Sports Car Feature Importance Data Analysis

James David and Zannate Malik
ENPM808W

Goal

What makes a sports car a sports car?

Utilizing concepts learned throughout the course to conduct a thorough comparative analysis of various components/features within sports cars in order to determine their respective contributions to the overall price of these types of vehicles.

Background

Reasoning of two data sets from Kaggle

- Utilizing two distinct datasets : one comprised of both numeric and categorical features and the other centered around categorical features
- Attempts to provide a holistic understanding of the interplay between quantitative and design aspects.
- Examining both numeric and categorical data can discern correlations and draw insights into the factors that (do or don’t) significantly influence the pricing of sports cars.
- Through different techniques such as feature engineering, minor data integration, and modeling,

Dataset #1 : Sports-car-prices-dataset

```
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Car Make                 1007 non-null   object
 1   Car Model                1007 non-null   object
 2   Year                     1007 non-null   int64
 3   Engine Size (L)          997 non-null    object
 4   Horsepower               1007 non-null   object
 5   Torque (lb-ft)           1004 non-null   object
 6   0-60 MPH Time (seconds)  1007 non-null   object
 7   Price (in USD)           1007 non-null   object
dtypes: int64(1), object(7)
memory usage: 63.1+ KB
```

	Car Make	Car Model	Year	Engine Size (L)	Horsepower	Torque (lb-ft)	0-60 MPH Time (seconds)	Price (in USD)
0	porsche	911	2022	3	379	331	4	101,200
1	lamborghini	huracan	2021	5.2	630	443	2.8	274,390
2	ferrari	488 gtb	2022	3.9	661	561	3	333,750
3	audi	r8	2022	5.2	562	406	3.2	142,700
4	mclaren	720s	2021	4	710	568	2.7	298,000

https://www.kaggle.com/datasets/rkiattisak/sports-car-prices-dataset

Dataset #2 : Sports car choice

	resp_id	ques	alt	segment	seat	trans	convert	price	choice
0	1	1	1	basic	2	manual	yes	35	0
1	1	1	2	basic	5	auto	no	40	0
2	1	1	3	basic	5	auto	no	30	1
3	1	2	1	basic	5	manual	no	35	0
4	1	2	2	basic	2	manual	no	30	1

Field	Description
resp_id	The identifier of each individual in the dataset
ques	The identifier of each specific purchase scenario
alt	The identifier of each alternative choice within a question
segment	The commercial segment of a sportscar model ('basic', 'fun', 'racer')
seat	The number of seats in the vehicle (2, 4, 5)
trans	The transmission type of the vehicle ('auto','manual')
convert	Whether or not the vehicle has a convertible top
price	The sportscar price (in thousands/$)
choice	Dummy indicator of the decision made. (1 = car chosen, 0 = alternative cars chosen from)

https://www.kaggle.com/datasets/vspencer88/sports-car-choice-data?select=sportscar_choice_long.csv

Data Cleaning

Within Dataset # 1 :

- Features that required Manipulation
    - Any Features containing string characters (car make, car model)
    - Engine Size (L)
    - Horsepower
    - Torque (lb-ft)
    - Price (in USD)
    - Engine Type

Horsepower, Torque (lb-ft), 0-60 MPH Time (seconds) Data Manipulation

These features required minor manipulation in terms of special characters:

- Dropped characters
    - ‘+’, ‘<’, ‘>’ , ‘,’

One special example was a Car Model Provided a special case where the Horsepower value was 10,000+

Engine Type

Created through feature engineering based off Engine Size (L) (explained in future slide)

This feature reads the values, based on the string it reads, it gets assigned a Engine Type of either: gas, electric, or hybrid

```python
def assign_engine_type(value):

  # if this string is not not found
  if (str(value).find("1.5 + elect") != -1):
    return 'hybrid'

  # search for string 'hybrid'
  elif re.search(r'\bhybrid\b', str(value)):
    return 'hybrid'

  # search for string 'electric'
  elif re.search(r'\belectric\b', str(value)):
    return 'electric'

  # assign remaining "non unique" cases to gas
  else:
    return 'gas'
```

Engine Size (L) Data Manipulation

The goal was to make this an all numeric feature of type float.

With the majority of the data already listing out values for the engine size, this feature contained special cases:

- ‘NaN’ values which was resolved with Data Integration and manipulation
- ‘electric’ - since electric cars contain motors and not engines, it was given value 0 (this also matched existing electric cars that were properly valued 0 within the raw data)
  - ‘electric (93 kWh)’
  - ‘electric (tri-motor)’
  - electric motor
- One car contained ‘1.5 + electric’ - given the information above about ‘electric’, this sports car was given 1.5 since it was hybrid and contained partial engine size + 0
- ‘hybrid’ - the sports car that contained this was the same as the special case above so it made sense to assign the ones involved with hybrid to 1.5 and group them together

1	Car Make	Car Model	Year	Engine Size	Horsepower	Torque (lb-ft)	0-60 MPH Time	Price (in USD)
44	BMW	i8	2020	1.5 + Electric	369	420	4.2	148,500
734	BMW	i8	2022	Hybrid	369	184	4.2	148,500
969	Porsche	Panamera Turbc	2021	Hybrid	689	642	3	190,000

Data Integration

Dataset #1 contained some cells that were NaN values. Instead of removing “incomplete” rows containing NaN, research of each car was needed to fill with accurate data.

```
# Lists out the specific rows within the Engine Size (L) column contain value 'NaN'
sports_car_df[sports_car_df["Engine Size (L)"].isna()]

```

	Car Make	Car Model	Year	Engine Size (L)	Horsepower	Torque (lb-ft)	0-60 MPH Time (seconds)	Price (in USD)
168	rimac	c_two	2022	NaN	1914	1696	1.9	2400000
171	tesla	model s plaid	2021	NaN	1020	1050	1.98	131190
222	porsche	taycan turbo s	2021	NaN	750	774	2.6	185000
247	tesla	model s plaid	2022	NaN	1020	1050	1.9	131190
387	rimac	c_two	2022	NaN	1888	1696	1.8	2400000
389	tesla	roadster	2022	NaN	10000+	0	1.9	200000
686	rimac	c_two	2022	NaN	1914	1696	1.85	2400000
697	lotus	evija	2022	NaN	1972	1254	2.5	2700000
752	porsche	taycan	2022	NaN	469	479	3.8	79900
916	tesla	roadster	2022	NaN	10,000+	NaN	1.9	200000

NaN was replaced with 0 since they were all electric cars

Feature Engineering

- Engine Type:
  - Based off ‘Engine Size (L)
  - Gas, electric, hybrid
  - Purpose : create this feature to see if there are any trends related to price and the type of engine. “Does the type of car have an impact on the sports car price”
- $ per Horsepower
  - Purpose : find any relationship between the cost of horsepower
- Origin
  - Through Data Integration (explained later on)
  - Purpose : find another feature that may contribute to the price of sports cars
- Engine Size (L) Range
  - Small sample size per unique Engine Size (L) so decided to make ranges
  - Purpose : increase count per range to see if there are more obvious trends
  - Last second decision to analyze
- Score
  - Give the existing numeric features weight
  - Purpose : see which cars have the highest and lowest scores and see if they have relationship to highest or lowest price
  - The weight distribution will be different depending who you ask. Everyone will have a different opinion on what feature is most important to their car

Data Integration

Origin

- Searched the internet of each Car Model

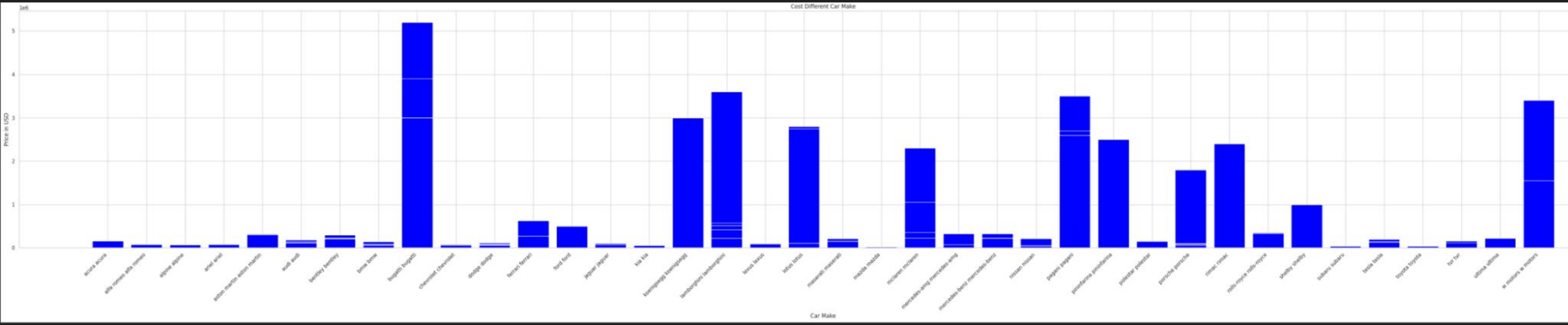
Total Count of Origins	
germany	287
england	229
america	185
italy	176
japan	72
france	24
sweden	15
croatia	14
lebanon	3
china	1
south korea	1

