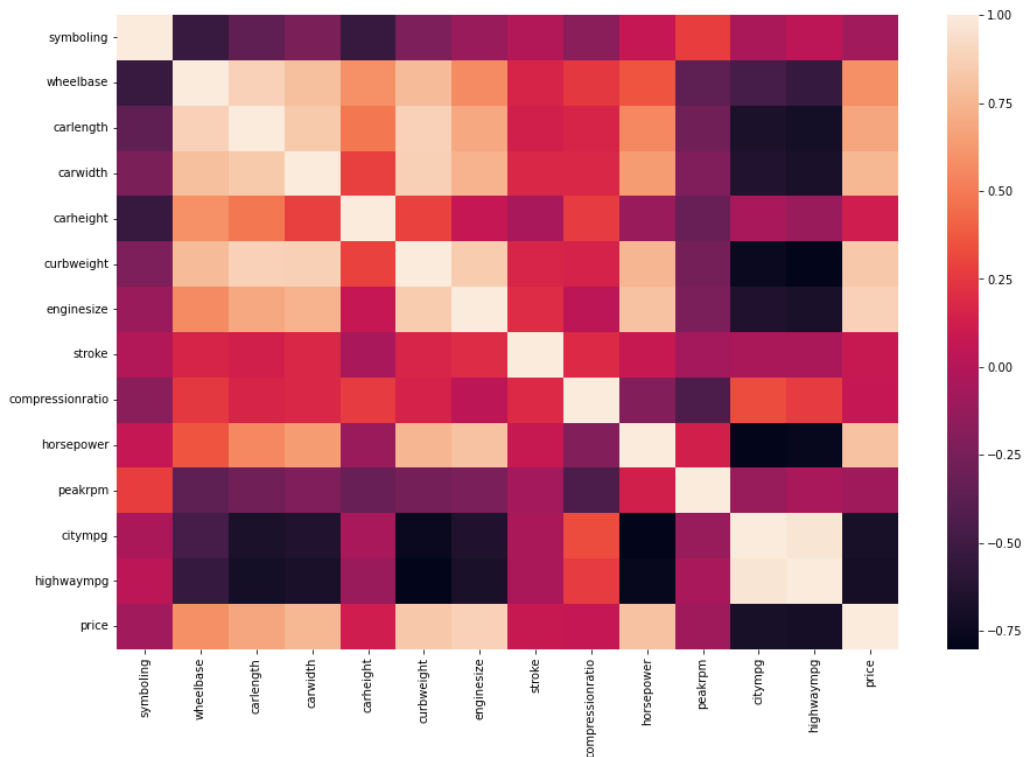


Evidencia 4

A00572429

Para determinar los factores más importantes en el precio de un automóvil es importante entender los comportamientos de compra del mercado americano. Una predicción correcta del costo de su automóvil según sus variables mas importantes puede ser clave para entrar con mayor impacto.



Primero un análisis de correlación entre las variables para detectar cuales son importantes entre ellas y no solo el precio, de esta manera podemos entender un poco mejor como se estructura la base de datos. Una vez que tenemos las variables numéricas con mayor correlación, buscamos variables categóricas que tengan más importancia según el análisis.

Variables Numéricas:

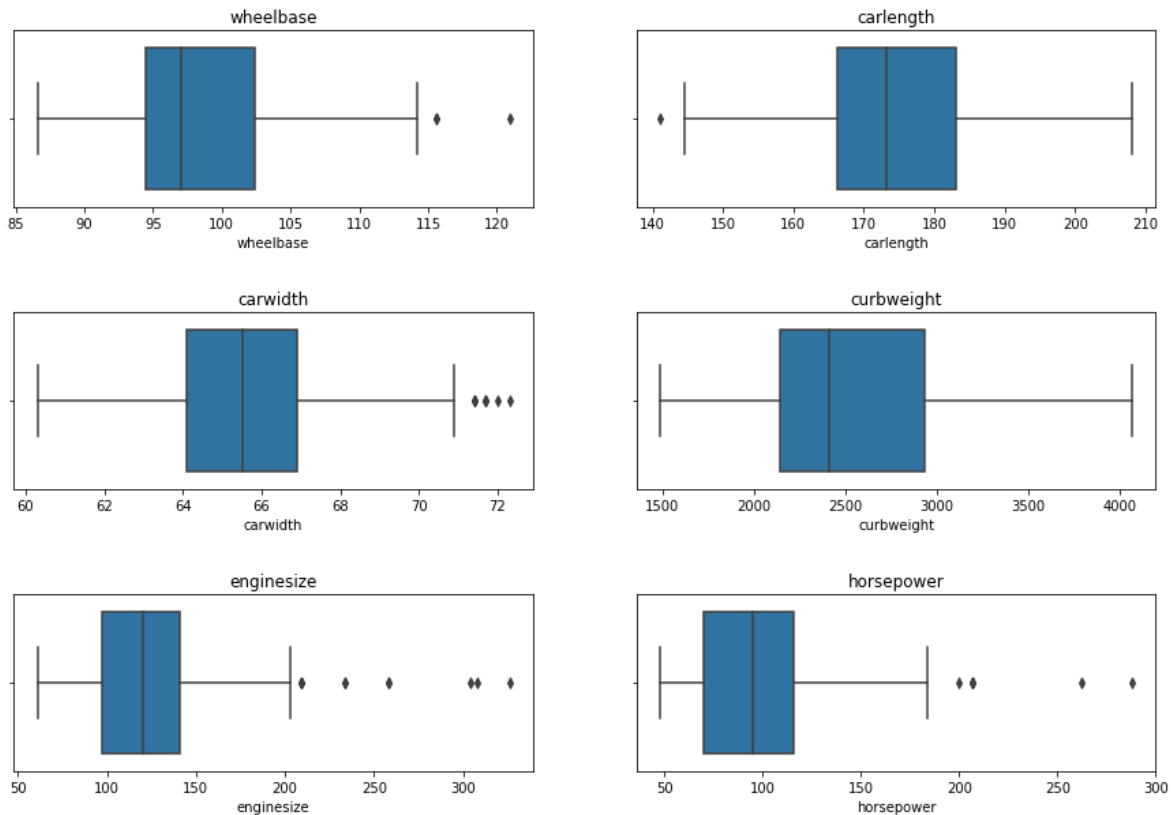
- Wheelbase
- Carlength
- Carwidth
- Carweight
- Enginesize
- Horsepower

Variables Categoricals:

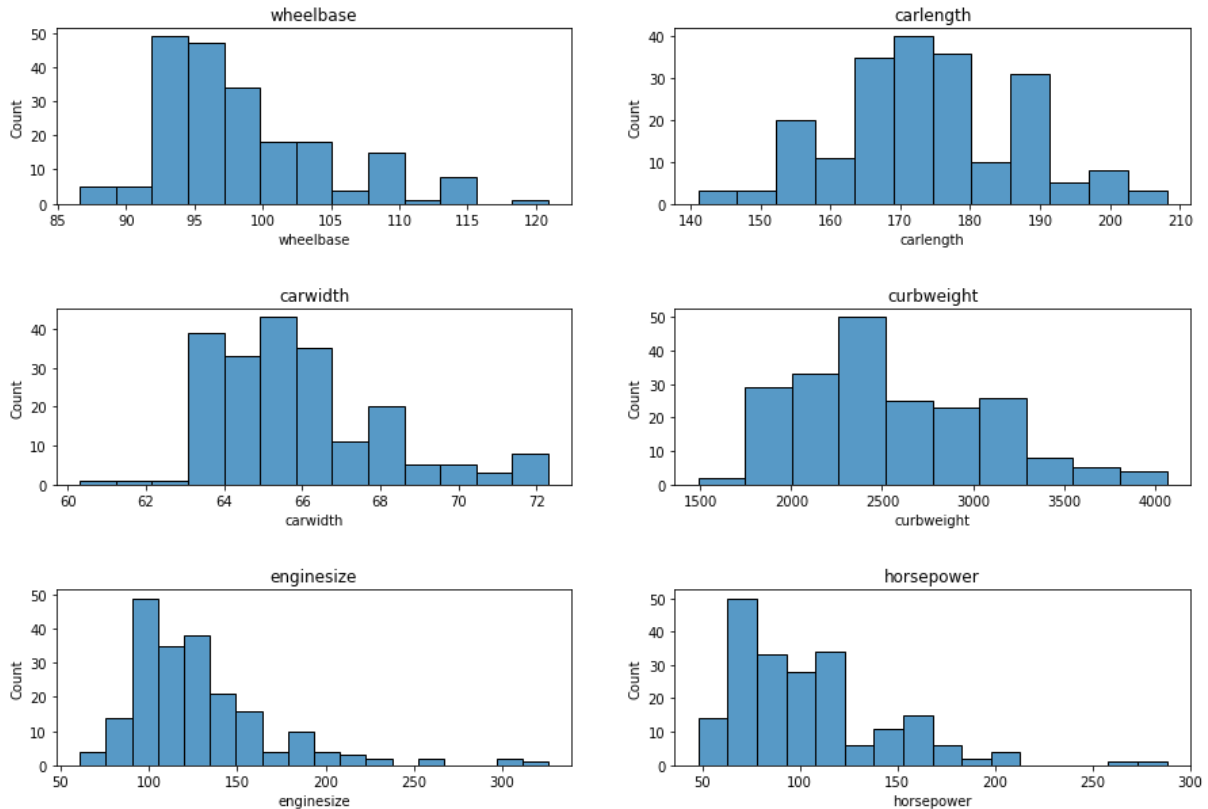
- Cylindernumber
- Carbody
- Fuletype

Entender el comportamiento de estas mismas es muy importante por lo que hice box plots para entender sus distribuciones. Después un histograma para poder tener una visión amplia y correcta.

Box-Plot Variables Numéricas



Histograma Variables Numéricas



El histograma me ayuda para entender sus distribuciones, y cómo se comporta. Son visualizaciones que el box-plot no me ayuda tanto a ver. Esto solo nos da una visión mas amplia de la distribución de nuestros datos.

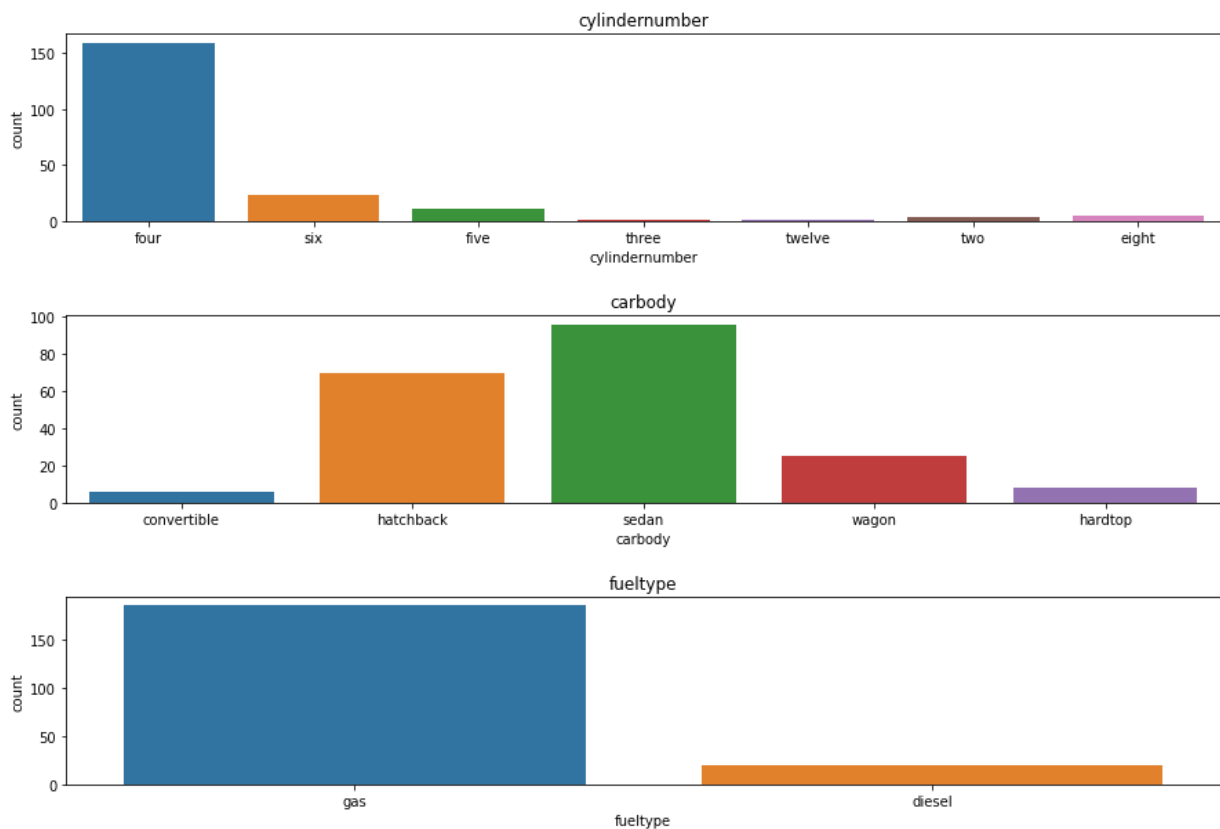
Para tener una mayor comprensión de las variables seleccionadas se utilizaron medidas básicas como desviación estándar, moda y promedio. Las cuales ayudan a entender mejor aun sus distribuciones de cada variable.

