Used-Car Analysis

Volga Shaji

Background

We will soon open a new used car sales branch in the city of Mississauga, Ontario.

To prepare, we have obtained data from the Toronto branch and analyze queries using SPSS Modeler and Cognos

https://www.kaggle.com/datasets/gagandeep16/car-sales?resource=download

Data queries

A. Data visualization

Market and data general understanding

B. Data modeling

1. What are the key factors influencing volume of sales? (decision tree)
2. What are the key factors predicting volume of sales?(linear regression model)
3. Model for resale price calculation (linear regression model)

Data Pre-processing

Data wrangling

```
data = pd.read_csv("D://car_sales.csv")
```

```
data.isnull().sum()
```

```
Manufacturer            0
Model                   0
Sales_in_thousands      0
__year_resale_value    36
Vehicle_type            0
Price_in_thousands      2
Engine_size             1
Horsepower              1
Wheelbase               1
Width                   1
Length                  1
Curb_weight             2
Fuel_capacity           1
Fuel_efficiency         3
Latest_Launch           0
Power_perf_factor       2
dtype: int64
```

```
data.shape
```

```
(157, 16)
```

```
data = data.dropna()
```

```
data.shape
```

Remove null values

```
data['Price_differences'] = data['Price_in_thousands'] - data['__year_resale_value']
print(data)
```

```
    Manufacturer    Model  Sales_in_thousands  __year_resale_value  \
0          Acura  Integra              16.919               16.360
1          Acura       TL              39.384               19.875
3          Acura       RL               8.588               29.725
4           Audi       A4              20.397               22.255
5           Audi       A6              18.780               23.555
..           ...      ...                 ...                  ...
145   Volkswagen     Golf               9.761               11.425
146   Volkswagen    Jetta              83.721               13.240
147   Volkswagen   Passat              51.102               16.725
148   Volkswagen   Cabrio               9.569               16.575
149   Volkswagen      GTI               5.596               13.760

    Vehicle_type  Price_in_thousands  Engine_size  Horsepower  Wheelbase  \
0      Passenger               21.50          1.8       140.0      101.2
1      Passenger               28.40          3.2       225.0      108.1
3      Passenger               42.00          3.5       210.0      114.6
4      Passenger               23.99          1.8       150.0      102.6
5      Passenger               33.95          2.8       200.0      108.7
..           ...                 ...          ...         ...        ...
145    Passenger               14.90          2.0       115.0       98.9
146    Passenger               16.70          2.0       115.0       98.9
147    Passenger               21.20          1.8       150.0      106.4
148    Passenger               19.99          2.0       115.0       97.4
149    Passenger               17.50          2.0       115.0       98.9
...
148          48.907372             3.415
149          47.946841             3.740

[117 rows x 17 columns]
```

```
#data['capacity'] = data['Length'] * data['Width']
data.Width
```

```
0      67.3
1      70.3
3      71.4
4      68.2
5      76.1
       ...
145    68.3
146    68.3
147    68.5
148    66.7
149    68.3
Name: Width, Length: 117, dtype: float64
```

Add column: price differences & capacity

Metadata

SALES INFO

Sales_in_thousands: total count of units sold (thousands)
__year_resale_value: average value of resale (thousands $)
Price_in_thousands: average value of new car (thousands $)
Price_differences = Price_in_thousands - __year_resale_value: price difference between new and old car

