

# Los Juegos Olímpicos desde una perspectiva de género

Entendamos el papel de las mujeres en la historia del deporte



# Contenido

「01

**Contexto y  
pregunta  
disparadora**

「02

**Historia de los  
JJOO**

Análisis exploratorio,  
obtención de insights y  
primer acercamiento a un  
modelo predictivo

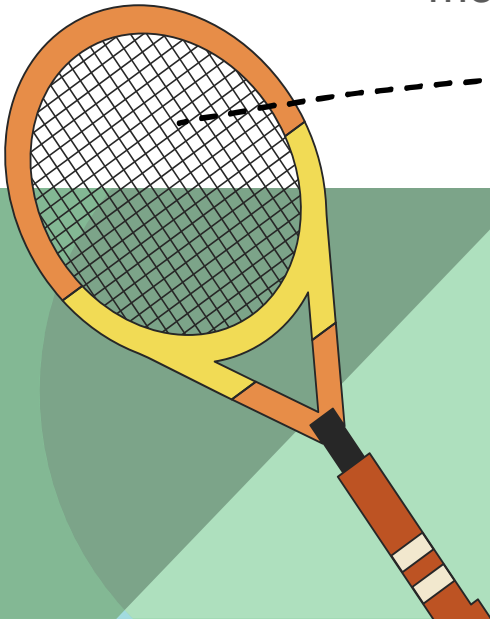
「03

**Contexto  
socio-económico**

Data Wrangling y  
obtención de insights  
Implementación y  
mejora de modelo  
predictivo

「04

**Conclusiones**



# 01 Contexto y preguntas disparadoras



## **Evolución del evento**

- Evolución de cantidad de medallas entregadas
- Incorporación de disciplinas
- Cantidad de países participantes



## **Papel de participantes femeninas**

- Participación de las mujeres al inicio de los JJOO y su crecimiento en la historia



## **Predicción de probabilidad de obtener una medalla**

- Análisis en función de la historia



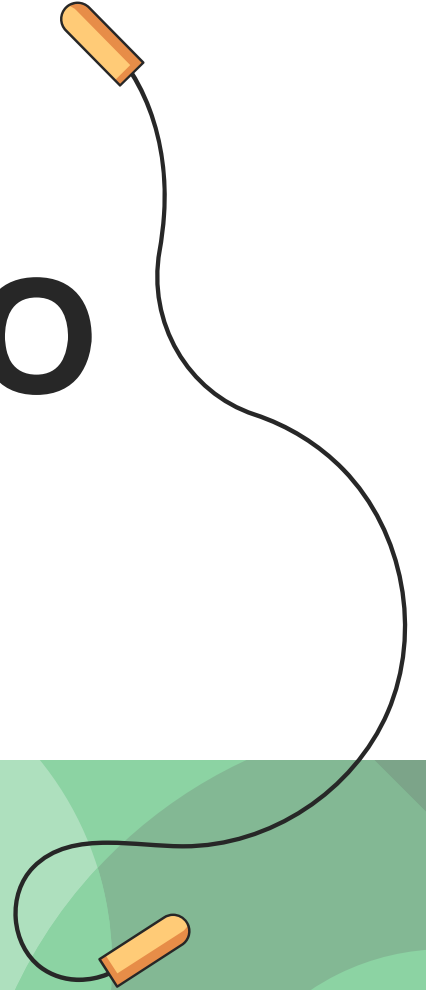
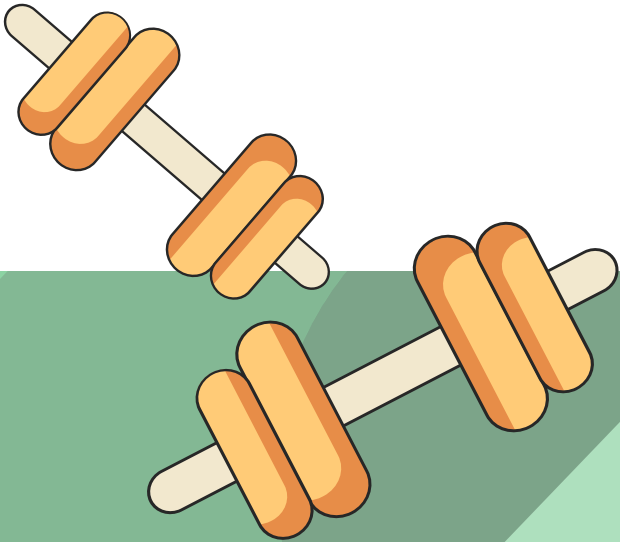
## **Condiciones socio-económicas**

