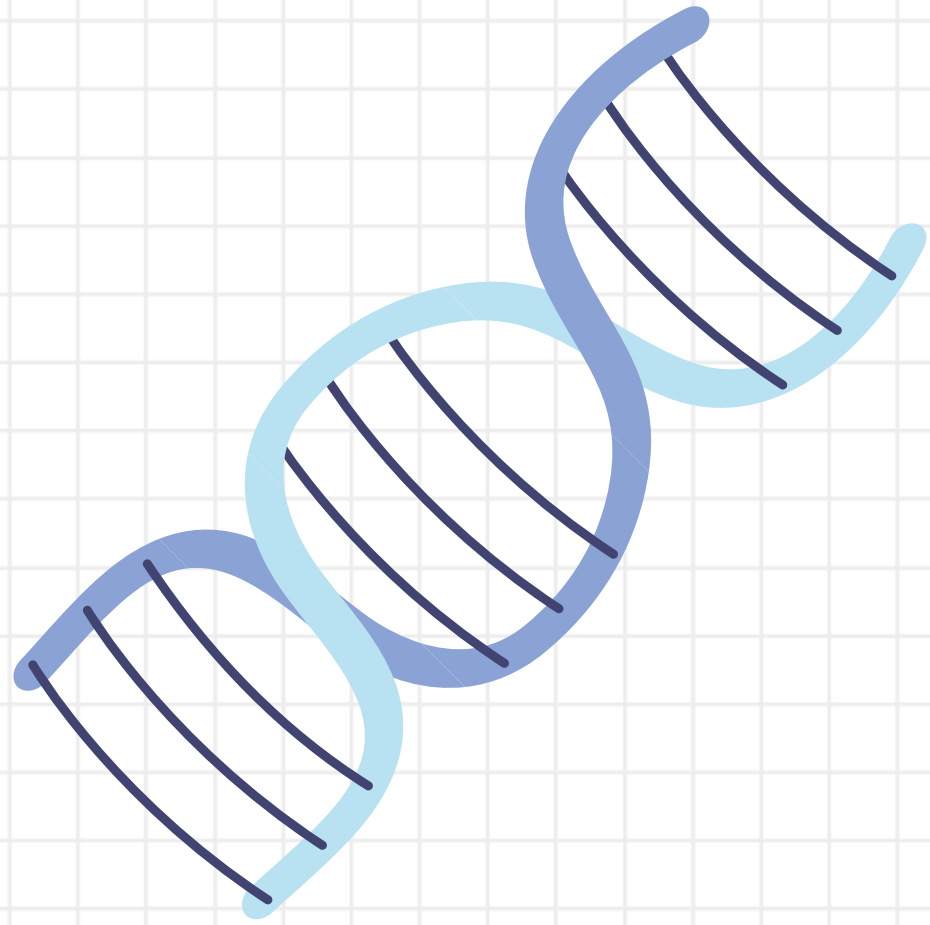




NACIDOS VIVOS EN HOSPITAL MANUEL URIBE ANGEL & NACIDOS HOSPITAL SAN JUAN DE DIOS RIONEGRO

Sofia Galindo, Anamaria Leguizamón,
Diego Herrera



Contenido

- | | |
|-----------------|-----------------|
| 1. Introducción | 4. Objetivos |
| 2. Equipo | 5. Conclusiones |
| 3. Conceptos | |



Contexto del estudio

- Importancia del seguimiento de datos de nacimientos para la salud materna e infantil en Colombia
- Portal de datos.gov.co como fuente de datos
- Implicaciones de los datos de nacimientos para el bienestar y desarrollo futuro



Selección de Conjuntos de Datos

- Conjuntos de datos seleccionados: "Nacidos Vivos en Hospital Manuel Uribe Ángel" (Envigado) y "Nacidos Hospital San Juan de Dios Rionegro"
- Comparación dentro del mismo contexto regional (Antioquia) para controlar variables externas



Conjunto de Datos

Nacidos Vivos en Hospital Manuel Uribe Angel

¿Qué hay en este conjunto de datos?

Filas	Columnas
17,3K	31

Nacidos Hospital San Juan de Dios Rionegro

¿Qué hay en este conjunto de datos?

Filas	Columnas
2.328	13

Conjunto de Datos

Nombre de la columna	Descripción	Tipo
DEPARTAMENTO		Texto simple T
MUNICIPIO		Texto simple T
AREA NACIMIENTO		Texto simple T
SEXO		Texto simple T
PESO (Gramos)		Número #
TALLA (Centímetros)		Número #
FECHA NACIMIENTO		Texto simple T
TIEMPO DE GESTACIÓN		Número #
NÚMERO CONSULTAS PRENATALES		Número #
TIPO PARTO		Texto simple T
MULTIPLICIDAD EMBARAZO		Texto simple T
EDAD MADRE		Texto simple T
EDAD PADRE		Texto simple T

Gobierno de Colombia. (2011). Nacidos Hospital San Juan de Dios Rionegro. [Datos Abiertos Colombia]. [URL del sitio web.](#)

Preguntas

¿Existen diferencias significativas en las características de los nacimientos entre los hospitales de diferentes municipios, como tasas de nacimientos prematuros, peso al nacer, complicaciones, etc.? Si es así, ¿qué factores podrían explicar estas diferencias?

¿Cómo se distribuyen las cesáreas entre programadas y no programadas, y cuál es la incidencia de nacimientos prematuros en cesáreas no programadas en comparación con las programadas?

Desarrollo

3.A

https://colab.research.google.com/drive/1tsFiMLsrBUsZxRtQ5As__7BTwUgNK6xS?usp=sharing

3.B

https://colab.research.google.com/drive/1E1XDHN8rzdA_u0nroqzeWZw-qUUJd_GX?usp=sharing

3.C

https://colab.research.google.com/drive/1FrEgcUGNDcFOhBGGihB6VMiuBy41gr_8?usp=sharing



Exploracion de los Datos.



Nacidos en el Hospital San Juan de Dios Rionegro

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2328 entries, 0 to 2327
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   DEPARTAMENTO                          2328 non-null   object
1   MUNICIPIO                            2328 non-null   object
2   AREA NACIMIENTO                      2328 non-null   object
3   SEXO                                 2328 non-null   object
4   PESO (Gramos)                       2328 non-null   int64
5   TALLA (Centímetros)                 2328 non-null   int64
6   FECHA NACIMIENTO                    2328 non-null   object
7   TIEMPO DE GESTACIÓN                 2328 non-null   int64
8   NÚMERO CONSULTAS PRENATALES         2328 non-null   int64
9   TIPO PARTO                          2328 non-null   object
10  MULTIPLICIDAD EMBARAZO              2328 non-null   object
11  EDAD MADRE                          2328 non-null   object
12  EDAD PADRE                          2328 non-null   object
dtypes: int64(4), object(9)
memory usage: 236.6+ KB
```

- Dataset con 2.328 Filas.
- Afortunadamente no hay valores nulos.
- Hay que investigar porque las edades de los padres son objetos y no numeros.

Entendamos un poco más sobre la columna de EDAD MADRE Y EDAD PADRE

EDAD MADRE	EDAD PADRE
29(4)	28(4)
26(4)	30(4)
25(4)	32(4)
16(4)	31(4)
22(4)	47(4)

→ Podemos ver que estan almacenados de forma extraña conteniendo un “(4)” extra:

Para transformar estas variables utilizamos el siguiente código:

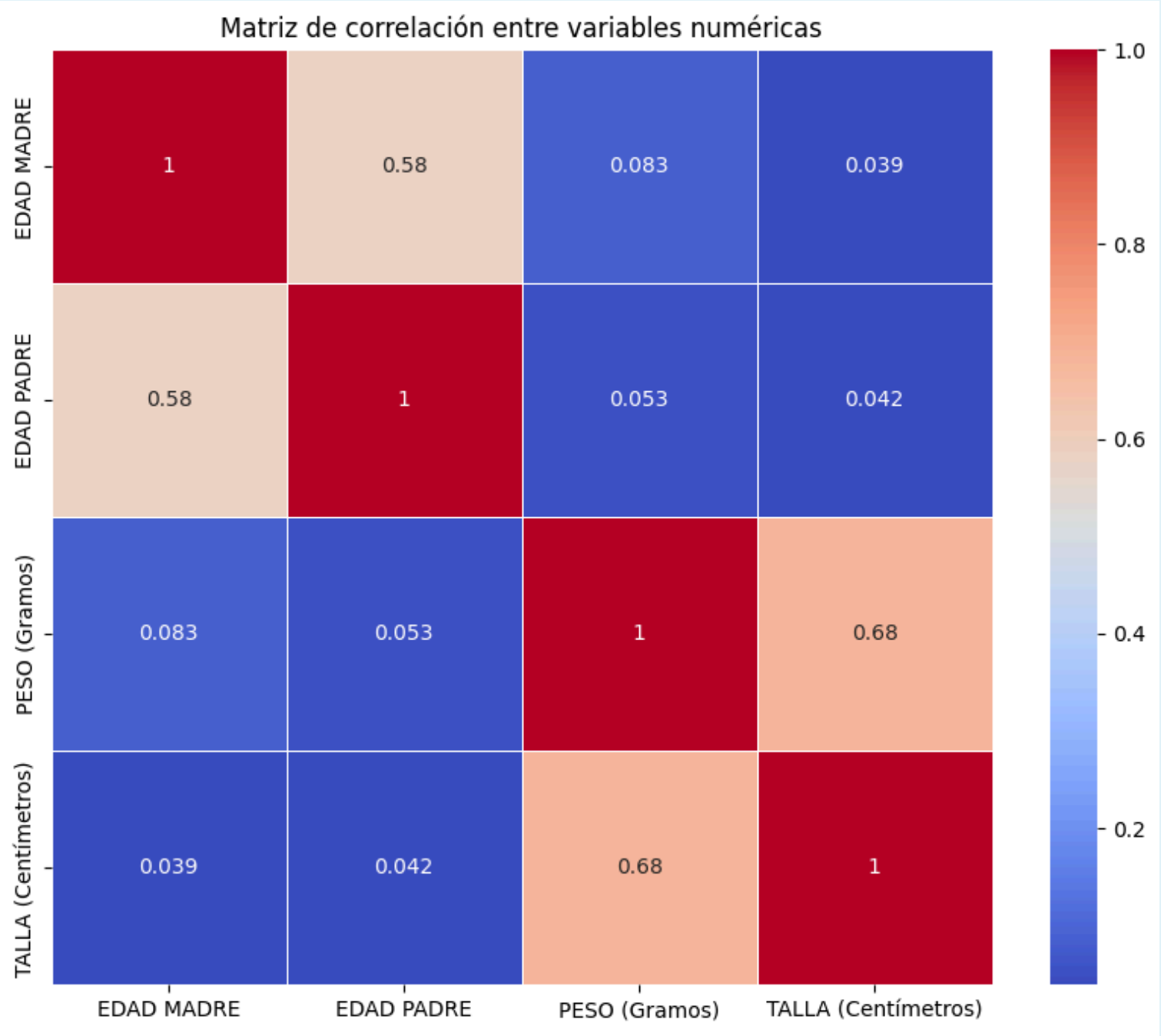
```
# Transformación de las edades de los padres a formato numérico
# Asegurarse que la columna es de tipo string antes de aplicar .str
if san_juan_hospital_data['EDAD MADRE'].dtype == 'object':
    san_juan_hospital_data['EDAD MADRE'] = san_juan_hospital_data['EDAD MADRE'].str.extract('(\d+)').astype(int)