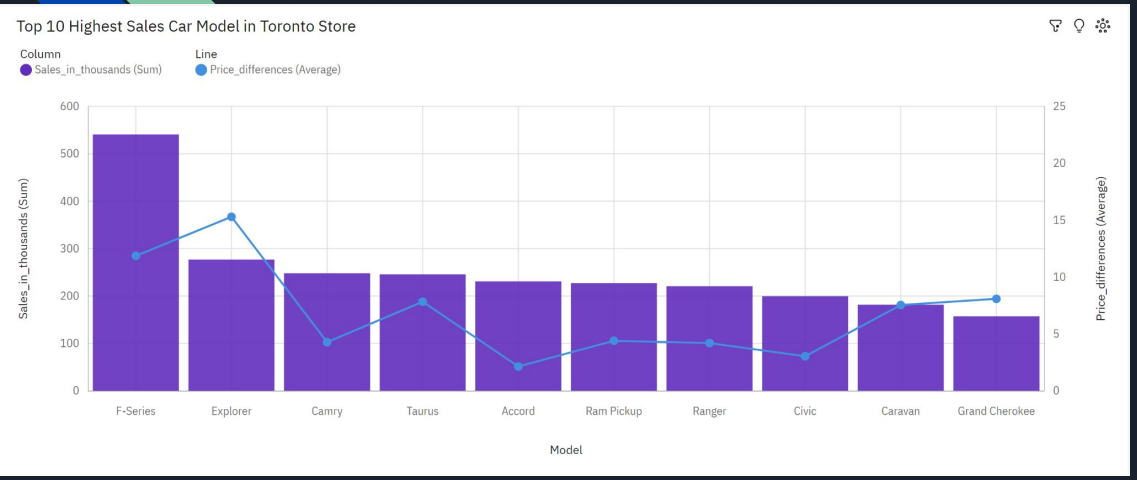
CAR INFO

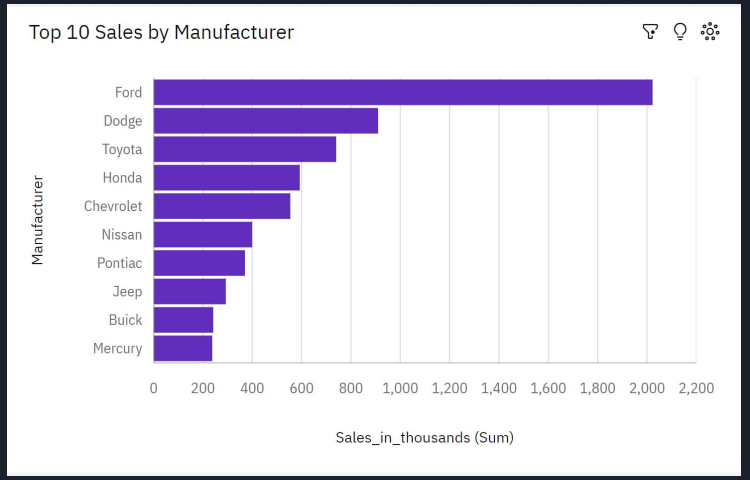
Manufacturer :-- car companies
Model: Car model
Vehicle_type: Car (2-4 seats) or Passenger (4-9 seats)
Engine_size:- size of engine in liters
Horsepower: measure of an engine's power, often measured in units like horsepower (hp) or kilowatts (kW).
Wheelbase: distance between the centers of the front and rear wheels of a vehicle, usually measured in inches or centimeters.
Width: wide a vehicle is from side to side
Length: how long a vehicle is from front to rear, usually measured in inches
Curb_weight: weight of a vehicle when it's empty and ready to drive, including all fluids and a full tank of fuel, but without any passengers or cargo.
Fuel_capacity: maximum amount of fuel a vehicle's fuel tank can hold, often measured in gallons or liters.
Fuel_efficiency: how far a vehicle can travel on a certain amount of fuel, often expressed in miles per gallon (mpg) or liters per 100 kilometers (L/100km).
Latest_Launch: Date launched
Power_perf_factor: car performance uses these factors, racing weight, engine, aerodynamic, transmission and chassis parameters, to create the "Pf" value.
Capacity: length * width

