

# Gender Recognition by Voice

## Ponentes:

Edwin Molina

Walter Pineda

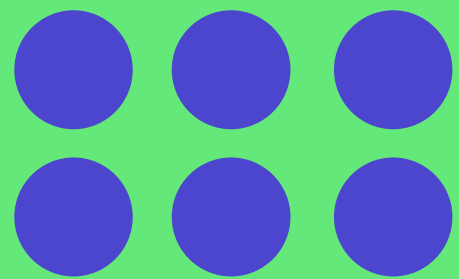


Proyecto final Curso Big Data & BI  
Intermediate

Universidad Veritas



# RECONOCIMIENTO DEL GENERO POR VOZ



Utilizando características acústicas de la voz (frecuencia media, skew, kurtosis, entre otros, para identificar el género de una persona.

Realizar comparaciones de modelos para encontrar patrones de reconocimiento precisos.

¿Se puede determinar el género de una persona mediante la voz?

¿Que características de la voz son las más eficientes o precisas para determinar el género de una persona?

# Dataset & Características

Este trabajo retoma una investigación realizada previamente acerca del diseño de un modelo computacional programado para determinar el género de una persona utilizando características de la voz.

El modelo se construye a partir de 3.168 muestras grabadas de voces, discursos y expresiones masculinas y femeninas. Las muestras se procesan mediante un análisis acústico.

# Características utilizadas

Para analizar el género de la voz y el habla, se necesita de una base de datos. Se utilizó una base de datos con miles de muestras de voces masculinas y femeninas, cada una de ellas etiquetada por su género de hombre o mujer.

# Características utilizadas

- **meanfreq**: frecuencia media (kHz)
- **sd**: desviación estandar de la frecuencia
- **median**: frecuencia mediana (kHz)
- **Q25**: primer cuantil (kHz)
- **Q75**: tercer cuantile (kHz)
- **IQR**: rango intercuantile (kHz)
- **skew**: sesgo/asimetría
- **kurt**: kurtosis
- **sp.ent**: entropía espectral
- **sfm**: planitud espectral
- **mode**: frecuencia moda
- **centroid**: centroide de la frecuencia
- **peakf**: frecuencia pico
- **meanfun**: promedio de la frecuencia fundamental medida a través de la señal acústica.
- **minfun**: frecuencia mínima fundamental medida a través de la señal acústica.
- **maxfun**: máxima frecuencia fundamental medida a través de la señal acústica.
- **meandom**: promedio de la frecuencia dominante medida a través de la señal acústica.
- **mindom**: mínimo de la frecuencia dominante medida a través de la señal acústica.
- **maxdom**: máximo de la frecuencia dominante medida a través de la señal acústica.
- **dfrange**: rango de frecuencia dominante medido a través de la señal acústica.
- **genero**: Masculino o femenino

# Preparación de los datos

06

## Metodología para limpieza de datos

Se realizó una limpieza y transformación de datos, el dataset original para este trabajo tenía cambios considerables comparados con el dataset original de la investigación.

*Dataset original*

	IQR	skew	kurt	sp.ent	sfm
193	0.075122	12.863462	274.402906	0.893369	0.491919
666	0.073252	22.423285	634.613855	0.892193	0.513726
908	0.123207	30.757155	1024.927705	0.846389	0.478900
955	0.111374	1.232831	4.177296	0.963322	0.727235
045	0.127325	1.101174	4.333713	0.971955	0.783564

*Dataset*

	IQR	skew	kurt	sp.ent
193	0.075122	128.634618	371.626	274.402905
666	0.073252	224.232853	628.204	634.613854
908	0.123207	307.571545	800.584	1.024927
955	0.111374	123.283127	554.372	417.729621
045	0.127325	110.117366	551.545	43.337131

