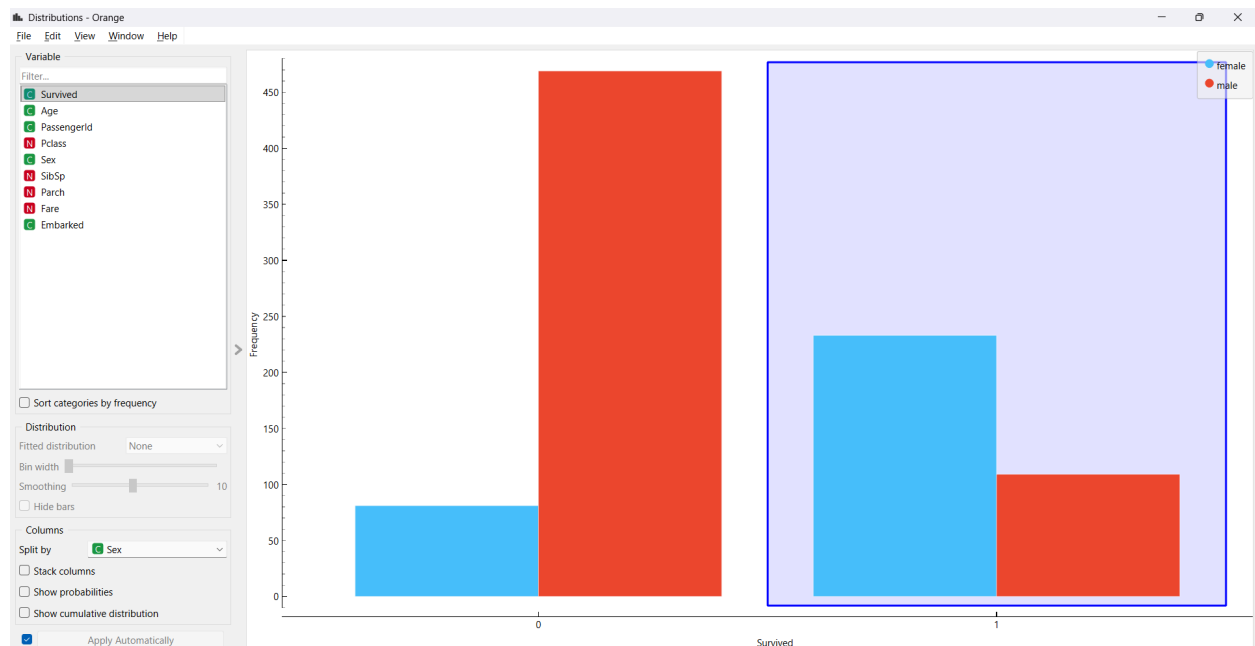


ANÁLISIS DATASET TITANIC

EN ORANGE

La primer división importante que miramos en el dataset fue la variable Supervivencia segmentada por el sexo de los pasajeros, los rojos siendo sexo masculino y los azules siendo sexo femenino. Está dividido en sobrevivieron (1) y no sobrevivieron (0)



Analizando el gráfico de distribución se observan dos puntos principales. En primer lugar: entre los que no sobrevivieron (0) predomina ampliamente el sexo masculino. De esta manera, también podemos ver que entre quienes sí sobrevivieron (1), la mayoría fueron mujeres.

Si además de contemplar la información del dataset analizamos lo acontecido, los resultados tienen confirmación el patrón histórico del Titanic. Donde las mujeres tuvieron prioridad en el rescate. La diferencia es muy marcada visualmente, ya que la barra roja (hombres) es mayor en fallecidos y la barra celeste (mujeres) es mayor en sobrevivientes.

De esta manera, este resultado indica que el sexo es una variable fuertemente asociada al resultado de supervivencia.

Otro aspecto que podemos ver y analizar en el modelo son los falsos positivos (donde la predicción decía que iba a sobrevivir y realmente no sobrevivió), en este caso 14 personas fueron un falso positivo. En esta variable, lo que tenemos que pensar es: ¿Qué es peor? que te digan que va a morir y que sobreviva o que no va a morir y que se termine muriendo.