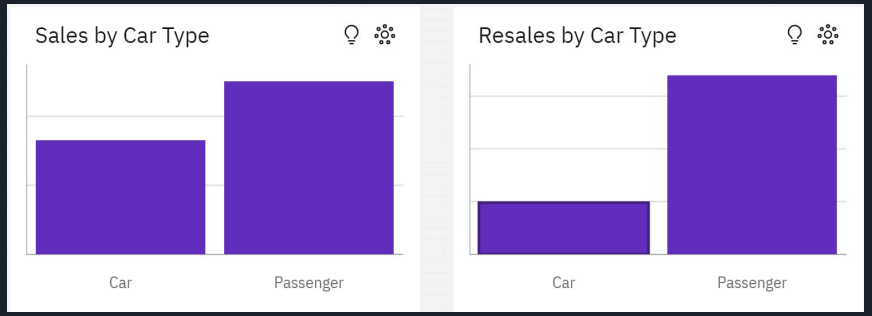
Input Data

	Manufactu	Model	Sales_in_th	__year_res	Vehicle_ty	Price_in_th	Engine_size	Horsepow	Wheelbase	Width	Length	Curb_weig	Fuel_capac	Fuel_efficie	Latest_Launch	Power_perf_factor
0	Acura	Integra	16.919	16.36	Passenger	21.5	1.8	140	101.2	67.3	172.4	2.639	13.2	28	2/2/2012	58.28015
1	Acura	TL	39.384	19.875	Passenger	28.4	3.2	225	108.1	70.3	192.9	3.517	17.2	25	6/3/2011	91.37078
3	Acura	RL	8.588	29.725	Passenger	42	3.5	210	114.6	71.4	196.6	3.85	18	22	3/10/2011	91.38978
4	Audi	A4	20.397	22.255	Passenger	23.99	1.8	150	102.6	68.2	178	2.998	16.4	27	10/8/2011	62.77764
5	Audi	A6	18.78	23.555	Passenger	33.95	2.8	200	108.7	76.1	192	3.561	18.5	22	8/9/2011	84.56511
6	Audi	A8	1.38	39	Passenger	62	4.2	310	113	74	198.2	3.902	23.7	21	2/27/2012	134.6569
8	BMW	328i	9.231	28.675	Passenger	33.4	2.8	193	107.3	68.5	176	3.197	16.6	24	1/29/2012	81.87707
9	BMW	528i	17.527	36.125	Passenger	38.9	2.8	193	111.4	70.9	188	3.472	18.5	25	4/4/2011	83.99872
10	Buick	Century	91.561	12.475	Passenger	21.975	3.1	175	109	72.7	194.6	3.368	17.5	25	11/2/2011	71.18145
11	Buick	Regal	39.35	13.74	Passenger	25.3	3.8	240	109	72.7	196.2	3.543	17.5	23	9/3/2011	95.6367
12	Buick	Park Avenu	27.851	20.19	Passenger	31.965	3.8	205	113.8	74.7	206.8	3.778	18.5	24	3/23/2012	85.82841
13	Buick	LeSabre	83.257	13.36	Passenger	27.885	3.8	205	112.2	73.5	200	3.591	17.5	25	7/23/2011	84.25453
14	Cadillac	DeVille	63.729	22.525	Passenger	39.895	4.6	275	115.3	74.5	207.2	3.978	18.5	22	2/23/2012	113.8546
16	Cadillac	Eldorado	6.536	25.725	Passenger	39.665	4.6	275	108	75.5	200.6	3.843	19	22	11/27/2011	113.7659
17	Cadillac	Catera	11.185	18.225	Passenger	31.01	3	200	107.4	70.3	194.8	3.77	18	22	9/28/2011	83.48309
19	Chevrolet	Cavalier	145.519	9.25	Passenger	13.26	2.2	115	104.1	67.9	180.9	2.676	14.3	27	8/17/2011	46.36335
20	Chevrolet	Malibu	135.126	11.225	Passenger	16.535	3.1	170	107	69.4	190.4	3.051	15	25	3/19/2012	67.31446
21	Chevrolet	Lumina	24.629	10.31	Passenger	18.89	3.1	175	107.5	72.5	200.9	3.33	16.6	25	5/24/2011	69.9914
22	Chevrolet	Monte Car	42.593	11.525	Passenger	19.39	3.4	180	110.5	72.7	197.9	3.34	17	27	12/22/2011	72.03092
23	Chevrolet	Camaro	26.402	13.025	Passenger	24.34	3.8	200	101.1	74.1	193.2	3.5	16.8	25	10/23/2011	81.11854
24	Chevrolet	Corvette	17.947	36.225	Passenger	45.705	5.7	345	104.5	73.6	179.7	3.21	19.1	22	5/12/2012	141.1412
25	Chevrolet	Prizm	32.299	9.125	Passenger	13.96	1.8	120	97.1	66.7	174.3	2.398	13.2	33	9/11/2011	48.29764
26	Chevrolet	Metro	21.855	5.16	Passenger	9.235	1	55	93.1	62.6	149.4	1.895	10.3	45	4/13/2012	23.27627
28	Chrysler	Sebring Co	7.854	12.36	Passenger	19.84	2.5	163	103.7	69.7	190.9	2.967	15.9	24	1/16/2012	65.95718
29	Chrysler	Sebring Co	32.775	14.18	Passenger	24.495	2.5	168	106	69.2	193	3.332	16	24	11/17/2011	69.52136
30	Chrysler	Concorde	31.148	13.725	Passenger	22.245	2.7	200	113	74.4	209.1	3.452	17	26	6/6/2012	80.02378
31	Chrysler	Cirrus	32.306	12.64	Passenger	16.48	2	132	108	71	186	2.911	16	27	10/6/2011	53.5662
32	Chrysler	LHS	13.462	17.325	Passenger	28.34	3.5	253	113	74.4	207.7	3.564	17	23	5/8/2012	101.3293
35	Dodge	Neon	76.034	7.75	Passenger	12.64	2	132	105	74.4	174.4	2.567	12.5	29	12/12/2011	52.0849
36	Dodge	Avenger	4.734	12.545	Passenger	19.045	2.5	163	103.7	69.1	190.2	2.879	15.9	24	7/1/2012	65.65051
37	Dodge	Stratus	71.186	10.185	Passenger	20.23	2.5	168	108	71	186	3.058	16	24	10/31/2011	67.87611
39	Dodge	Viper	0.916	58.47	Passenger	69.725	8	450	96.2	75.7	176.7	3.375	19	16	8/7/2011	188.1443