# Lo mismo para 'EDAD PADRE' si es necesario
if san_juan_hospital_data['EDAD PADRE'].dtype == 'object':
    san_juan_hospital_data['EDAD PADRE'] = san_juan_hospital_data['EDAD PADRE'].str.extract('(\d+)').astype(int)
```

Utilizamos el método de extract con “**regex**” para sacar solo el número que está fuera de los paréntesis.

Transformamos el numero de **string** a **int**

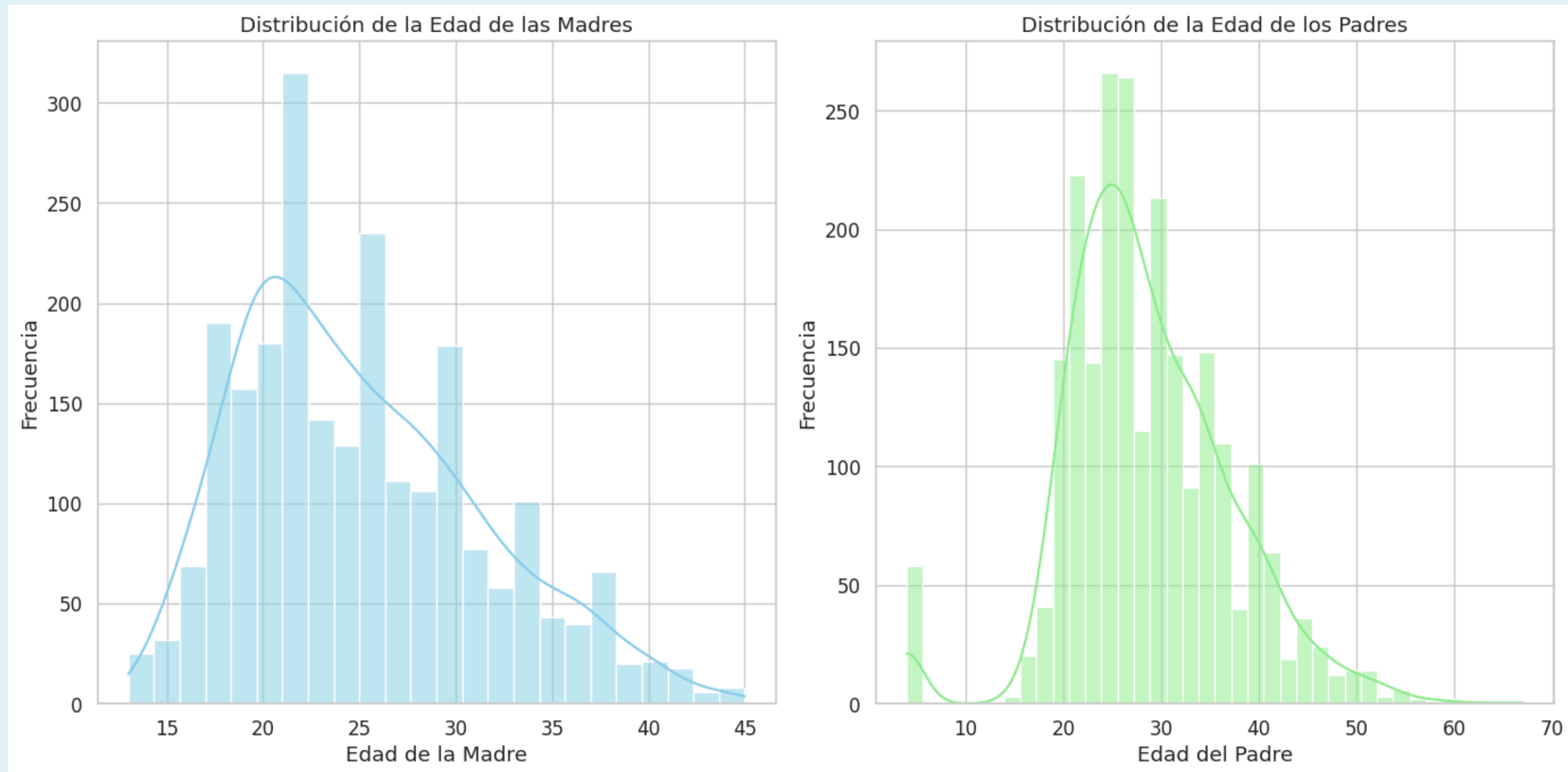
Correlacion entre variables Numericas:



- **Edad de los padres:** Hay una **correlación moderada (0.58)** entre la edad de la madre y del padre, lo que podría indicar que las parejas **tienden a ser de edades similares**.
- **Peso y Talla de los bebés:** Se observa una correlación notable **(0.68)** entre el peso y la talla de los bebés, lo cual es esperado ya que **bebés más grandes tienden a pesar más**.
- Las correlaciones entre las edades de los padres y las dimensiones físicas de los bebés son muy bajas, lo que sugiere que **no hay una relación directa entre la edad de los padres y el peso o talla de los bebés al nacer**.

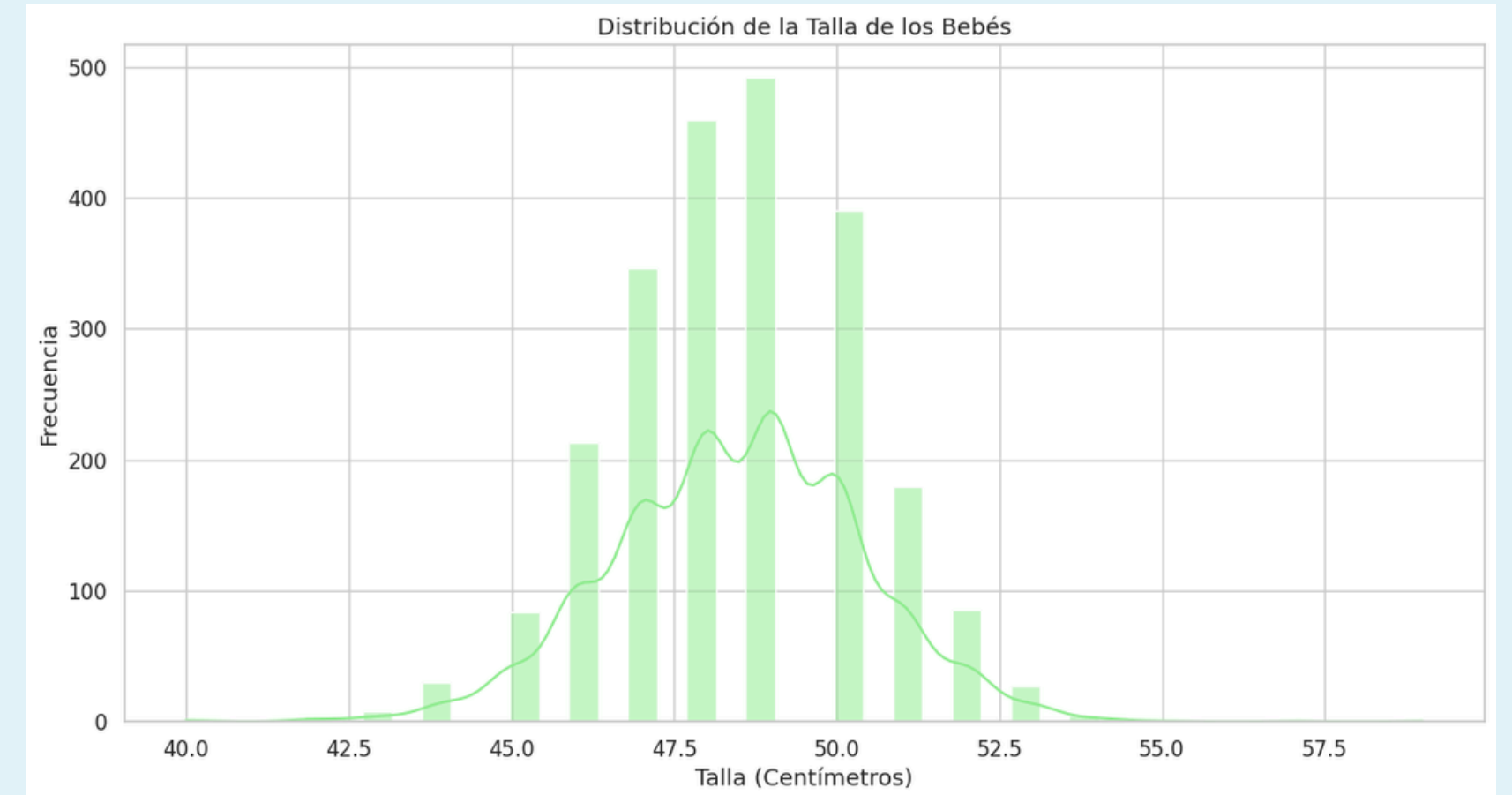
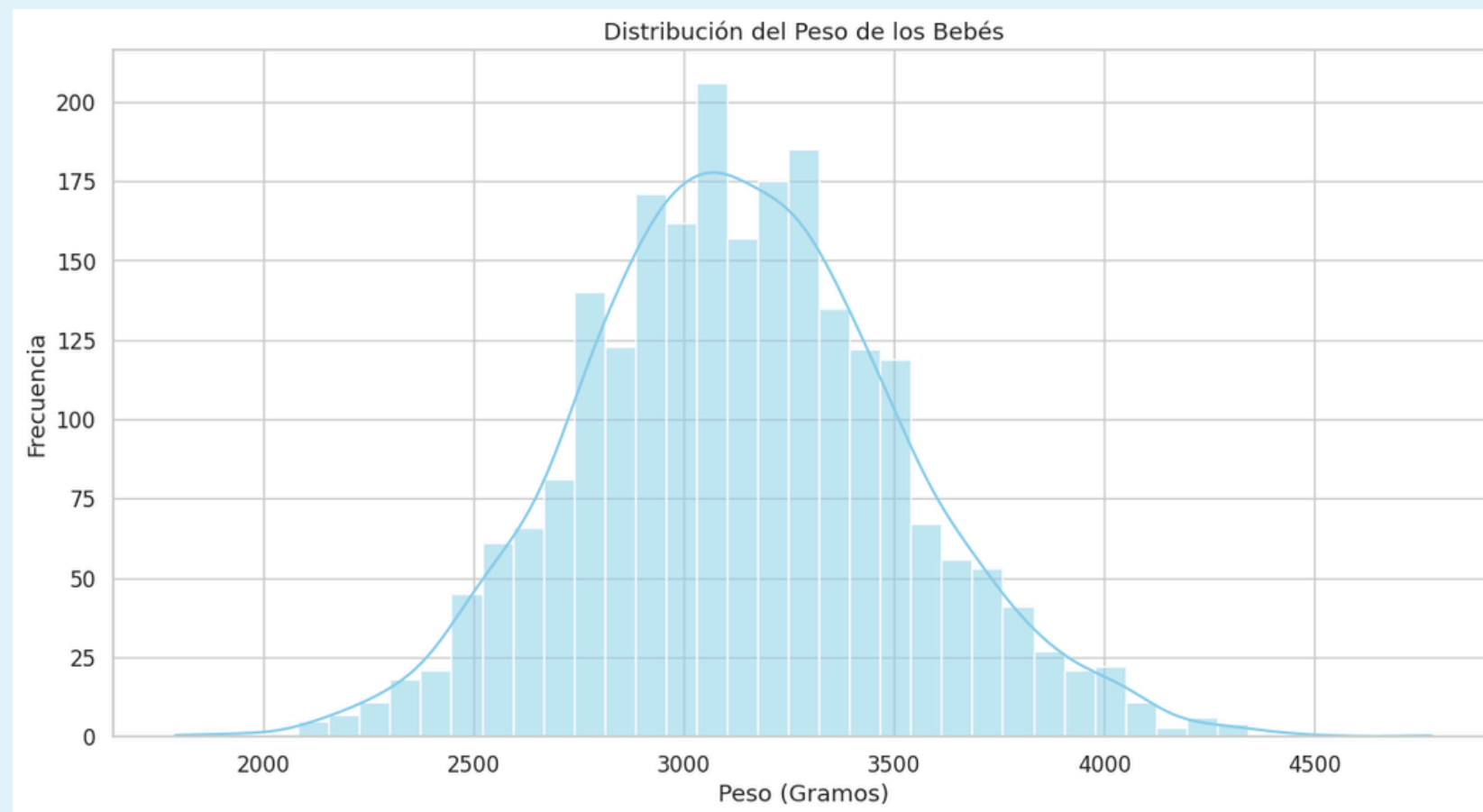
Distribucion de edades de los padres en Rionegro:

Ambos gráficos (madre y padre) muestran distribuciones con sesgo positivo, indicando una **concentración de padres más jóvenes**. La edad de la madre presenta un **pico entre los 20 y 23 años**, mientras que la de los padres muestra una distribución más uniforme pero todavía concentrada en edades más jóvenes.



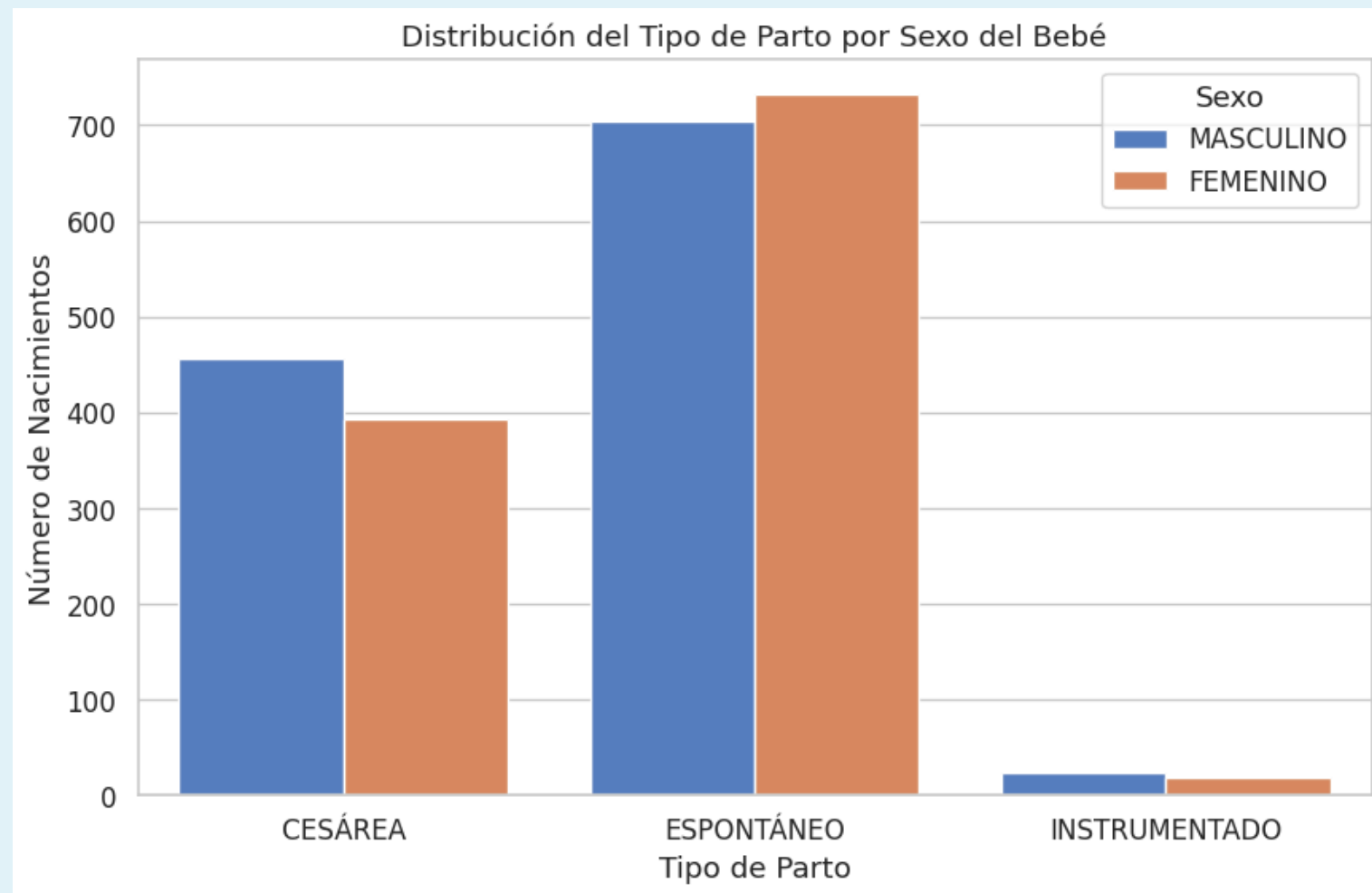
Peso y talla de los bebés en Rionegro:

Ambas variables muestran distribuciones **aproximadamente normales**, con el **peso centrado alrededor de 3100 gramos** y la **talla alrededor de 49** centímetros.



Este tipo de distribución es común en **variables humanas**.

Distribucion del Tipo de Parto por Sexo del Beb  en Rionegro:



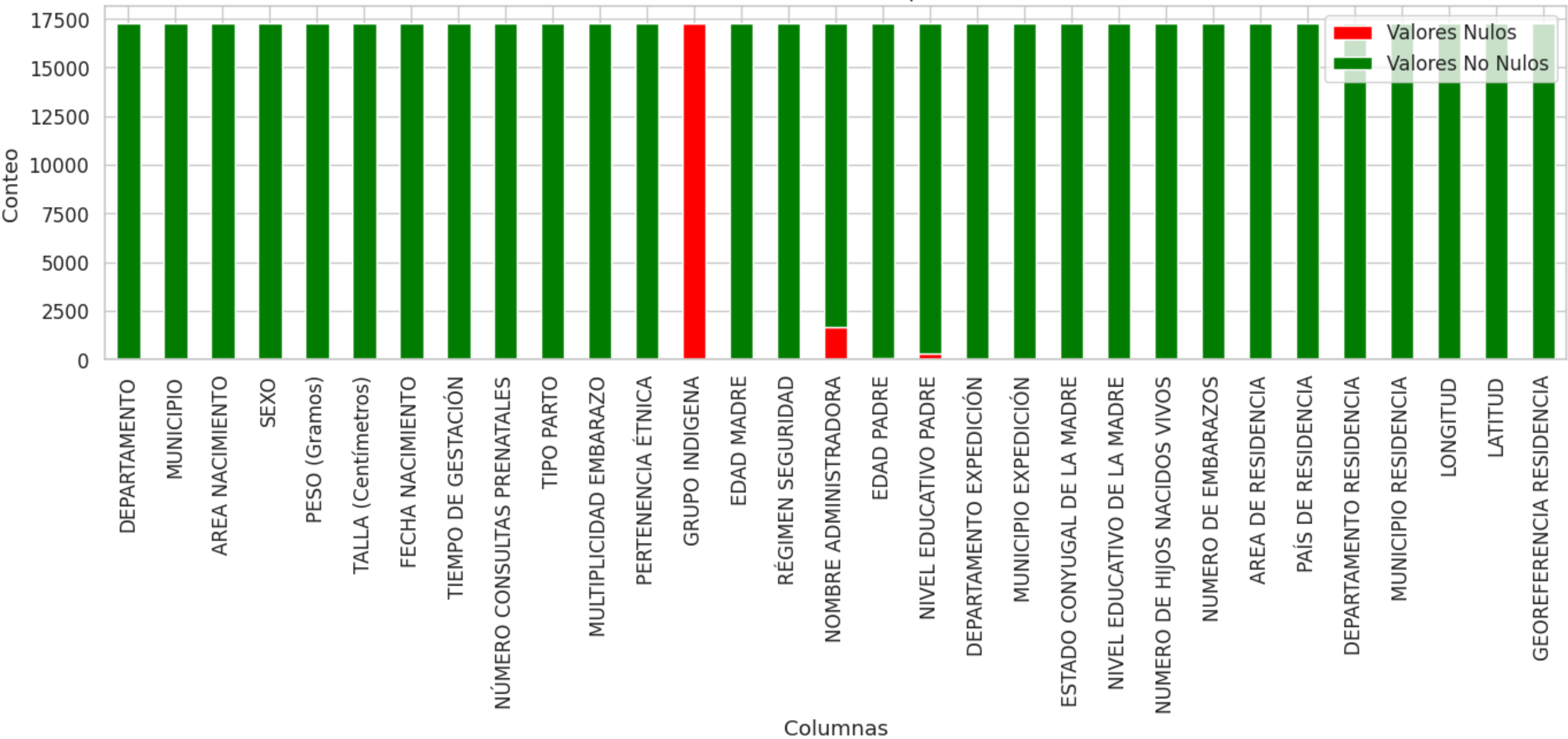
- Los nacimientos por **ces rea** son ligeramente m s comunes en **beb s masculinos**.
- Los **partos espont neos** son m s comunes en **beb s femeninos**. Esto podr a sugerir una tendencia m dica hacia ciertos tipos de parto basados en el sexo del beb .

