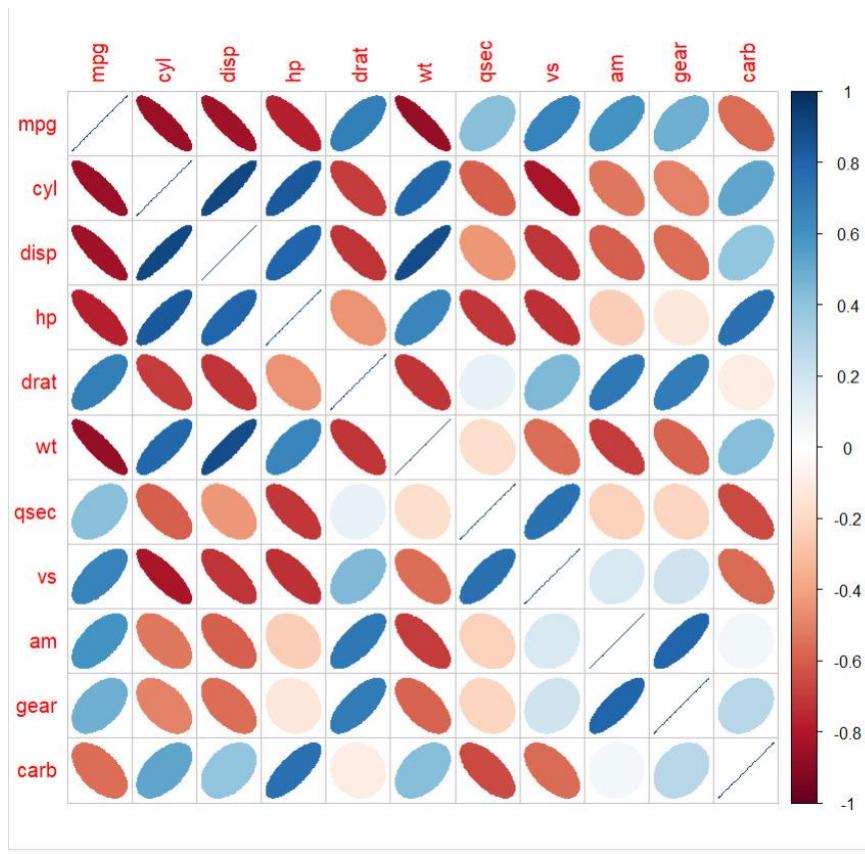


Correlación



Al examinar el diagrama de correlación, se observan conexiones significativas entre las siguientes variables:

Desplazamiento (disp) y Cilindros (cyl): Se evidencia que a medida que el número de cilindros aumenta, generalmente el desplazamiento del motor tiende a ser mayor. Esto sugiere una relación positiva entre estas dos variables, donde motores con más cilindros tienden a tener un desplazamiento más grande.

Cilindros (cyl) y Millas por galón (mpg): Se encuentra una correlación negativa entre estas dos variables, indicando que a medida que el número de cilindros aumenta, el consumo de combustible tiende a ser mayor. En otras palabras, los vehículos con más cilindros suelen ser menos eficientes en términos de consumo de combustible.

Caballos de fuerza (hp) y Cilindros (cyl): Esta relación puede variar, pero en general, se espera una correlación positiva. Es decir, motores con más caballos de fuerza tienden a tener más cilindros, ya que estos motores suelen ser más potentes y requieren más cilindros para generar esa potencia adicional.

