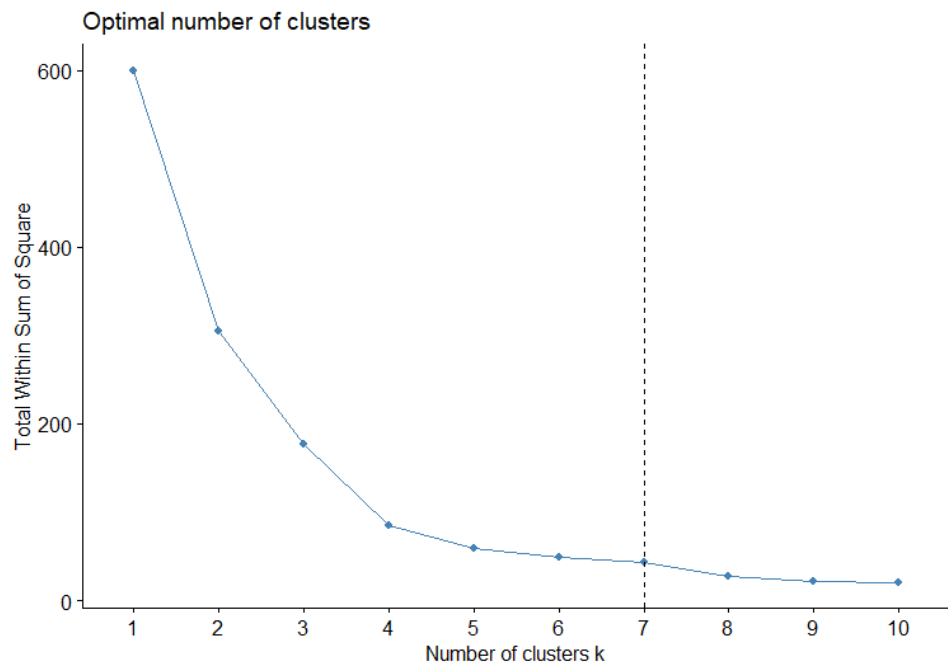


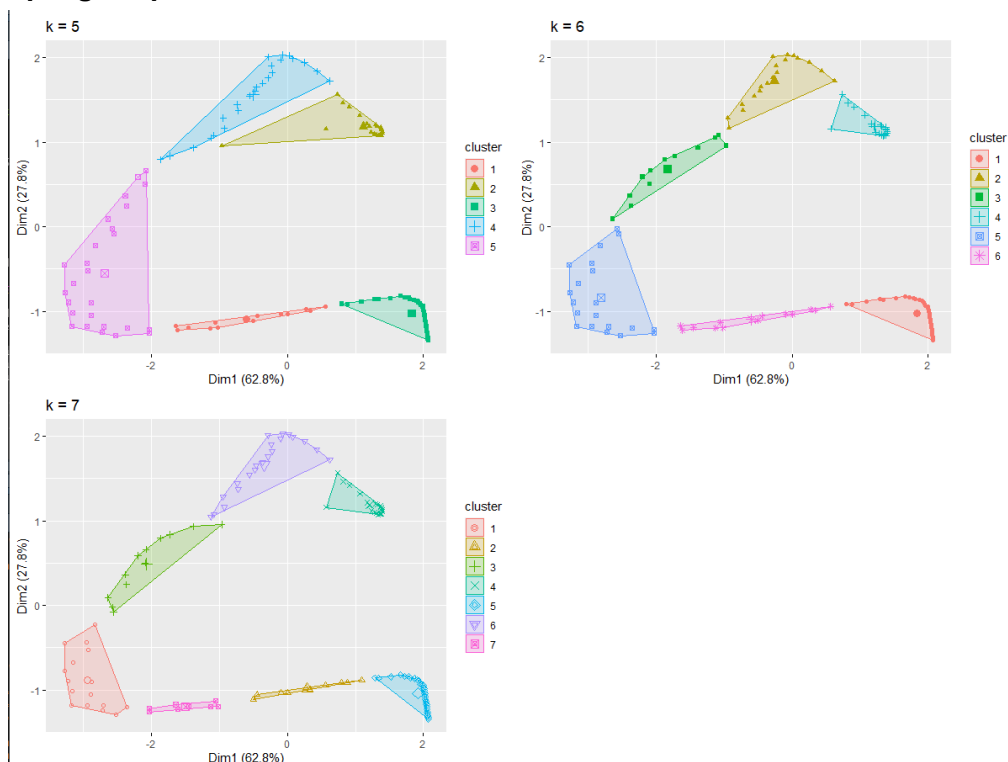
Parte II:

Justifique porque eligió esos tres valores de K

Previo a la creación de gráficas para analizar el comportamiento de los clusters, se realizó un análisis acerca de las cantidades de Clusters y su efecto significativo al agregar incrementalmente estos clusters, en el cual decidimos seleccionar a base de esto, una vista más a fondo entre 5 a 7 clusters ya que en la siguiente gráfica el punto de inflexión se encuentra entre el rango del 5 al 7.