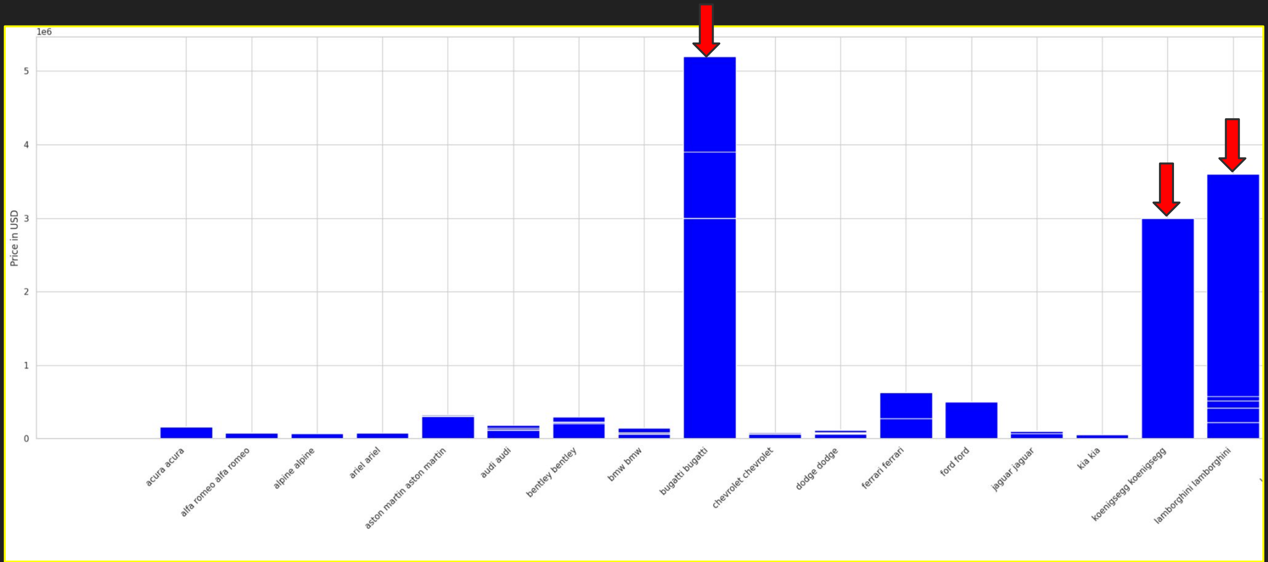
[38]

```python
# mapping of each unique Sports Car Make
car_origin_mapping = {

'acura'          : 'america',
'alfa romeo'     : 'italy',
'alpine'         : 'france',
'ariel'          : 'england',
'aston martin'   : 'england',
'audi'           : 'germany',
'bentley'        : 'england',
'bmw'            : 'germany',
'bugatti'        : 'france',
'chevrolet'      : 'america',
'dodge'          : 'america',
'ferrari'        : 'italy',
'ford'           : 'america',
'jaguar'         : 'england',
'kia'            : 'south korea',
'koenigsegg'     : 'sweden',
'lamborghini'    : 'italy',
'lexus'          : 'japan',
'lotus'          : 'england',
'maserati'       : 'italy',
'mazda'          : 'japan',
'mclaren'        : 'england',
'mercedes-amg'   : 'germany',
'mercedes-benz'  : 'germany',
'nissan'         : 'japan',
'pagani'         : 'italy',
'pininfarina'    : 'italy',
'polestar'       : 'china',
'porsche'        : 'germany',
'rimac'          : 'croatia',
'rolls-royce'    : 'england',
'shelby'         : 'america',
'subaru'         : 'japan',
'tesla'          : 'america',
'toyota'         : 'japan',
'tvr'            : 'england',
'ultima'         : 'england',
'w motors'       : 'lebanon',
}

sports_car_df['Origin'] = sports_car_df['Car Make'].map(car_origin_mapping)
```

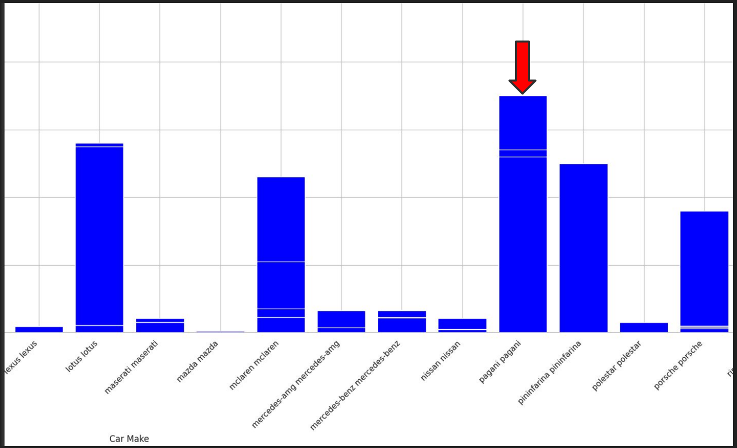
	Car Make	Car Model	Year	Engine Size (L)	Horsepower	Torque (lb-ft)	0-60 MPH Time (seconds)	Price (in USD)	Engine Type
36	nissan	370z nismo	2021	3.7	350	276	4.5	45690	gas
37	porsche	taycan 4s	2022	0.0	562	479	3.8	104000	electric
38	lamborghini	urus	2021	4.0	641	626	3.5	218000	gas
39	ferrari	roma	2021	3.9	611	561	3.3	222000	gas
40	audi	rs3	2022	2.5	394	369	3.9	57000	gas
41	mclaren	gt	2021	4.0	612	465	3.1	210000	gas
42	bmw	i8	2020	1.5	369	420	4.2	148500	hybrid
43	mercedes-benz	cls63 amg	2019	4.0	603	627	3.4	132000	gas

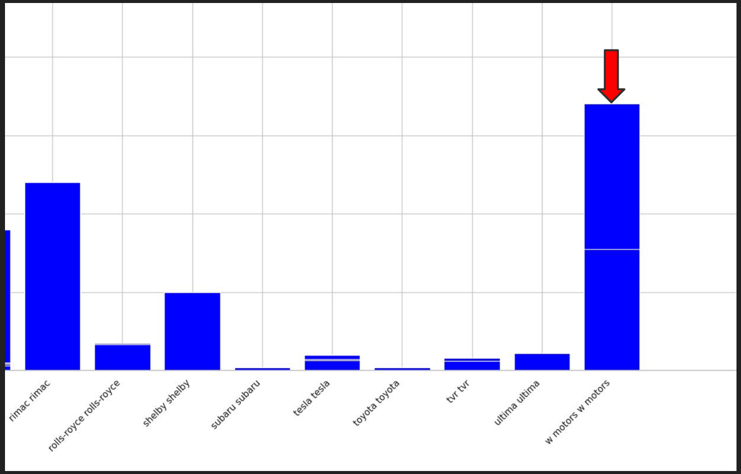
Car Make	Car Model	Year	Engine Size (L)	Horsepower	Torque (lb-ft)	0-60 MPH Time	Price (in USD)	Engine Type	$ per Horsepowe	Origin	Score
porsche	911	2022	3	379	331	4	101200	gas	267.0184697	germany	186.4
lamborghini	huracan	2021	5	630	443	2.8	274390	gas	435.5396825	italy	298.36
ferrari	488 gtb	2022	3	661	561	3	333750	gas	504.9167927	italy	322
audi	r8	2022	5	562	406	3.2	142700	gas	253.9145907	germany	267.54
mclaren	720s	2021	4	710	568	2.7	298000	gas	419.7183099	england	342.54
bmw	m8	2022	4	617	553	3.1	130000	gas	210.6969206	germany	303.92
mercedes-benz	amg gt	2021	4	523	494	3.8	118500	gas	226.5774379	germany	260.56
chevrolet	corvette	2021	6	490	465	2.8	59900	gas	122.244898	america	244.86

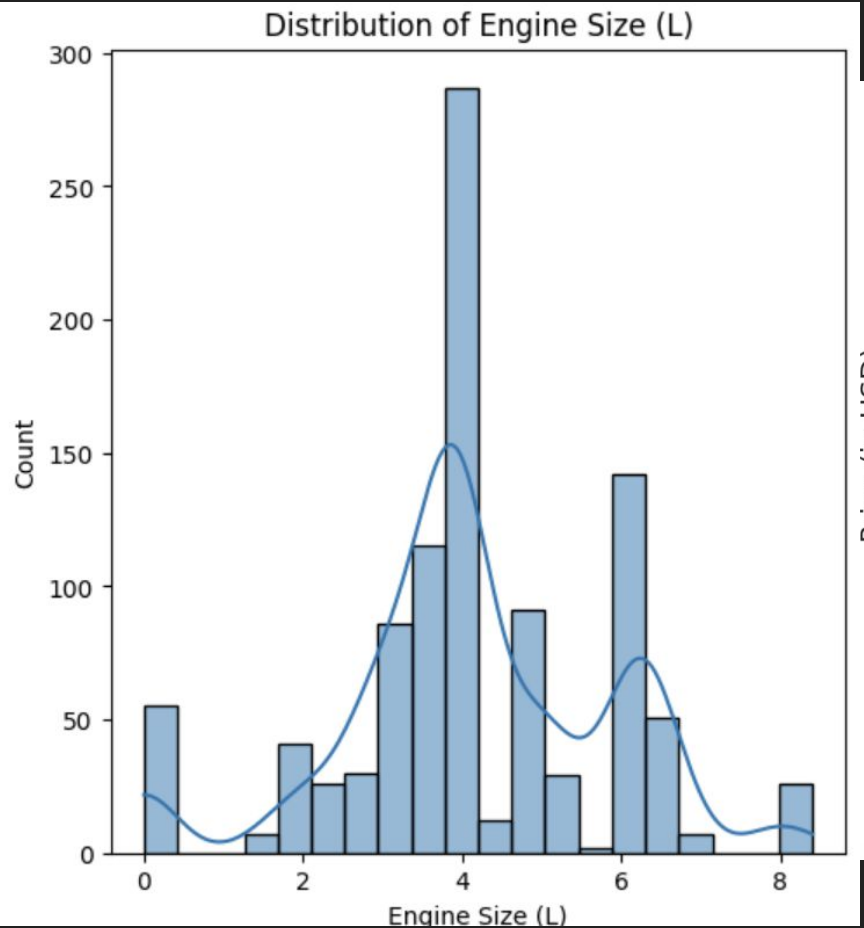

Cost Different Car Make
1e6
5
4
3
2
1
0
Price in USD
acura acura
alfa romeo alfa romeo
alpine alpine
ariel ariel
aston martin aston martin
audi audi
bentley bentley
bmw bmw
bugatti bugatti
chevrolet chevrolet
dodge dodge
ferrari ferrari
ford ford
jaguar jaguar
kia kia
koenigsegg koenigsegg
lamborghini lamborghini
lexus lexus
lotus lotus
maserati maserati
mazda mazda
mclaren mclaren
mercedes-amg mercedes-amg
mercedes-benz mercedes-benz
nissan nissan
pagani pagani
pininfarina pininfarina
polestar polestar
porsche porsche
rimac rimac
rolls-royce rolls-royce
shelby shelby
subaru subaru
tesla tesla
toyota toyota
tvr tvr
ultima ultima
w motors w motors
Car Make