Nacidos Vivos en Hospital Manuel Uribe Ángel

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17299 entries, 0 to 17298
Data columns (total 31 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
0   DEPARTAMENTO                          17299 non-null  object
1   MUNICIPIO                           17299 non-null  object
2   AREA NACIMIENTO                     17299 non-null  object
3   SEXO                                17299 non-null  object
4   PESO (Gramos)                       17299 non-null  int64
5   TALLA (Centímetros)                 17299 non-null  int64
6   FECHA NACIMIENTO                    17299 non-null  object
7   TIEMPO DE GESTACIÓN                 17299 non-null  int64
8   NÚMERO CONSULTAS PRENATALES         17299 non-null  int64
9   TIPO PARTO                          17299 non-null  object
10  MULTIPLICIDAD EMBARAZO              17299 non-null  object
11  PERTENENCIA ÉTNICA                  17299 non-null  object
12  GRUPO INDIGENA                      9 non-null      object
13  EDAD MADRE                          17299 non-null  int64
14  RÉGIMEN SEGURIDAD                  17299 non-null  object
15  NOMBRE ADMINISTRADORA               15644 non-null  object
16  EDAD PADRE                          17241 non-null  float64
17  NIVEL EDUCATIVO PADRE               16989 non-null  object
18  DEPARTAMENTO EXPEDICIÓN              17299 non-null  object
19  MUNICIPIO EXPEDICIÓN                17299 non-null  object
20  ESTADO CONYUGAL DE LA MADRE         17299 non-null  object
21  NIVEL EDUCATIVO DE LA MADRE         17299 non-null  object
22  NUMERO DE HIJOS NACIDOS VIVOS       17299 non-null  int64
23  NUMERO DE EMBARAZOS                 17299 non-null  int64
24  AREA DE RESIDENCIA                  17299 non-null  object
25  PAÍS DE RESIDENCIA                  17299 non-null  object
26  DEPARTAMENTO RESIDENCIA              17299 non-null  object
27  MUNICIPIO RESIDENCIA                17299 non-null  object
28  LONGITUD                            17295 non-null  object
29  LATITUD                             17295 non-null  object
30  GEOREFERENCIA RESIDENCIA            17294 non-null  object
dtypes: float64(1), int64(7), object(23)
memory usage: 4.1+ MB
```