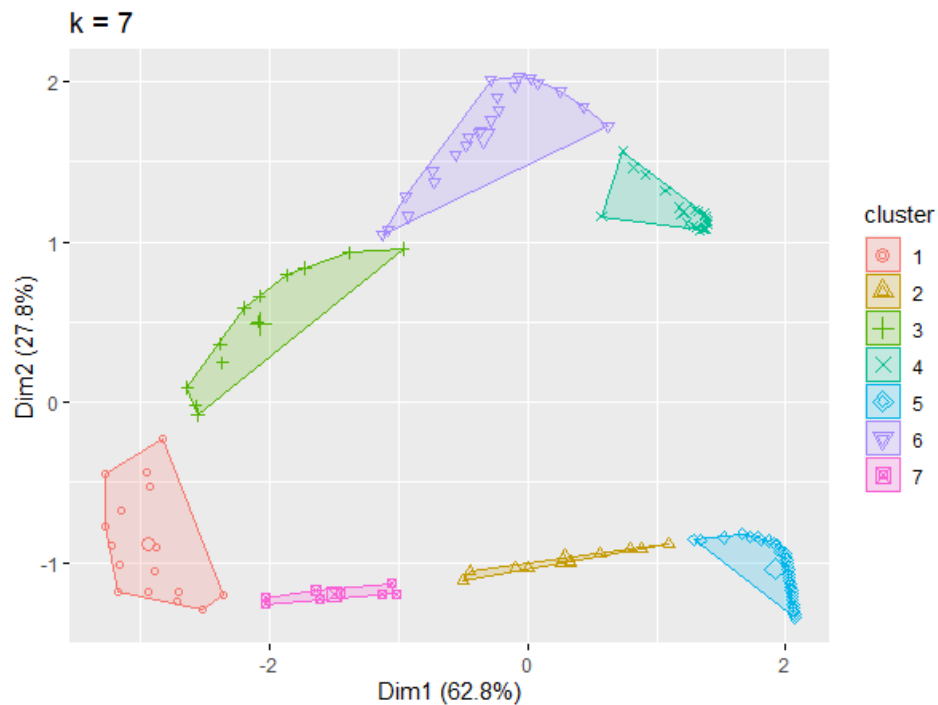


Un plot que grafiqué los tres diferentes modelos de k means

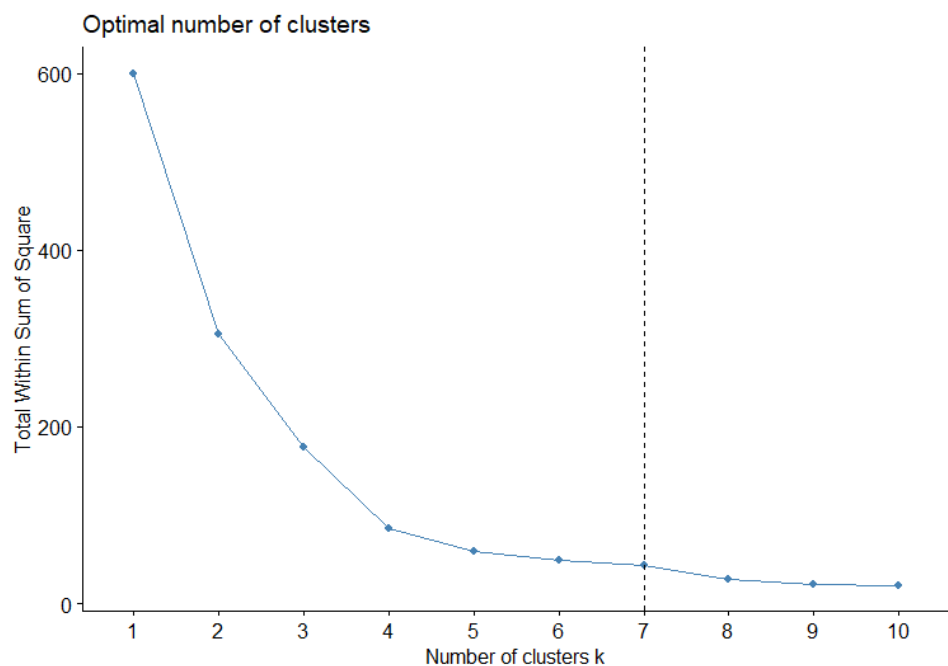


¿Cuál es el mejor valor para “K” para este problema? ¿Por qué?