- Situaciones que condicionan la vida y el desarrollo de las mujeres alrededor del mundo que pueden ser condicionantes a la hora de perseguir una carrera como deportistas de alto rendimiento

02

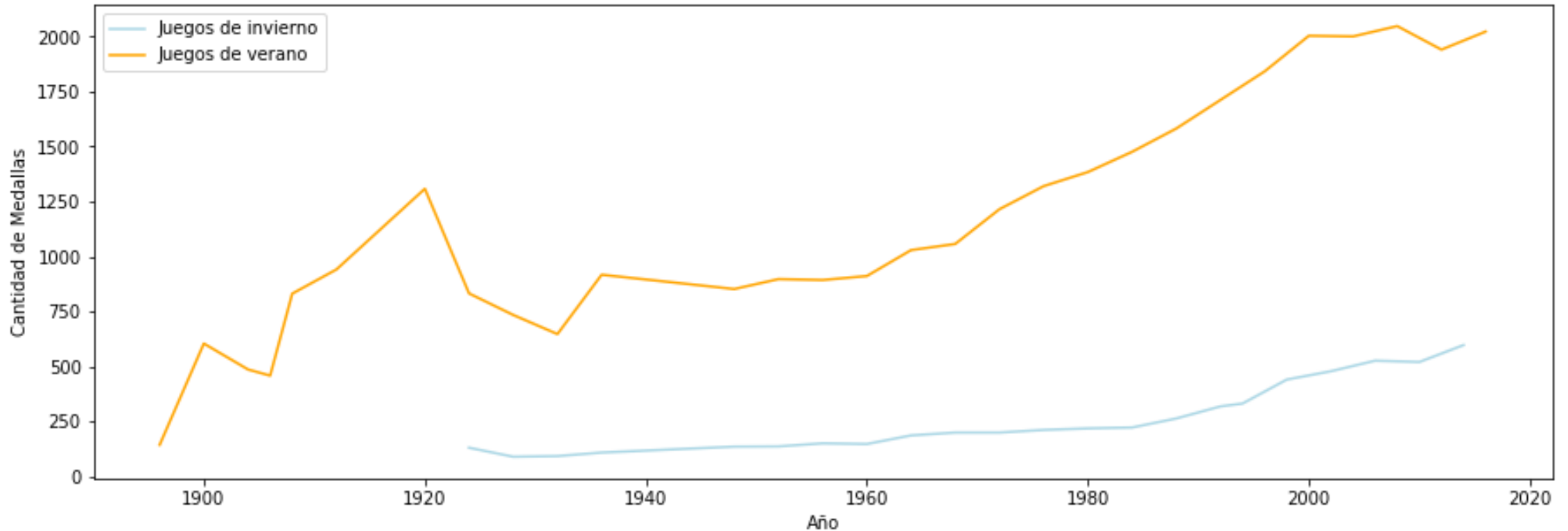
# Historia de los JJOO

Análisis Exploratorio y obtención de insights



# Evolución del evento

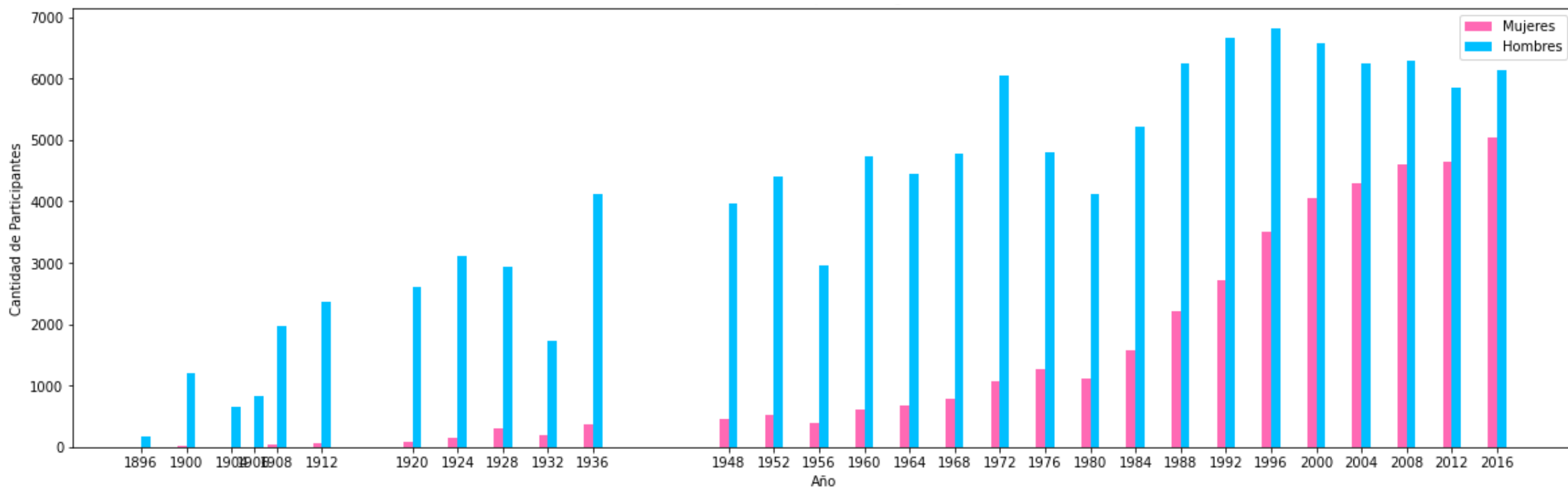
