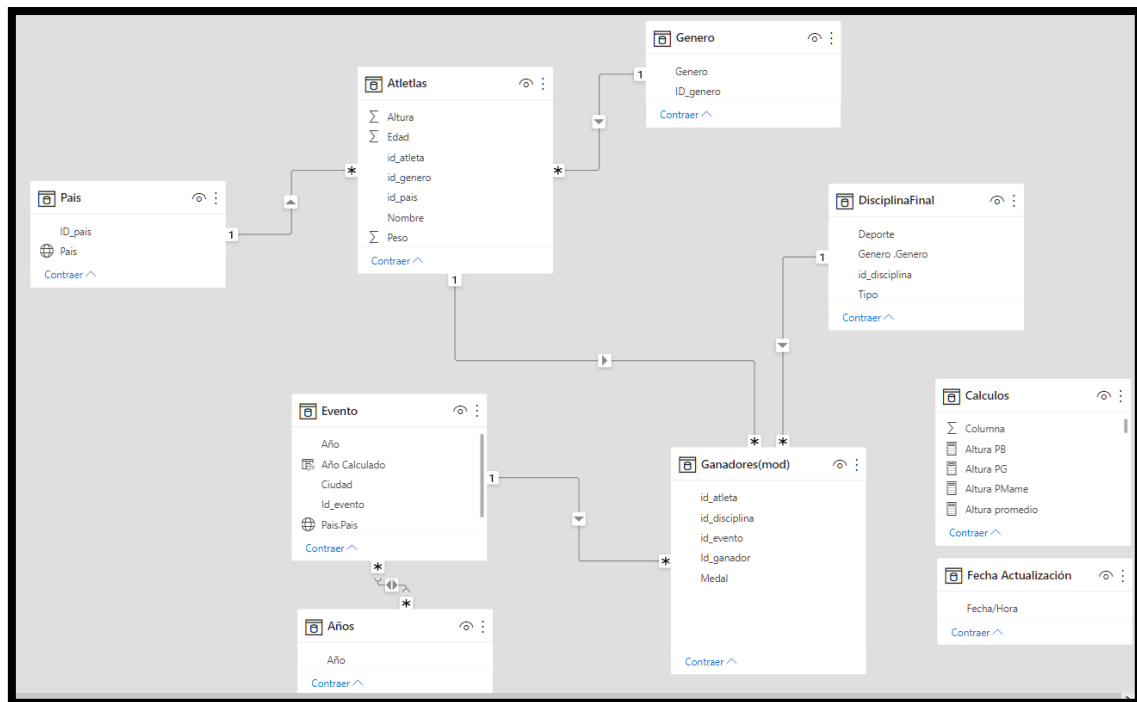
Cálculos:

- Se agregó una nueva tabla con todas las medidas realizadas para una visualización más adecuada.

Años:

- Con la información de la fecha obtenida en “Evento” se procedió a crear una tabla calendario que nos servirá para la segmentación de datos de algunas paginas siguientes.

Todo esto se realizó para que el diagrama entidad relación quedara de la siguiente manera en Power BI:



Segmentaciones elegidas

En el presente trabajo, se utilizaron diversas segmentaciones, dependiendo la página:

1. Desplegables: Se utilizaron en 3 páginas para filtrar por deporte y por tipo de competición en las páginas Altura, Edad y Peso, para mostrar las características de los participantes;

2. Barras Deslizantes: Esta barra fue utilizada únicamente para filtrar las ediciones de los Juegos Olímpicos por año;

Medidas Calculadas

Estas fueron las medidas que creamos para mejorar la visualización de los datos.

1. # Medallas
2. 1-altura_ganador
3. Altura PMame
4. Altura promedio
5. Altura_F
6. Altura_ganadores_promedio



7. Altura_M
8. Altura_no_ganadores_promedio
9. AlturaF_ganadores_promedio
10. AlturaF_P_Ganadores
11. AlturaM_ganadores_promedio
12. AlturaM_P_Ganadores
13. Atletas con Bronce
14. Atletas con Oro
15. Atletas con Plata
16. Bronce
17. Bronze_Summer
18. Bronze_Winter
19. Edad promedio
20. Edad_F
21. Edad_iconos
22. Edad_M
23. EdadF_ganadores_promedio
24. EdadF_P_Ganadores
25. EdadM_ganadores_promedio
26. EdadM_P_Ganadores
27. evento_invierno
28. Gold_Summer
29. Gold_Winter
30. Máxima Medalla*
31. Oro
32. Peso promedio
33. Peso_F
34. Peso_M
35. Peso_promedio_iconos
36. PesoF_ganadores_promedio
37. PesoF_P_Ganadores
38. PesoM_ganadores_promedio
39. PesoM_P_Ganadores
40. Plata
41. Silver_Summer
42. Silver_winter
43. Suma_medallas_summer
44. Suma_medallas_winter
45. Tiene Bronce
46. Tiene Medalla
47. Tiene Oro
48. Tiene Plata
49. Total AtletaO
50. Total Atletas
51. Total Ciudades
52. Total Deportes
53. Total Disciplinas
54. Total Eventos
55. Total_invierno





56. Total_medallas_sin_na*

57. Total_Verano

58. Fecha Actualización

Detalles:

- Se crearon medidas para hacer énfasis en la altura promedio diferentes formas;
 - Para los atletas se hicieron cálculos para encontrar cuál es la media de la altura de los atletas ganadores por género y en qué porcentaje están los atletas por arriba de la media por género.
 - `AlturaM_P_Ganadores = CALCULATE(AVERAGE(Atletlas[Altura]),Atletlas[medalla]>0,Atletlas[id_genero]="M")`
 - `AlturaF_P_Ganadores = CALCULATE(AVERAGE(Atletlas[Altura]),Atletlas[medalla]>0,Atletlas[id_genero]="F")`
 - `AlturaM_ganadores_promedio = ((CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[id_genero]="M",Atletlas[medalla]>0,Atletlas[Altura]>179.44))/CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[medalla]>0,Atletlas[id_genero]="M"))`
 - `AlturaF_ganadores_promedio = ((CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[id_genero]="F",Atletlas[medalla]>0,Atletlas[Altura]>168.93))/CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[medalla]>0,Atletlas[id_genero]="F"))`
 - Después se calculó el % de la altura de los medallistas que sea mayor a la media con "Altura_ganadores_promedio" y el % de los que atletas cuya altura sea menor a la media.
 - `Altura_ganadores_promedio = ((CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[medalla]>1,Atletlas[Altura]>176.32))/CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[medalla]>1))`
 - `1-altura_ganador = 1-Calculos[Altura_ganadores_promedio]`
 - `Altura_no_ganadores_promedio = CALCULATE(1-Calculos[Altura_ganadores_promedio])`

Promedio Ganadores	Promedio Ganadoras
181.12	170.62
Ganadores por Encima Promedio	Ganadoras por Encima Promedio
48.19 %	58.40 %

- Se calcularon los promedios de la altura de todos los atletas participantes con "Altura promedio" y con "Altura PMame" se calculó a los atletas que estaban por encima "▲" o por debajo de la media de la altura "▼".
 - `Altura promedio = AVERAGE(Atletlas[Altura])`





- Altura PMame = `if([Altura promedio]>176.32,"▲","▼")`

Pais	Promedio ...	Altura
Afghanistan	170.13	▼
Albania	173.74	▼
Algeria	175.25	▼
American Samoa	175.71	▼
Andorra	173.54	▼
Angola	178.73	▲
Antigua and Barbuda	175.30	▼
Argentina	177.05	▲
Armenia	172.05	▼
Aruba	174.52	▼
Australasia	177.53	▲
Australia	177.76	▲
Austria	177.39	▲
Azerbaijan	173.93	▼
Bahamas	178.68	▲
Bahrain	172.04	▼
Bangladesh	166.26	▼
Barbados	176.21	▼
Belarus	176.04	▼
Belgium	176.97	▲
Belize	174.34	▼
Benin	173.86	▼
Bermuda	176.82	▲
Bhutan	166.12	▼
Total	176.32	▼

- Se calculó la Altura por Género donde "Altura_M" es la altura promedio del género masculino y "Altura_F" del Femenino, con el fin de dar a conocer las diferencias entre la estatura por género.

- Altura_M = `CALCULATE(AVERAGE(Atletlas[Altura]),Atletlas[id_genero]="M")`
- Altura_F = `CALCULATE(AVERAGE(Atletlas[Altura]),Atletlas[id_genero]="F")`



- Se crearon las medidas para mostrar visualmente el total de ciertas características que describen a los Juegos Olímpicos.
 - Se calcularon las medidas para contar el total de atletas, disciplinas, deportes, ediciones y ciudades que han participado en toda la historia de los Juegos Olímpicos desde Atenas 1896.
- Total Atleta0 = `DISTINCTCOUNT('Atletlas'[id_atleta])`
 - Total Disciplinas = `DISTINCTCOUNT(DisciplinaFinal[Tipo])`
 - Total Deportes = `DISTINCTCOUNT(DisciplinaFinal[Deporte])`
 - Total Eventos = `DISTINCTCOUNT(Evento[Id_evento])`
 - Total Ciudades = `DISTINCTCOUNT(Evento[Ciudad])`





Atletas Olímpicos	135571
Disciplinas Olímpicas	510
Deportes Olímpicos	66
Ediciones Olímpicas	51
Ciudades Olímpicas	42

- Se calculó el total de medallas que se han entregado en todas las ediciones de los Juegos Olímpicos.

- Oro = `CALCULATE(Calculos[Total Medallas], 'Ganadores(mod)'[Medal] = "Gold")`
- Plata = `CALCULATE(Calculos[Total Medallas], 'Ganadores(mod)'[Medal] = "Silver")`
- Bronce = `CALCULATE(Calculos[Total Medallas], 'Ganadores(mod)'[Medal] = "Bronze")`

- Se calculó el total de Atletas que consiguieron medallas en todas las ediciones de los Juegos Olímpicos.

- Atletas con Oro = `CALCULATE([Total Atletas], 'Ganadores(mod)'[Medal] = "Oro")`
- Atletas con Plata = `CALCULATE([Total Atletas], 'Ganadores(mod)'[Medal] = "Silver")`
- Atletas con Bronce = `CALCULATE([Total Atletas], 'Ganadores(mod)'[Medal] = "Bronze")`

Atletas con Medallas
10425
11441
11879

- Se crearon las medidas para hacer énfasis en la edad promedio de los atletas diferentes formas;

- Se calcularon medidas de la misma forma que altura, pero para las especificaciones de la edad de los atletas.