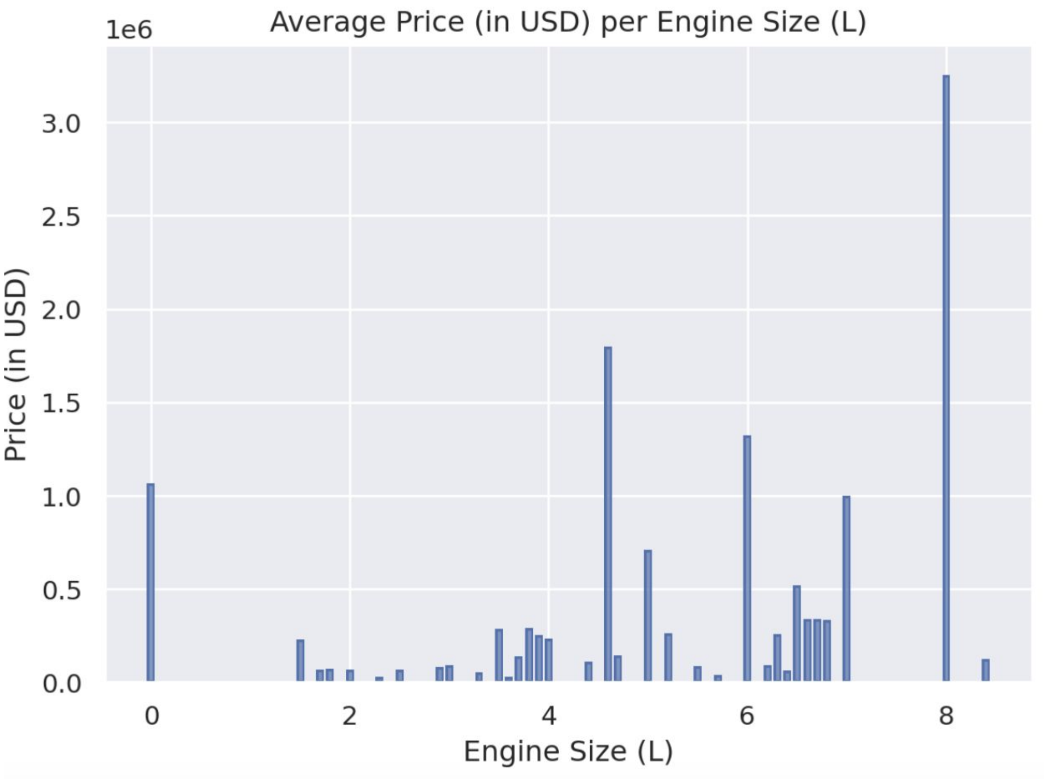


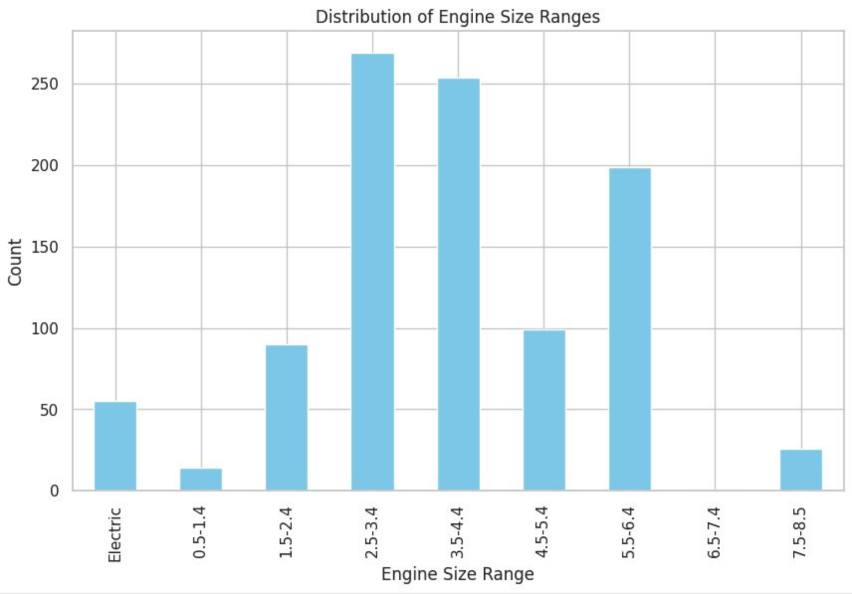
Each unique Car Make within the data set and the associated cost. We can analyze that the Bugatti sports car results in the most expensive car. There are other cars like the Lamborghini, Pagani makes, Koenigsegg, W Motors, that stands out in regard to the upper expensive Car Makes

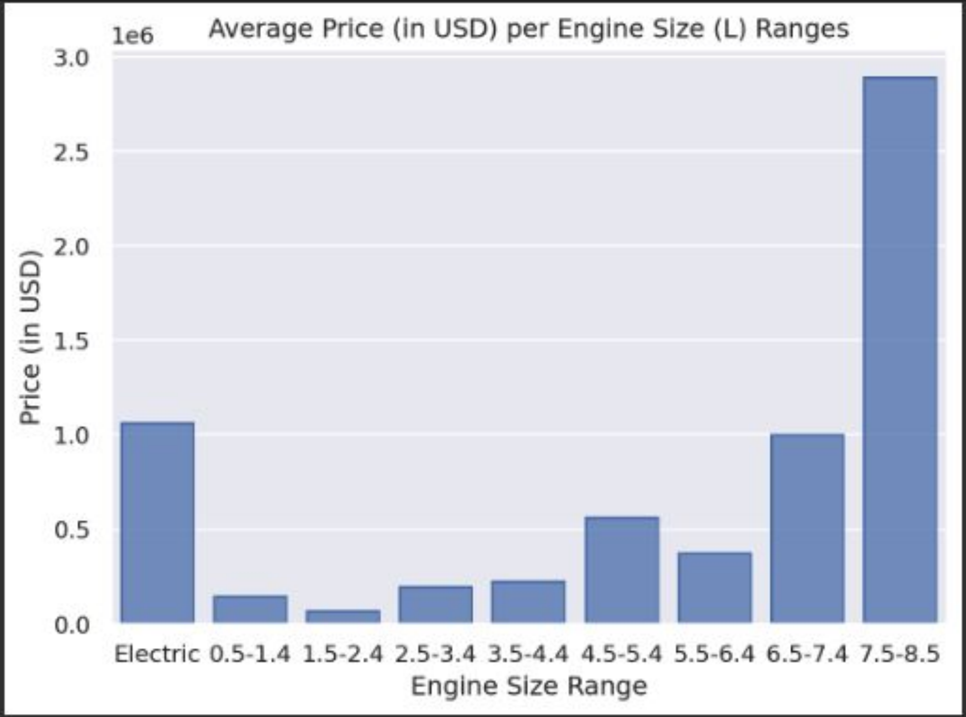





Distribution of Engine Size (L)
300
250
200
150
100
50
0
Count
0
2
4
6
8
Engine Size (L)


Average Price (in USD) per Engine Size (L)
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Price (in USD)
0
2
4
6
8
Engine Size (L)





Total Count of Each Engine Size based on Ranges

2.5-3.4	269
3.5-4.4	254
5.5-6.4	199
4.5-5.4	99
1.5-2.4	90
Electric	55
7.5-8.5	26
0.5-1.4	14
6.5-7.4	1

```
Highest Prices vs Engine Size (L) Values:
```

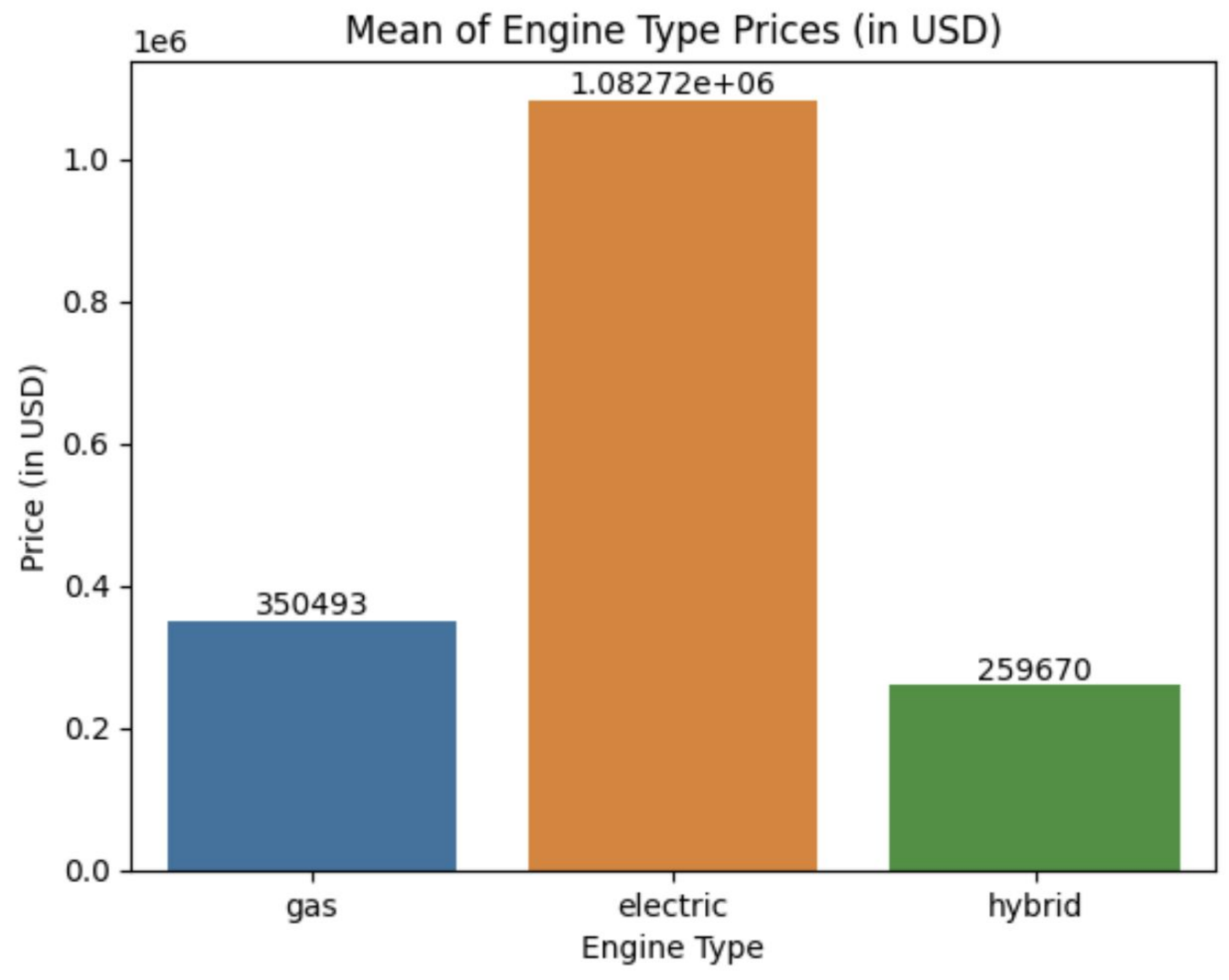
	Car Make	Car Model	Price (in USD)	Engine Size (L)
159	dodge	viper	120000	8.4
369	dodge	viper	118795	8.4
405	dodge	viper acr	126190	8.4
11	bugatti	chiron	3000000	8.0
85	bugatti	chiron	3000000	8.0
113	bugatti	chiron	3000000	8.0
158	bugatti	chiron	3000000	8.0
206	bugatti	chiron	3000000	8.0
274	bugatti	chiron	2998000	8.0
303	bugatti	chiron	2998000	8.0
341	bugatti	chiron	3000000	8.0
376	bugatti	chiron	3000000	8.0
434	bugatti	chiron	3000000	8.0
499	bugatti	chiron	3000000	8.0
519	bugatti	chiron	3000000	8.0
541	bugatti	chiron super sport 300+	5200000	8.0
571	bugatti	chiron	3000000	8.0
624	bugatti	chiron pur sport	3599000	8.0