## Cantidad de medallas entregadas



Los juegos aumentaron en magnitud a lo largo de los años.  
En los juegos de verano se entrega una cantidad mucho mayor de medallas,  
por lo tanto, mayor cantidad de disciplinas y deportes.

# Evolución del evento

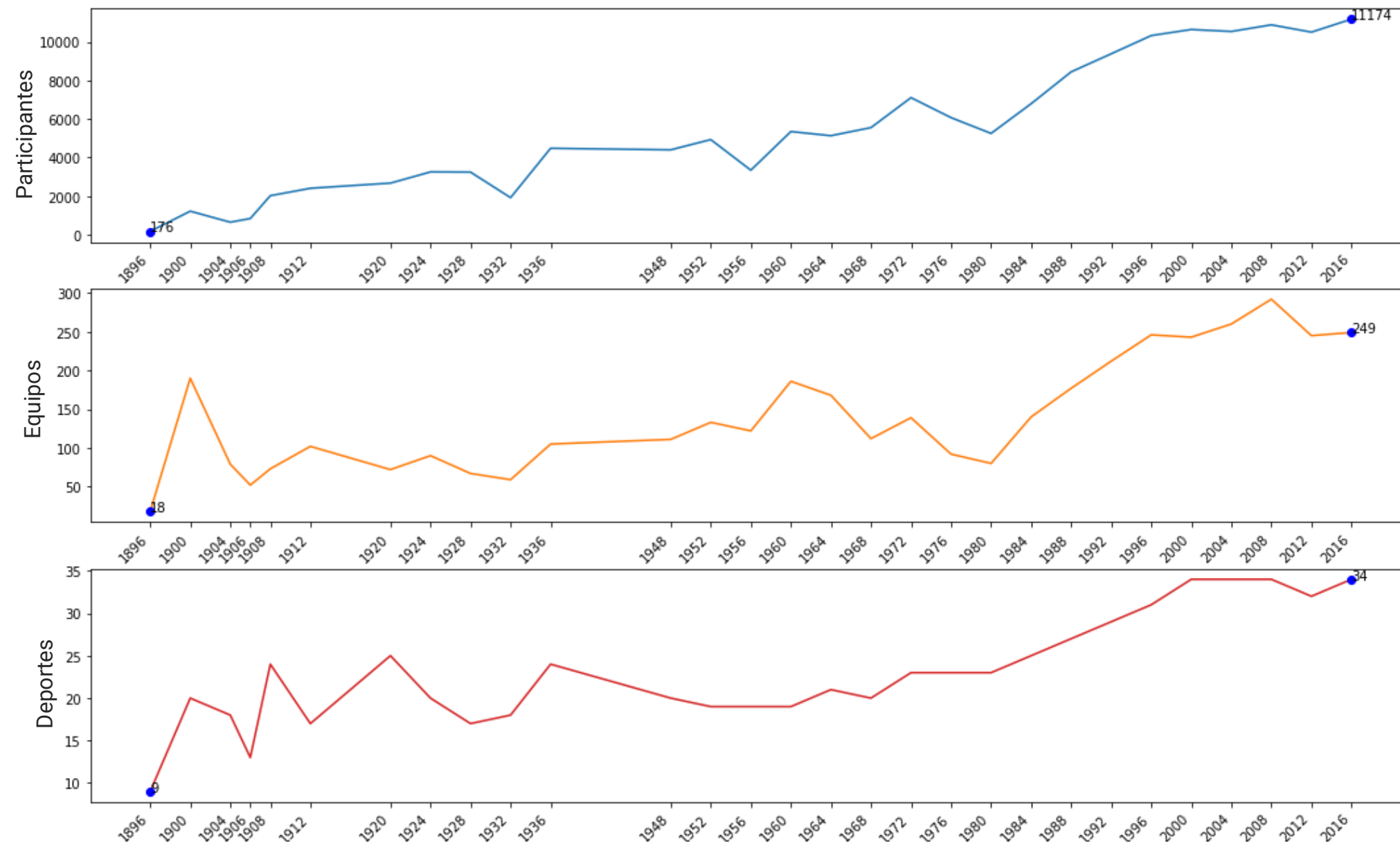
## Cantidad de participantes según género