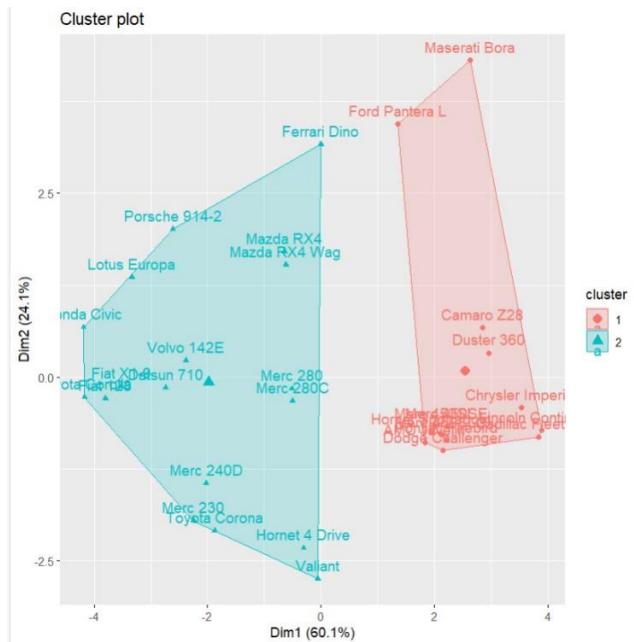
Caballos de fuerza (hp) y Desplazamiento (disp): También se espera una correlación positiva entre estas dos variables. Los motores con mayor desplazamiento tienden a producir más caballos de fuerza, ya que tienen más espacio para la combustión y, por lo tanto, pueden generar más potencia.

K-MEAN

Después de extraer y escalar la información de los autos, se generaron tres gráficas que dividen los datos en 2, 3 y 4 grupos, respectivamente. Estas gráficas proporcionan una visualización de la segmentación de los datos en diferentes números de clusters, lo que ayuda a comprender mejor las relaciones y patrones presentes en los datos.

2 Clusters

Se puede interpretar la segmentación de los vehículos en el diagrama como una distinción entre vehículos ligeros (representados por el grupo azul) y pesados (representados por el grupo rojo). Los vehículos más pesados necesitan más potencia para moverse eficientemente. Esta necesidad de mayor potencia se traduce en la presencia de un mayor desplazamiento y/o un mayor número de cilindros en los motores de los vehículos más pesados.



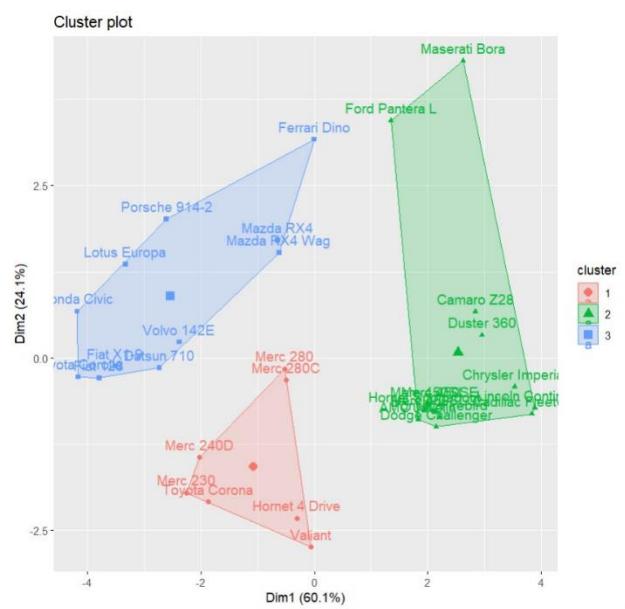
3 Clusters

En este caso, se puede interpretar que el diagrama de correlación refleja las prestaciones de los automóviles, dividiendo los datos en tres áreas distintas:

Cuadrante inferior: Este cuadrante engloba a los automóviles con prestaciones modestas. Estos vehículos pueden caracterizarse por tener menor desplazamiento, menos cilindros y, en consecuencia, menor potencia y eficiencia en el consumo de combustible.

Superior izquierdo: En esta área se encuentran los vehículos de mayor rendimiento sin llegar a la deportividad extrema. Tienden a ser automóviles premium que ofrecen un equilibrio entre potencia, desplazamiento y eficiencia. A menudo, estos vehículos tienen una combinación de cilindros y caballos de fuerza que les permite ofrecer un rendimiento superior sin comprometer demasiado la eficiencia en el consumo de combustible.