# Transformación & Limpieza

07

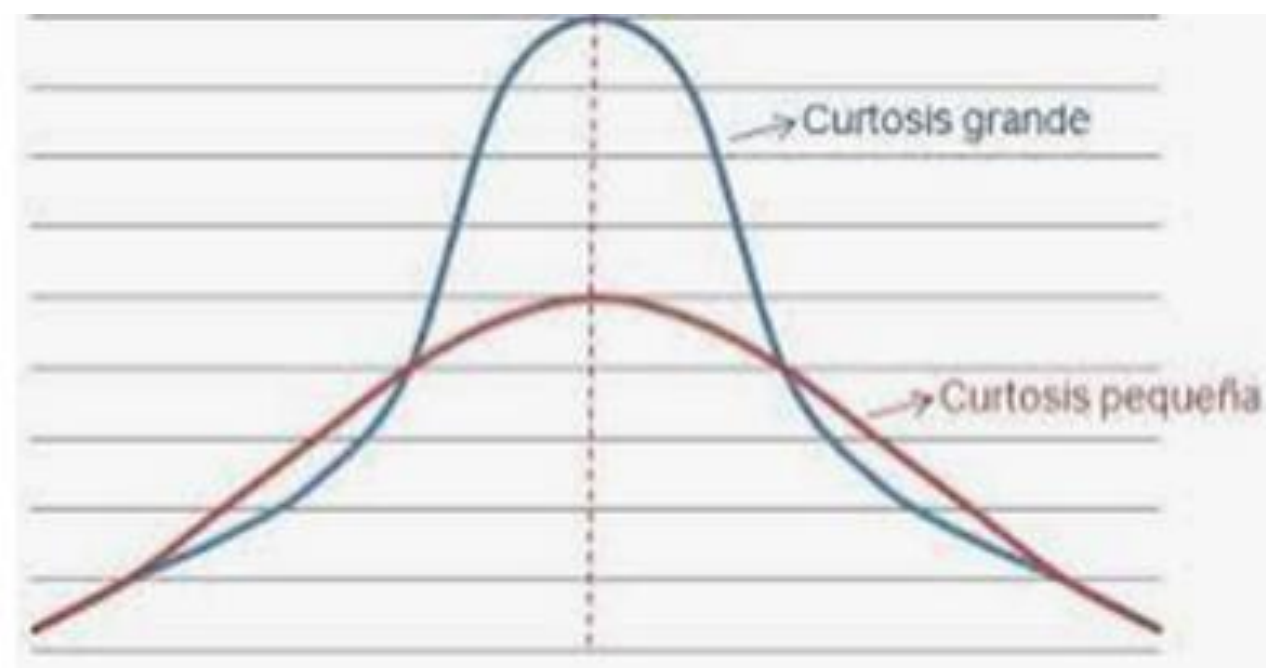
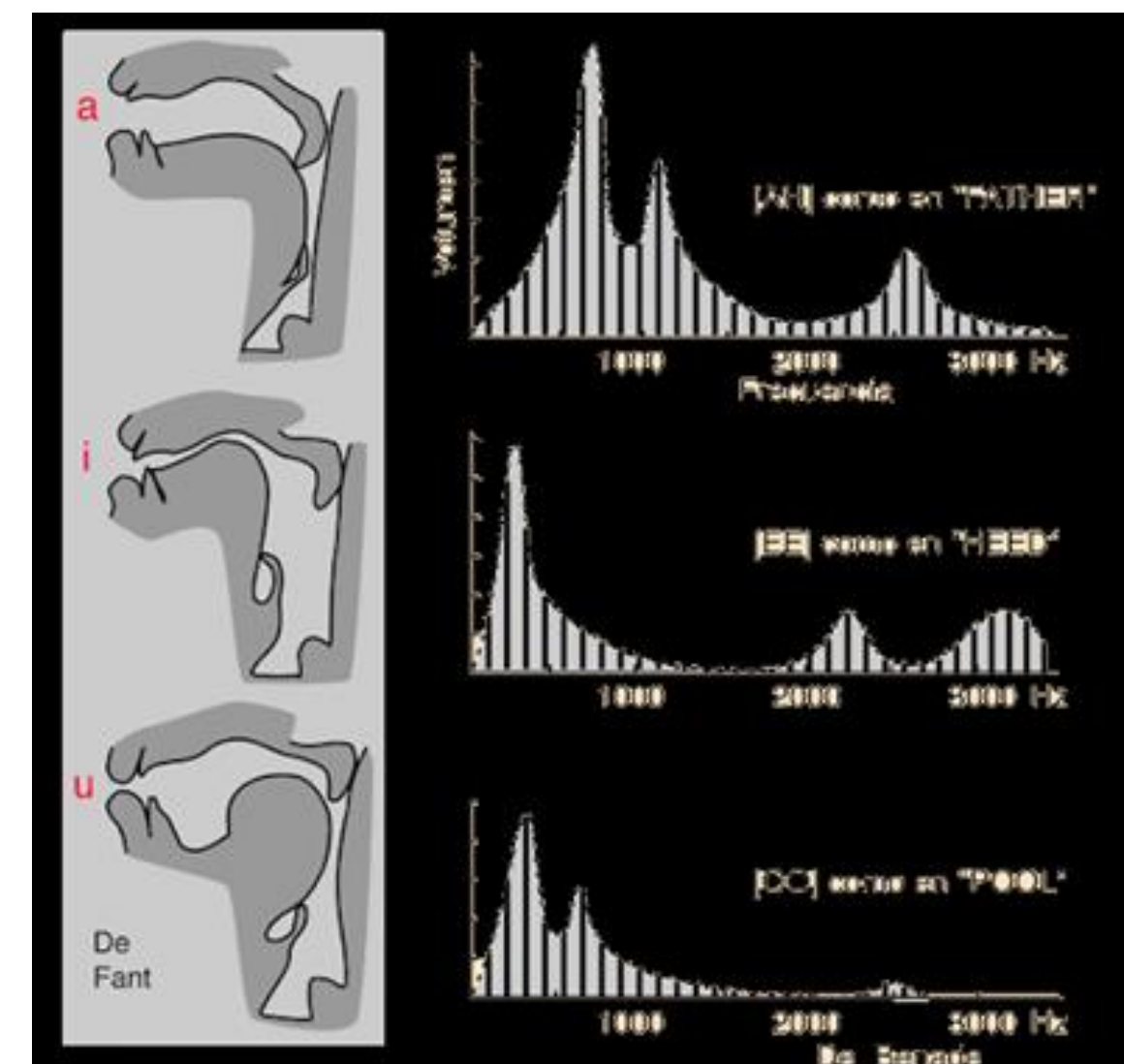
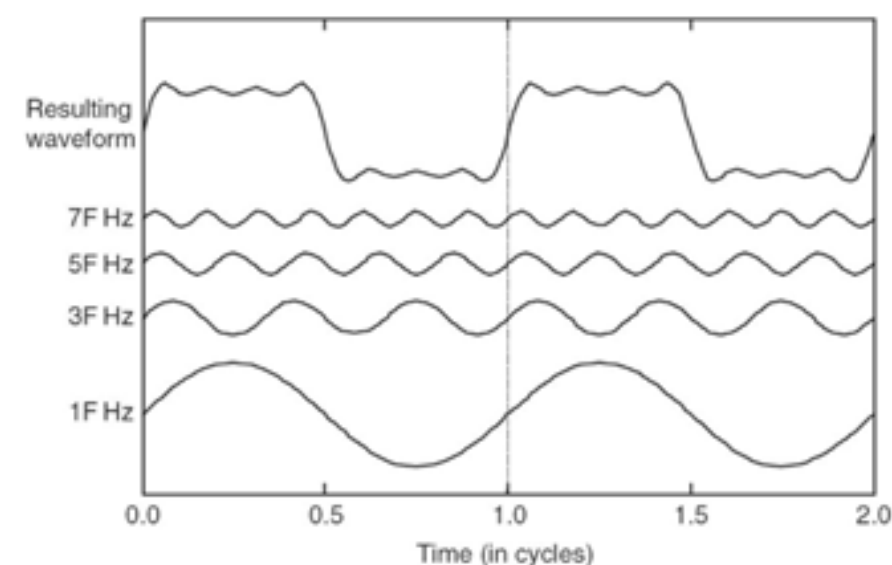
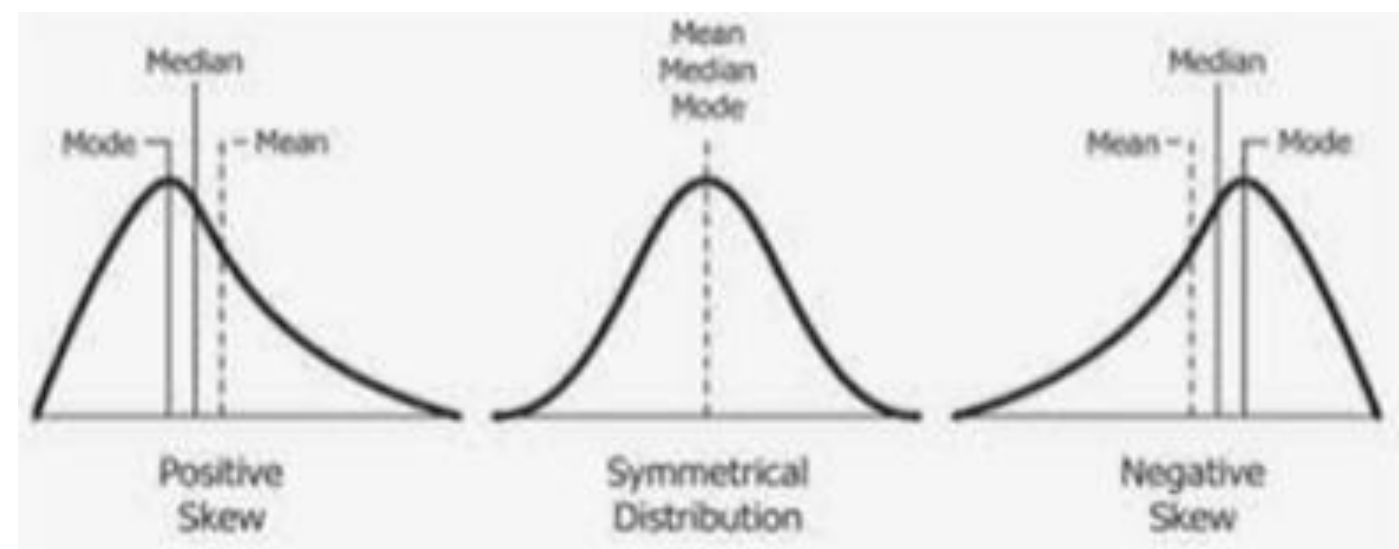
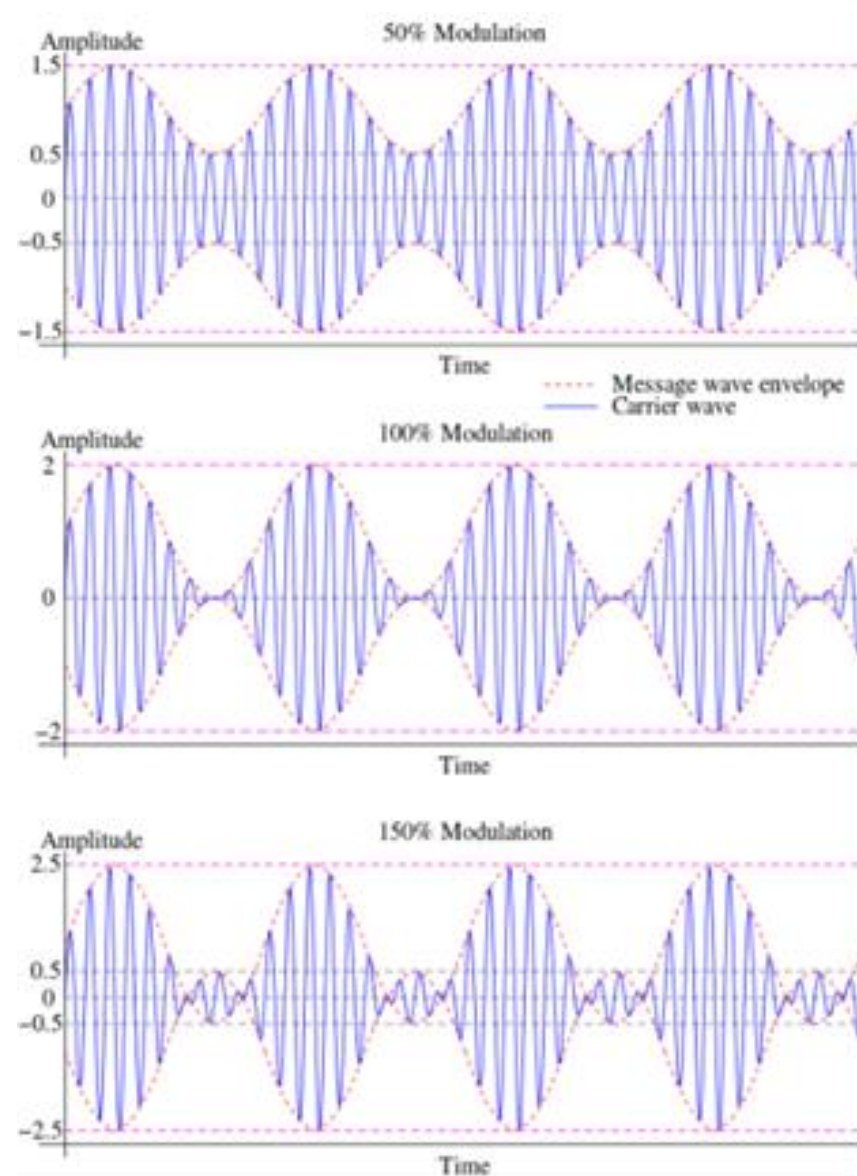
## Problema