- Dataset con 17.299 Filas.
- En este si podemos encontrar bastantes valores nulos.

Valores Nulos vs No Nulos por Columna



Tratamiento para valores nulos importantes

Cambiamos todos los nulos de la columna indígena por “no indígena”

```
# Rellenar valores faltantes para 'GRUPO INDIGENA' con 'No Indígena'
uri_hospital_data['GRUPO INDIGENA'].fillna('No Indígena', inplace=True)
```

Utilizamos la “moda” (valor más común) para completar los datos faltantes del **Nombre de la Administradora**.

```
# Rellenar valores faltantes para 'NOMBRE ADMINISTRADORA' con el valor más común (moda)
uri_hospital_data['NOMBRE ADMINISTRADORA'].fillna(uri_hospital_data['NOMBRE ADMINISTRADORA'].mode()[0], inplace=True)
```

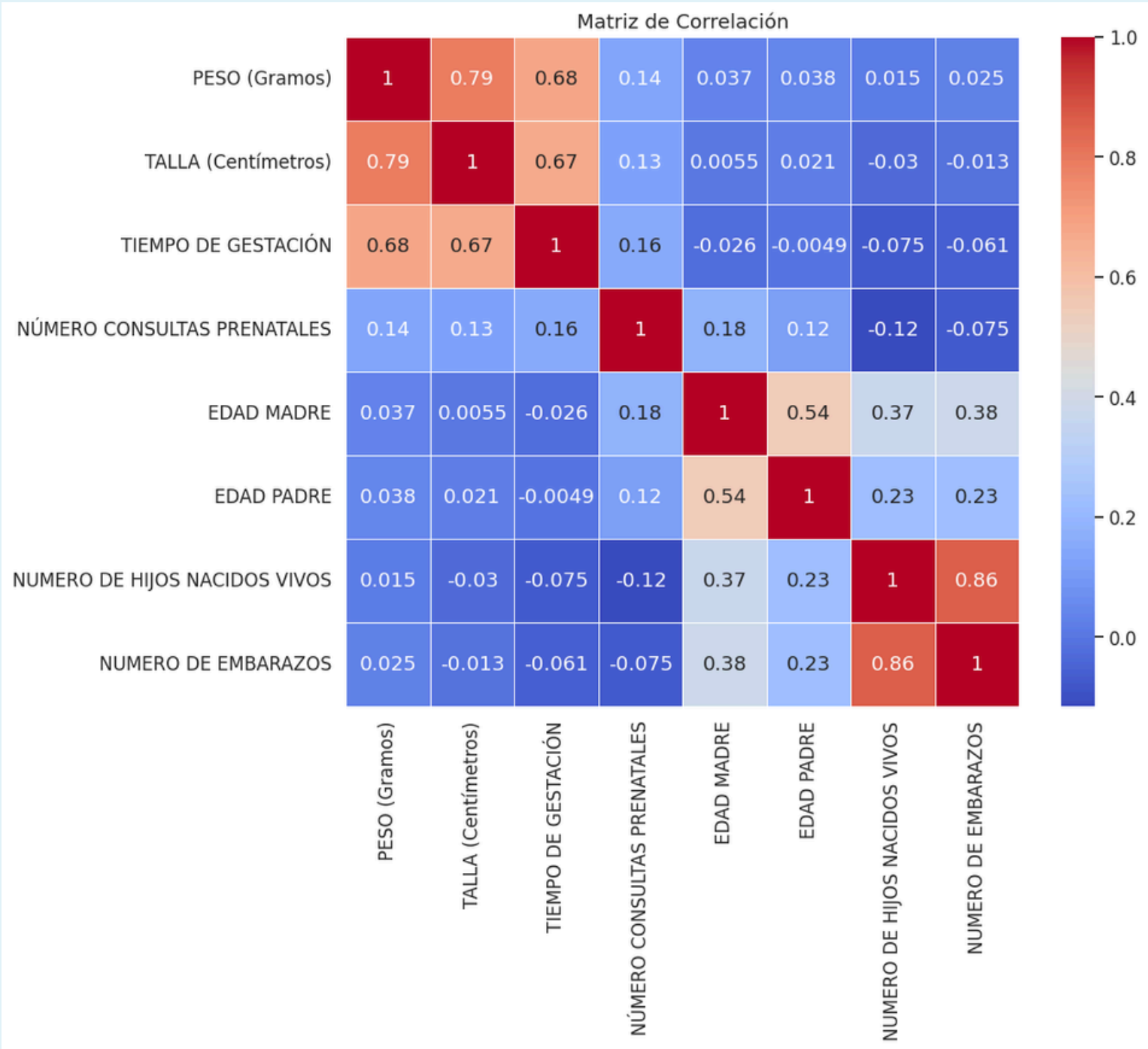
Utilizamos la mediana para llenar los datos nulos en la **edad del padre**.

```
# Convertir 'EDAD PADRE' a numérico, asegurando que todos los valores sean números y rellenar los valores faltantes con la mediana
uri_hospital_data['EDAD PADRE'] = pd.to_numeric(uri_hospital_data['EDAD PADRE'], errors='coerce')
uri_hospital_data['EDAD PADRE'].fillna(uri_hospital_data['EDAD PADRE'].median(), inplace=True)
```

