

Reporte final de "El precio de los autos"

Resumen

Se comparte un set de datos con el que se requiere saber el impacto de las variables en el precio final del producto, para realizar esto primero se tuvieron que elegir 6 variables, las cuales fueron horsepower, enginesize, curbweight, carwidth, carlength y wheelbase, esta elección fue en base al impacto que tenían en el precio con ayuda de un mapa de correlación. Después de esto se eligieron dos métodos estadísticos; Anova y prueba de hipótesis para ver la relación entre las variables, con un umbral y alpha de 0.05 para ambos, los resultados que se obtuvieron fueron que si existen diferencias significativas en ambas pruebas, después de esto se realizó un proceso de normalización para ver su peso en estas, y a pesar de que si hubo un cambio las conclusiones, los resultados continuaron siendo los mismo para ambas pruebas.

Introducción

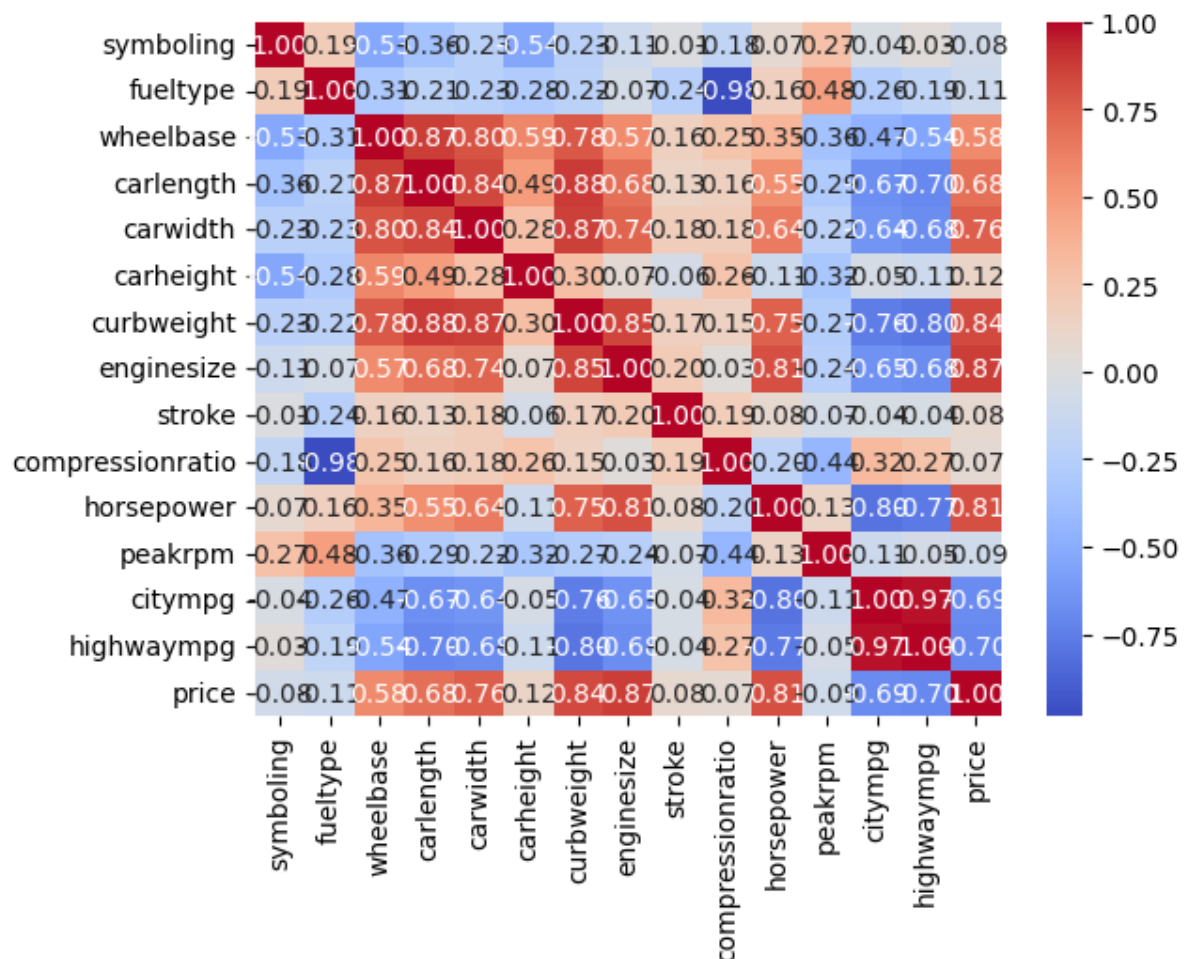
Una empresa automovilística nos comparte un set de datos, dentro del cual se aprecia el precio de un automóvil en conjunto con otros datos de este, el nombre, medidas, cilindros, entre otros, la empresa cuenta con dos cuestiones principales, la primera es que variables son significativas para la predicción de un automóvil y la segunda estas variables que tan bien describen el precio.