Desde el comienzo de la historia de los JJOO hubo participación de mujeres, pero en muy baja proporción. Por suerte, esta realidad comenzó a cambiar a partir de los años 90 y la participación de mujeres muestra una clara tendencia a continuar aumentando.

# Evolución del evento

Se observa que la magnitud del evento creció a lo largo de la historia, amesetándose a partir del año 2000. Si se compara el primer registro de 1896 contra el 2016, se concluye que las cantidades de Participantes, Deportes, Eventos y Equipos aumentaron considerablemente.





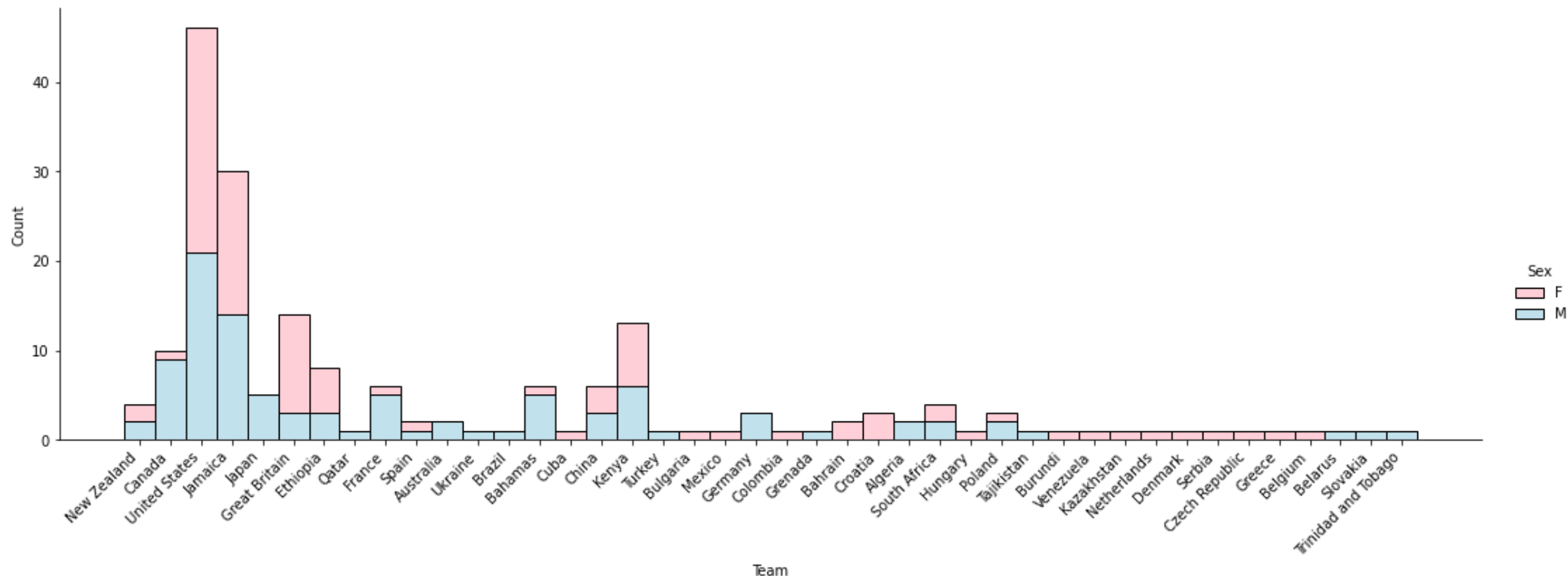
# Analicemos el atletismo femenino

Y de paso, derribemos algunos mitos



# Atletismo femenino en 2016

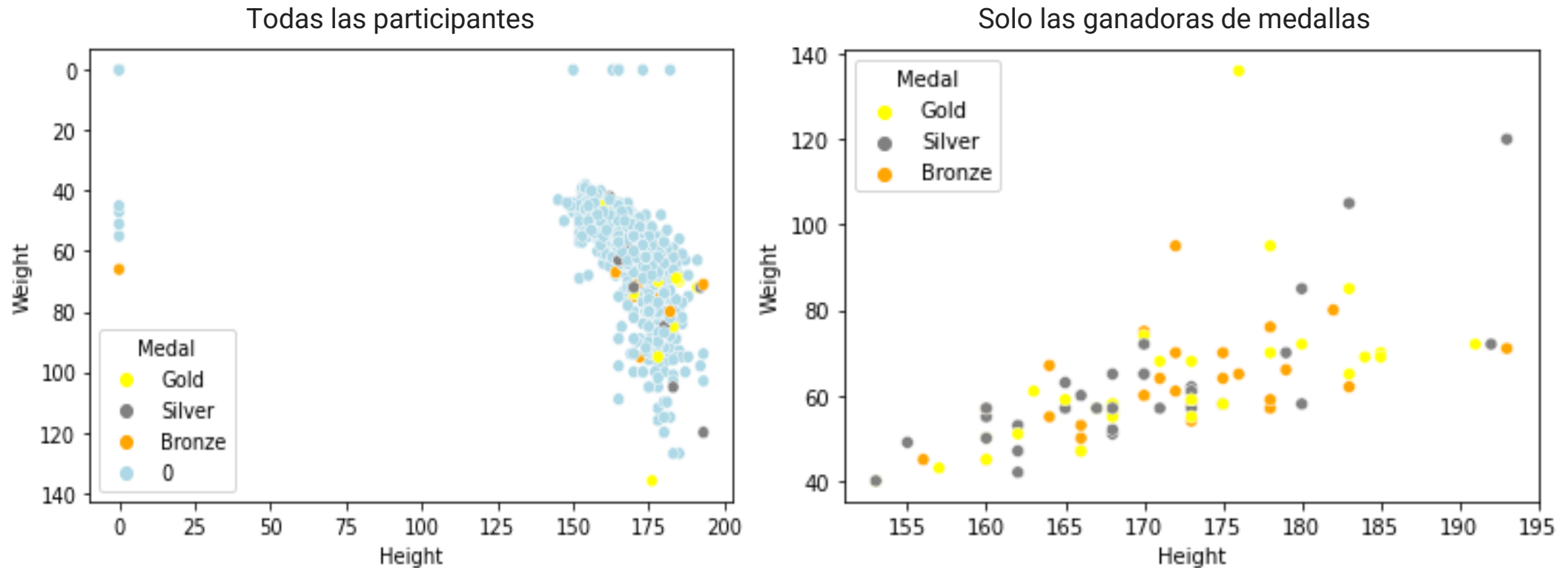
Cantidad de participantes por equipo que obtuvieron medallas en Atletismo en 2016 - Dividido por género



Los países con mayor cantidad de medallas fueron Estados Unidos, Jamaica, Gran Bretaña, Kenya y Canadá