- Transformación de variables de object a float
- Verificación y limpieza de datos nulos
- Imputación de datos faltantes

## Resultado

- Todas las columnas con variables correctas
- No hay datos nulos ni faltantes
- Dataset igual al dataset original (Kaggle)



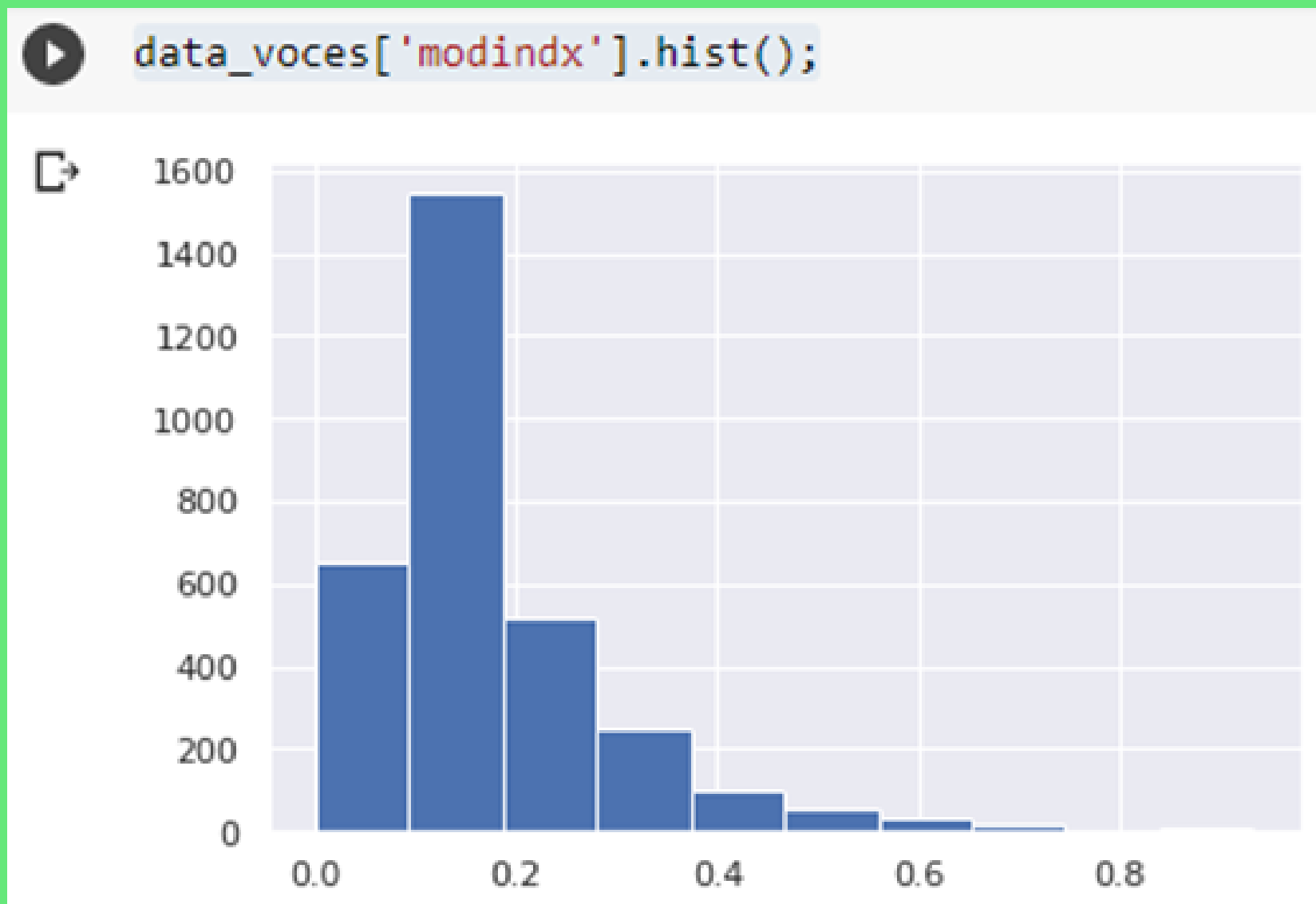


**Analisis y modelado de las variables del dataset**



# Evaluación de Variables

Utilización del describe para identificar datos estadísticos.



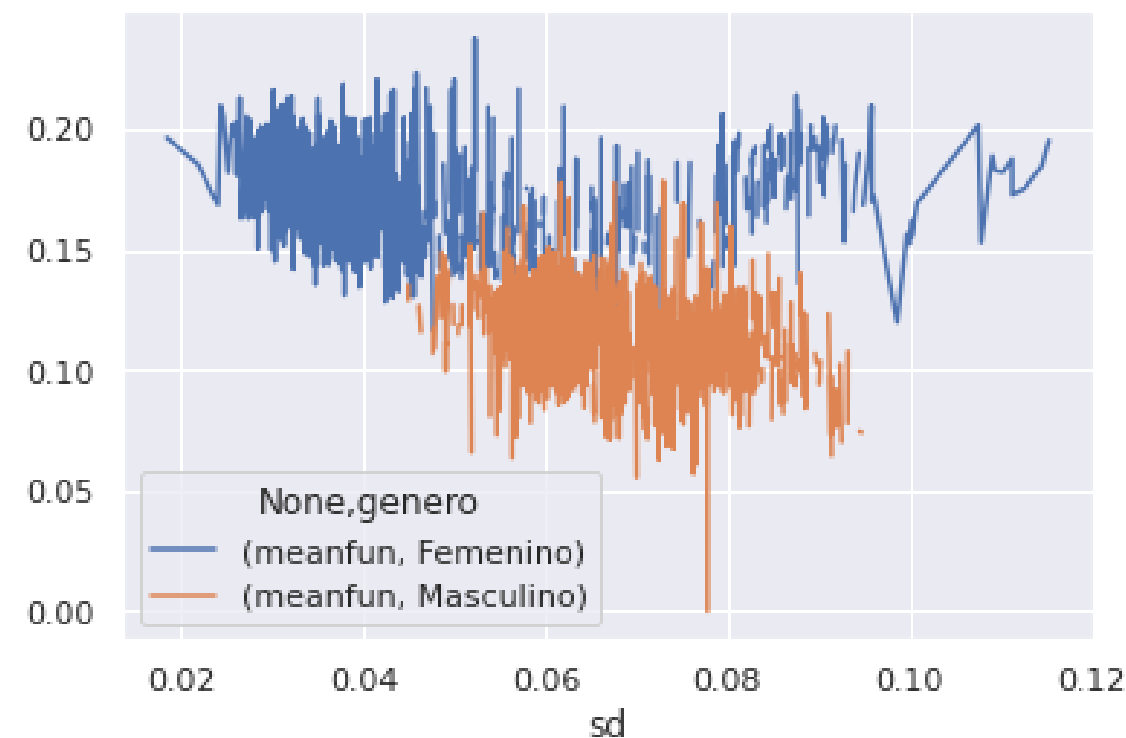
```
[153] data_voces['meanfreq'].describe()
```

count	3168.000000
mean	0.180907
std	0.029918
min	0.039363
25%	0.163662
50%	0.184838
75%	0.199146
max	0.251124
Name: meanfreq, dtype: float64	

A primera vista la variable modindx (índice de modulación) parece ser bastante importante por su definición y teórica dentro de la identificación de voces. Esto se definirá posteriormente en el análisis de relación de datos. Por el momento encontramos una distribución asimétrica positiva bastante marcada, teniendo prácticamente la mitad de los datos entra entre 0.1 y 0.2.