Es importante para la empresa conocer este tipo de información debido a que de esta forma sabrán que el precio que se le está dando a cada automóvil es el adecuado tomando en cuenta todos los factores que se nos compartieron, todo esto debido principalmente por la diferencia de precios en los diferentes países, ya que ellos necesitan información sobre el mercado estadounidense.

Análisis de resultados

Elección de variables

Lo primero que se realizó fue ver las variables con las que contábamos así como el tipo de datos que era cada una de estas, para saber cómo manejar cada valor de la manera correcta, después de esto se hizo un análisis de las variables cuantitativas como de las variables cualitativas, la manera principal que se tuvo de ver que tanta relación tenía cada variable con el precio fue con ayuda de un mapa de correlación. En el cual se puede observar que las variables que más impacto tienen sobre el precio son horsepower, enginesize, curbweight, carwidth, carlength y wheelbase.



De igual manera, con el análisis se realizó el uso de boxplot, en el que pudimos observar que sobre las seis variables existen valores atípicos, la única variable que no contaba con ningún valor atípico es curbweight, sin embargo antes de tomar una decisión en cuanto a que hacer

con estas variables es ver cómo nos afectan en un futuro, y a partir de ahí tomar una decisión, para ver si es necesario escalarlos o normalizarlos.

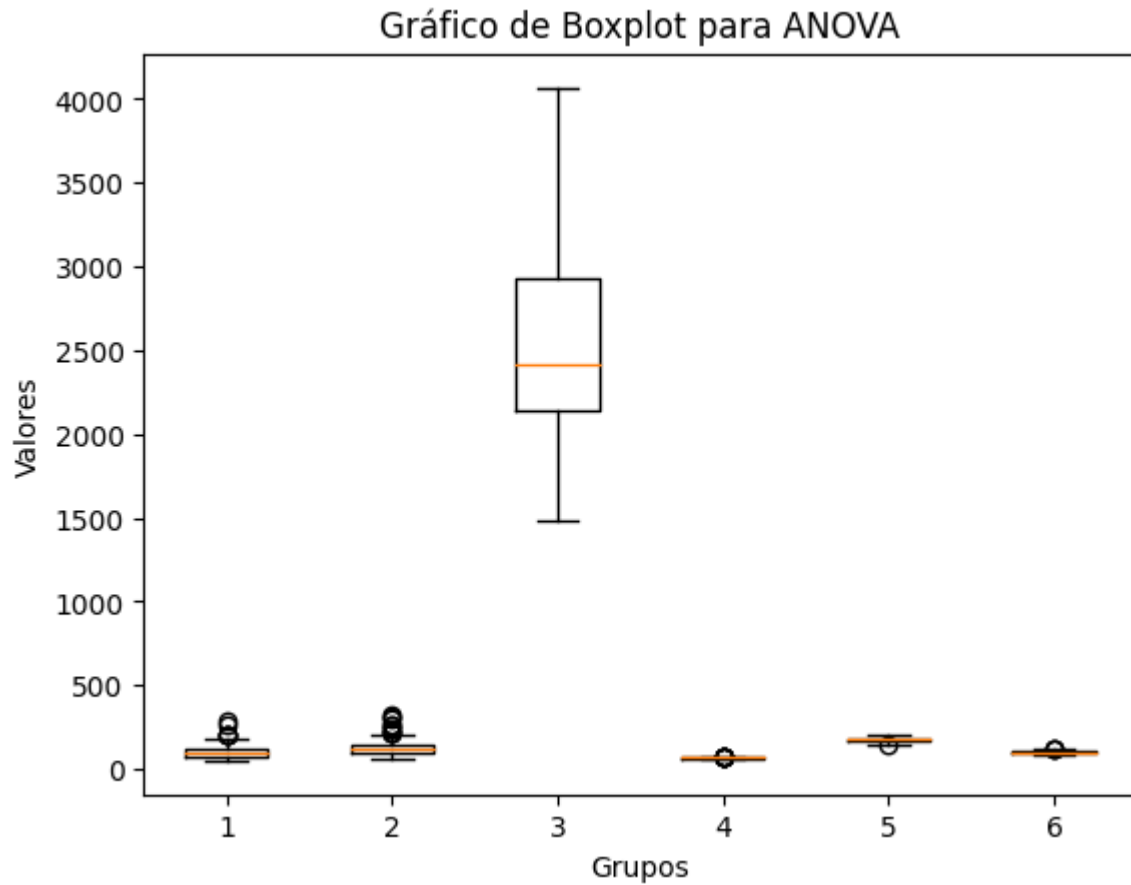
Construcción de un modelo estadístico

Nuestra pregunta base para esto fue que tanto se diferencian las variables elegidas una de la otra (la dependencia entre estas) afectando el precio del automóvil. Para esto existen diferentes tipos de herramientas estadísticas que podrían sernos de ayuda, en este caso usaremos ANOVA y pruebas de hipótesis. La prueba de hipótesis es un enfoque general que nos servirá para evaluar afirmaciones mientras que ANOVA será para comparar en múltiples grupos.

Para realizar la prueba de hipótesis se hizo uso de una prueba t la cual está relacionada con las medias, realizamos las pruebas con las muestras de nuestras variables elegidas anteriormente haciendo uso de un alpha de 0.05, sin embargo todas nuestras hipótesis fueron rechazadas debido a la diferencia entre nuestras variables.

