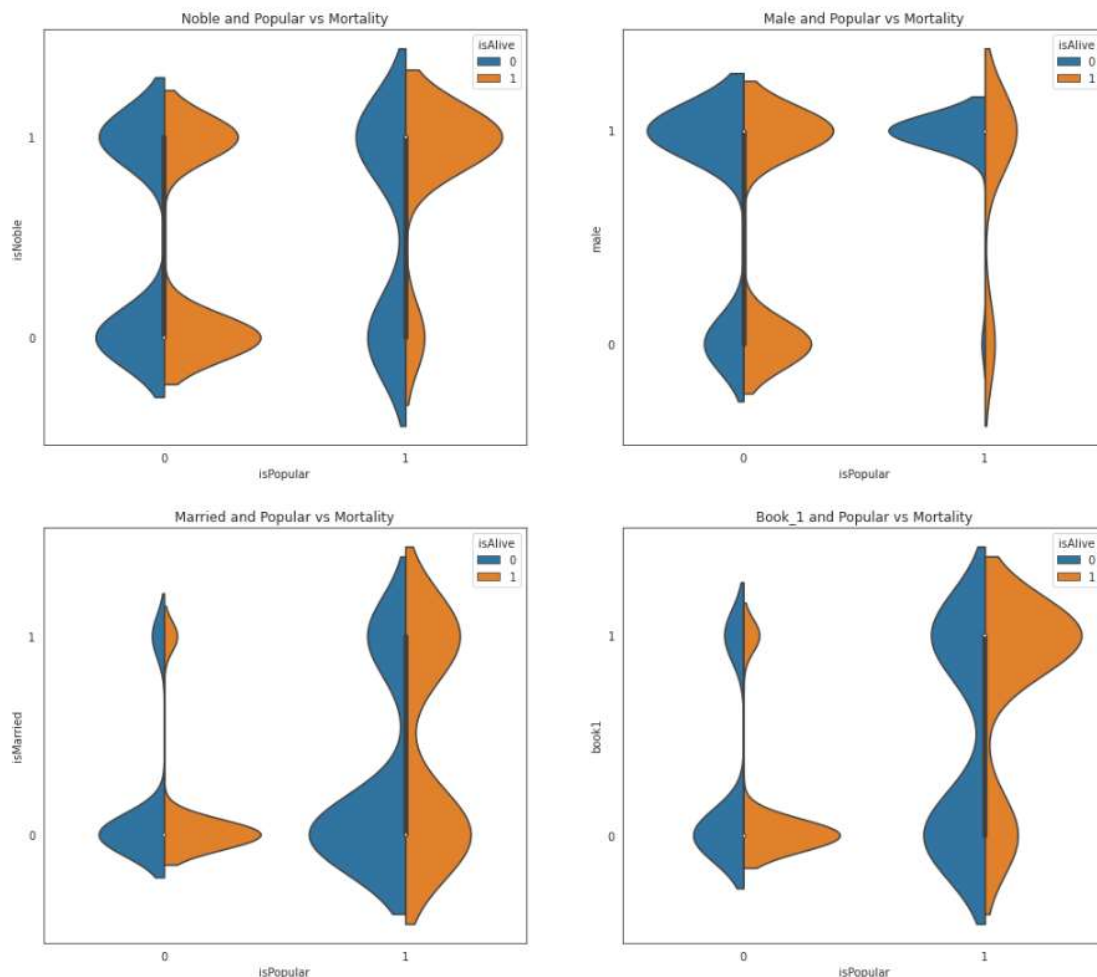


Examen SAA

Nombre: _____

Para el examen vamos a utilizar “carácter-predictions.csv” donde aparecen 1946 personajes de la serie "A Song of Fire and Ice", que consta de 5 libros, aunque parece ser que ya se está escribiendo un sexto "The Winds of Winter" y que se piensa en un séptimo "A Dream of Spring". El primero es "A Game of Thrones".

- 1) (0.5 ptos.) En el supuesto se usa *pandas.factorize()* para convertir atributos categóricos no numéricos a numéricos, escribe otra forma de hacer lo mismo.
- 2) (1 pto.) ¿Consideras que los datos de entrada están balanceados? Si no es así, ¿qué harías para que estuviesen balanceados?
- 3) (0.5 ptos.) ¿Se ha aplicado One-Hot Encoding en el supuesto? Escríbelo.
- 4) (0.5 ptos.) En un análisis previo, ¿qué podrías inferir del siguiente gráfico?



- 5) (0.5 ptos.) ¿Crees que se puede aplicar PCA con dos componentes? Analiza el resultado.
- 6) (1.5 ptos.) Aplica correlación para reducción de atributos, ¿con qué atributos te quedarías?

Examen SAA

Nombre: _____

- 7) (0.5 ptos.) En el supuesto no se ha aplicado regresión, en el caso de utilizar regresión, ¿qué sería mejor regresión lineal o logística? Explícalo
- 8) (0.5 ptos.) En Clustering se han considerado 5 clusters, ¿cómo se haya el número de clusters óptimo?
- 9) (1 pto.) ¿Por qué se ha aplicado RandomForestClassifier en vez de RandomForestRegressor?
- 10) (1.5 ptos.) ¿Cómo interpretas la matriz de confusión en RandomForestClassifier? Explícalo.
- 11) (0.5 ptos.) En redes neuronales ¿por qué se usa una función sigmoide como función de activación para la última capa?
- 12) (1 pto.) ¿Qué fallos hay en la red neuronal? Corrígelos y explica el problema.
- 13) (0.5 ptos.) De todos los métodos, ¿cuál es el que da mejor resultado?