- Edad promedio = `AVERAGE(Atletas[Edad])`

Promedio de Edad
24





- `Edad_F = CALCULATE(AVERAGE(Atletlas[Edad]), 'Genero '[ID_genero]="F")`
- `Edad_M = CALCULATE(AVERAGE(Atletlas[Edad]), 'Genero '[ID_genero]="M")`

Edad Promedio Varonil	Edad Promedio Femenil
24.98	22.94

- `Edad_icons = if([Edad promedio]>24.46,"▲","▼")`

Promedio Ganadores	Promedio Ganadoras
24.67	22.62
Ganadores por Encima Promedio	Ganadoras por Encima Promedio
35.90 %	41.24 %

- De la misma manera se hicieron cálculos para encontrar cuál es la media de la edad de los atletas ganadores por género y en qué porcentaje están los atletas por arriba de la media por género.

- `EdadM_P_Ganadores = CALCULATE(AVERAGE(Atletlas[Edad]),Atletlas[medalla]>0,Atletlas[id_genero]="M")`
- `EdadF_P_Ganadores = CALCULATE(AVERAGE(Atletlas[Edad]),Atletlas[medalla]>0,Atletlas[id_genero]="F")`
- `EdadM_ganadores_promedio = ((CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[id_genero]="M",Atletlas[medalla]>0,Atletlas[Edad]>24.98))/CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[medalla]>0,Atletlas[id_genero]="M"))`
- `EdadF_ganadores_promedio = ((CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[id_genero]="F",Atletlas[medalla]>0,Atletlas[Edad]>22.94))/CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[medalla]>0,Atletlas[id_genero]="F"))`

Pais	Promedio ...	Edad
Zimbabwe	25.74	▲
Zambia	23.09	▼
Yugoslavia	23.64	▼
Yemen	21.14	▼
West Germany	23.74	▼
Vietnam	23.34	▼

- Se crearon las medidas hacer énfasis en el peso promedio de los atletas diferentes formas;

- Se calcularon medidas de la misma forma que altura, pero para las especificaciones del peso de los atletas.

- `Peso promedio = AVERAGE(Atletlas[Peso])`

Promedio de Peso
71.96

- `Peso_F = CALCULATE(AVERAGE(Atletlas[Peso]), 'Genero '[ID_genero]="F")`
- `Peso_M = CALCULATE(AVERAGE(Atletlas[Peso]), 'Genero '[ID_genero]="M")`





Peso Promedio Varonil	Peso Promedio Femenil
76.47	61.28

- `Peso_promedio_iconos = if([Peso promedio]>71.96,"▲","▼")`

Pais	Promedio ...	Peso
Afghanistan	66.38	▼
Albania	73.62	▲
Algeria	68.97	▼
American Samoa	85.76	▲
Andorra	69.28	▼
Anqola	74.62	▲

- Para los atletas se hicieron cálculos para encontrar cuál es la media de la altura de los atletas ganadores por género y en qué porcentaje están los atletas por arriba de la media por género.

- `PesoM_P_Ganadores = CALCULATE(AVERAGE(Atletlas[Peso]),Atletlas[medalla]>0,Atletlas[id_genero]="M")`
- `PesoF_P_Ganadores = CALCULATE(AVERAGE(Atletlas[Peso]),Atletlas[medalla]>0,Atletlas[id_genero]="F")`
- `PesoM_ganadores_promedio = ((CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[id_genero]="M",Atletlas[medalla]>0,Atletlas[Peso]>76.47))/CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[medalla]>0,Atletlas[id_genero]="M"))`
- `PesoF_ganadores_promedio = ((CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[id_genero]="F",Atletlas[medalla]>0,Atletlas[Peso]>61.28))/CALCULATE(COUNT('Ganadores(mod)'[id_atleta]),Atletlas[medalla]>0,Atletlas[id_genero]="F"))`

Promedio Ganadores	Promedio Ganadoras
79.18	63.50
Ganadores por Encima Promedio	Ganadoras por Encima Promedio
42.99 %	55.36 %

- Se crearon las medidas para visualizar por sedes lo que ha sucedido con las medallas.
 - Para la temporada de verano se creó el total de atletas que han ganado medallas y el medallero total de todas las ediciones.
 - `Suma_medallas_summer = [Silver_Summer] + 'Calculos'[Gold_Summer] + [Bronze_Summer]`
 - `Gold_Summer = CALCULATE(COUNT('Ganadores(mod)'[Id_ganador]),'Ganadores(mod)'[Medal]="Gold",Evento[Temporada]="Summer")`
 - `Silver_Summer = CALCULATE(COUNT('Ganadores(mod)'[Id_ganador]),'Ganadores(mod)'[Medal]="Silver",Evento[Temporada]="Summer")`
 - `Bronze_Summer = CALCULATE(COUNT('Ganadores(mod)'[Id_ganador]),'Ganadores(mod)'[Medal]="Bronze",Evento[Temporada]="Summer")`





Pais	Oro	Plata	Bronce	# Medallas
United States	4999	4999	4999	4999
Soviet Union	2189	2189	2189	2189
Great Britain	1992	1992	1992	1992
Germany	1702	1702	1702	1702
France	1623	1623	1623	1623
Italy	1436	1436	1436	1436