Utilizamos la “moda” (valor más común) para llenar los datos nulos en el **nivel educativo del padre**.

```
# Rellenar valores faltantes para 'NIVEL EDUCATIVO PADRE' con el valor más común (moda)
uri_hospital_data['NIVEL EDUCATIVO PADRE'].fillna(uri_hospital_data['NIVEL EDUCATIVO PADRE'].mode()[0], inplace=True)
```

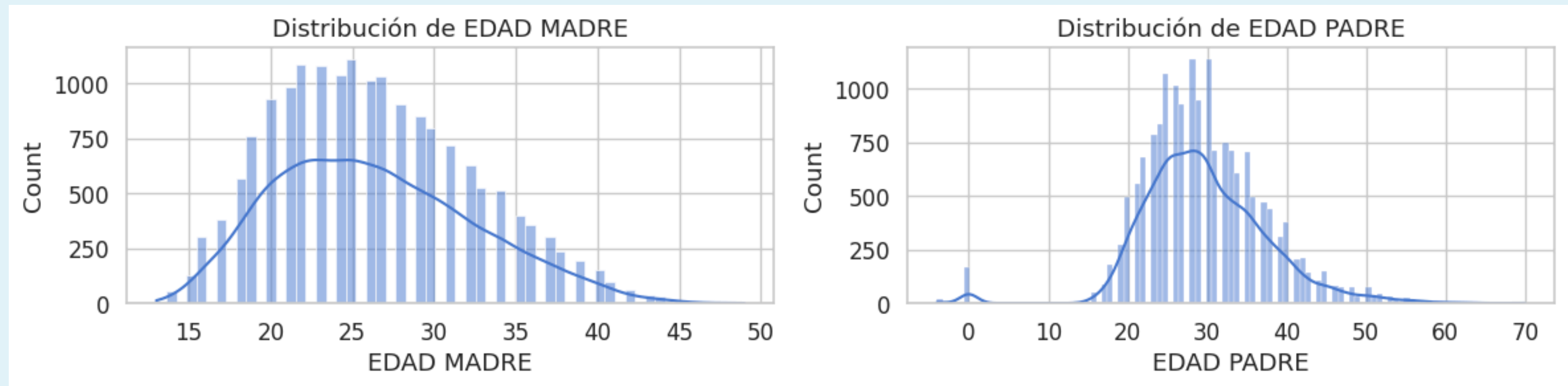
Correlacion entre variables Numericas:



- Fuerte correlación positiva entre el **peso y la talla del bebé (0.79)**:
 - Los bebés que pesan más también tienden a ser más altos.
- Correlación moderada positiva entre el tiempo de **gestación y el peso del bebé (0.68)**:
 - **Un mayor tiempo de gestación** está asociado con un mayor peso al nacer.

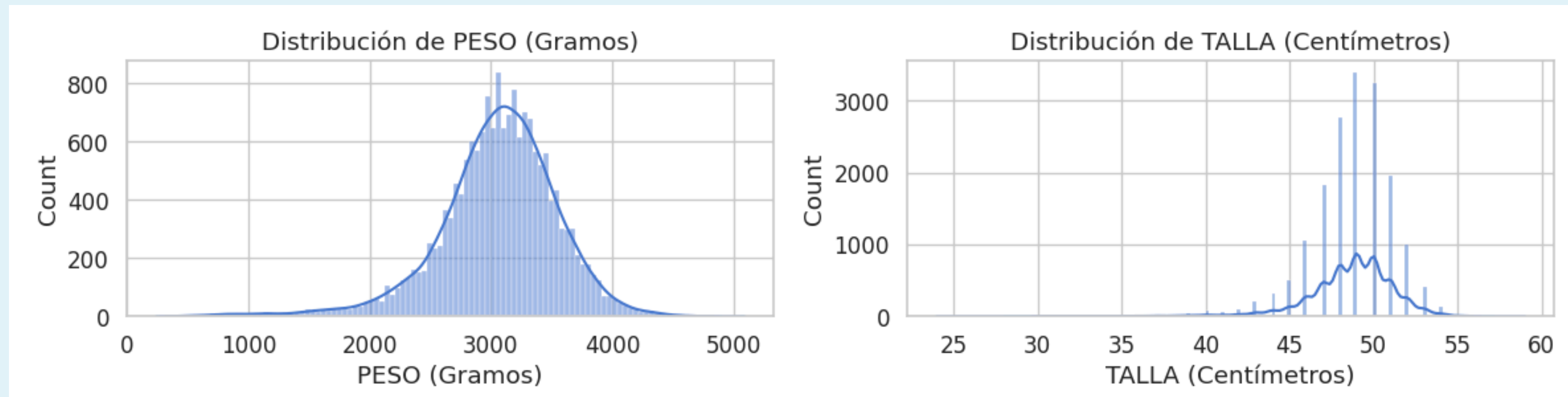
Distribucion de edades de los padres en Envigado:

- La mayoría de las madres están concentradas en el **rango de edad de 20 a 35 años**, con un pico alrededor de los 25 años, indicando que la mayoría de las madres son jóvenes.
- Los padres muestran una **distribución de edad más amplia**, con un pico alrededor de los 30 años y presencia significativa de padres más jóvenes y mayores, **extendiendo la edad hasta los 60 años**.

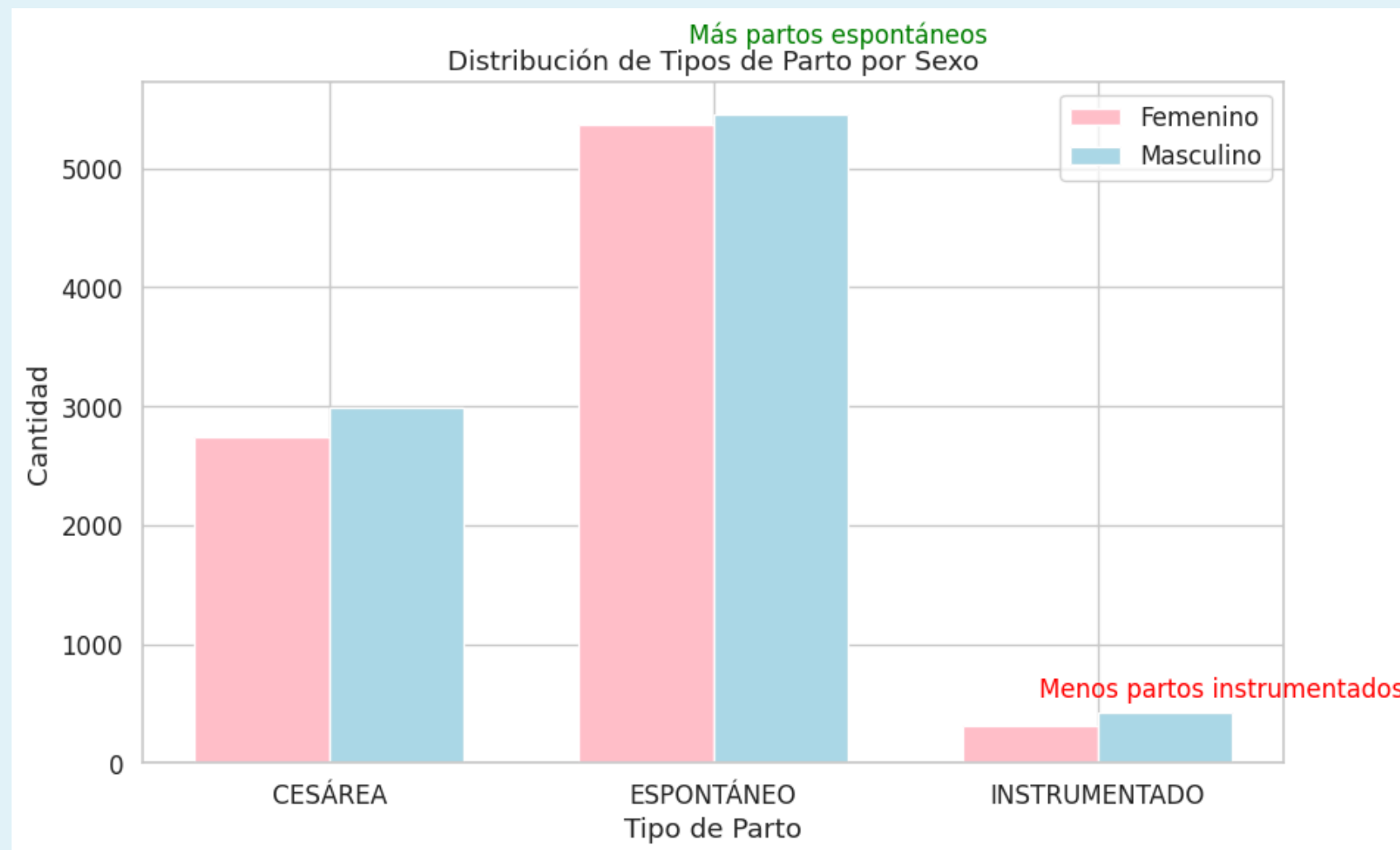


Peso y talla de los bebés en Envigado:

- **Distribución de Peso:** La mayoría de los bebés pesan entre 2500 y 3500 gramos, con el peso más común alrededor de 3000 gramos.
- **Distribución de Talla:** La mayoría de los bebés miden entre 45 y 50 centímetros, con la talla más común alrededor de 48 centímetros.



Distribución del Tipo de Parto por Sexo del Bebé en Envigado:



- **Partos Espontáneos Dominan:** La mayoría de los nacimientos son espontáneos, superando significativamente a las cesáreas e instrumentados.
- **Casi Igual Número de Cesáreas y Partos Espontáneos en Ambos Sexos:** Los nacimientos por cesárea y espontáneos muestran cantidades similares entre bebés femeninos y masculinos.
- **Partos Instrumentados Raros:** Los partos instrumentados son significativamente menos comunes en comparación con los otros tipos de parto.

3.B. - Problema

"Modelos de Machine Learning seleccionados"

- Regresión Logística:

Adecuada para problemas de clasificación binaria y multiclase.