Podemos ver como Wheel base tiene una tendencia de cola derecha, lo cual nos puede indicar un Wheel base mas chico de 93 – 95 o mejor dicho el segundo cuartil es el mas recurrido. Al igual en horsepower tenemos un histograma con cola derecha indicando que la mayoría de los coches caen en el segundo y tercer cuartil, dando un histograma bimodal. En engine size podemos ver como cae igual una cola derecha que nos indica que la mayoría de los coches tienen tamaño chico dado una gran variación en los valores. Pocos coches muestran valores fuertes respecto a esta variable, pero sin duda existen. Los promedios nos indican igualmente donde se pudieran encontrar la mayoría de los picos.

| | wheelbase | carlength | carwidth | curbweight | enginesize | horsepower |
|------|-----------|------------|-----------|-------------|------------|------------|
| std | 6.021776 | 12.337289 | 2.145204 | 520.680204 | 41.642693 | 39.544167 |
| mean | 98.756585 | 174.049268 | 65.907805 | 2555.565854 | 126.907317 | 104.117073 |
| mode | 94.500000 | 157.300000 | 63.800000 | 2385.000000 | 92.000000 | 68.000000 |

Después de tener una visión amplia sobre las variables numéricas y sus distribuciones podemos ver que existen pocos datos no verídicos o mejor dicho outliers. Los cuales no ayudan mucho para nuestro modelo de predicción.

Seguimos con las variables categóricas, las cuales fueron elegidas anteriormente. Estas nos pueden ayudar a entender de una manera sencilla como se comporta el mercado. Empezamos por un countplot para entender que valores de las variables se repiten mas y entender el mecanismo del mercado.



Como se puede ver existe una preferencia de mercado por los coches de 4 cilindros siendo los mas demandados del mercado. Nos muestra una mayor cantidad por mucho a los demás. Después podemos ver como los sedan son igual mas famosos entre los consumidores, siendo una distribución casi pareja con los hatchbacks y wagons, dando poco mercado a los convertibles. Por ultimo ver como el mercado es dominado por los coches con funcionamiento de gasolina, siendo el diesel una opción menos favorita del mercado.

Haremos un Cross table para entender qué tipo de coches tiene preferencias según estas mismas variables categóricas.

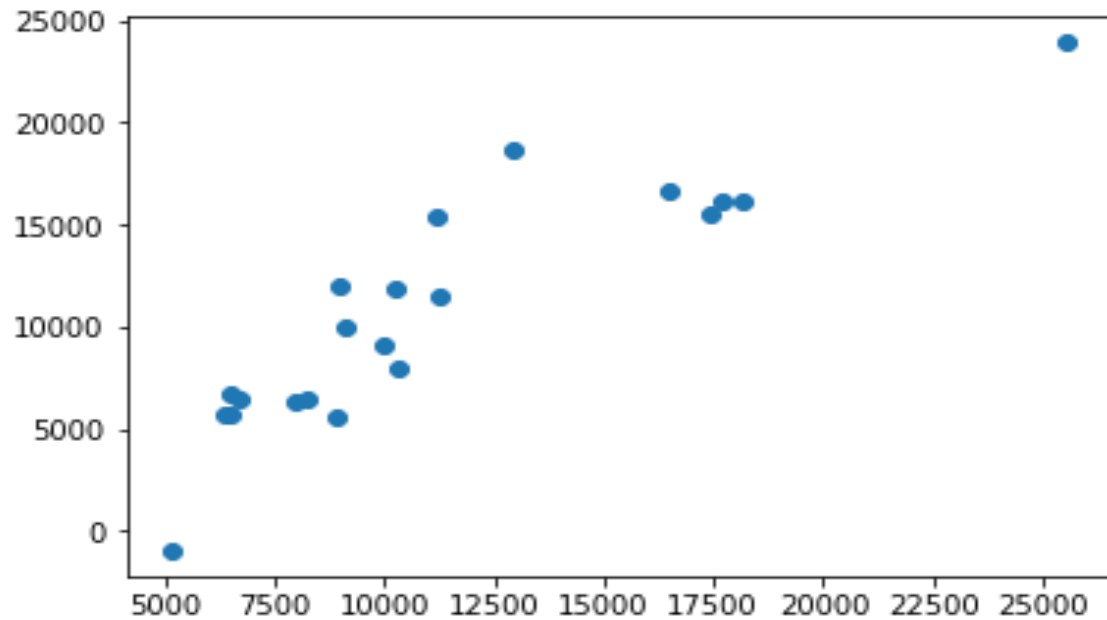
| carbody | convertible | hardtop | hatchback | sedan | wagon |
|----------------|-------------|---------|-----------|-------|-------|
| cylindernumber | | | | | |
| eight | 1 | 1 | 1 | 2 | 0 |
| five | 0 | 1 | 1 | 7 | 2 |
| four | 4 | 4 | 57 | 73 | 21 |
| six | 1 | 2 | 6 | 13 | 2 |
| three | 0 | 0 | 1 | 0 | 0 |
| twelve | 0 | 0 | 0 | 1 | 0 |
| two | 0 | 0 | 4 | 0 | 0 |