List of top sports cars with largest engine size

List of top sports cars based on price

```
Highest priced cars
```

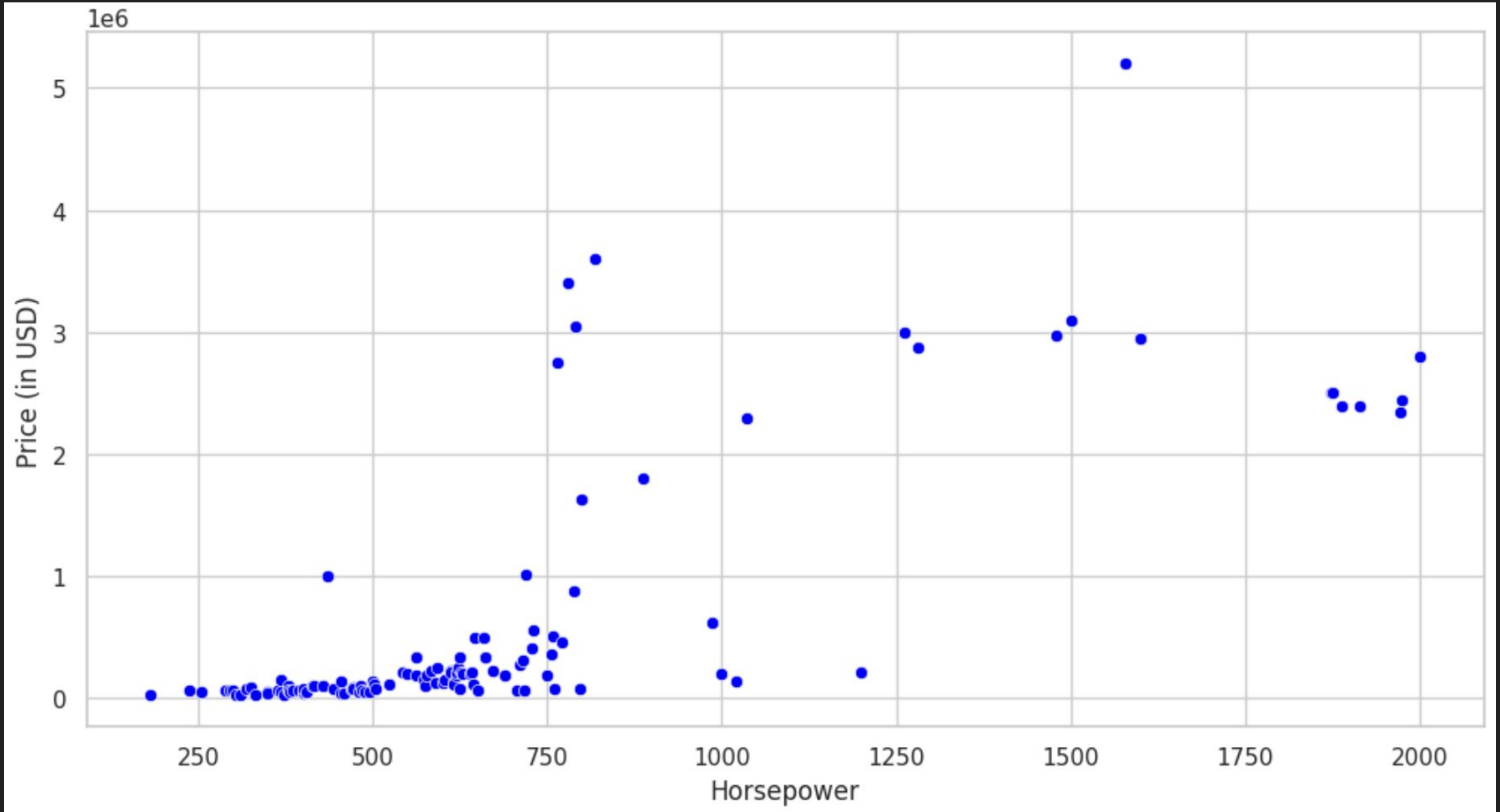
	Car Make	Car Model	Price (in USD)	Engine Size (L) \
541	bugatti	chiron super sport 300+	5200000	8.0
823	bugatti	chiron super sport 300+	5200000	8.0
983	bugatti	chiron	3900000	8.0
438	lamborghini	sián	3600000	6.5
624	bugatti	chiron pur sport	3599000	8.0
279	pagani	huayra roadster bc	3500000	6.0
385	pagani	huayra	3500000	6.0
174	w motors	lykan hypersport	3400000	3.7
11	bugatti	chiron	3000000	8.0
85	bugatti	chiron	3000000	8.0
88	koenigsegg	jesko	3000000	5.0
113	bugatti	chiron	3000000	8.0
158	bugatti	chiron	3000000	8.0
161	koenigsegg	jesko	3000000	5.0
206	bugatti	chiron	3000000	8.0
275	koenigsegg	jesko	3000000	5.0
328	koenigsegg	jesko	3000000	5.0
341	bugatti	chiron	3000000	8.0
376	bugatti	chiron	3000000	8.0

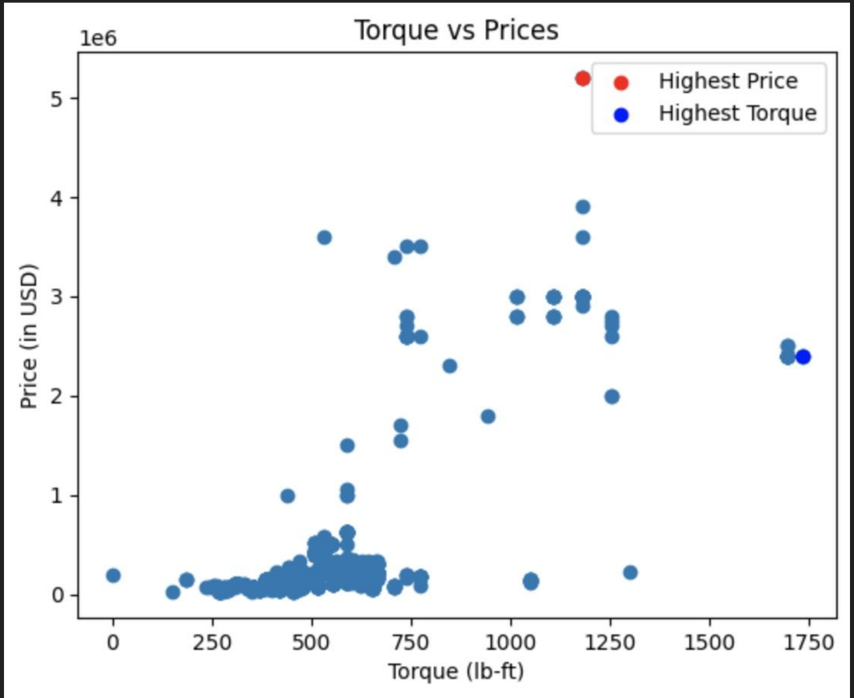



```
Total Count of Each Engine Type
gas          958
electric      44
hybrid         5
```

- Problem : due to the lack of data within dataset #1, it was hard to make a conclusion
    - Majority was gas making that average price of a gas car more accurate then the average of electric and hybrid since there was a smaller sample size of electric and hybrid cars

Horsepower vs Price in USD

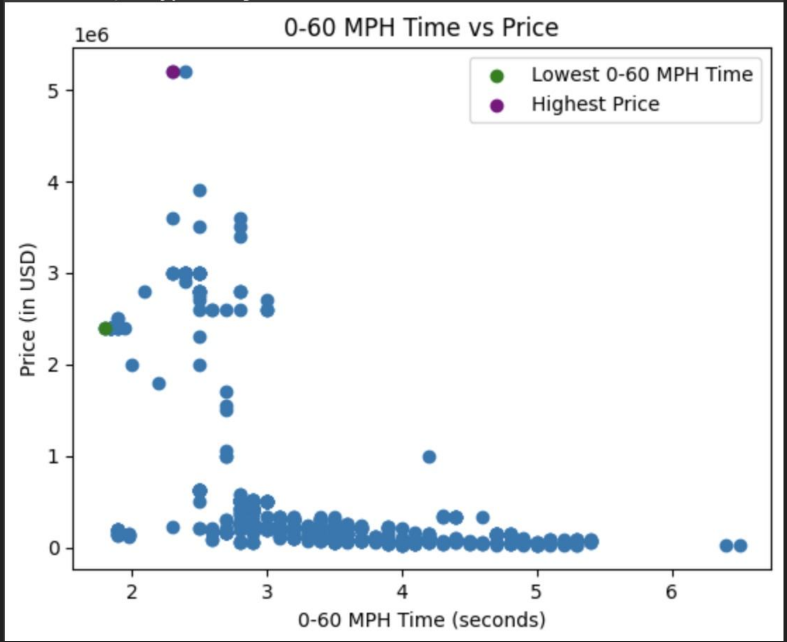





```
Car with the Highest Price:
Car Make                              Bugatti
Car Model                Chiron Super Sport 300+
Year                                     2022
Engine Size (L)                             8
Horsepower                               1578
Torque (lb-ft)                         1180.0
0-60 MPH Time (seconds)                   2.3
Price (in USD)                        5200000
Name: 541, dtype: object

Car with the Highest Torque:
Car Make                     Rimac
Car Model                    C_Two
Year                          2022
Engine Size (L)           Electric
Horsepower                    1914
Torque (lb-ft)              1732.0
0-60 MPH Time (seconds)       1.85
Price (in USD)             2400000
Name: 278, dtype: object
```

```
Highest Torque vs Prices Values:
    Car Make Car Model  Price (in USD)  Torque (lb-ft)
278    Rimac     C_Two         2400000          1732.0
439    Rimac     C_Two         2400000          1732.0
26     Rimac    Nevera         2400000          1696.0
97     Rimac    Nevera         2400000          1696.0
168    Rimac     C_Two         2400000          1696.0
     Price (in USD)  Torque (lb-ft)
278         2400000          1732.0
439         2400000          1732.0
26          2400000          1696.0
97          2400000          1696.0
168         2400000          1696.0
```

```
Highest Prices vs Torque Values:
       Car Make                Car Model  Price (in USD)  Torque (lb-ft)
541     Bugatti  Chiron Super Sport 300+         5200000          1180.0
823     Bugatti  Chiron Super Sport 300+         5200000          1180.0
983     Bugatti                   Chiron         3900000          1180.0
438  Lamborghini                    Sián         3600000           531.0
624     Bugatti         Chiron Pur Sport         3599000          1180.0
     Price (in USD)  Torque (lb-ft)
541         5200000          1180.0
823         5200000          1180.0
983         3900000          1180.0
438         3600000           531.0
624         3599000          1180.0
```




```
Highest Price vs 0-60 MPH Time Values:
        Car Make              Car Model  0-60 MPH Time (seconds)
541      Bugatti  Chiron Super Sport 300+                     2.3
823      Bugatti  Chiron Super Sport 300+                     2.4
983      Bugatti                   Chiron                     2.5
438  Lamborghini                     Sián                     2.8
624      Bugatti         Chiron Pur Sport                     2.3

     Price (in USD)
541        5200000
823        5200000
983        3900000
438        3600000
624        3599000
     0-60 MPH Time (seconds)  Price (in USD)
541                      2.3         5200000
823                      2.4         5200000
983                      2.5         3900000
438                      2.8         3600000
624                      2.3         3599000
```