# Atletismo femenino en 2016

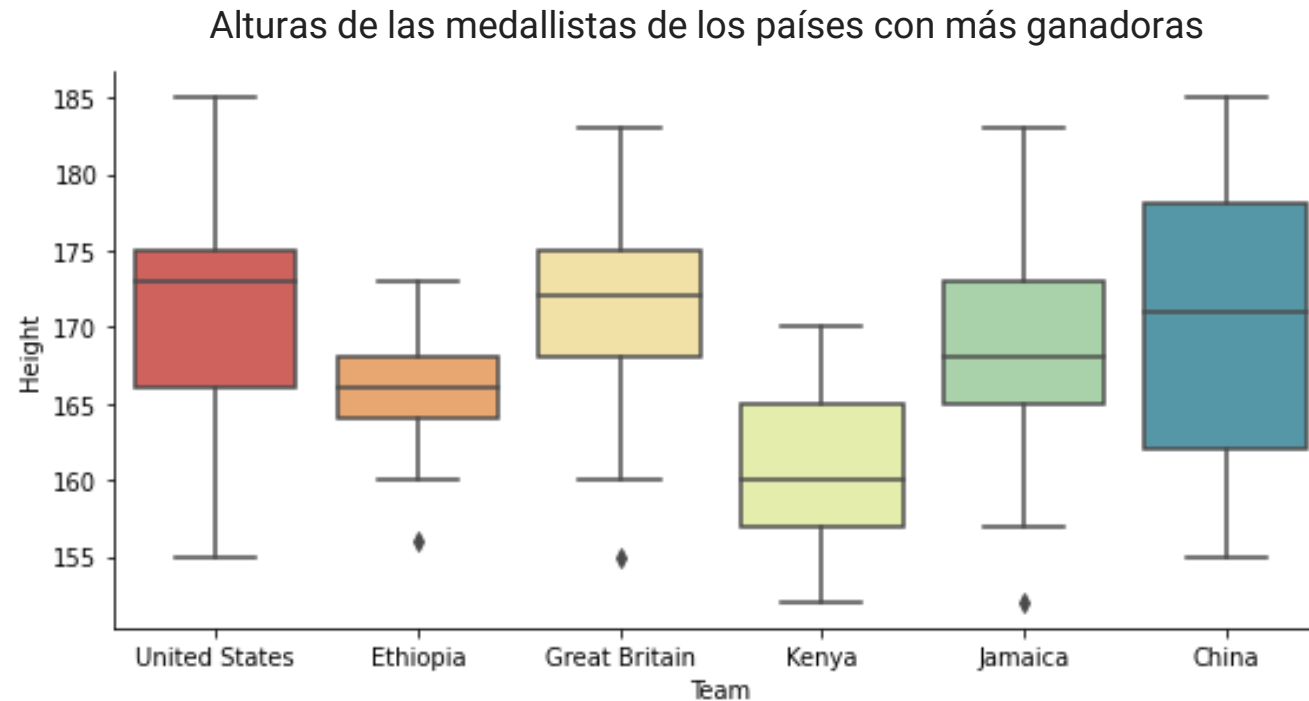
**MITO:** Existe una relación entre el índice de masa corporal (ratio entre su altura y su peso) y la capacidad de destacarse en atletismo



**FALSO!** Las ganadoras de los podios presentan un rango muy variado de alturas y pesos.

# Atletismo femenino en 2016

**MITO:** Como la altura de las deportistas representa una ventaja para ellas, los países más exitosos tienen distribución de alturas similar



**FALSO!** Vemos distribuciones de alturas muy diversas e incluso algunas medallistas muy alejadas del promedio de su país

**Modelo predictivo a partir de la historia**

# **Modelo de Clasificación con Árbol de Decisión**

Se busca predecir si una participante ganará o no una medalla en base a su país, su altura, su peso, el año y el evento.

Se obtuvo un modelo con Underfitting, por lo que se decide buscar otros factores ajenos a la historia de los JJOO que puedan ayudar a predecir los resultados que obtendrá una corredora.

\*En notebook adjunto pueden observarse las métricas que llevaron a dicha conclusión

03



# Incorporación del contexto socio-económico

Datos del Banco Mundial  
Data Wrangling y obtención de insights  
Implementación y mejora de modelo predictivo