- Para la temporada de invierno se creó lo mismo que verano.
 - `Suma_medallas_winter =`
`[Silver_winter]+[Bronze_winter]+[gold_winter]`
 - `gold_winter =`
`CALCULATE(COUNT('Ganadores(mod)'[Id_ganador]), 'Ganadores(mod)'[Medal]="Gold", Evento[Temporada]="Winter")`
 - `Silver_winter =`
`CALCULATE(COUNT('Ganadores(mod)'[Id_ganador]), 'Ganadores(mod)'[Medal]="Silver", Evento[Temporada]="Winter")`
 - `Bronze_winter =`
`CALCULATE(COUNT('Ganadores(mod)'[Id_ganador]), 'Ganadores(mod)'[Medal]="Bronze", Evento[Temporada]="Winter")`

Medallero Olímpico (Invierno)

Pais	Oro	Plata	Bronce	# Medallas
United States	634	634	634	634
Canada	610	610	610	610
Soviet Union	471	471	471	471
Norway	443	443	443	443
Sweden	428	428	428	428
Finland	426	426	426	426
Germany	347	347	347	347
Austria	275	275	275	275
Switzerland	274	274	274	274
Italy	191	191	191	191

- Se crearon las medidas para visualizar la cantidad de medallas que ha conseguido cada atleta y/o la máxima medalla que consiguió en todas sus participaciones.
 - Primero se calcularon medidas para saber que medallas posee cada atleta
 - `Tiene Oro = CALCULATE([Total Medallas], 'Ganadores(mod)'[Medal]="Gold")`
 - `Tiene Plata = CALCULATE([Total Medallas], 'Ganadores(mod)'[Medal]="Silver")`
 - `Tiene Bronce = CALCULATE([Total Medallas], 'Ganadores(mod)'[Medal]="Bronze")`



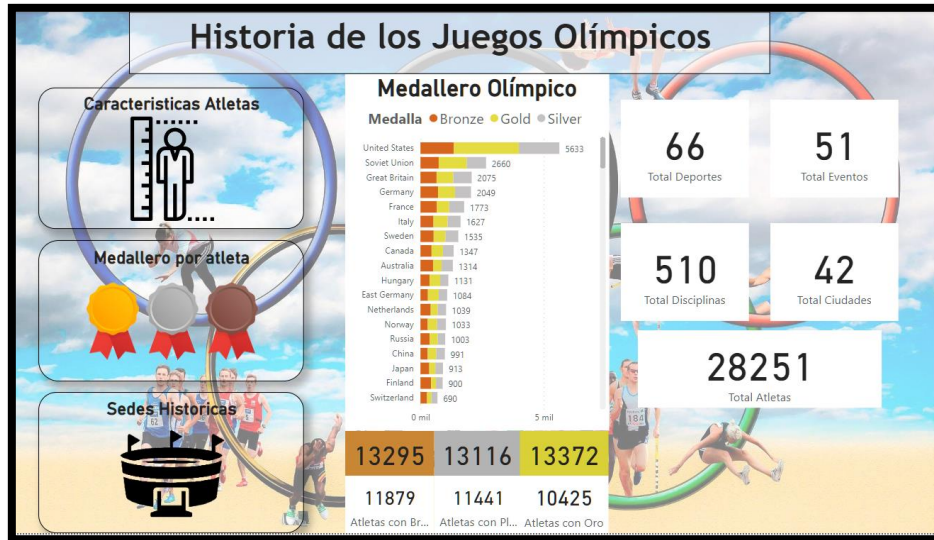


Visualización de los datos:

La visualización se llevó a cabo mediante la aplicación Power Bi desarrollado por Microsoft.

Página Portada

Esta página fue diseñada con el propósito de tener un acceso rápido a cada una de las siguientes tablas. En un comienzo se planeó que esta página tuviera cierta información que finalmente se terminaron llevando a otras página.



Primera versión con datos que dieron lugar a la página “medallero”

Para la versión final se quitó la mayoría de los datos para hacer una versión más simplificada con botones. Esta página se creó a partir de una imagen del logo de los juegos olímpicos y a partir de esta se crearon óvalos con una forma similar y con sus colores correspondientes que sirvan como botones para acceder a cada tabla. Las imágenes usadas se tuvieron que quitar el fondo con programas externos para poder insertarle un fondo de un color del aro en cual está ubicado.

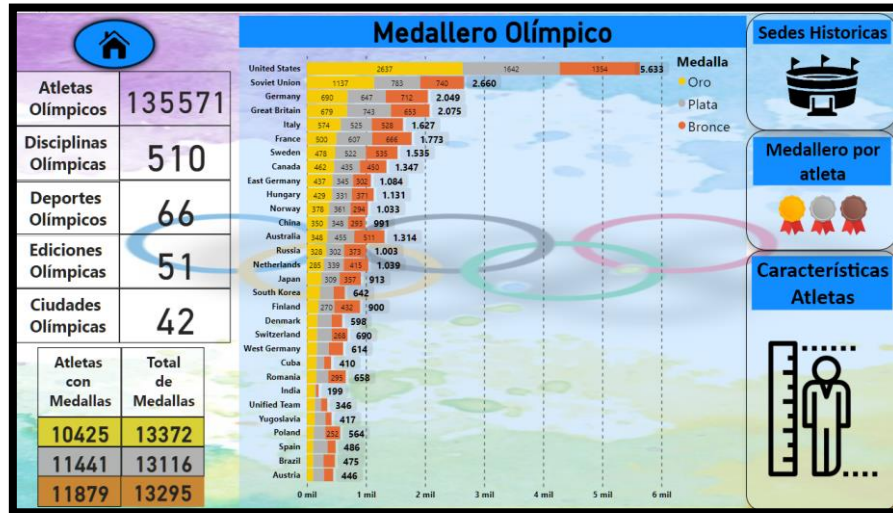


Versión Final

Por último, se agregó la información de la última actualización que recibió la visualización y por quienes fue creado.