Lado derecho: Este sector alberga a los automóviles con una inclinación totalmente deportiva. Son vehículos diseñados principalmente para ofrecer altas prestaciones y rendimiento deportivo. Tienen un alto desplazamiento, más cilindros y una potencia considerable, lo que los hace ideales para aquellos conductores que buscan una experiencia de conducción emocionante y dinámica.



4 Clusters

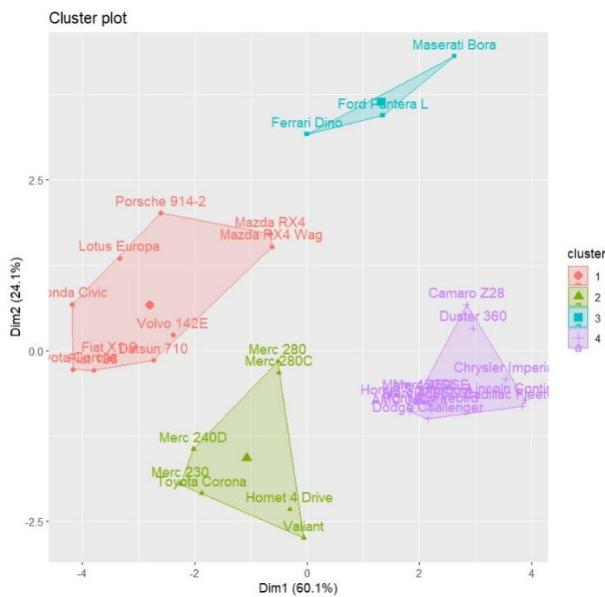
Con la segmentación en 4 clusters, se puede realizar un análisis más detallado basado en las prestaciones de los autos, lo que permite identificarlos en las siguientes categorías:

Superdeportivos (azul): Estos vehículos destacan por combinar los motores más potentes con pesos óptimos y tiempos de cuarto de milla más rápidos. Son la encarnación del rendimiento extremo y suelen ser los más deseables para los entusiastas de la velocidad y el alto rendimiento.

Drag/Muscle (morado): Este grupo comprende vehículos con motores grandes y potentes, pero que no necesariamente consideran factores como el peso o el consumo de combustible. Por lo general, tienen las peores millas por galón (MPG). Estos automóviles están diseñados para la potencia bruta y el rendimiento en línea recta, siendo populares en carreras de arrastre y entre los aficionados a los muscle cars.

Sport/Premium (rojo): Estos autos se caracterizan por su bajo peso y motores eficientes, que no necesariamente son pequeños. Suelen tener un buen rendimiento en términos de MPG y ofrecen un equilibrio entre potencia y eficiencia. Son populares entre aquellos que valoran tanto el rendimiento como la calidad de conducción y la eficiencia de combustible.

Genéricos (verde): Este grupo incluye autos que priorizan el mejor MPG, con motores pequeños y de baja cilindrada. Son más ligeros y compactos en su mayoría, lo que los hace ideales para aquellos que buscan eficiencia en el consumo de combustible y una conducción más económica y amigable con el medio ambiente.



En este caso particular, se puede concluir que la segmentación óptima de los datos se logra al dividirlos en 4 clusters. Esta elección se fundamenta en la observación de que al agrupar en 2 clusters, no se logra una especificidad suficientemente alta. Por otro lado, al emplear 3 clusters, se identifica una agrupación con una distancia considerable entre Ford Pantera y Camaro Z28. Sin embargo, al aumentar a 4 clusters, este grupo se divide en dos, lo que añade más precisión y por ende mejora la calidad de la segmentación.

Conclusión

Entendiendo la correlación entre las variables de los automóviles, hace más sentido a la hora de segmentar la información en k-mean. Se pueden agrupar los automóviles por categorías, siendo una de estas la de los autos de mayor desempeño, misma que se caracteriza por tener motores de mayor desplazamiento y cilindraje a la vez que un rendimiento de combustible bastante pobre (Ferrari Dino, Maserati Bora y Ford Pantera). Siguiendo la misma convención se pudieron identificar grupos como autos genéricos que incluyen menor cilindraje y un alto rendimiento de combustible, así como grupos intermedios que contienen una mezcla de las variables.