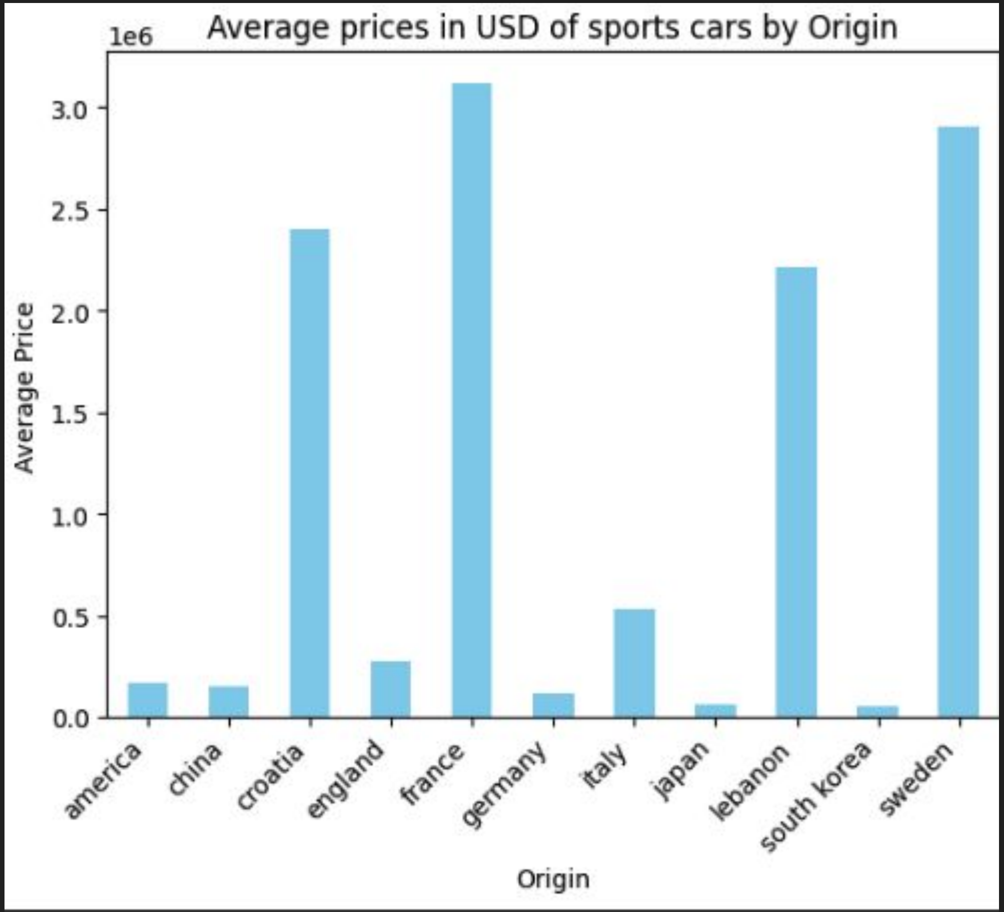
```
'0-60 MPH Time' and 'Price' Categories:
      0-60 MPH Time (seconds)  Price (in USD)
0                        4.00          101200
1                        2.80          274390
2                        3.00          333750
3                        3.20          142700
4                        2.70          298000
...                       ...             ...
1002                     2.50         3000000
1003                     2.00         2000000
1004                     2.70         1000000
1005                     3.00         2600000
1006                     1.85         2400000
```

```
Lowest 0-60 MPH Time vs Highest Price Values:
    Car Make Car Model  0-60 MPH Time (seconds)  Price (in USD)
387    Rimac     C_Two                     1.80         2400000
439    Rimac     C_Two                     1.80         2400000
26     Rimac    Nevera                     1.85         2400000
278    Rimac     C_Two                     1.85         2400000
352    Rimac    Nevera                     1.85         2400000
     0-60 MPH Time (seconds)  Price (in USD)
387                     1.80         2400000
439                     1.80         2400000
26                      1.85         2400000
278                     1.85         2400000
352                     1.85         2400000
```

```
Car with the Lowest 0-60 MPH Time:
Car Make                     Rimac
Car Model                    C_Two
Year                          2022
Engine Size (L)                NaN
Horsepower                    1888
Torque (lb-ft)                1696
0-60 MPH Time (seconds)        1.8
Price (in USD)             2400000
Name: 387, dtype: object

Car with the Highest Price:
Car Make                                 Bugatti
Car Model                Chiron Super Sport 300+
Year                                        2022
Engine Size (L)                                8
Horsepower                                  1578
Torque (lb-ft)                              1180
0-60 MPH Time (seconds)                      2.3
Price (in USD)                           5200000
Name: 541, dtype: object
```

```
Total Count of Origins
germany        287
england        229
america        185
italy          176
japan           72
france          24
sweden          15
croatia         14
lebanon          3
china            1
south korea      1
Name: Origin, dtype: int64


Average Prices of Sports Cars By Origin:
Origin
america        165096
china          155000
croatia       2400000
england        273701
france        3119438
germany        117987
italy          534350
japan           64701
lebanon       2216667
south korea     52200
sweden        2906667
```

```
       Car Make   Origin  Price (in USD)
541     bugatti   france         5200000
823     bugatti   france         5200000
983     bugatti   france         3900000
438  lamborghini   italy         3600000
624     bugatti   france         3599000
279      pagani    italy         3500000
385      pagani    italy         3500000
174    w motors  lebanon         3400000
11      bugatti   france         3000000
85      bugatti   france         3000000
88   koenigsegg   sweden         3000000
113     bugatti   france         3000000
158     bugatti   france         3000000
161  koenigsegg   sweden         3000000
206     bugatti   france         3000000
275  koenigsegg   sweden         3000000
328  koenigsegg   sweden         3000000
341     bugatti   france         3000000
376     bugatti   france         3000000
434     bugatti   france         3000000
435  koenigsegg   sweden         3000000
499     bugatti   france         3000000
```



Highest Scored Sports Car Based of theoretical Weighted Features:

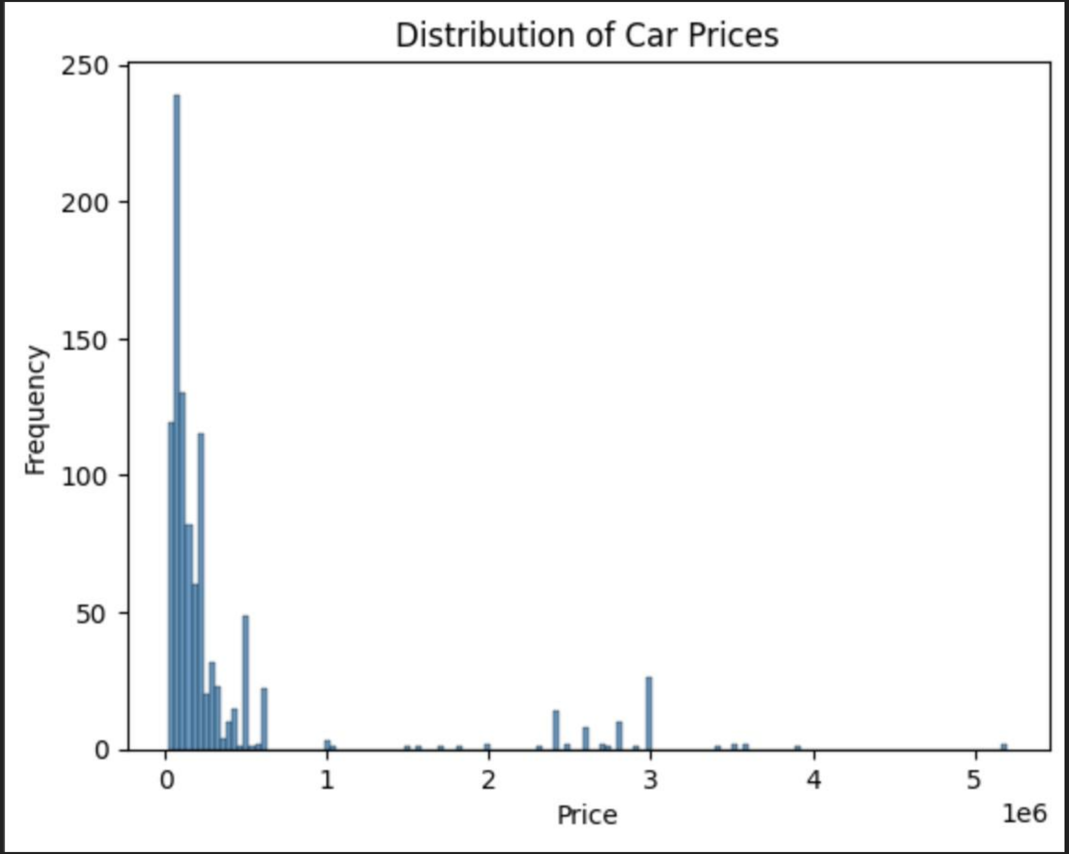
	Car Make	Car Model	Year	Engine Size (L)	Horsepower	Torque (lb-ft)	0-60 MPH Time (seconds)	Price (in USD)	Score
885	tesla	roadster	2022	0	10000	7376	1.9	200000	4738
389	tesla	roadster	2022	0	10000	0	1.9	200000	4000
354	tesla	roadster	2022	0	1000	10000	1.9	200000	1400
278	rimac	c_two	2022	0	1914	1732	1.85	2400000	939
439	rimac	c_two	2021	0	1914	1732	1.8	2400000	939
97	rimac	nevera	2022	0	1914	1696	1.95	2400000	936
168	rimac	c_two	2022	0	1914	1696	1.9	2400000	936
509	rimac	c_two	2021	0	1914	1696	1.9	2400000	936
526	rimac	c_two	2022	0	1914	1696	1.9	2400000	936
640	rimac	nevera	2021	0	1914	1696	1.9	2400000	936
26	rimac	nevera	2022	0	1914	1696	1.85	2400000	936
352	rimac	nevera	2022	0	1914	1696	1.85	2400000	936
686	rimac	c_two	2022	0	1914	1696	1.85	2400000	936
824	rimac	nevera	2021	0	1914	1696	1.85	2400000	936
986	rimac	nevera	2022	0	1914	1696	1.85	2400000	936
877	lotus	evija	2021	0	2000	1254	2.8	2800000	926
1006	rimac	nevera	2021	0	1888	1696	1.85	2400000	925
387	rimac	c_two	2022	0	1888	1696	1.8	2400000	925
280	pininfarina	battista	2022	0	1874	1696	1.9	2500000	920
988	pininfarina	battista	2021	0	1872	1696	1.9	2500000	919
420	lotus	evija	2022	0	1973	1254	2.5	2750000	915
523	lotus	evija	2022	0	1973	1254	2.5	2600000	915
987	lotus	evija	2022	0	1973	1254	2.5	2000000	915
697	lotus	evija	2022	0	1972	1254	2.5	2700000	915
1003	lotus	evija	2021	0	1972	1254	2	2000000	915
88	koenigsegg	jesko	2022	5	1600	1106	2.5	3000000	753
161	koenigsegg	jesko	2022	5	1600	1106	2.5	3000000	753
822	koenigsegg	jesko	2022	5	1600	1106	2.5	3000000	753
418	koenigsegg	jesko absolut	2022	5	1600	1106	2.1	2800000	753
823	bugatti	chiron super sport 300+	2021	8	1578	1180	2.4	5200000	752
541	bugatti	chiron super sport 300+	2022	8	1578	1180	2.3	5200000	752
631	bugatti	chiron	2021	8	1500	1180	2.5	3000000	721
983	bugatti	chiron	2022	8	1500	1180	2.5	3900000	721
11	bugatti	chiron	2021	8	1500	1180	2.4	3000000	721
85	bugatti	chiron	2022	8	1500	1180	2.4	3000000	721