Pagina Medallero

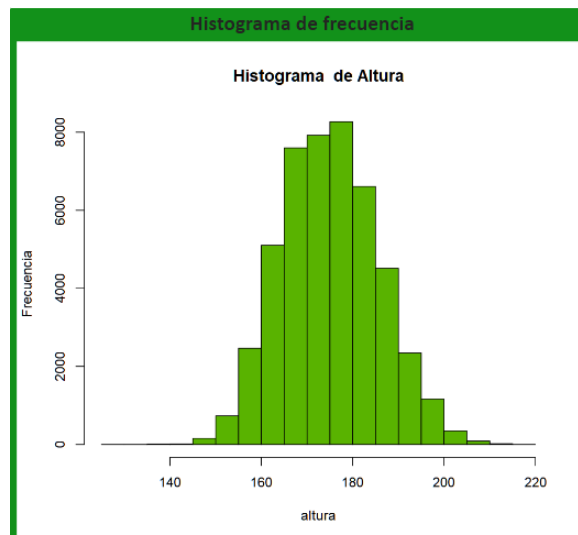
Esta página contiene el medallero olímpico formado por un top 30 de los países participantes ordenados por la mayor cantidad de medallas de oro. Esta página fue creada a partir de la anterior versión de la página “portada”.



En la esquina inferior izquierda se muestra la cantidad de medallas repartidas en todos los eventos dependiendo de su valor, junto con los atletas únicos que consiguieron una medalla. En el sector derecho se colocaron 3 accesos directos a otras páginas con iconos que sirvan para identificarlos. Por último, a la izquierda del medallero tenemos información variada, y un botón home para volver a la página de inicio.

Pagina Atletas, Edad y Peso

En estas 3 páginas, la visualización principal está dada por un gráfico histograma de frecuencias, se eligió este tipo de grafico ya que es el más visual a la hora de mostrar la mayor concentración de frecuencias, además de mostrarnos la moda (el valor que más se repite) de los datos.



Ya que Power Bi no cuenta con un gráfico que nos permite mostrar la distribución en forma de campana, se recurrió al lenguaje R, ya que este viene con gran cantidad de gráficos de distribución que nos servirán para este caso.



Lo primero que se tuvo que hacer fue cargar los datos a través de un data set, este último se crea gracias a los datos de la columna “peso/edad/altura” de la tabla “Atletas”. Una vez cargado el dataset correspondiente, la función utilizada fue “HIST”:

```
alt=dataset$Peso[!is.na(dataset$Peso)]  
  
den=density(alt)  
  
as=hist(alt , prob=FALSE ,las=2, main = "Histograma del Peso",ylab =  
"Frecuencia", col="#CDAD00",xlab="Peso")
```

Los argumentos de la función HIST son los siguientes:

- Alt: Este es el nombre que se le dio al dataset para poder incluirla en la función.
- Prob=FALSE: Sirve si deseamos incluir la frecuencia relativa a través de porcentaje o simplemente mostrar el total de dato.
- Las=2: Es para rotar la información del eje Y.
- Main: Sirve para colocar el título al histograma.
- Ylab: Nombre al eje Y.
- Xlab: Nombre al eje X.
- Col: Código del color que se desea poner para el histograma.

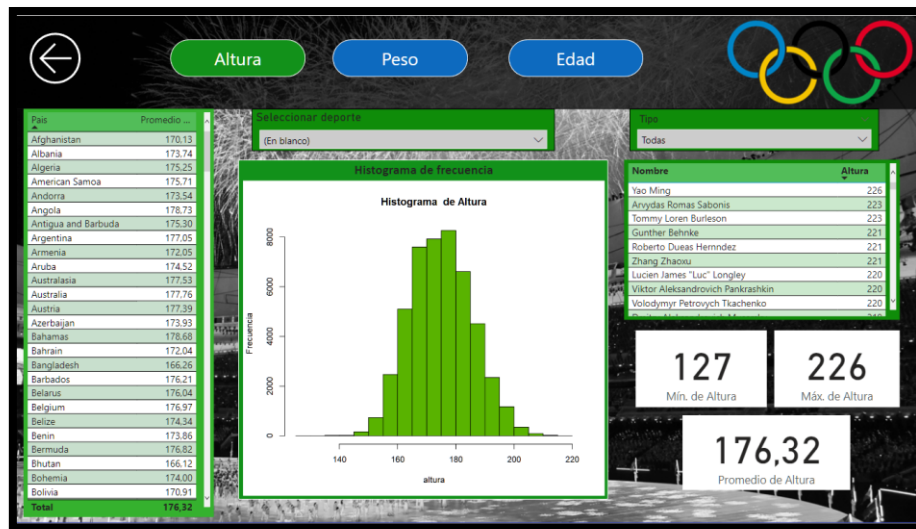
Una vez finalizado el histograma, la información más importante para mostrar fue el promedio en general y los valores máximos y mínimos. En la tabla izquierda aparece el promedio de todos los atletas por su nacionalidad, además de agregarle un icono para indicar visualmente si el país estaba por encima o por debajo del promedio general.