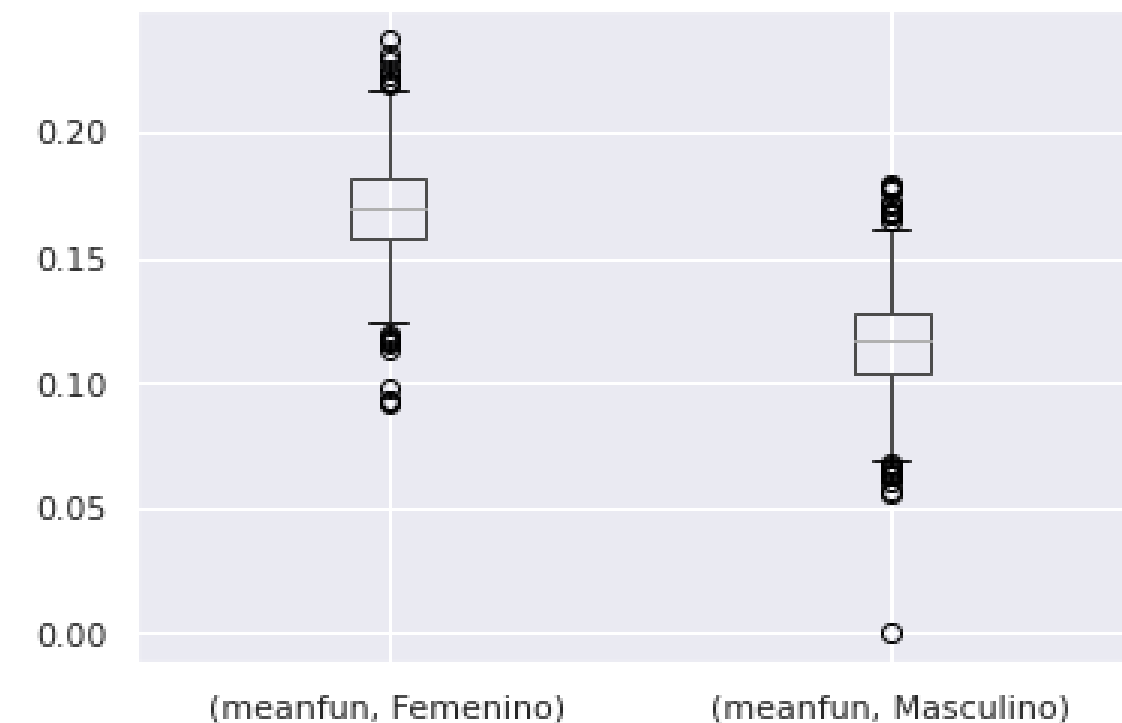
```
✓ [179] data_voces.pivot_table(index= ['genero'], values = ['meanfreq'], aggfunc='mean')
```

meanfreq	
genero	
Femenino	0.191000
Masculino	0.170813



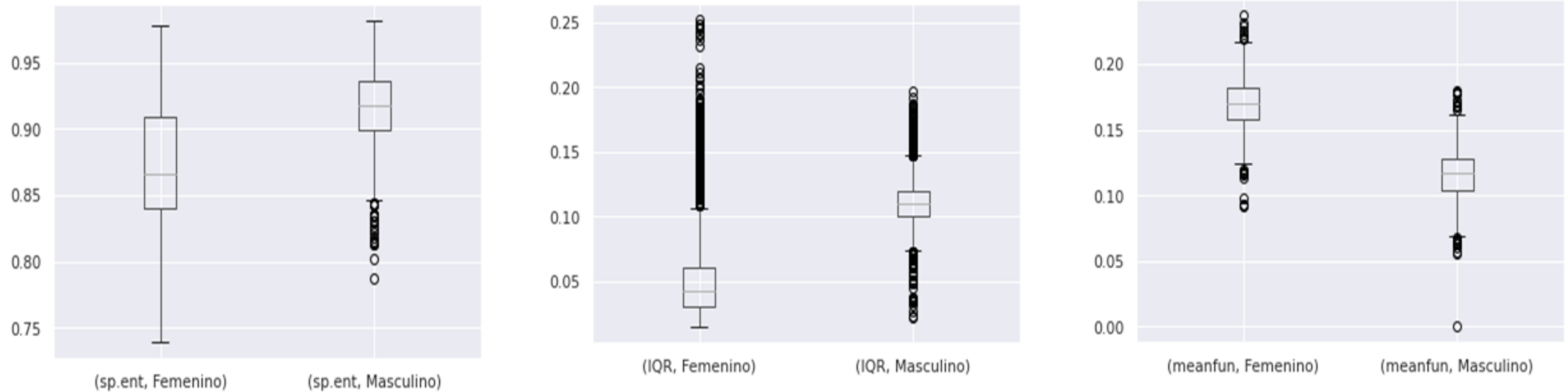
```
✓ [180] data_voces.pivot_table(index= ['genero'], values = ['median'], aggfunc='std')
```

median	
genero	
Femenino	0.032149
Masculino	0.037392



Meanfun VRS sd: la desviación estandar tiende a disminuir cuando el promedio de frecuencia fundamental aumenta, de manera parecida a la relación con meanfreq pero de una manera menos pronunciada. Podemos confirmar que los datos tienden a estabilizarse cuanto mas altos son (meanfun). Este gráfico muestra la importancia mencionada de esta variable (meanfun), las variables de hombres y mujeres estan en su gran mayoría separadas y se relacionan traslapan de manera casi inexistente. Estos ultimos datos mencionados podemos inclusive llegarlos a tomar como datos outliers.

# Metodología establecida de % de detección



Este tipo de tablas ya se realizan con un trasfondo de analisis para fundamentar un modelo CART. Podemos observar que la variable IQR, tiene una cantidad de datos importantes que solo las mujeres y hombres pueden llegar a tener, siendo un 65.52 % probable el analizar e identificar a un genero con esta información.

Este tipo de tablas ya se realizan con un trasfondo de analisis para fundamentar un modelo CART. Podemos observar que la variable sp.ent, tiene una cantidad de datos importantes que solo las mujeres pueden llegar a tener, siendo un 51.67 % probable el analizar e identificar a un genero con esta información.

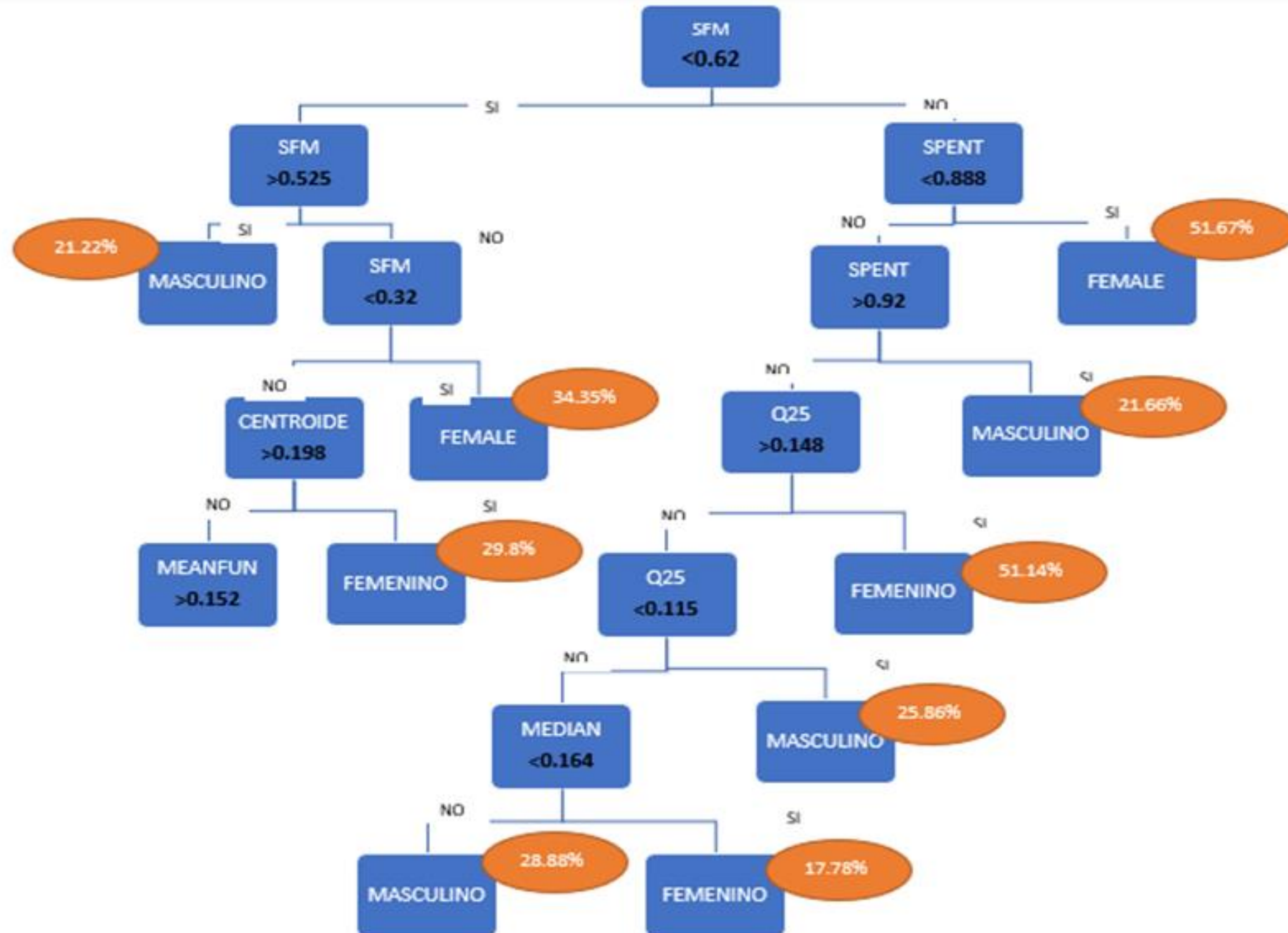
# Tabla de decisiones y probabilidades de detección