Highest Priced Sports Car

	Car Make	Car Model	Year	Engine Size (L)	Horsepower	Torque (lb-ft)	0-60 MPH Time (seconds)	Price (in USD)	Score
541	bugatti	chiron super sport 300+	2022	8	1578	1180	2.3	5200000	752
823	bugatti	chiron super sport 300+	2021	8	1578	1180	2.4	5200000	752
983	bugatti	chiron	2022	8	1500	1180	2.5	3900000	721
438	lamborghini	sián	2021	6	819	531	2.8	3600000	383
624	bugatti	chiron pur sport	2021	8	1500	1180	2.3	3599000	721
279	pagani	huayra roadster bc	2021	6	791	774	2.5	3500000	396
385	pagani	huayra	2021	6	764	738	2.8	3500000	382
174	w motors	lykan hypersport	2015	3	780	708	2.8	3400000	384
11	bugatti	chiron	2021	8	1500	1180	2.4	3000000	721
85	bugatti	chiron	2022	8	1500	1180	2.4	3000000	721
88	koenigsegg	jesko	2022	5	1600	1106	2.5	3000000	753
113	bugatti	chiron	2021	8	1500	1180	2.4	3000000	721
158	bugatti	chiron	2021	8	1500	1180	2.4	3000000	721
161	koenigsegg	jesko	2022	5	1600	1106	2.5	3000000	753
206	bugatti	chiron	2021	8	1500	1180	2.3	3000000	721
275	koenigsegg	jesko	2021	5	1280	1015	2.5	3000000	616
328	koenigsegg	jesko	2022	5	1280	1015	2.5	3000000	616
341	bugatti	chiron	2021	8	1500	1180	2.4	3000000	721
376	bugatti	chiron	2022	8	1500	1180	2.4	3000000	721
434	bugatti	chiron	2022	8	1500	1180	2.4	3000000	721
435	koenigsegg	jesko	2021	5	1262	1106	2.5	3000000	617
499	bugatti	chiron	2022	8	1500	1180	2.3	3000000	721
519	bugatti	chiron	2021	8	1500	1180	2.3	3000000	721
571	bugatti	chiron	2021	8	1479	1180	2.5	3000000	712
631	bugatti	chiron	2021	8	1500	1180	2.5	3000000	721
683	bugatti	chiron	2022	8	1500	1180	2.4	3000000	721
782	bugatti	chiron	2021	8	1479	1180	2.4	3000000	712
822	koenigsegg	jesko	2022	5	1600	1106	2.5	3000000	753
898	bugatti	chiron	2021	8	1500	1180	2.4	3000000	721
984	koenigsegg	jesko	2022	5	1280	1015	2.5	3000000	616
1001	bugatti	chiron	2021	8	1479	1180	2.4	3000000	712
1002	koenigsegg	jesko	2022	5	1280	1106	2.5	3000000	625
274	bugatti	chiron	2021	8	1500	1180	2.4	2998000	721
303	bugatti	chiron	2021	8	1479	1180	2.3	2998000	712
864	bugatti	chiron	2022	8	1479	1180	2.4	2900000	712
14	koenigsegg	jesko	2021	5	1280	1015	2.5	2800000	616
24	pagani	huayra	2021	6	720	737	2.8	2800000	364

- Sports car prices are right-skewed distribution indicating that there are relatively fewer sports cars with the extreme price points, which leads to longer right tail

- You can see the asymmetry in the distribution of prices, from this dataset we can say that sports car are relatively affordable, however is price point what really makes a sports car a sports car? It is definitely not the sole determinant.

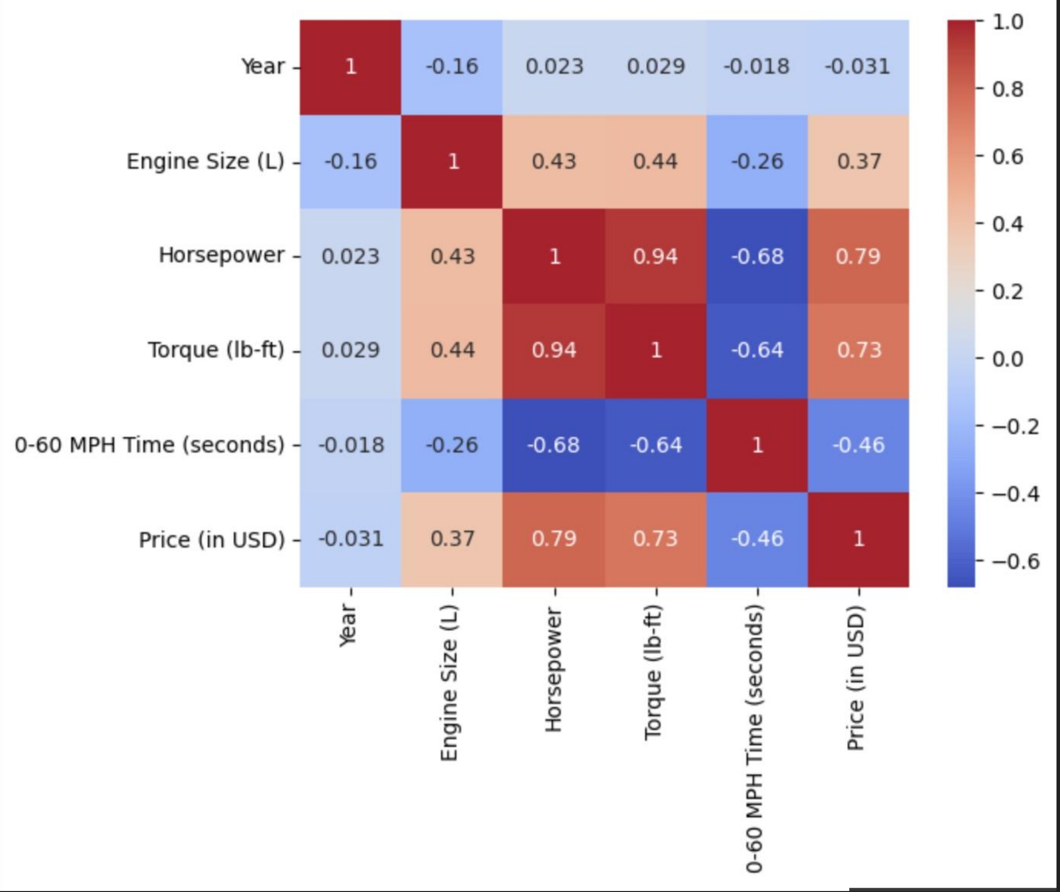


```
Crosstabulation for segment and price:

price      30    35    40  Total
segment
basic    1288  1280  1272   3840
fun       514   520   496   1530
racer     206   211   213    630
Total    2008  2011  1981   6000

Feature preference for each element in segment and price:

price            30          35          40  Total
segment
basic     64.143426   63.649925   64.209995   64.0
fun       25.597610   25.857782   25.037860   25.5
racer     10.258964   10.492292   10.752145   10.5
Total    100.000000  100.000000  100.000000  100.0

Crosstabulation for seat and price:

price    30    35    40  Total
seat
2       667   668   678   2013
4       672   674   660   2006
5       669   669   643   1981
Total  2008  2011  1981   6000
```

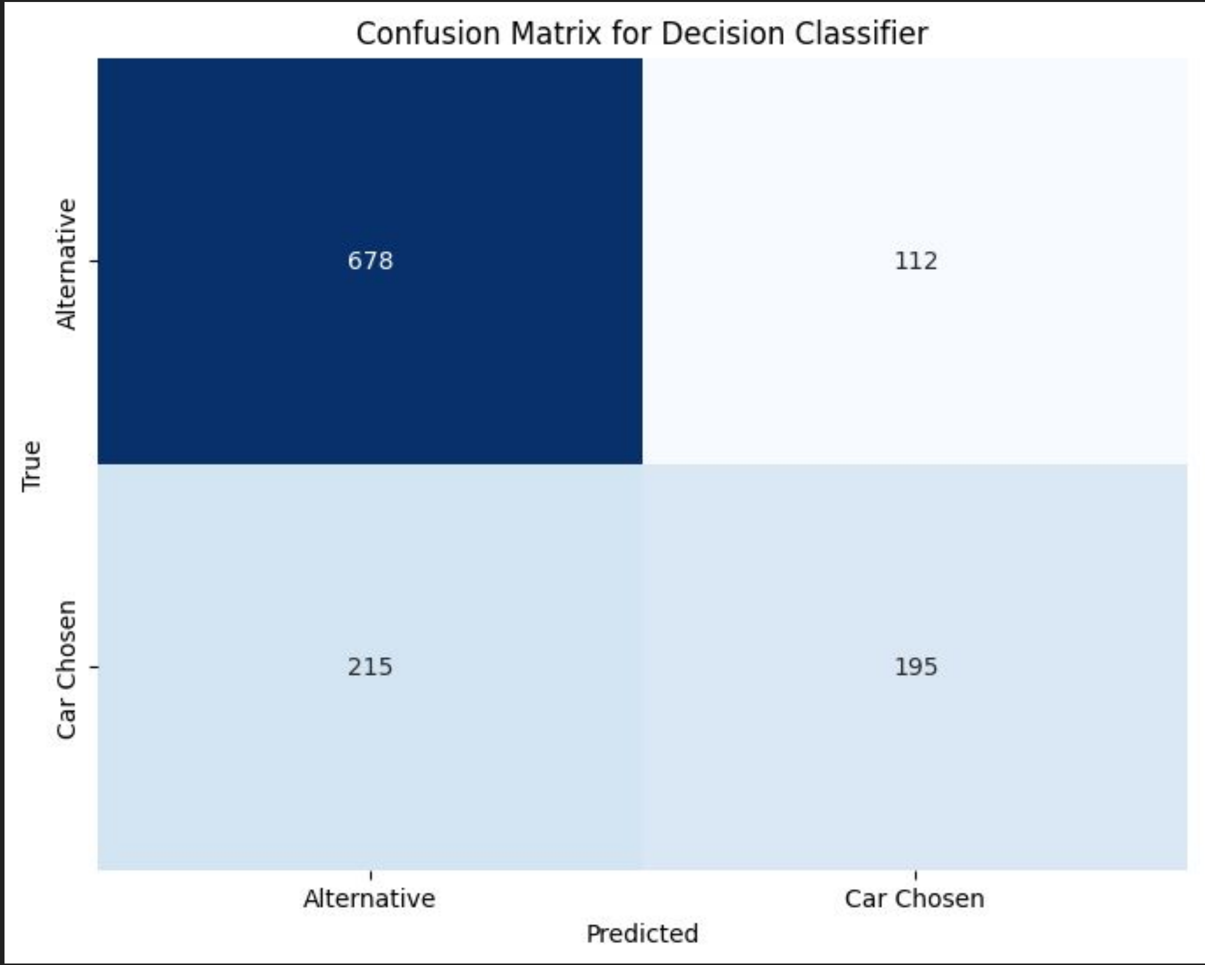
```
no         993  1012   983   2988
yes       1015   999   998   3012
Total     2008  2011  1981   6000

Feature preference for each element in convert and price:

price            30          35          40  Total
convert
no        49.452191   50.323222   49.621403   49.8
yes       50.547809   49.676778   50.378597   50.2
Total    100.000000  100.000000  100.000000  100.0

Crosstabulation for choice and price:

price     30    35    40  Total
choice
0        998  1345  1657   4000
1       1010   666   324   2000
Total   2008  2011  1981   6000

Feature preference for each element in choice and price:

price           30          35          40       Total
choice
0        49.701195   66.882148   83.644624   66.666667
1        50.298805   33.117852   16.355376   33.333333
Total   100.000000  100.000000  100.000000  100.000000
```

```
Crosstabulation for segment and choice:

choice      0     1  Total
segment
basic    2560  1280   3840
fun      1020   510   1530
racer     420   210    630
Total    4000  2000   6000

Percentage of decision-making for each element in segment:

choice       0      1  Total
segment
basic     64.0   64.0   64.0
fun       25.5   25.5   25.5
racer     10.5   10.5   10.5
Total    100.0  100.0  100.0

Crosstabulation for seat and choice:

choice     0     1  Total
seat
2       1405   608   2013
4       1390   616   2006
5       1205   776   1981
Total   4000  2000   6000

Percentage of decision-making for each element in seat:

choice        0      1       Total
seat
2        35.125   30.4   33.550000
4        34.750   30.8   33.433333
5        30.125   38.8   33.016667
Total   100.000  100.0  100.000000

Crosstabulation for trans and choice:

choice     0     1  Total
trans
auto    1673  1328   3001
manual  2327   672   2999
Total   4000  2000   6000
```

```
Percentage of decision-making for each element in trans:

choice        0      1       Total
trans
auto     41.825   66.4   50.016667
manual   58.175   33.6   49.983333
Total   100.000  100.0  100.000000

Crosstabulation for convert and choice:

choice      0     1  Total
convert
no       2047   941   2988
yes      1953  1059   3012
Total    4000  2000   6000

Percentage of decision-making for each element in convert:

choice         0       1  Total
convert
no        51.175   47.05   49.8
yes       48.825   52.95   50.2
Total    100.000  100.00  100.0

Crosstabulation for price and choice:

choice     0     1  Total
price
30       998  1010   2008
35      1345   666   2011
40      1657   324   1981
Total   4000  2000   6000

Percentage of decision-making for each element in price:

choice        0      1       Total
price
30       24.950   50.5   33.466667
35       33.625   33.3   33.516667
40       41.425   16.2   33.016667
Total   100.000  100.0  100.000000
```