Relaciones lineales entre variables independientes y dependiente.

Probabilidades para interpretación directa de pertenencia a clases.

- Árbol de Decisión

No requiere relaciones lineales entre variables

Captura interacciones complejas y no lineales

Interpretabilidad y facilidad de visualización

Útil con variables categóricas y numéricas

3.B. - Problema

“Rangos de peso al nacer”

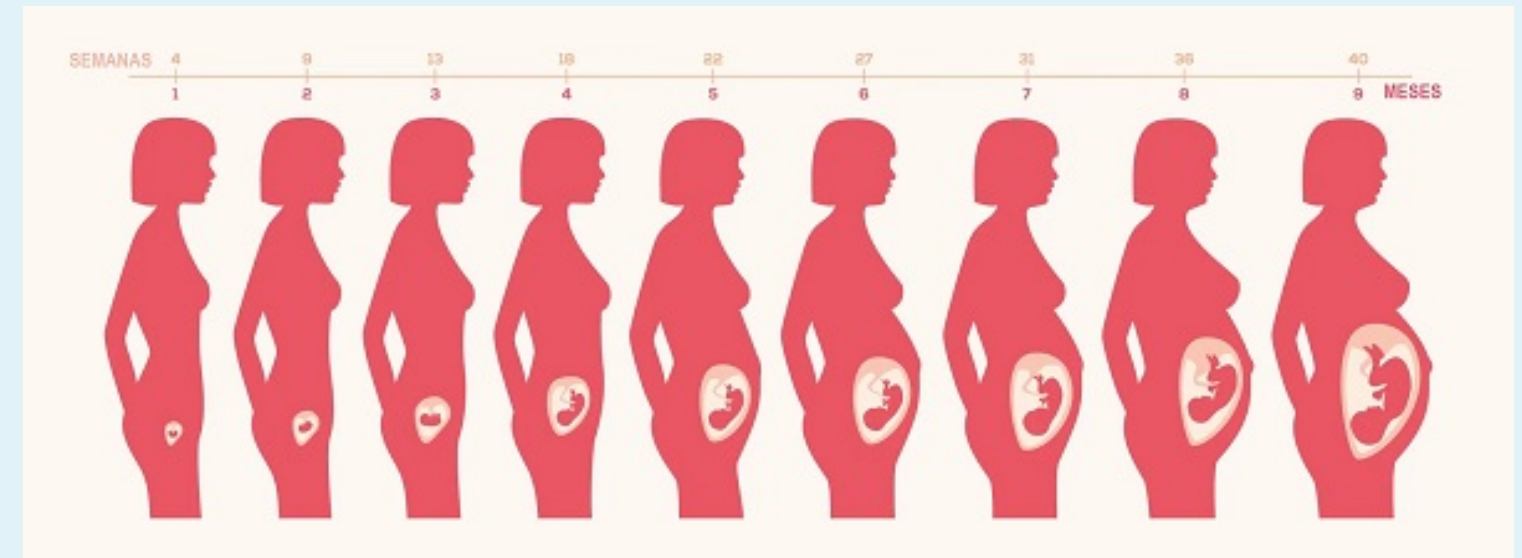
- Fuente: Reproducción Asistida ORG
- Rangos:
 - Peso normal: 2500 - 4000 gramos
 - Bajo peso: < 2500 gramos
 - Peso alto: > 4000 gramos
- Función personalizada para clasificar y mapear numéricamente las categorías



3.B. - Problema

"Semanas de gestación"

- Fuente: Organización Mundial de la Salud (OMS)
- Categorías:
 - Prematuro extremo: < 28 semanas
 - Muy prematuro: 28 - 32 semanas
 - Prematuro moderado a tardío: 32 - 37 semanas
 - A término: 37 - 42 semanas
 - Embarazo prolongado: > 42 semanas
- Función personalizada para categorizar y mapear numéricamente

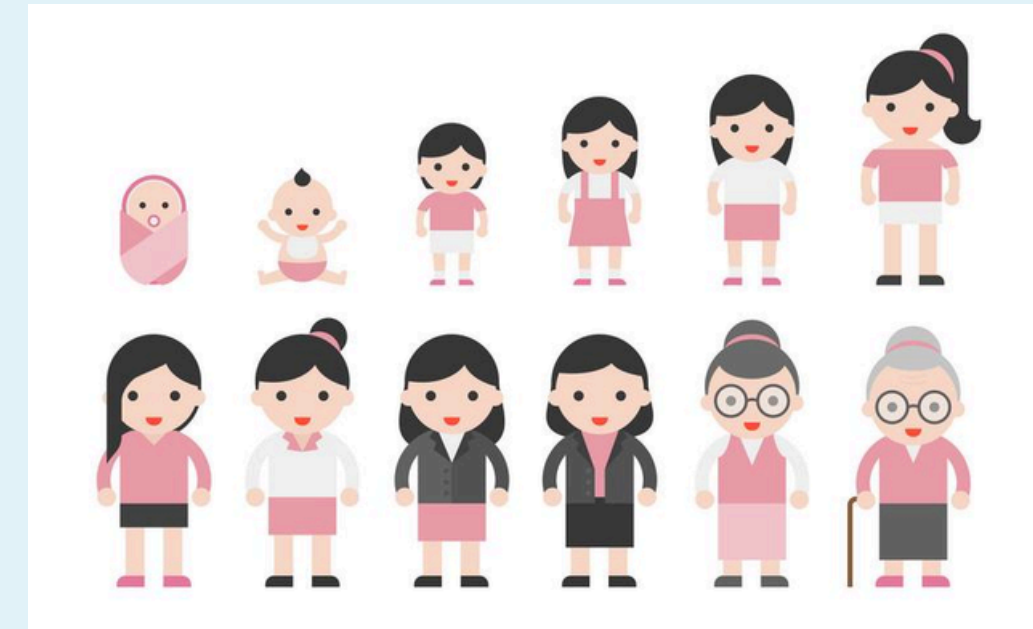


(“Se considera prematuro un bebé”, 2023). Nacimientos Prematuors. [Organización Mundial de la Salud]. [URL del sitio web.](#)

3.B. - Problema

"Edad materna"

- Categorías:
 - Adolescencia: 12 - 17 años
 - Adultas jóvenes: 18 - 28 años
 - Adulthood: 29 - 59 años
- Función personalizada para categorizar y mapear numéricamente



3.C.

Análisis de Clasificación de Categoría de Edad de Madres en Antioquia

- Introducción
- Este análisis utiliza datos de nacimientos en Antioquia para predecir la categoría de edad de las madres al momento del parto. Se emplean dos modelos de clasificación: Regresión Logística y Árbol de Decisión.
- Preprocesamiento de Datos
- Los datos se preprocesan aplicando:
 1. Codificación One-Hot para variables categóricas como SEXO y TIPO PARTO.
 2. Normalización de características numéricas usando StandardScaler para asegurar que todas las variables contribuyan equitativamente al modelo.