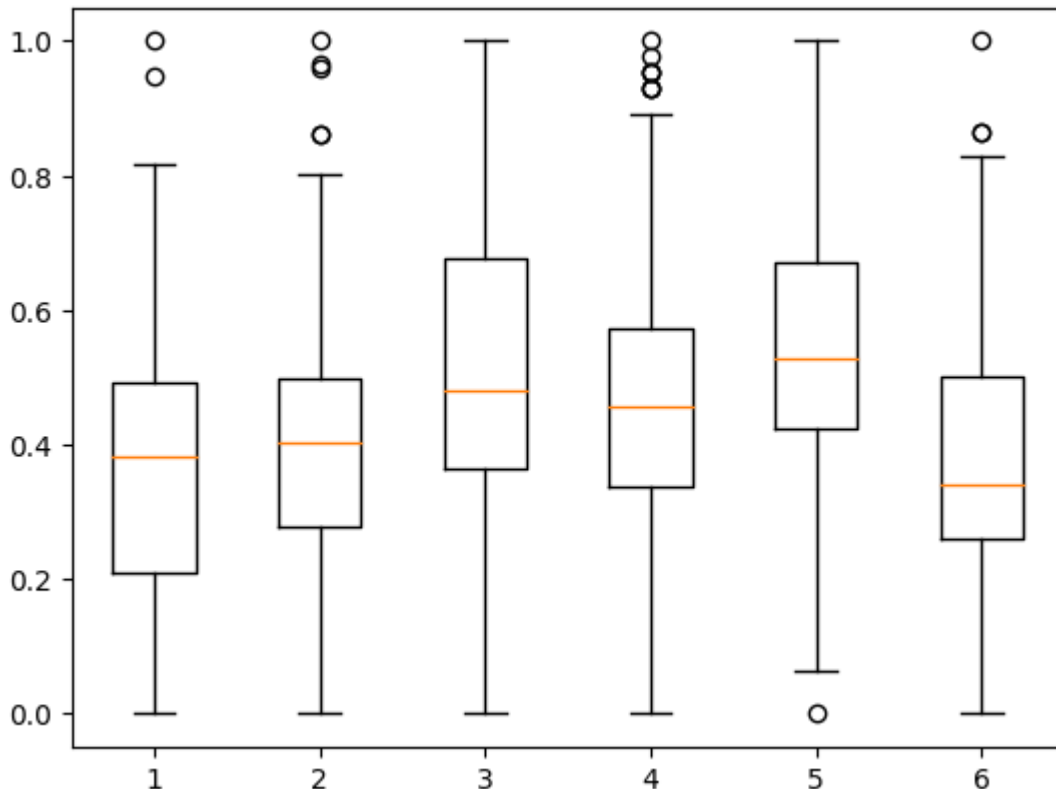
Con ayuda de diferentes librerías realizamos el cálculo del ANOVA entre nuestras variables elegidas, el valor de umbral elegido para nuestro ANOVA fue de 0.05, lo cual nos indica que estamos dispuestos a aceptar un riesgo de 5% de cometer un error, el primer ANOVA que obtuvimos como resultado fue de $2.214091958718416e-248$, por lo que podríamos indicar que si existen diferencias significativas sobre los grupos.

Después de esto realizamos una Boxplot para ver nuestros valores de manera gráfica, lo cual quedaría de la siguiente manera:



En la imagen podemos ver como primero nuestras variables están en diferentes medidas, lo cual, en caso de que en un futuro se pretenda realizar una regresión no sería de ayuda.

En este caso realizamos una normalización para igualar las escalas, que si bien no tendría un impacto o diferencias en cuanto a nuestro ANOVA nos ayudará a que las diferencias de nuestras variables sean menos significativas que como eran antes.



En esta imagen podemos observar como ya todas nuestras variables se manejan con el mismo mínimo y el mismo máximo, con lo cual no esperamos que los resultados del ANOVA cambien, sin embargo lo calculamos para ver las diferencias y el resultado fue de $1.1817110158648524e-24$, que si bien tuvo una diferencia a comparación de nuestra primera prueba podemos llegar a las mismas conclusiones que al inicio.

Conclusión

Durante el proceso de realización vimos que había variables que tenían un mayor impacto que otras en el precio del auto, la variable que más relación tiene con el precio del auto es enginesize, lo cual pudimos observar en el mapa de correlación, las variables elegidas para realizar todo el proceso fueron las siguientes en tener el mayor impacto con el precio del auto, y también se encontraron variables que tienen el mínimo impacto en el precio, información que será de mucha ayuda para la empresa.

En cuanto al proceso de modelo, antes de la normalización las variables se diferenciaban mucho de las otras en cuanto al rango de medida. Y como respuesta a la pregunta generada es que si, las variables en conjunto se diferencian mucho de ellas mismas por lo que afectan

al precio en sí, y se podría intentar calcular el precio de un automóvil con cada variable individualmente o encontrando el conjunto de variables que tengan la relación suficiente para obtener los mejores resultados, ya que las variables no muestran ningún tipo de dependencia o relación las unas con las otras.

Anexos

Código-<https://drive.google.com/file/d/1C8kwTRmNu6ISW9Ca-q7WMTMZlyveLhOe/view?usp=sharing>