Luego de este análisis, se concluyó que la cantidad óptima de Clusters es de 7, ya que este separa de forma más segmentada minimizando las distancias entre datos pero sin sobreponer clusters ni crear clusters con una cantidad de datos insignificante. Además, en cantidad de clusters menores, los grupos eran muy grandes y abarcan mucha distancia entre datos.



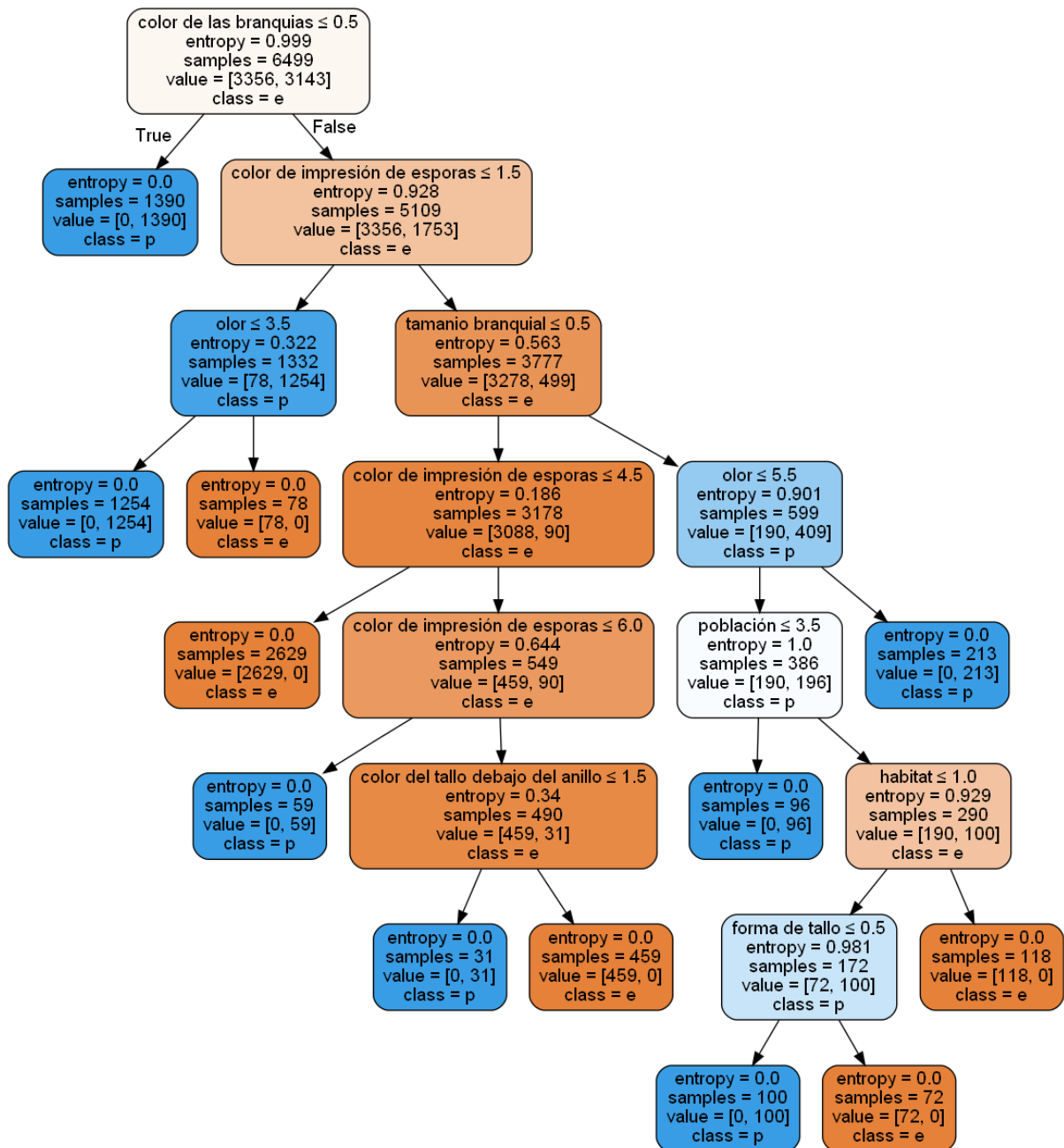
Un plot de wss



Explique que interpreta del plot WSS

Por medio del plot WSS, se interpreta que el aumentar cantidad de clusters no tiene mayor efecto en la distancia entre datos en un mismo Cluster, siendo el punto óptimo entre distancias de 7 clusters. Luego de esto, no hay mayor cambio significativo entre distancias.