12

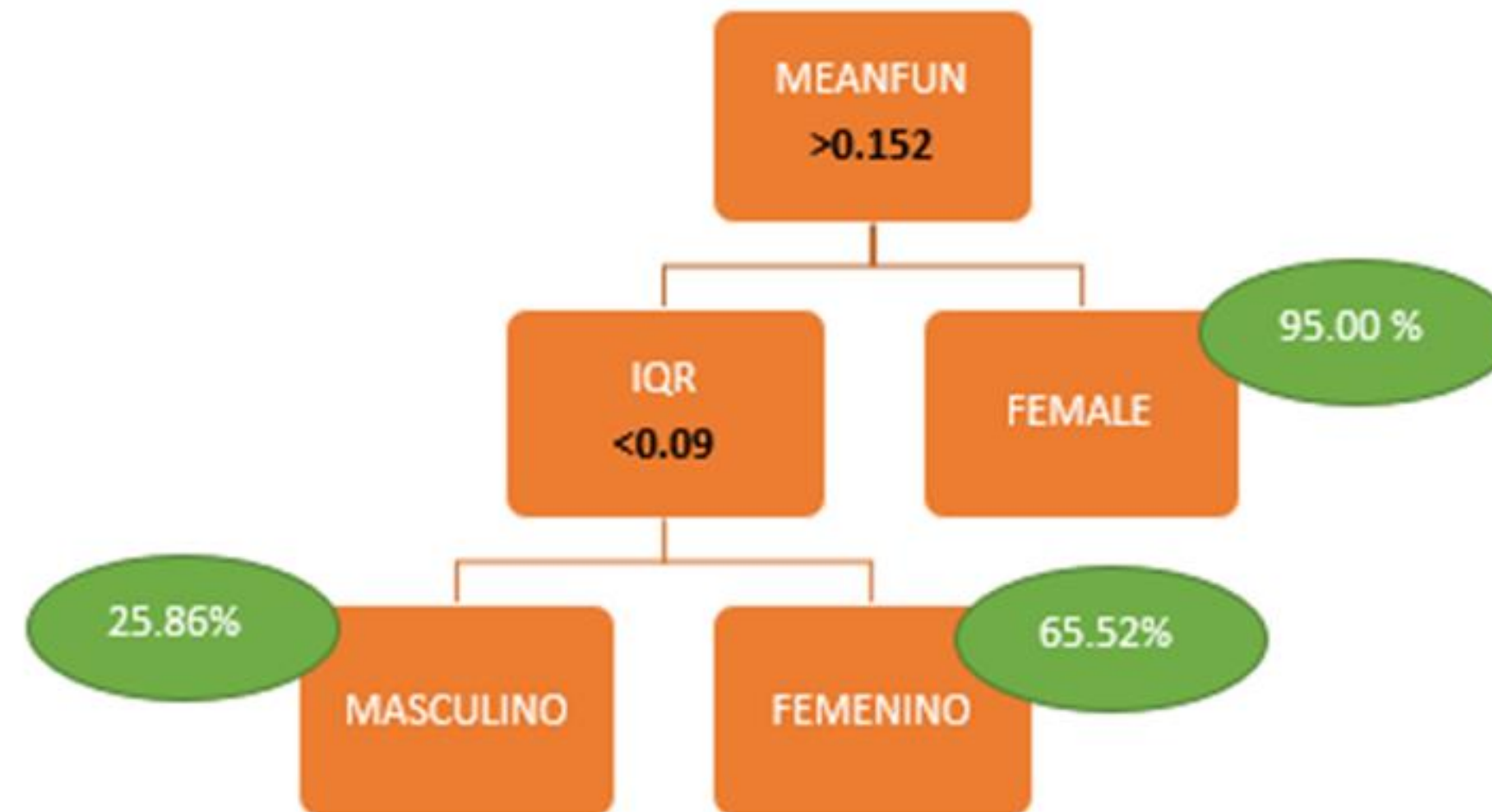
VARIABLES	MEAN		SD		MEAN+-SD				CONCLUSIONES		% DE DETECCIÓN		TOTAL % DETECCIÓN
	MASCULINO	FEMENINO	MASCULINO	FEMENINO	MASCULINO		FEMENINO		MASCULINO	FEMENINO	MASCULINO	FEMENINO	
MEANFREQ	0.17	0.19	0.026	0.03	0.144	0.196	0.16	0.22	<0.16	>0.196	21.05	31.58	52.63
SD	0.06	0.05	0.009	0.018	0.051	0.069	0.032	0.068	>0.068	<0.051	2.7	51.35	54.05
MEDIAN	0.175	0.196	0.037	0.032	0.138	0.212	0.164	0.228	<0.164	>0.212	28.88	17.78	46.66
Q25	0.116	0.165	0.032	0.05	0.084	0.148	0.115	0.215	<0.115	>0.148	23.66	51.14	74.8
Q75	0.226	0.223	0.024	0.023	0.202	0.25	0.2	0.246	>0.246	<0.202	8	4	12
IQR	0.111	0.058	0.02	0.043	0.091	0.131	0.015	0.101	>0.101	<0.091	25.86	65.52	91.38
SKEW	22.23	24.923	17.547	14.762	4.683	39.777	10.161	39.685	<10.161	0	15.61	0	15.61
KURT	435.443	419.067	262.8	305.97	172.643	698.243	113.097	725.037	0	698.243<X<172.643	0	14.44	14.44
SP.ENT	0.917	0.873	0.029	0.047	0.888	0.946	0.826	0.92	>0.92	<0.888	21.66	51.67	73.33
SFM	0.472	0.345	0.15	0.18	0.322	0.622	0.165	0.525	>0.525	<0.322	21.22	34.35	55.57
MODE	0.152	0.178	0.084	0.067	0.068	0.236	0.111	0.245	<0.111	>0.236	24.29	5.08	29.37
CENTROIDE	0.171	0.191	0.027	0.03	0.144	0.198	0.161	0.221	<0.161	>0.198	22.08	29.87	51.95
PEAKF													0
MEANFUN	0.11	0.17	0.017	0.018	0.093	0.127	0.152	0.188	TOTAL	TOTAL	TOTAL	TOTAL	#¡VALOR!
MINFUN	0.034	0.039	0.016	0.022	0.018	0.05	0.017	0.061	0	>0.05	0	25	25
MAXFUN	0.253	0.264	0.039	0.022	0.214	0.292	0.242	0.286	0.286<X<0.242	0	43.58	0	43.58
MEANDOM	2.59E+14	3.16E+14	3.18E+14	2.78E+14	0	5.769E+14	3.81E+13	5.939E+14	<3.81E+13	>5.76E+14	6.42	2.86	9.28
MINDOM	0.04	0.065	0.049	0.073	0	0.089	0	0.138	0	>0.089	0	35.51	35.51
MAXDOM	82.417	2852.58	434.254	44864.353	0	516.671	0	47716.933	0	516.671	0	0	0
DFRANGE	84.22	3592.97	449.737	66819.585	0	533.957	0	70412.555	0	>533.957	0	0	0
MODINDX	0.17	0.177	0.13	0.108	0.04	0.3	0.069	0.285	0.285<X<0.069	0	16.92	0	16.92