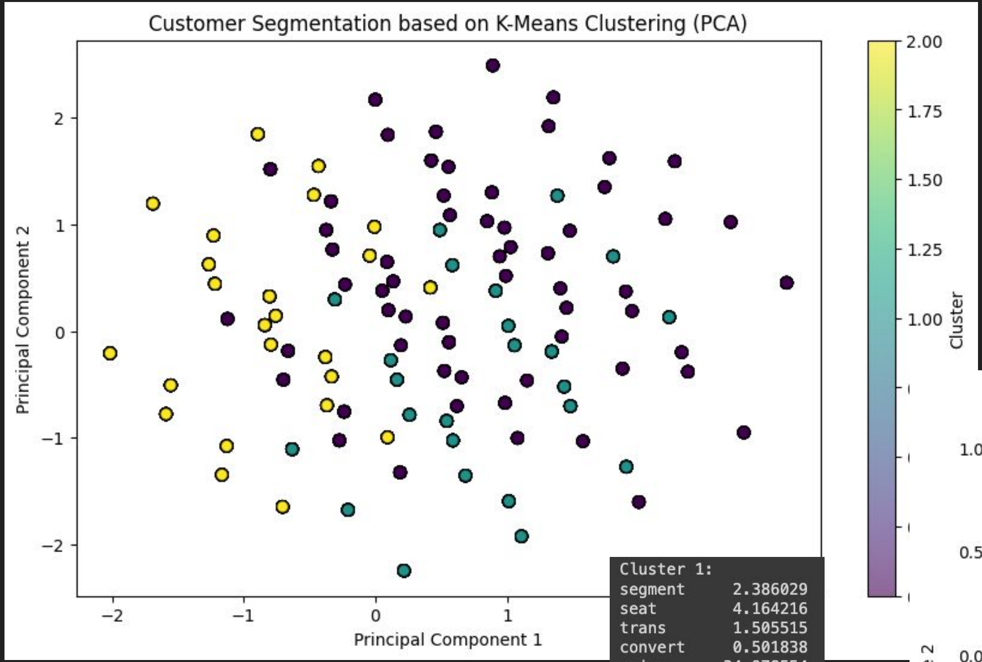
Year
Engine Size (L)
Horsepower
Torque (lb-ft)
0-60 MPH Time (seconds)
Price (in USD)
1 -0.16 0.023 0.029 -0.018 -0.031
-0.16 1 0.43 0.44 -0.26 0.37
0.023 0.43 1 0.94 -0.68 0.79
0.029 0.44 0.94 1 -0.64 0.73
-0.018 -0.26 -0.68 -0.64 1 -0.46
-0.031 0.37 0.79 0.73 -0.46 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6



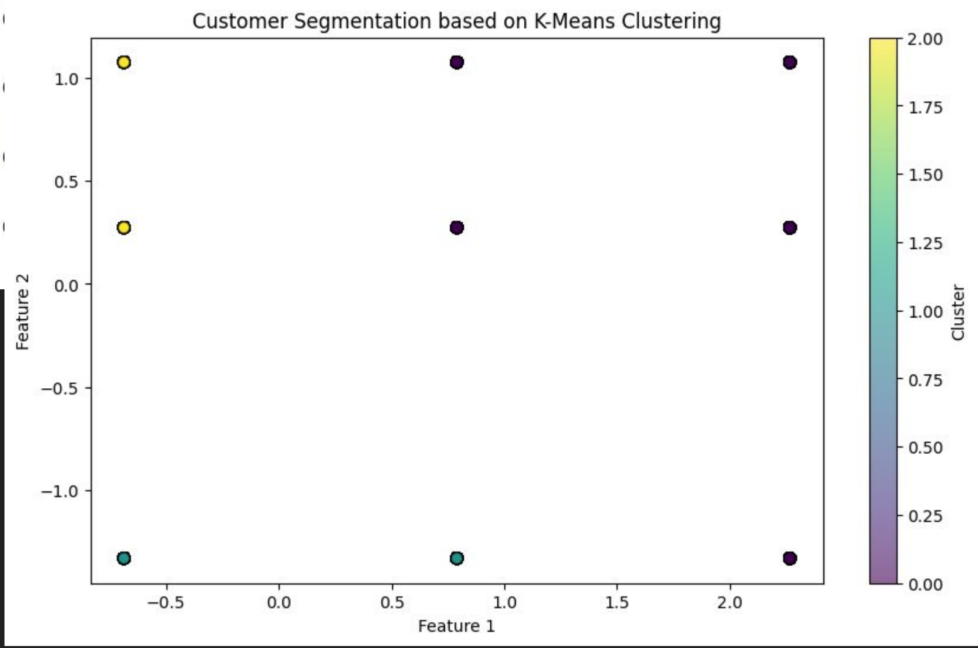
Confusion Matrix for Decision Classifier
True
Alternative
Car Chosen
678
112
215
195
Alternative
Car Chosen
Predicted

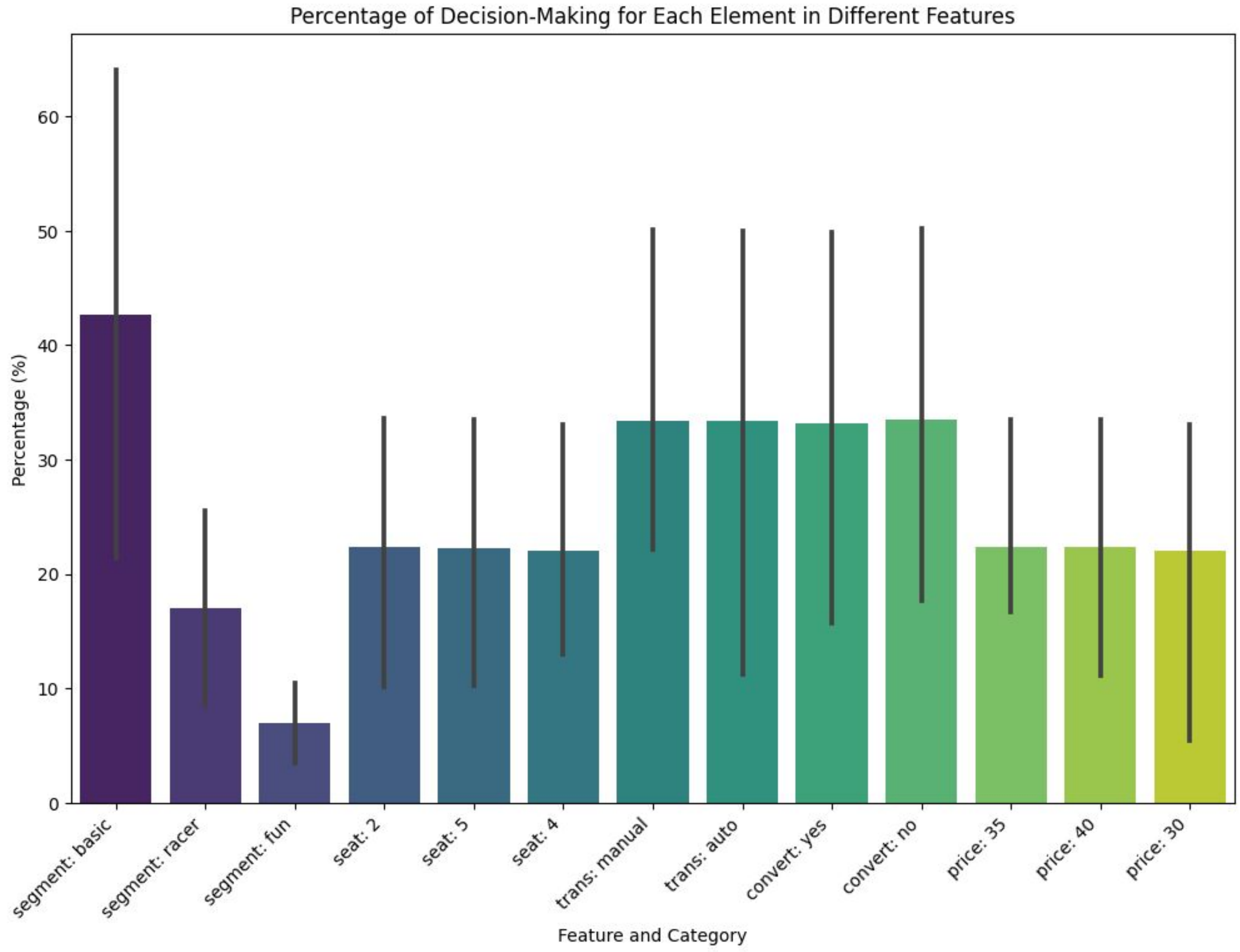
```
Metrics for Dataset 1:
MAE: 38867.66752479322, MSE: 20783945379.814144, R-squared: 0.9582037343230088

Metrics for Dataset 2:
MAE: 1.000000000000001e-05, MSE: 1.3333333333333357e-07, R-squared: 0.9999995953868389
Best hyperparameters for Dataset 1: {'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50}
Cross-validation scores for Dataset 1: [-45150.45127416 -31079.63919398 -54360.26598507 -25147.09363467
 -47186.30520951]
Cross-validation scores for Dataset 2: [-7.83333333e-05 -2.21666667e-04 -0.00000000e+00 -1.66666667e-05
 -1.33333333e-05]
Updated Metrics for Dataset 1:
MAE: 38867.66752479322, MSE: 20783945379.814144, R-squared: 0.9582037343230088

Updated Metrics for Dataset 2:
MAE: 1.000000000000001e-05, MSE: 1.3333333333333357e-07, R-squared: 0.9999995953868389
```

```
Mean Squared Error (MSE) for Price Prediction: 248313548283.21426
R-squared: 0.5906905608619119
Mean Squared Error (MSE) for Price Prediction: 26059083644.368526
R-squared: 0.9461832545187393
Drive already mounted at /content/drive; to attempt to forcibly remount,
   resp_id  ques  alt segment  seat   trans convert  price  choice
0        1     1    1   basic     2  manual     yes     35       0
1        1     1    2   basic     5    auto      no     40       0
2        1     1    3   basic     5    auto      no     30       1
3        1     2    1   basic     5  manual      no     35       0
4        1     2    2   basic     2  manual      no     30       1
Accuracy for Decision Classifier: 0.7275
```