Análisis de Clasificación de Categoría de Edad de Madres en Antioquia

Entrenamiento de los modelos

```
# Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# Entrenar los modelos
model_logreg.fit(X_train, y_train)
model_tree.fit(X_train, y_train)

# Evaluar los modelos
y_pred_logreg = model_logreg.predict(X_test)
y_pred_tree = model_tree.predict(X_test)

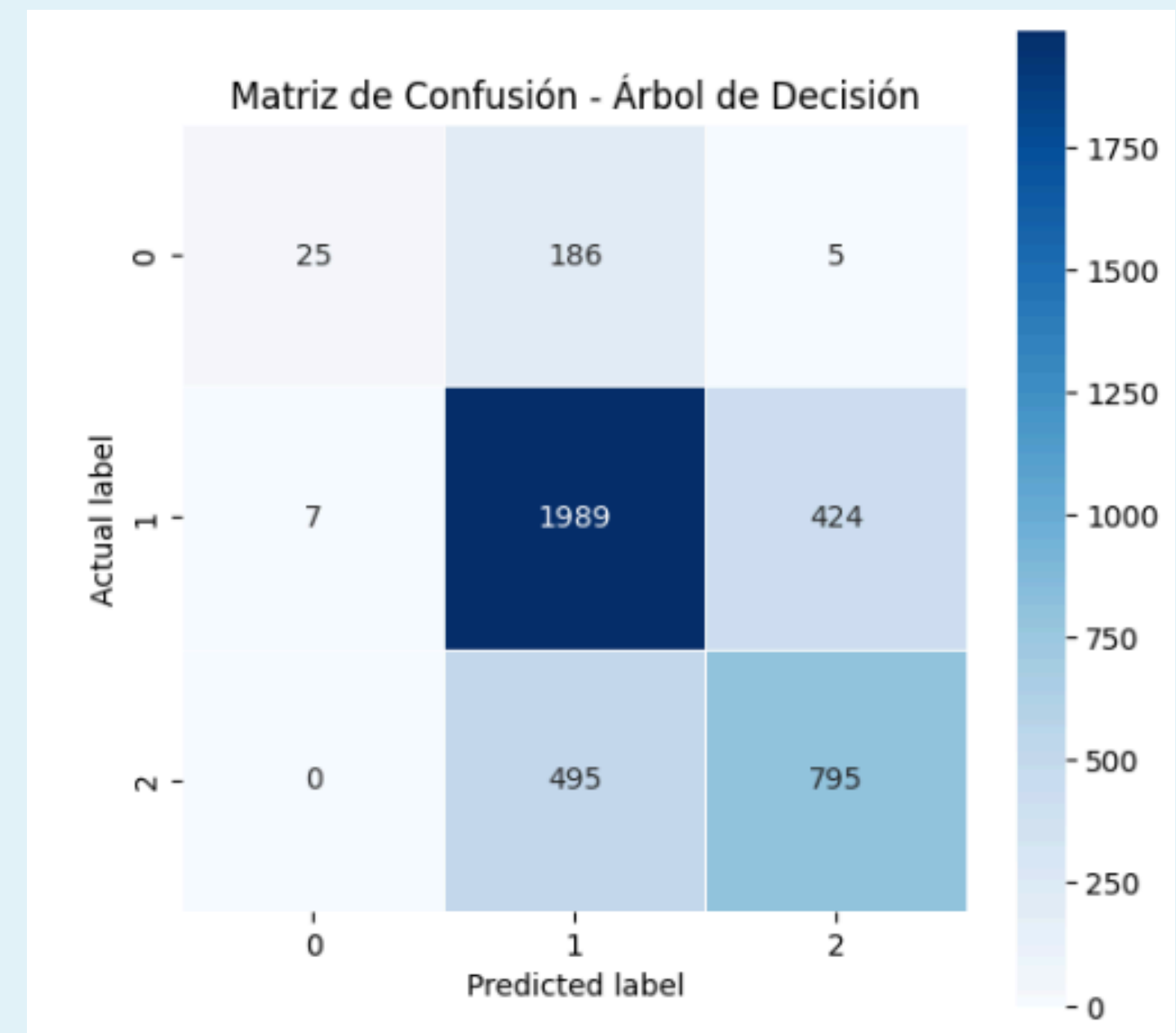
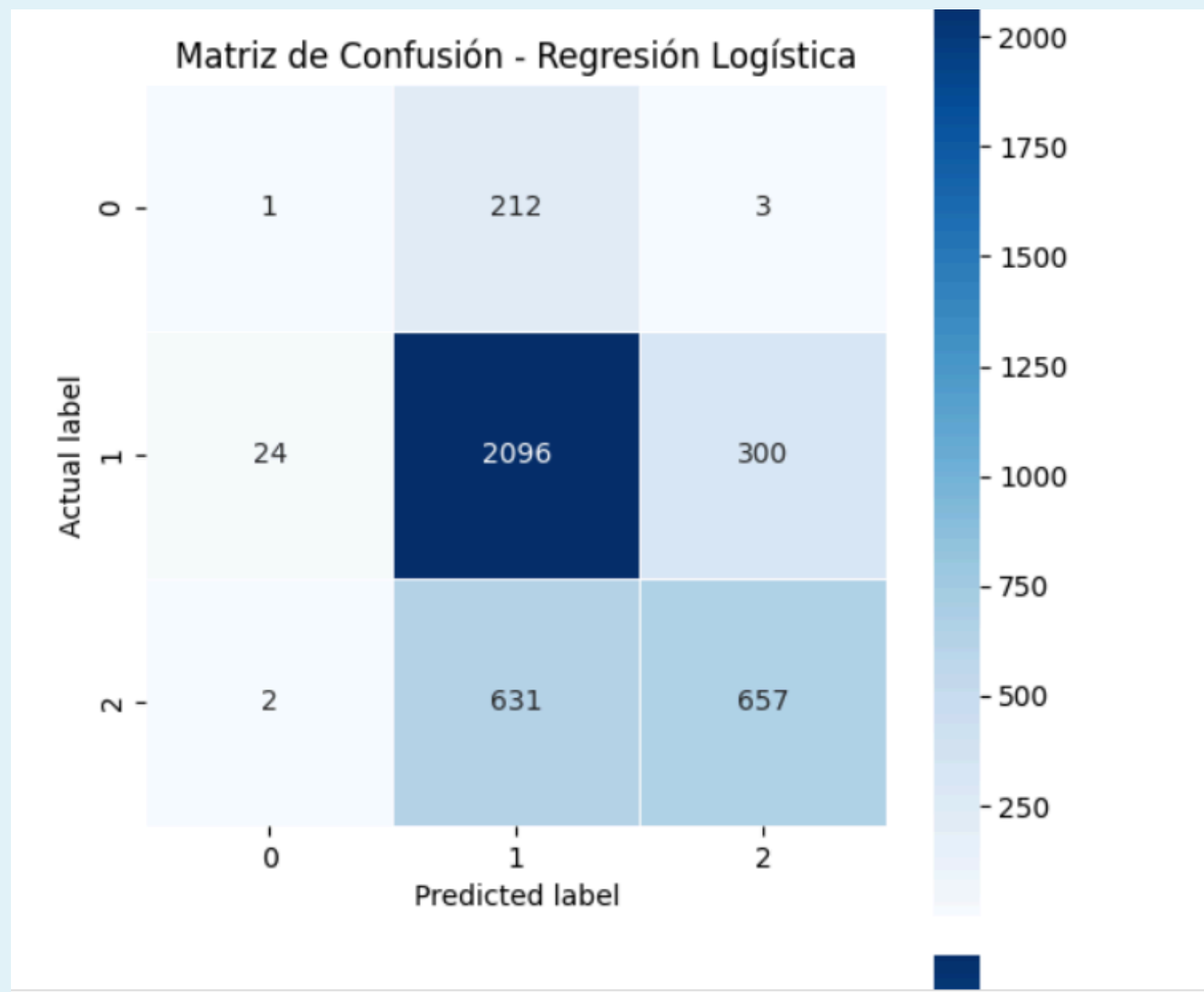
accuracy_logreg = accuracy_score(y_test, y_pred_logreg)
f1_score_logreg = f1_score(y_test, y_pred_logreg, average='weighted')
accuracy_tree = accuracy_score(y_test, y_pred_tree)
f1_score_tree = f1_score(y_test, y_pred_tree, average='weighted')

report_logreg = classification_report(y_test, y_pred_logreg)
report_tree = classification_report(y_test, y_pred_tree)

# Matrices de confusión
cm_logreg = confusion_matrix(y_test, y_pred_logreg)
cm_tree = confusion_matrix(y_test, y_pred_tree)

# Función para visualizar la matriz de confusión
def plot_confusion_matrix(cm, title):
    plt.figure(figsize=(6,6))
```

Resultados matrices de confusión



Resultados métricas y conclusión

Accuracy Logistic Regression: 0.7014773306164035
F1 Score Logistic Regression: 0.6745154527974048
Classification Report Logistic Regression:

	precision	recall	f1-score	support
1	0.04	0.00	0.01	216
2	0.71	0.87	0.78	2420
3	0.68	0.51	0.58	1290
accuracy			0.70	3926
macro avg	0.48	0.46	0.46	3926
weighted avg	0.67	0.70	0.67	3926

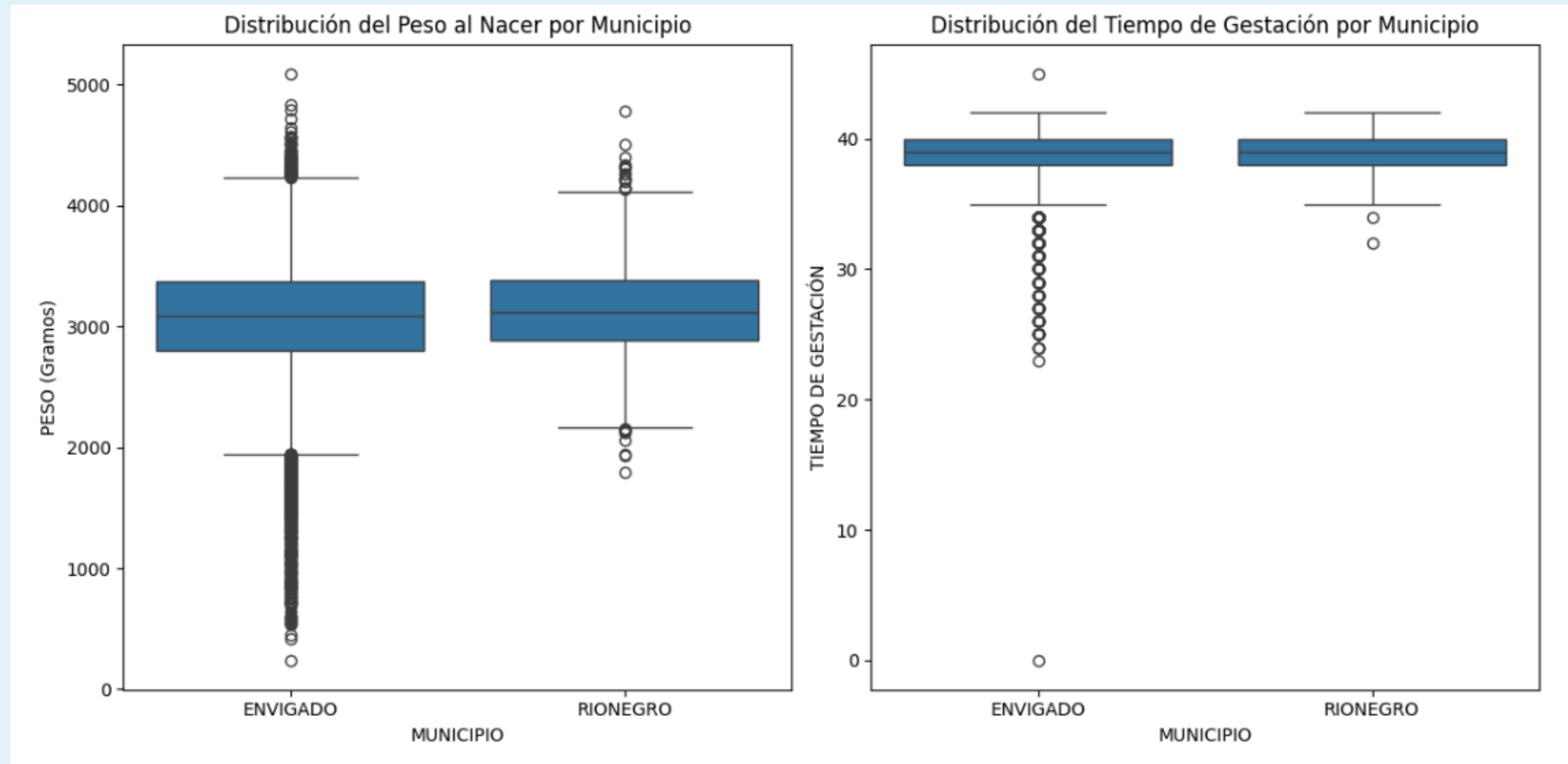
Accuracy Decision Tree: 0.7154865002547122
F1 Score Decision Tree: 0.7006439013632916
Classification Report Decision Tree:

	precision	recall	f1-score	support
1	0.78	0.12	0.20	216
2	0.74	0.82	0.78	2420
3	0.65	0.62	0.63	1290
accuracy			0.72	3926
macro avg	0.73	0.52	0.54	3926
weighted avg	0.72	0.72	0.70	3926

- a. Equilibrio entre Precisión y Recall: El Árbol de Decisión demostró un mejor equilibrio entre precisión y recall en comparación con la Regresión Logística. Esta característica es crucial para asegurar que el modelo clasifique correctamente las categorías de edad, mientras minimiza los errores de clasificación.
- b. Desempeño en Categorías Desbalanceadas: Aunque ambas técnicas mostraron limitaciones en la categoría de "Adolescencia" debido a su menor representación, el Árbol de Decisión logró mejores resultados en términos de precisión para esta categoría específica.
- c. Interpretabilidad: Los Árboles de Decisión son intrínsecamente más fáciles de interpretar y visualizar. Esta ventaja es significativa en contextos donde se requiere explicar y entender cómo el modelo realiza sus predicciones, lo cual es común en estudios demográficos y de salud pública.
- d. Superioridad en F1 Score: El modelo de Árbol de Decisión también presentó un F1 Score superior, indicando un balance más eficaz entre precisión y recall, particularmente en categorías con un número mayor de muestras.