| fueltype | diesel | gas |
|----------------|--------|-----|
| cylindernumber | | |
| eight | 0 | 5 |
| five | 4 | 7 |
| four | 15 | 144 |
| six | 1 | 23 |
| three | 0 | 1 |
| twelve | 0 | 1 |
| two | 0 | 4 |

| fueltype | diesel | gas |
|-------------|--------|-----|
| carbody | | |
| convertible | 0 | 6 |
| hardtop | 1 | 7 |
| hatchback | 1 | 69 |
| sedan | 15 | 81 |
| wagon | 3 | 22 |

Podemos ver como la mayoría de los coches con cuatro cilindros son modelos hatchback, sedan y wagon, siendo un posible segmento de mercado. O siendo solo los hatchbacks los únicos con motores de dos cilindros. Igualmente, los coches de cuatro cilindros tienden a ser mas de gasolina que de diesel, siendo un mercado mas grande. Siendo los de cinco cilindros los mas propensos a usar diesel. Por ultimo los modelos que mas usan gasolina serían los hatchback, y sedan, los sedan muestran también una corriente por el uso de diesel, siendo una posible apertura de mercado.

Por último, creamos un modelo de regresión lineal el cual nos ayudara a predecir el precio según las variables mas importantes que tenemos, usamos las variables anteriormente mencionadas. Nuestro modelo es eficiente pero realmente no es lo mejor que se puede obtener, pero dado las circunstancias y mi conocimiento creo que puede servir para la solución de problema. Con un resultado de 0.73 se considera un modelo positivo mediano.



Creo que es un mercado muy centrado en coches de uso para clase media según las graficas anteriormente mostradas, lo cual me pone a recomendar que existe una gran brecha de oportunidad tanto en coches de menor cilindraje y precio, creo que estos pueden ser una gran oportunidad de negocio. No existen muchos modelos lo cual puede crear una oportunidad para crear mas de estos de estilo hatchback a un precio accesible.

| | cylindernumber | carbody | price |
|----|-----------------------|----------------|--------------|
| 18 | three | hatchback | 5151.000000 |
| 10 | four | hatchback | 9044.245614 |
| 9 | four | hardtop | 9384.000000 |
| 11 | four | sedan | 10837.479452 |
| 12 | four | wagon | 11046.761905 |
| 20 | two | hatchback | 13020.000000 |
| 8 | four | convertible | 14814.750000 |
| 17 | six | wagon | 15074.500000 |
| 15 | six | hatchback | 17392.166667 |
| 5 | five | hatchback | 17859.167000 |
| 6 | five | sedan | 20676.000000 |
| 7 | five | wagon | 23584.000000 |
| 16 | six | sedan | 25387.538462 |
| 4 | five | hardtop | 28176.000000 |
| 2 | eight | hatchback | 31400.500000 |
| 14 | six | hardtop | 33278.000000 |
| 0 | eight | convertible | 35056.000000 |
| 19 | twelve | sedan | 36000.000000 |
| 13 | six | convertible | 37028.000000 |
| 3 | eight | sedan | 37572.000000 |
| 1 | eight | hardtop | 45400.000000 |