# Modelo de árbol de decisión de clasificación y regresión (CART)

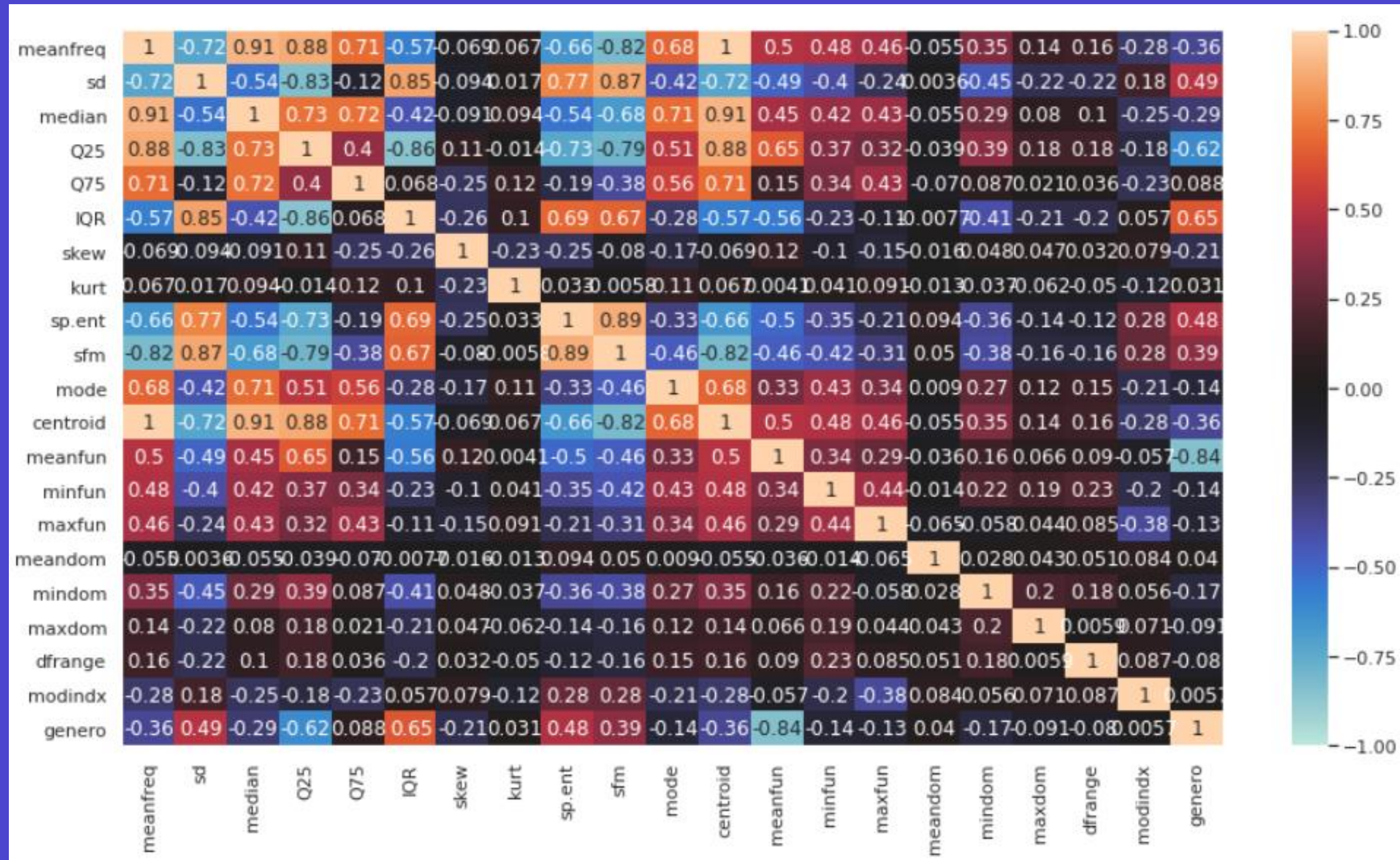


# Modelo de árbol de decisión de clasificación y regresión 2 (CART)





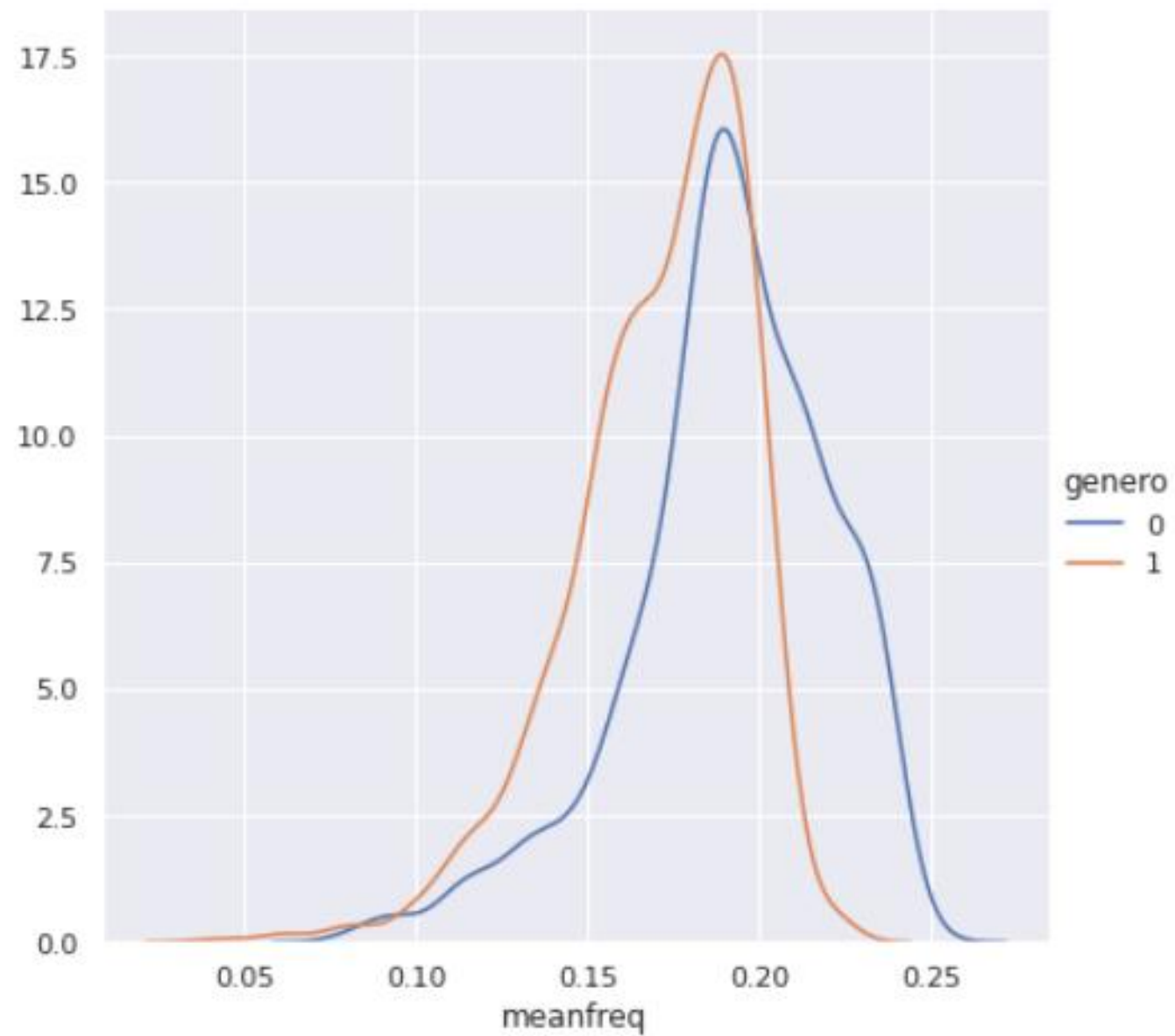
# Matriz de correlación



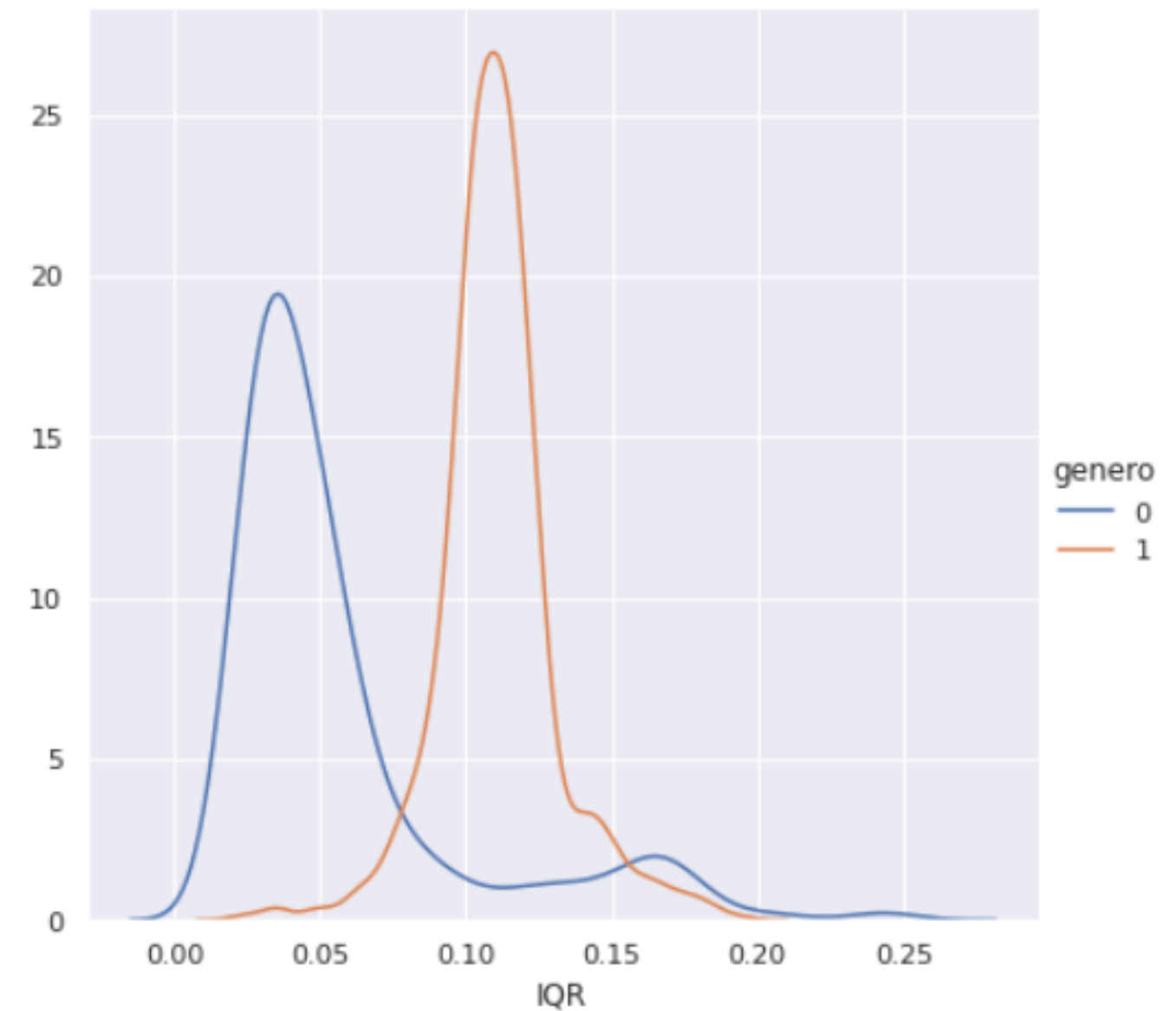


# Variables para clusterización

Frecuencia media

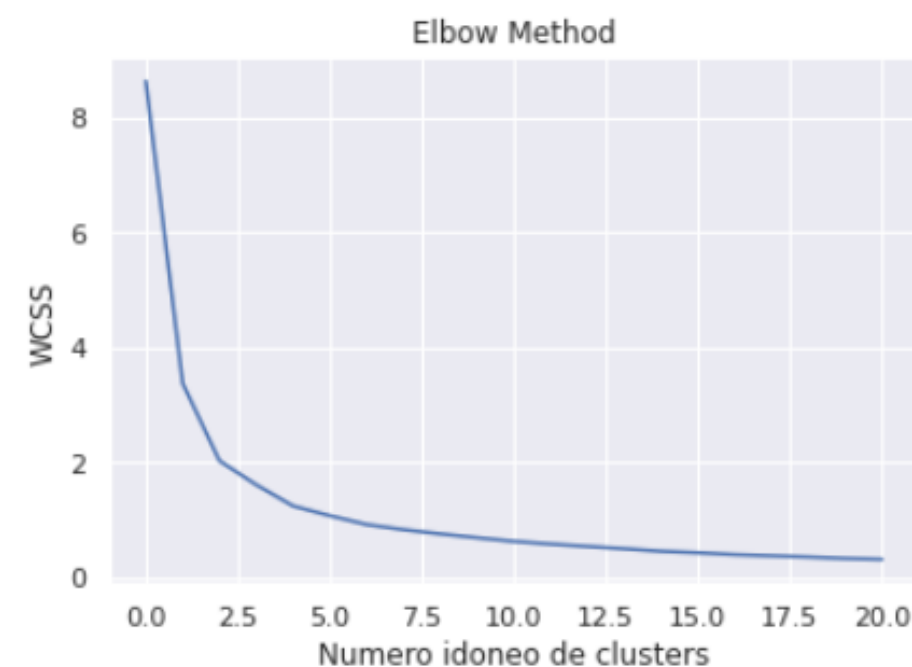


Rango intercuantil



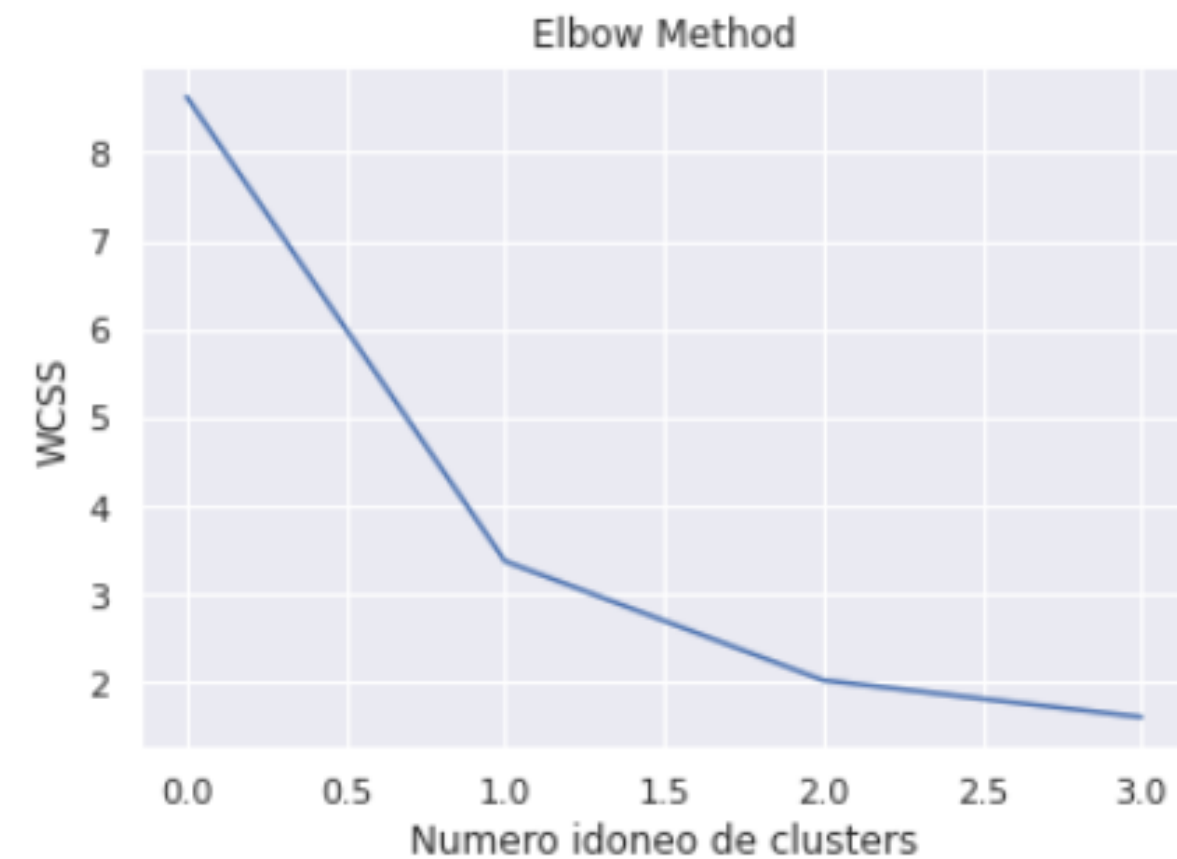
# Clusterización

```
Iteracion N Clusters: k: 1  
Iteracion N Clusters: k: 2  
Iteracion N Clusters: k: 3  
Iteracion N Clusters: k: 4  
Iteracion N Clusters: k: 5  
Iteracion N Clusters: k: 6  
Iteracion N Clusters: k: 7  
Iteracion N Clusters: k: 8  
Iteracion N Clusters: k: 9  
Iteracion N Clusters: k: 10  
Iteracion N Clusters: k: 11  
Iteracion N Clusters: k: 12  
Iteracion N Clusters: k: 13  
Iteracion N Clusters: k: 14  
Iteracion N Clusters: k: 15  
Iteracion N Clusters: k: 16  
Iteracion N Clusters: k: 17  
Iteracion N Clusters: k: 18  
Iteracion N Clusters: k: 19  
Iteracion N Clusters: k: 20  
Iteracion N Clusters: k: 21
```



```
# k optima para cantidad de clusters a visualizar  
  
epsilon = 0.30  
_, k = elbow_method(epsilon, figure= True)
```