Parte III:



¿Qué atributos son más significativos?

- El color de las branquias, el color de impresión de esporas, el olor y el tamaño branquial.

¿Se descarto algún atributo? De ser así, ¿Cuál o cuáles?

- apego-branquial, color, color de velo, color del tallo debajo del anillo, color del tallo sobre el anillo, espaciamento branquial, figura, magulladuras, número de anillo,

superficie, tallo-superficie-debajo-anillo, tallo-superficie-sobre-anillo, tipo anillo y tipo velo

¿Qué clases (valores de los atributos) producen un mayor impacto en sí un hongo ¿es venenoso o no? Por ejemplo “los hongos que tienen figuras convexa, plana y hundida tienden a tener una probabilidad más alta de ser venenosos”

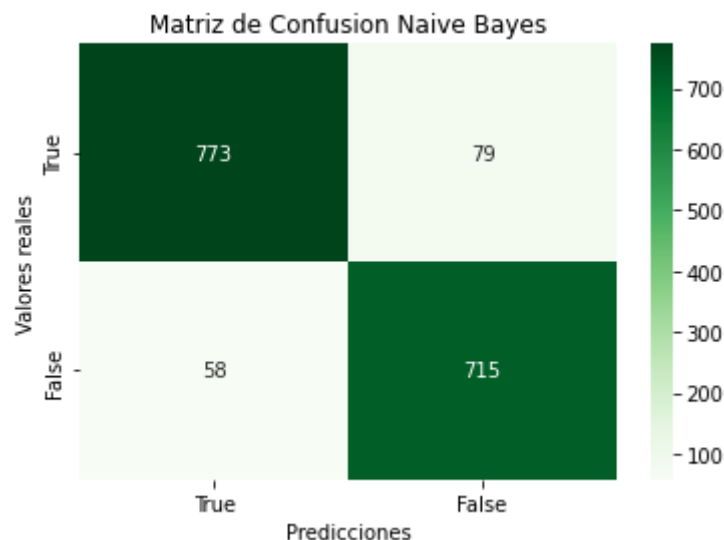
- Los hongos con color de branquias ante tienden a tener mayor probabilidad de ser venenosos.
- Los hongos con color de impresión de esporas blancas y chocolate tienen mayor probabilidad de ser venenosos.
- Los hongos con tamaño branquial estrecho tienden a tener mayor probabilidad de ser venenosos.
- Los hongos con olor a sucio tienden a tener mayor probabilidad de ser venenosos.

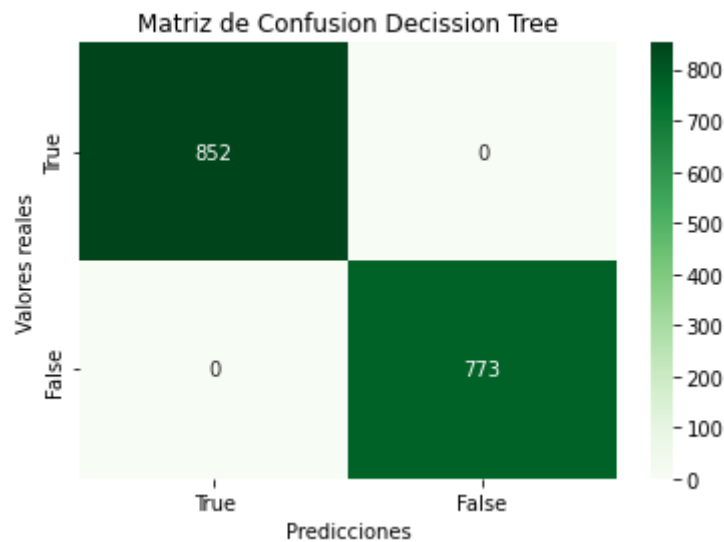
¿Cuáles son las probabilidades de los resultados de la clasificación de si un hongo es venenoso o no?

	clase	frecuencia	Probabilidad
No venenoso	e	852	52.430769
Venenoso	p	773	47.569231

Adjunte las tablas de probabilidad de cada atributo - esto no es posible con Naive Bayes

Usando los resultados de la predicción de ambos modelos, cree y adjunte las matrices de confusión





¿Qué modelo tiene más errores Tipo 1?

- Naive Bayes

¿Qué modelo tiene más errores Tipo 2?

- Naive Bayes

¿Qué modelo tiene una mejor precisión?

- Decission Tree

¿Cuál de los dos modelos elegiría y por qué?

- Decission Tree ya que ofrece mayor precisión para este set de datos y también no se poseen tantas clasificaciones por lo cual se ajusta bien al problema. Además, poder visualizar las probabilidades es parte de lo que lo hace más valioso para este problema.