PREGUNTAS

1. ¿Existen diferencias significativas en las características de los nacimientos entre los hospitales de diferentes municipios, como tasas de nacimientos prematuros, peso al nacer, complicaciones, etc.? Si es así, ¿qué factores podrían explicar estas diferencias?



PREGUNTAS

1. ¿Existen diferencias significativas en las características de los nacimientos entre los hospitales de diferentes municipios, como tasas de nacimientos prematuros, peso al nacer, complicaciones, etc.? Si es así, ¿qué factores podrían explicar estas diferencias?

Resultados de Prueba ANOVA

- **Peso al Nacer:**
 - Estadístico F: 55.64
 - Valor p: 9.05e-14
- **Tiempo de Gestación:**
 - Estadístico F: 199.76
 - Valor p: 3.92e-45

Interpretación

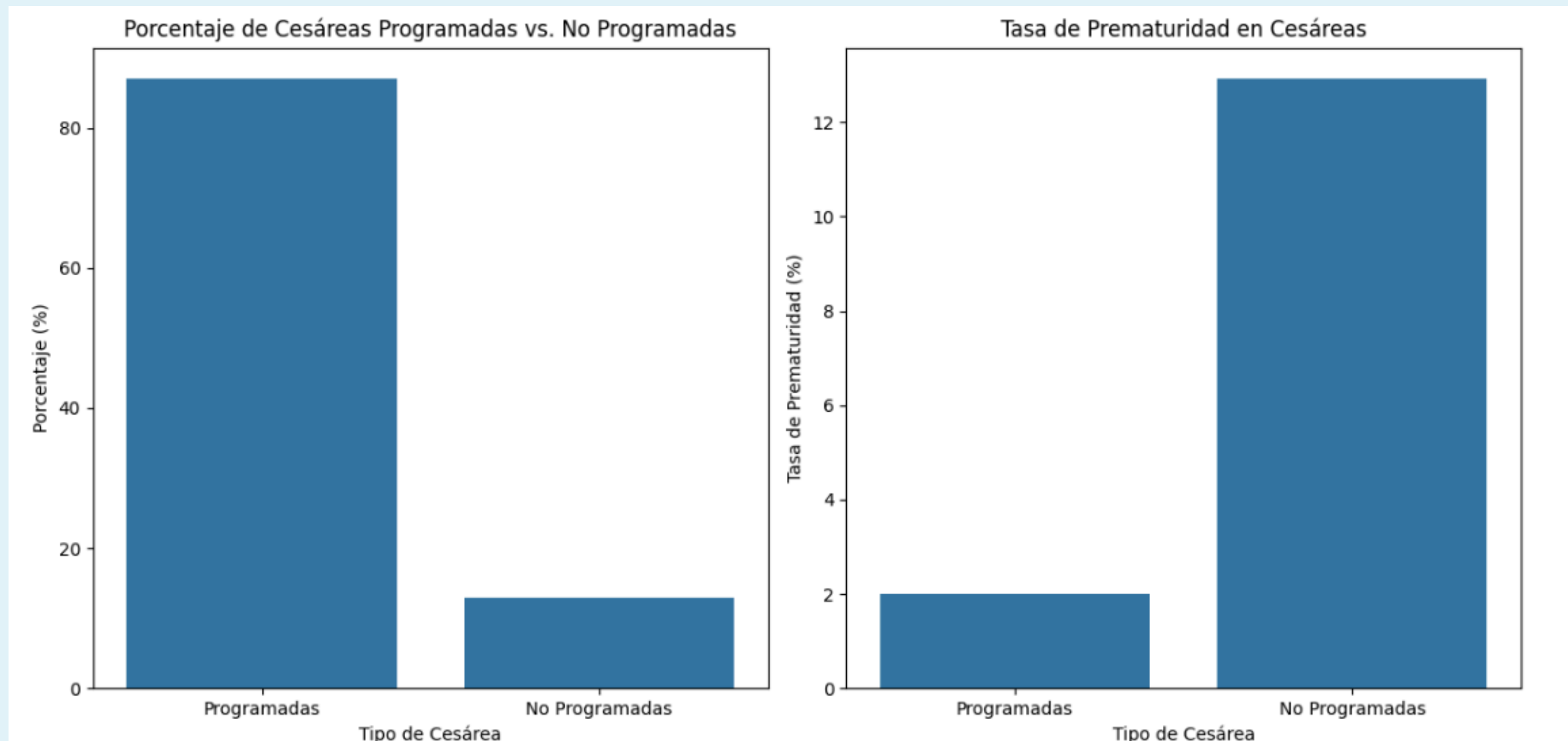
Los resultados de la prueba ANOVA indican diferencias estadísticamente significativas en el peso al nacer y el tiempo de gestación entre los municipios más grandes del dataset. El bajo valor p en ambos casos sugiere que las diferencias observadas en estas características no son aleatorias y pueden ser atribuibles a variaciones en factores ambientales, prácticas de atención prenatal o demográficas entre los municipios.

Factores Potenciales Explicativos

- **Acceso y Calidad de la Atención Prenatal:** Diferencias en la disponibilidad y calidad de la atención médica prenatal entre municipios.
- **Factores Socioeconómicos:** Variaciones en el nivel socioeconómico que pueden influir en la nutrición y la salud general de las madres.
- **Políticas de Salud Locales:** Diferentes políticas o programas de salud materno-infantil implementados a nivel municipal.

PREGUNTAS

2. ¿Cómo se distribuyen las cesáreas entre programadas y no programadas, y cuál es la incidencia de nacimientos prematuros en cesáreas no programadas en comparación con las programadas?



PREGUNTAS

2. ¿Cómo se distribuyen las cesáreas entre programadas y no programadas, y cuál es la incidencia de nacimientos prematuros en cesáreas no programadas en comparación con las programadas?

Hallazgos Clave del Análisis

De todos los partos por cesárea analizados, un 87.07% fueron programados y un 12.93% no programados. Este alto porcentaje de cesáreas programadas sugiere una preferencia o tendencia en la práctica obstétrica local hacia la programación de cesáreas, lo que puede reflejar tanto decisiones médicas informadas como posibles áreas de revisión en la gestión de partos.

Más preocupante, sin embargo, fue la tasa de prematuridad en las cesáreas no programadas, que también fue del 12.93%. Este dato es revelador, ya que sugiere que una proporción significativa de cesáreas de emergencia está directamente relacionada con nacimientos prematuros, una condición que aumenta considerablemente los riesgos de complicaciones neonatales y de salud a largo plazo en los niños.

Implicaciones y Reflexiones

El análisis ilustra cómo las políticas y prácticas de salud pueden influir en los resultados del parto. Aunque las cesáreas pueden ser herramientas vitales para preservar la salud y seguridad de madres y bebés durante complicaciones del parto, su alta frecuencia, particularmente en forma programada, plantea preguntas importantes sobre la necesidad y criterios con los que se toman estas decisiones.

Este fenómeno no es exclusivo de Colombia; países de todo el mundo han visto un aumento en las tasas de cesárea, muchas veces sin una justificación médica clara. Este patrón sugiere la posibilidad de una dependencia excesiva en intervenciones quirúrgicas que, aunque a veces necesarias, no están exentas de riesgos para madres y bebés, incluyendo tasas aumentadas de infecciones, recuperaciones más largas y mayores complicaciones en embarazos futuros.

Recomendaciones Basadas en el Análisis

1. **Revisión de Protocolos:** Es esencial que las instituciones médicas revisen continuamente sus protocolos de parto para asegurar que las decisiones de realizar cesáreas estén basadas en evidencia sólida y necesidades médicas reales.
2. **Educación y Capacitación:** Capacitar a los profesionales de la salud sobre alternativas a las cesáreas y fortalecer la educación prenatal puede ayudar a reducir las tasas de cesáreas no esenciales.
3. **Seguimiento a Largo Plazo:** Implementar un seguimiento más riguroso de las madres y los bebés después de una cesárea puede ayudar a identificar y mitigar cualquier complicación temprana derivada de la cirugía.

Bibliografía.

Gobierno de Colombia. (2011). Nacidos Hospital San Juan de Dios Rionegro. [Datos Abiertos Colombia]. [URL del sitio web.](#)

Gobierno de Colombia. (2011). Nacidos Vivos en Hospital Manuel Uribe Angel. [Datos Abiertos Colombia]. [URL del sitio web.](#)

Eureka Fertility S.L. (2008). ¿Cuál es el peso adecuado del bebé en el momento de nacer?. [Reproducción Asistida ORG]. [URL del sitio web.](#)

(“Se considera prematuro un bebé”, 2023). Nacimientos Prematuors. [Organización Mundial de la Salud]. [URL del sitio web.](#)

GRACIAS