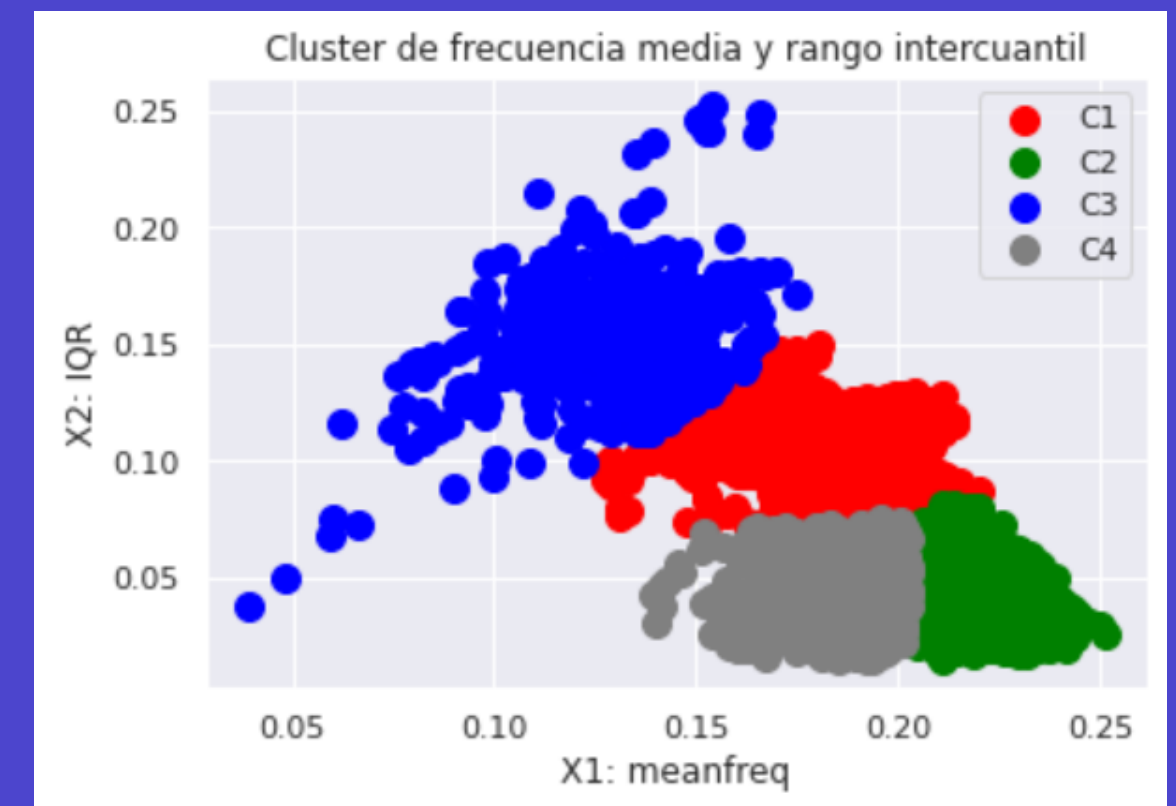
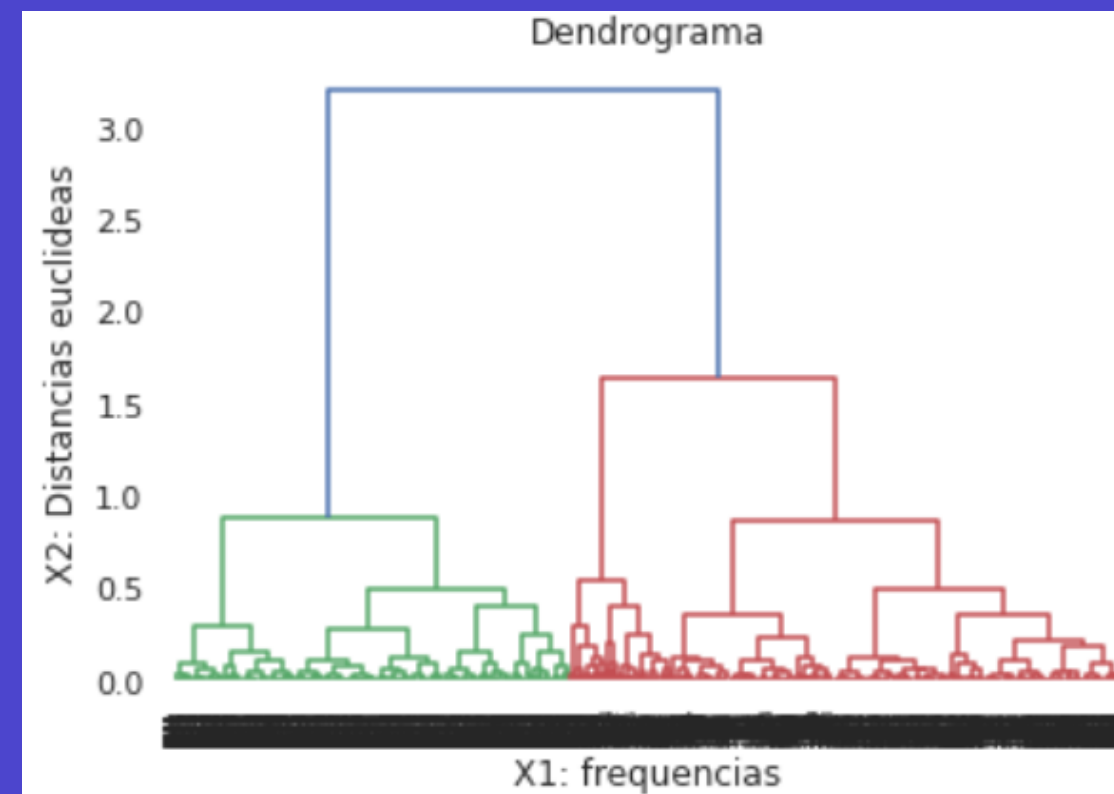
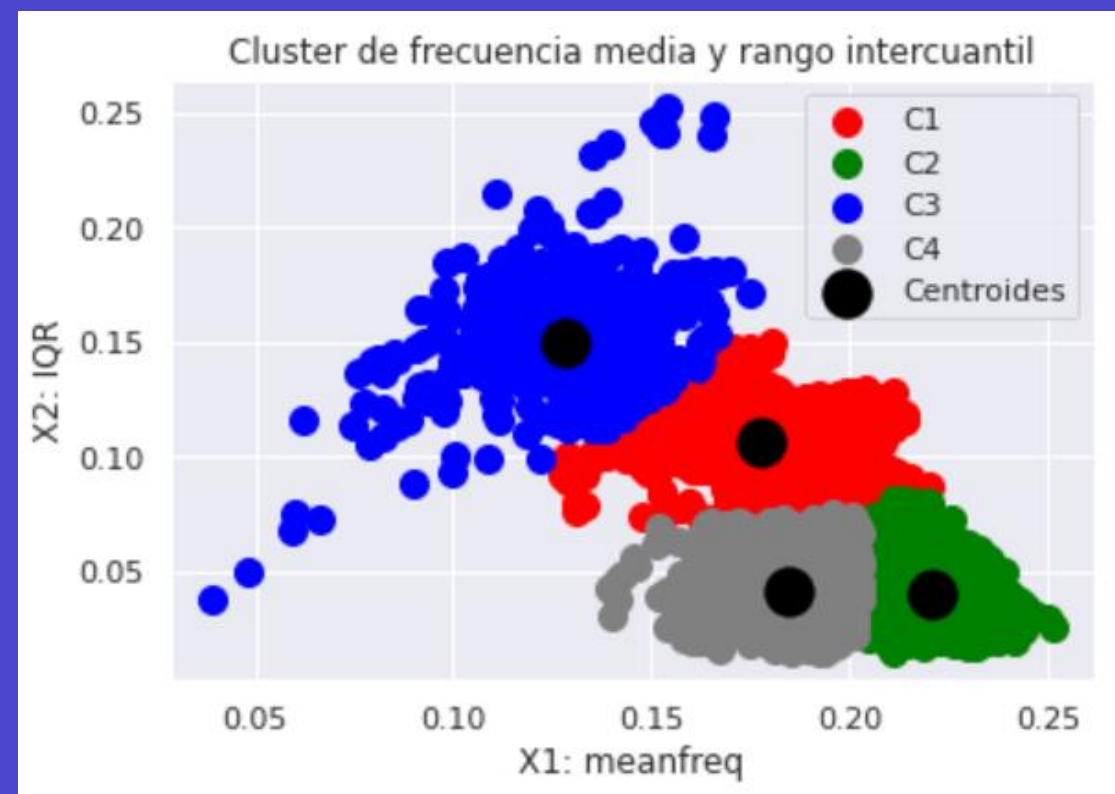
```
Iteracion N Clusters: k: 1  
Iteracion N Clusters: k: 2  
Iteracion N Clusters: k: 3  
Iteracion N Clusters: k: 4
```



Se obtiene una k optima de 4 clusters para este analisis.

# Machine Learning

## K-means (Elbow Method) - Dendrograma



# Resultados y conclusiones finales

- Se obtiene una precisión del modelo original con el modelo resuelto para este proyecto.
- Utilizar diferentes variables acusticas de la voz ayuda a que el modelo pueda ser exitoso en cuanto al reconocimiento de voz.
- La mayoría de datos tiene una relación directa con la variable género. Sin embargo, se deben buscar metodologías para poder seleccionar numéricamente las variables que se adecuen a un modelo de decisión. En este caso, la probabilidad de detección por género, media y desviación estándar. Definiendo un modelo adecuado, existe una alta probabilidad de tener éxito en la detección de voz.
- Los datos y análisis estadísticos (máximos, mínimos, media, histogramas, Anova, etc) nos sirven para comprender los datos en un dataset. La estrategia para resolver el problema debe ser definida por el analista, seleccionando la mejor opción para brindar datos óptimos al cliente.

# Sugerencias finales

- Para la resolución del caso “identificación de género por voz”, se sugiere la utilización de un filtro para aclarar el sonido detectado y disminuir los sonidos de ambiente. Esto generaría probabilidades más altas de detección.
- Realizar un análisis final de las variables, según las conclusiones, siempre es importante para brindar las mejores opciones al problema. Aunque numéricamente puede darnos una tendencia, depende de los costos de implementación o dificultad de realizar las propuestas.



**Muchas gracias.**