		Predicted		
		0	1	Σ
Actual	0	102	14	116
	1	19	43	62
Σ		121	57	178

El modelo tiene una precisión de 0,812, es decir que las predicciones que realiza el modelo son bastante buenas, aunque hay espacio de mejora ajustando algunos valores o probando con otros modelos. Igualmente, es una precisión bastante alta

Predictions - Orange													
File View Window Help													
Show probabilities for Classes in data <input checked="" type="checkbox"/> Show classification errors													
	Tree	error	Survived	Name	Ticket	Cabin	Age	Pclass	Sex	SibSp	Parch	Fare	Embarked
1	0.88 : 0.12 → 0	0.875	1	Barah, Mr. Han...	2663	?	20	3	male	0	0	7.2292	C
2	0.00 : 1.00 → 1	0.000	1	Harper, Mr. Ha...	PC 17572	D33	48	1	male	1	0	76.7292	C
3	1.00 : 0.00 → 0	1.000	1	Hamalainen, M...	250649	?	0.67	2	male	1	1	14.5000	S
4	0.01 : 0.99 → 1	0.012	1	Duff Gordon, La...	11755	A16	48	1	female	1	0	39.6000	C
5	0.83 : 0.17 → 0	0.167	0	Svensson, Mr. J...	347060	?	74	3	male	0	0	7.7750	S
6	0.01 : 0.99 → 1	0.012	0	Silvey, Mrs. Wil...	113507	?	39	1	female	1	0	55.9000	S
7	0.25 : 0.75 → 1	0.250	1	Backstrom, Mrs...	3101278	?	33	3	female	3	0	15.8500	S
8	0.87 : 0.13 → 0	0.133	0	Coleridge, Mr...	W./C. 14263	?	29	2	male	0	0	10.5000	S
9	0.67 : 0.33 → 0	0.333	0	Thayer, Mr. Joh...	17421	C68	49	1	male	1	1	110.8833	C
10	1.00 : 0.00 → 0	0.000	0	Sage, Master. T...	CA 2343	?	?	3	male	0	2	69.5500	S
11	0.88 : 0.12 → 0	0.125	0	Kassem, Mr. Far...	2700	?	?	3	male	0	0	7.2292	C
12	0.33 : 0.67 → 1	0.667	0	Holmerson, Mr...	113789	?	42	1	male	1	0	52.0000	S
13	0.01 : 0.99 → 1	0.012	1	Crosby, Miss. H...	WE/P 5735	B22	36	1	female	0	2	71.0000	S
14	1.00 : 0.00 → 0	0.000	0	Van der hoef,	111240	B19	61	1	male	0	0	33.5000	S
15	0.00 : 1.00 → 1	1.000	0	Hegarty, Miss...	365226	?	18	3	female	0	0	6.7500	Q
16	0.83 : 0.17 → 0	0.167	0	Eklund, Mr. Han...	347074	?	16	3	male	0	0	7.7750	S
17	0.75 : 0.25 → 0	0.250	0	Carlsson, Mr. A...	350042	?	28	3	male	0	0	7.7958	S
18	0.50 : 0.50 → 0	0.500	1	Marechal, Mr. P...	11774	C47	?	1	male	0	0	29.7000	C
19	0.81 : 0.19 → 0	0.194	0	Shorney, Mr. Ch...	374910	?	?	3	male	0	0	8.0500	S
20	0.81 : 0.19 → 0	0.194	0	Simmons, Mr. J...	SOTON/OQ 39...	?	?	3	male	0	0	8.0500	S
21	0.67 : 0.33 → 0	0.333	0	Zabour, Miss. T...	2665	?	?	3	female	1	0	14.4542	C
22	0.81 : 0.19 → 0	0.194	0	Rush, Mr. Alfre...	A/4. 20589	?	16	3	male	0	0	8.0500	S
23	1.00 : 0.00 → 0	1.000	1	Dick, Mr. Albert...	17474	?	31	1	male	1	0	37.0000	S
24	0.50 : 0.50 → 0	0.500	1	Anderson, Mr...	19952	E12	48	1	male	0	0	26.5500	S
25	1.00 : 0.00 → 0	1.000	1	Leeni, Mr. Fah...	2620	?	22	3	male	0	0	7.2250	C
Show performance scores Target class: (Average over classes)													
Model	AUC	CA	F1	Prec	Recall	MCC							
Tree	0.798	0.815	0.813	0.812	0.815	0.585							

A su vez, si miramos el CA nos da también alto, es 0.815, lo que quiere decir que el modelo acierta con las predicciones el 81% de veces, lo que es positivo y un indicador de un buen modelo

Para concluir, con este análisis pudimos ver como distintas variables influyen en la supervivencia de los pasajeros. El análisis dle modelo mostró que el sexo fue una de las variables más importantes. Pudimos ver que habría una gran diferencia entre las probabilidades de sobrevivir entre un sexo y el otro, siendo las mujeres quienes tuvieron una tasa de supervivencia mayor.