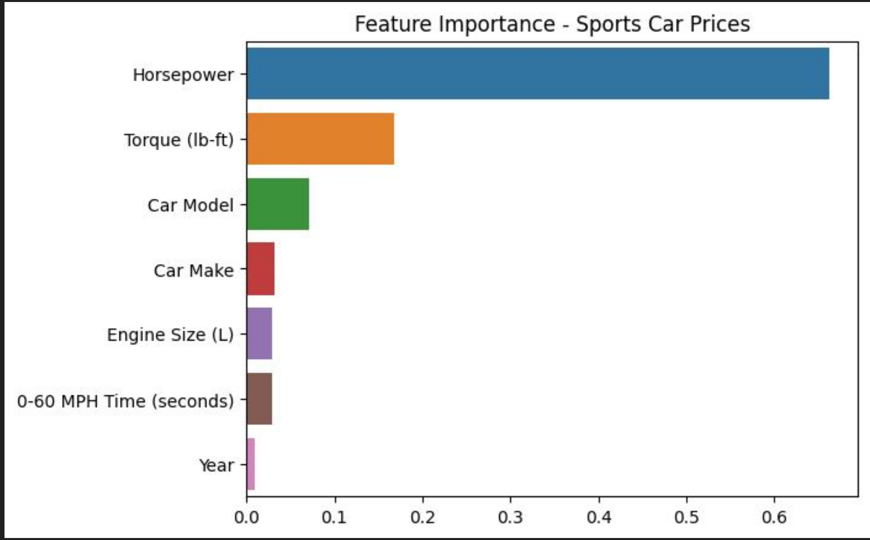


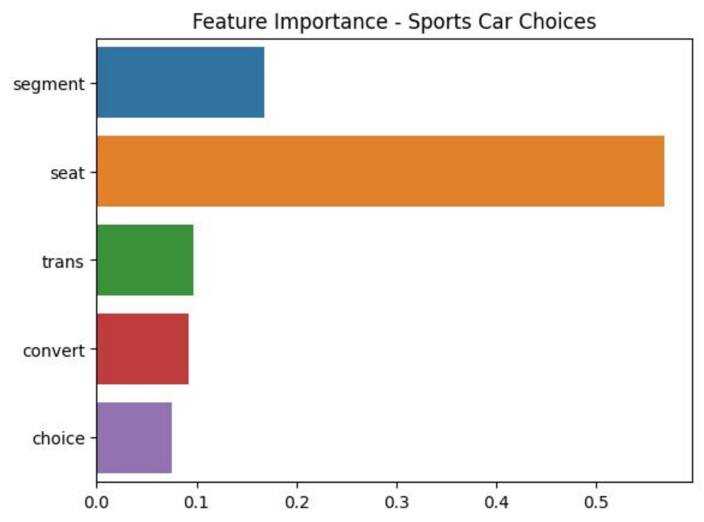

```
Cluster 1:
segment     2.386029
seat        4.164216
trans       1.505515
convert     0.501838
price      34.978554
dtype: float64

Cluster 2:
segment     1.293333
seat        2.000000
trans       1.496111
convert     0.497222
price      35.011111
dtype: float64

Cluster 3:
segment     1.000000
seat        4.501168
trans       1.499611
convert     0.505452
price      34.953271
dtype: float64
```



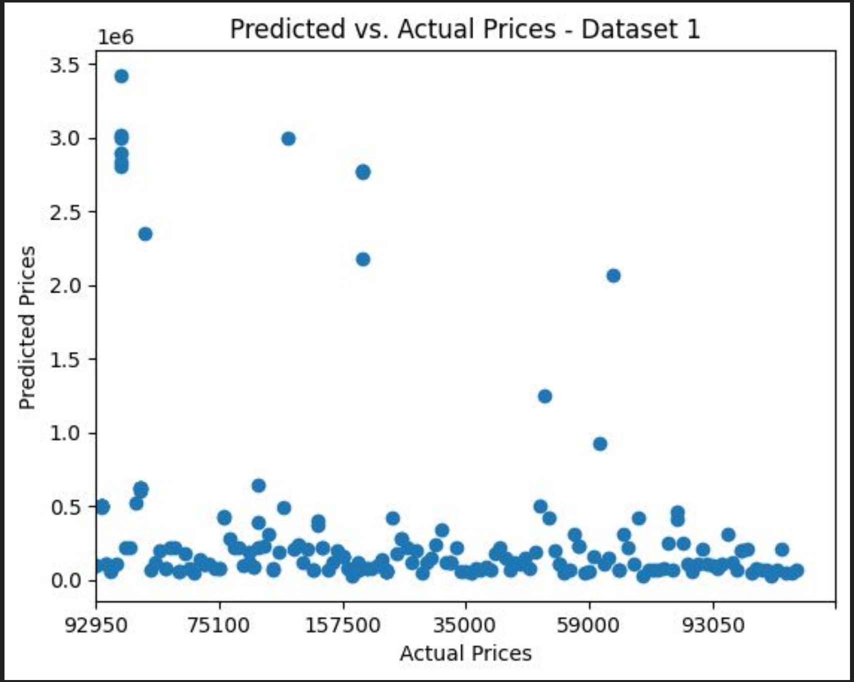


Percentage of Decision-Making for Each Element in Different Features
Percentage (%)
0
10
20
30
40
50
60
segment: basic
segment: racer
segment: fun
seat: 2
seat: 5
seat: 4
trans: manual
trans: auto
convert: yes
convert: no
price: 35
price: 40
price: 30
Feature and Category

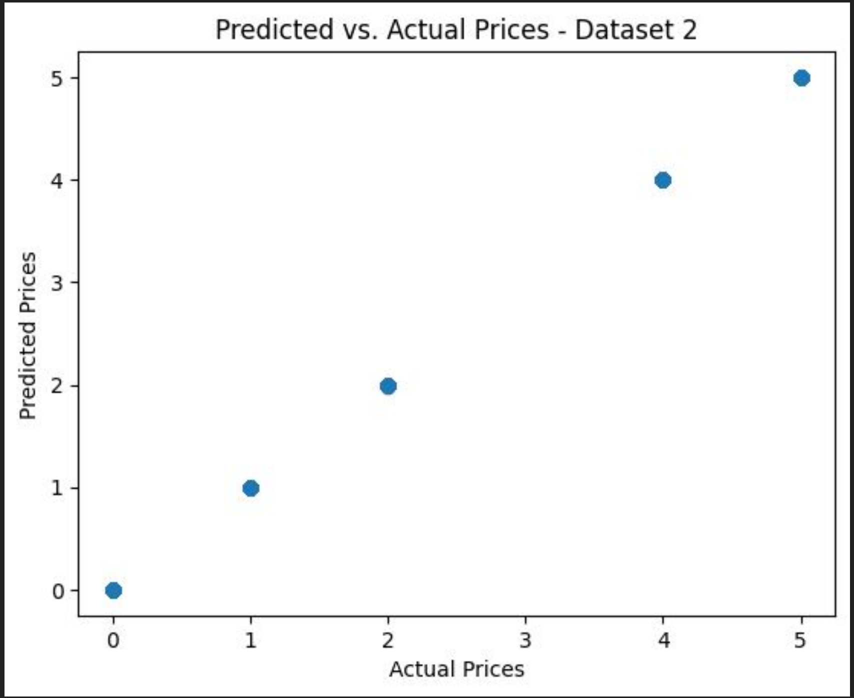




- Seat count:
  - Using a 2-seater as the baseline seat configuration, we can see that respondents were statistically more likely to choose 5-seater, with a confidence level of 99.9%.
- Transmission type:
  - Respondents were much more likely to choose automatic when asked to choose between cars with automatic and manual transmissions.
- Convertible tops:
  - Convertible-top car models were statistically more popular than those with standard roofs.
- Price:
  - Chosen cars were statistically cheaper than the alternatives
- Price interacted with Segment
  - When controlling for price, both the fun and racer segments were statistically more likely to be chosen at higher price points than their basic counterpart



Predicted vs. Actual Prices - Dataset 1
1e6
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Predicted Prices
92950
75100
157500
35000
59000
93050
Actual Prices



Predicted vs. Actual Prices - Dataset 2
5
4
3
2
1
0
Predicted Prices
0
1
2
3
4
5
Actual Prices

Conclusion

- Feature Importance of car does not pinpoint what actually makes a car valuable (rarity, popularity, etc.)
- A sports car can have affordable price points and still be considered valuable because it offer specifications, functionality, performance, etc.
- We explored in the secondary dataset how the decision making of a client results in different feature importance than the original dataset.
- We can confidently say through various pieces of evidence and data analysis that Horsepower is one of the prominent features when it comes to value of a sports car. For characteristics dataset we can say that basic commercial segmented model or auto transmission is more desired and considered more important. Note: this can be different for everyone