En el lado derecho se hizo una distinción por género, agregando el promedio de los atletas de cada género, el promedio de los que obtuvieron 1 medalla o más para cada género y el porcentaje de atletas que ganaron 1 medalla o más pero que superaban el promedio de su género.

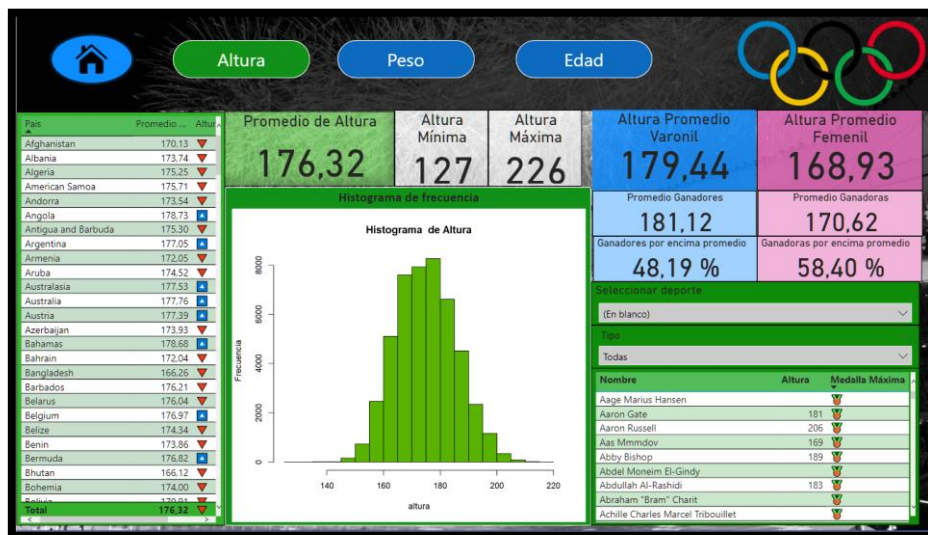
Por último, se agregó una tabla para mostrar los participantes, para poder ver más claramente quienes integran el top de los más altos, más ancianos y pesados, además de tener una columna que nos muestra con un icono cual fue la medalla de más valor que consiguió.

La versión inicial del mismo se visualizaba de la siguiente manera:





Para la versión final se modificó la ubicación de la mayoría de los datos y la implementación de más información, principalmente los datos de genero e iconos para mejorar la visualización:



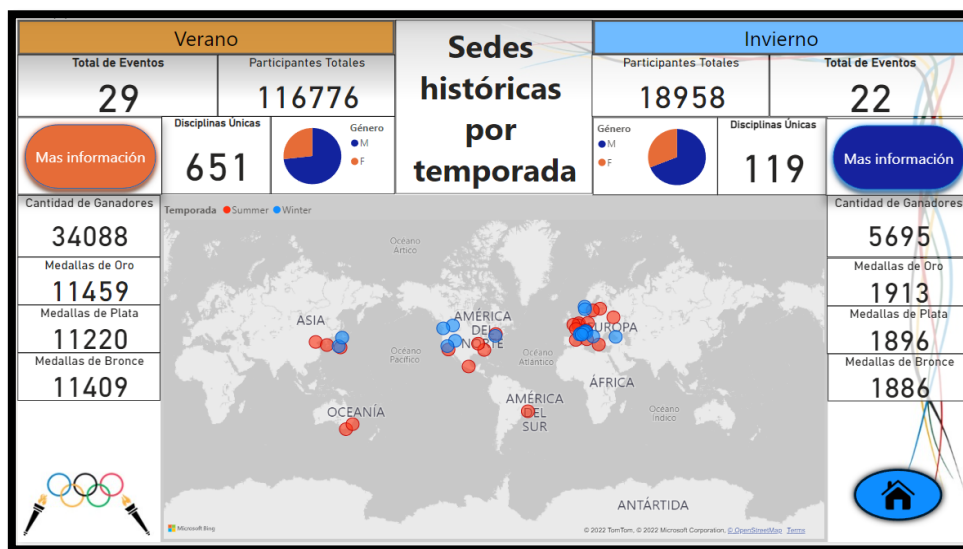
Página sedes

Esta es una de las páginas que más información contiene. El principal objetivo es mostrar todas las sedes olímpicas de la historia de los Juegos Olímpicos haciendo una distinción entre las competencias de invierno y de verano. En un comienzo tenía muy poca información, siendo simplemente una página de inicio para ir a las otras dos siguientes:



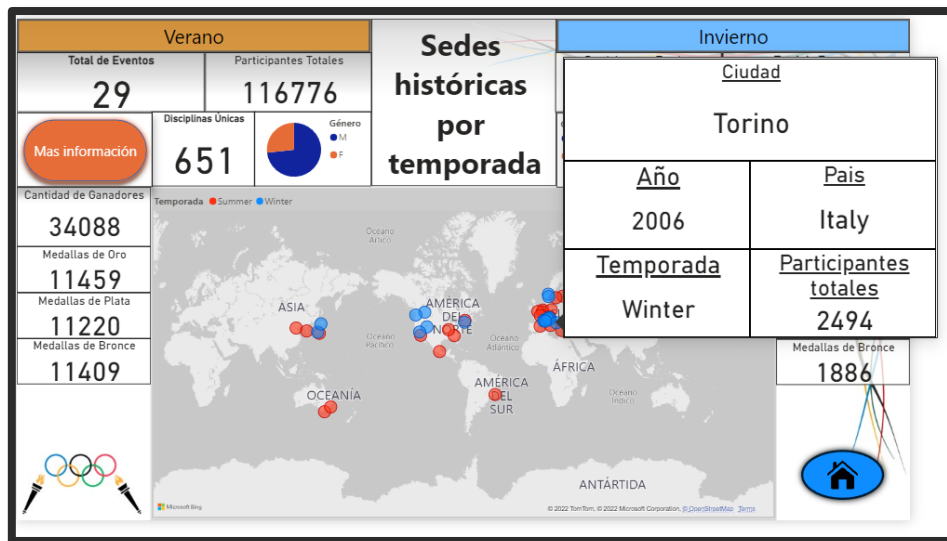


Para la versión final, el principal dato para mostrar fue la cantidad total de eventos que tuvieron, los participantes totales, con otra distinción por género, las disciplinas únicas y la cantidad de medallas repartidas.



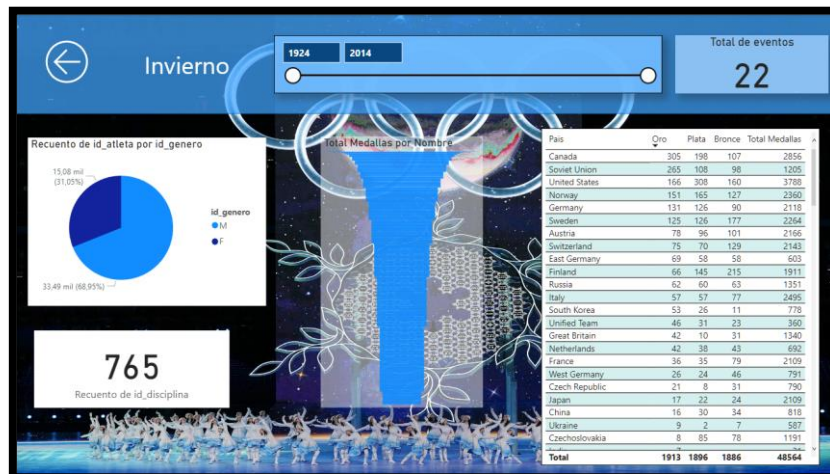
Lo que se lleva la principal atención es el grafico ubicando en el centro de la página, este es un mapa del mundo completo donde en cada ciudad donde alguna vez se realizó una edición de los juegos olímpicos aparece un círculo, con un color diferente diferenciando si fue una sede que alojo los JJOO de invierno o de verano. Este círculo a su vez tiene integrado un ToolTip para dar mejor información sobre la sede, agregando los atletas que participaron, el nombre de la ciudad, el año, entre otros.

Por último, se agregaron varios botones que llevan a una página específica con más información sobre las ediciones de invierno o de verano, y otro que nos conduce a la página principal (portada). Así como un tooltip a la hora de seleccionar alguna sede en el mapa para visualizar las características de esta.



Pagina Invierno/Verano

Estas dos páginas surgen de la anterior y sirven para ampliar la información que no se pudo enseñar en la página “sedes”. En un primer momento no podíamos encontrar un gráfico que de la mayor información:



Ya para la versión final se le dio más importancia al gráfico y la lista total de eventos:





En estas páginas se agregó un medallero y segmentaciones únicas, con colores distintos que represente cada edición. La información más destacada es el gráfico de total de participaciones por cada género, este nos proporciona bastante información, como el incremento exponencial de participantes para cada edición más moderna, la mayor participación femenina que para los últimos años suben de manera paralela, el descenso en eventos claves ya sea por motivos bélicos o políticos y la nula participación femenina del comienzo de los JJOO.

Para finalizar tenemos un gráfico de torta para mostrar más claramente la participación total de cada género, y una lista de todos los eventos ordenado por ciudades y años, mostrando el total de atletas y el total de medallas repartidas.

Página atletas

Como última página de la presentación se decidió agregar una tabla que nos enseñe la totalidad de los atletas participantes, además de un buscador para facilitar la tarea. Con esto se puede analizar al detalle cada uno de los atletas que alguna vez participaron en un juego olímpico, al comienzo era simplemente una lista de los participantes con un buscado por nombre:





Buscador por nombre		
<input type="text" value="Messi"/>		
Nombre	Total Medallas	Pais
Diogo Almeida Messias Ganchinho	2	Portugal
Francesco Maria Messina	2	Italy
Albert John "Bertie" Messitt	1	Ireland
Alessandro Messina	1	Canada
Faud Mauricio Kattan Shemmessian	1	Bolivia
Guido Messina	1	Italy
Lionel Andrs Messi Cucittini	1	Argentina
Messias Jos Baptista	1	Brazil
S. Messina	1	Italy
Sheppard Norman "Shep" Messing	1	United States
Silvia M. Morel Messina	1	Chile
Total	13	

Ya para la versión final, para segmentar los datos se agregaron distintos tipos de filtros, tales como buscador por deportes, por año de edición y por nacionalidad, un fondo y una mejor visualización de la información por atleta, además de agregar iconos que faciliten la apreciación.