# Selección de un Indicador y Proceso de Data Wrangling



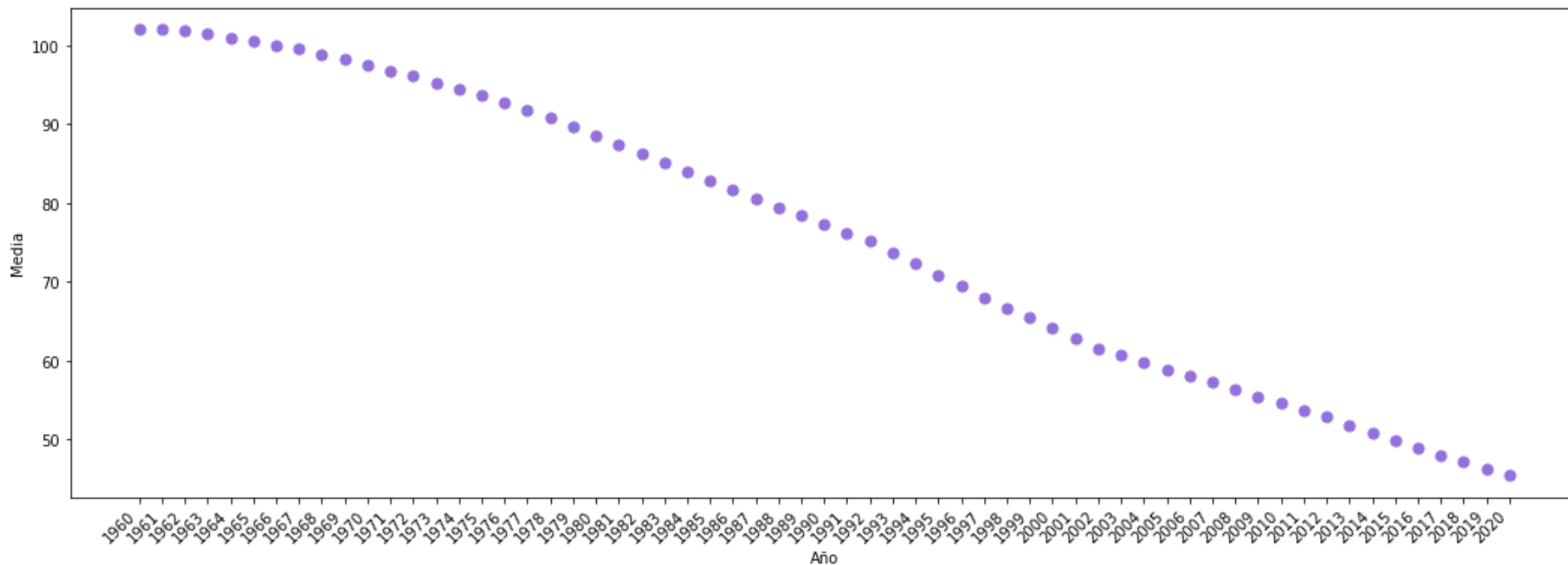
Base de datos número 17: "Gender"

Se seleccionó el indicador:  
"SP.ADO.TFRT - Adolescent fertility  
rate (births per 1,000 women ages 15-  
19)"