Edición

2016

Buscador por nombre

Search

Pais

Todas

Seleccionar deporte

☐ (En blanco)

☐ Aeronautics

☐ Alpine Skiing

☐ Alpinism

☐ Archery

☐ Au Chapelet, 33 metres

☐ Au Chapelet, 50 metres

☐ Au Cordon Dore, 33 metres

☐ Au Cordon Dore, 50 metres

☐ Championnat Du Monde

☐ Continental Style

☐ Double American Round

☐ Double Columbia Round

☐ Double National Round

☐ Double York Round

☐ Individual

☐ Pole Archery, Large Birds, Individual

☐ Pole Archery, Large Birds, Team

☐ Pole Archery, Small Birds, Individual

☐ Pole Archery, Small Birds, Team

☐ Sir La Perche a La Harce

Nombre	id_pais	Edad	Total de eventos	Disciplinas participante	Tiene Oro	Tiene Plata	Tiene Bronce	# Medallas	Total medallas	Máxima Medalla
Michael Fred Phelps, II	USA	15	1	6	5	1		6	6	...
Kathleen Genevieve "Katie" Ledecy	USA	15	1	5	4	1		5	5	...
Simone Arianne Biles	USA	19	1	6	4		1	5	5	...
Emma McKeon	AUS	22	1	5	1	2	1	4	4	...
Katinka Hossz	HUN	15	1	5	3	1		4	4	...
Madeline Jane "Maya" DiRado	USA	23	1	4	2	1	1	4	4	...
Nathan Ghar-Jun Adrian	USA	19	1	4	2		2	4	4	...
Penelope "Penny" Oleksiak	CAN	16	1	5	1	1	2	4	4	...
Simone Ashley Manuel	USA	20	1	4	2	2		4	4	...
Alexandra Rose "Aly" Raisman	USA	18	1	5	1	2		3	3	...
Aliya Parkhatovna Mustafina	RUS	17	1	5	1	1	1	3	3	...
Allyson Michelle Felix	USA	18	1	3	2	1		3	3	...
Andre De Grasse	CAN	21	1	3		1	2	3	3	...
Dana Whitney Volmer (-Grant)	USA	16	1	3	1	1	1	3	3	...
Danuta Kosz	HUN	21	1	3	3			3	3	...
Denis Mikhaylovich Aplyazin	RUS	19	1	4		2	1	3	3	...
Elaine Thompson	JAM	24	1	3	2	1		3	3	...
Freitorish "Tori" Bowie	USA	25	1	3	1	1	1	3	3	...
Isaquias Queiroz dos Santos	BRA	22	1	3		2	1	3	3	...

Conclusión

En este trabajo facilitamos el acceso a las estadísticas de los Juegos Olímpicos, en primera instancia nos vamos a un aspecto general, mostrando la cantidad de medallas por país, el total de eventos disputados, total de medallas entregadas y total de atletas que consiguieron medallas, así como cuantas sedes y cuantas ediciones han presentado a los JO.

Logramos mostrar características importantes de los atletas para que cualquier atleta que quiera participar, además de las habilidades técnicas de su deporte, considere la altura, la edad, el peso para facilitar ganar una medalla y que medalla ganaría.

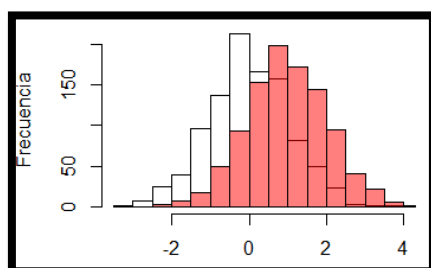


Armamos diferentes páginas para que la búsqueda pueda ser por diferentes vías y que con ello logren ver diferentes cosas pero que puedan apoyarse en la otra página.

Por último, logramos crear un buscador por atleta si es que se quiere buscar como le ha ido a cada atleta, cuantas medallas ha conseguido, de qué color, sus características físicas, con ello logramos que las personas que quieran buscar a los atletas puedan filtrar más sencillo la información.

Próximos pasos:

Para futuras versiones se planea modificar el gráfico de R que se encuentra en las páginas altura, edad y peso con dos gráficos superpuestos dando la información de frecuencia tanto para hombres como para los datos de las mujeres. Como boceto inicial tenemos la siguiente imagen:



donde tengamos dos colores diferentes para ambos datos.

También se planea agregar los datos de las ediciones más modernas, ya sea Tokio 2021 y Beijing 2022 una vez que se actualicen la base de dato iniciales.

Por otro lado, intentaremos visualizar si los atletas deben ser más altos, más grandes o pesados siendo influyente para conseguir una medalla.

Bibliografía

Chris McCrory (2022) *¿Qué representan los aros olímpicos?* 12 News Consultado en: <https://www.12news.com/article/syndication/spanish/qu-representan-los-aros-olmpicos-en-el-mundo/75-a565d225-3e68-4ff5-b6f6-c0537a8022e2#:~:text=Seg%C3%BAAn%20la%20Regla%208%20de,%2C%20Am%C3%A9rica%2C%20Europa%20y%20Ocean%C3%ADa.>

Herramientas utilizadas

- Diagrams para crear el modelo relacional.
- Microsoft Excel para la transformación de datos iniciales.
- Power Bi enfocado en la visualización de los datos.
-