No se detectaron valores atípicos

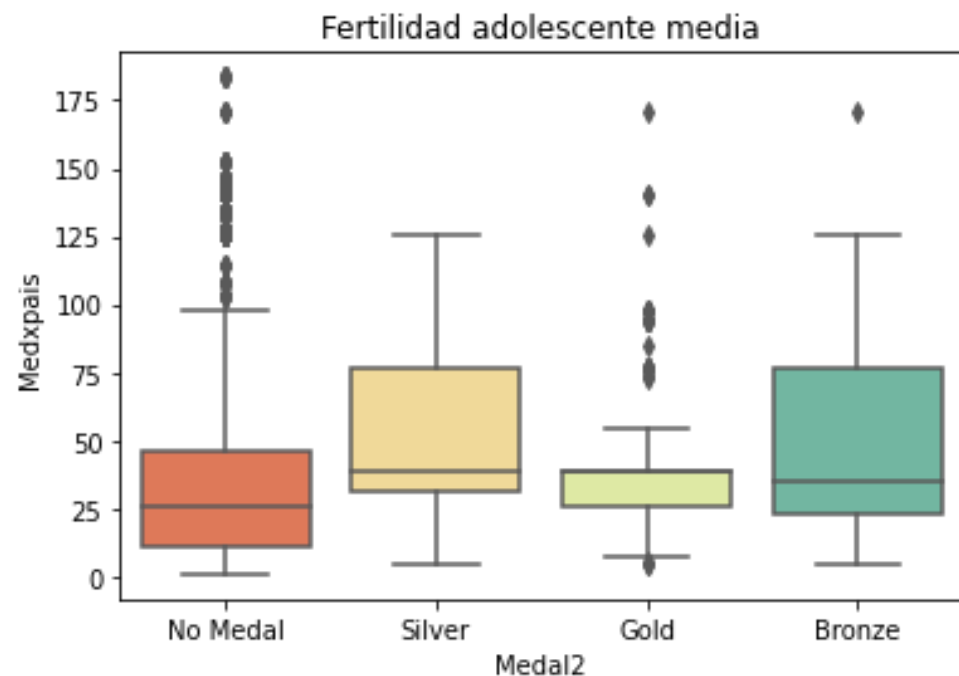
# Primer acercamiento al indicador

## Análisis Media vs. Tiempo



Se observa una tendencia de disminución constante, sin denotar variaciones drásticas año a año.

# Impacto del Indicador de Fertilidad Adolescente en performance



**Conclusiones:** esta característica en el background de las participantes no las condiciona a la hora de obtener medallas.

En el caso de las que no obtuvieron medallas, el promedio del indicador es menor, mientras que en las que sí obtuvieron medallas, los promedios son mayores y la distribución de valores es bastante amplia, contando incluso con algunos valores atípicamente altos en la categoría "Medalla de Oro".

	count	mean	std	min	25%	50%	75%	max
Medal2								
Bronze	200.0	44.969521	34.289965	4.480097	23.304426	35.021271	76.258045	170.766781
Gold	193.0	41.401843	28.084129	4.480097	25.961903	39.132161	39.132161	170.766781
No Medal	5879.0	36.501448	32.601313	0.872077	11.166368	25.961903	46.106832	183.536355
Silver	197.0	48.033352	30.106518	4.480097	31.254148	39.132161	76.258045	125.562116



# Optimización y mejora de modelo predictivo

Incorporación de variables socioeconómicas



Balanceo de Dataset – Random Oversampling



Feature Selection – Implementación de Standard Scaler y PCA



**Mejor Performance**

**Modelo con mejor Performance:** Random Forest Classifier

Métricas	Precision	Recall	F1 Score	Support
0 – Sin medalla	1.00	0.92	0.96	1555
1 – Bronce	0.90	0.94	0.92	1523
2 – Plata	0.93	0.91	0.92	1564
3 – Oro	0.90	0.95	0.93	1643

Accuracy: 93%

True label	No Medal	1424	33	42	56
	Bronze	0	1426	38	59
	Silver	0	79	1430	55
	Gold	0	44	31	1568
		No Medal	Bronze	Silver	Gold
		Predicted label			

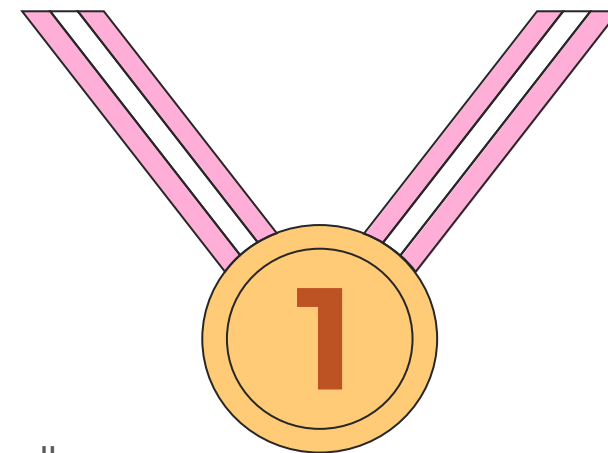
# Optimización de hiperparámetros

Se utiliza Randomized Search CV, perteneciente a la librería Sklearn



# 04

## Conclusiones



A lo largo del proyecto se buscó responder el problema comercial y desarrollar un algoritmo que pudiese clasificar a las participantes de atletismo en función de si obtendrán o no una medalla.

En primer lugar se hizo un acercamiento a los datos, obteniendo conclusiones valiosas que supieron orientar la búsqueda y entender el problema.

Luego, se trabajó sobre el dataset, imputando datos nulos, transformando variables categóricas en numéricas, balanceándolo y estandarizándolo para poder así realizar un feature selection y obtener mejores resultados.

Se procedió a aplicar 3 algoritmos de clasificación al dataset mejorado.

Random Forest Classifier fue el que obtuvo la mejor performance.

Por último, se realizó una optimización de hiperparámetros para obtener así los mejores resultados posibles.