A. Data visualization



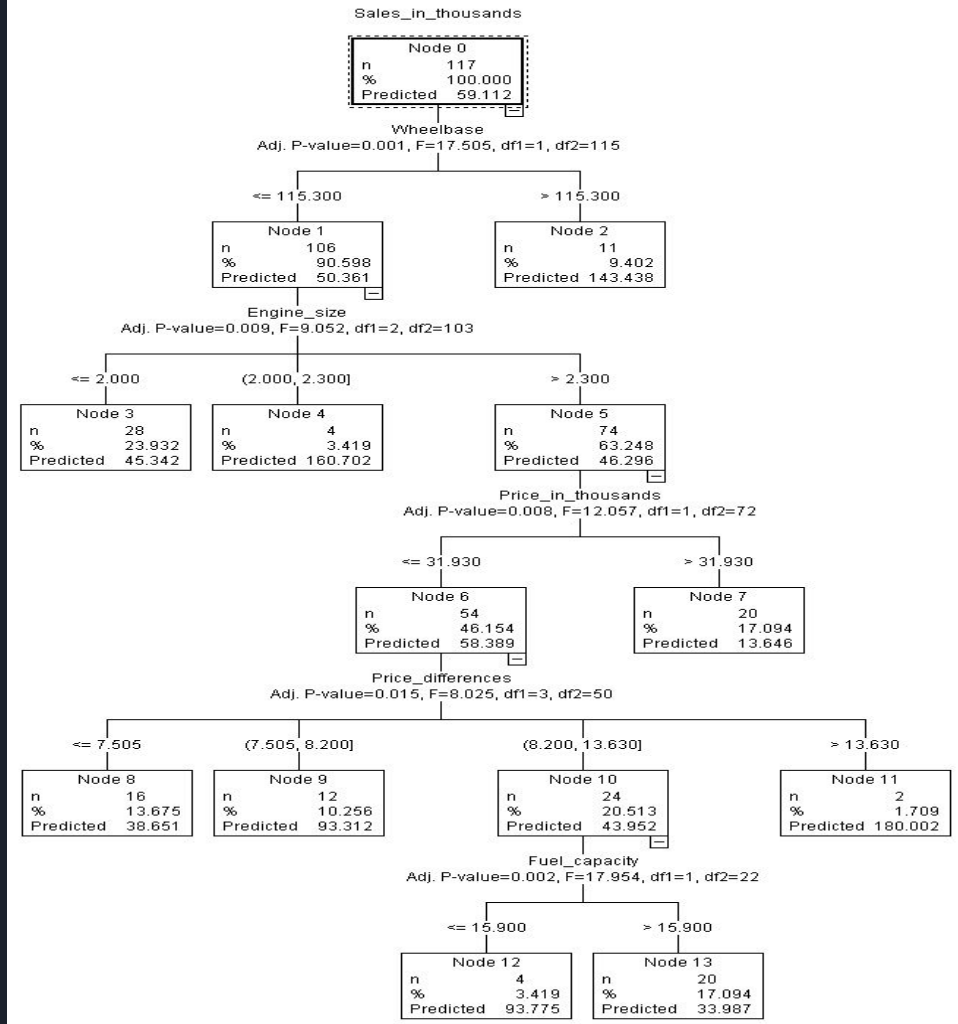




B. Data modeling

1. What are the key factors influencing the volume of sales?

Decision Tree



B. Data modeling

2. What are the key factors predicting volume of sales?(linear regression model)

Linear Regression

a. all variables

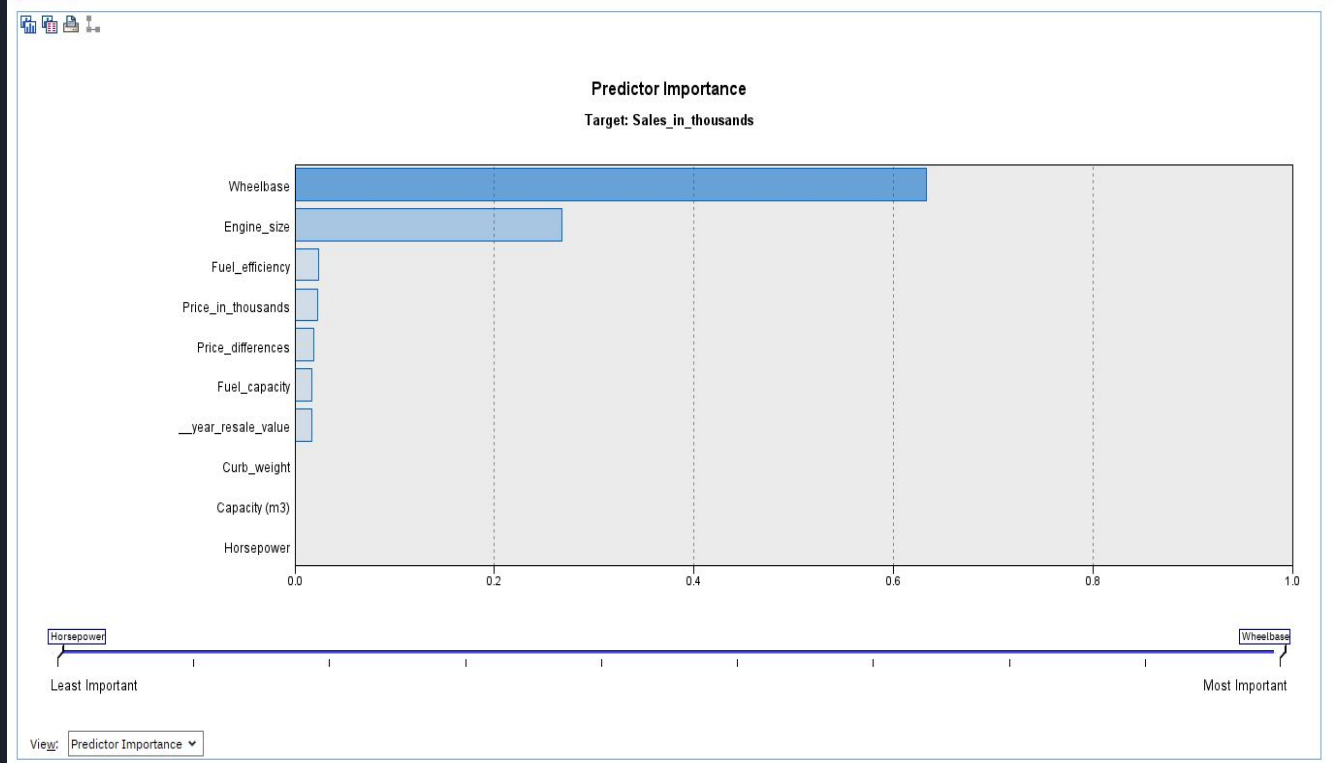
Model | Summary | Advanced | Settings | Annotations

Collapse All | Expand All

- Analysis
  - __year_resale_value * 0.0799 +
  - Price_in_thousands * -58.08 +
  - Engine_size * 31960.2 +
  - Horsepower * -510.4 +
  - Wheelbase * 5486.6 +
  - Capacity (m3) * -1.354 +
  - Curb_weight * -59504.9 +
  - Fuel_capacity * -285.6 +
  - Fuel_efficiency * -1802.3 +
  - Price_differences * 26.63 +
  - -274648.1
- Fields
- Build Settings
- Training Summary

File | Edit | Generate | View | Insert | Format | Preview

Model | Summary | Advanced | Settings | Annotations



B. Data modeling

2. What are the key factors predicting volume of sales?(linear regression model)

b. final result

Model | Summary | Advanced | Settings | Annotations

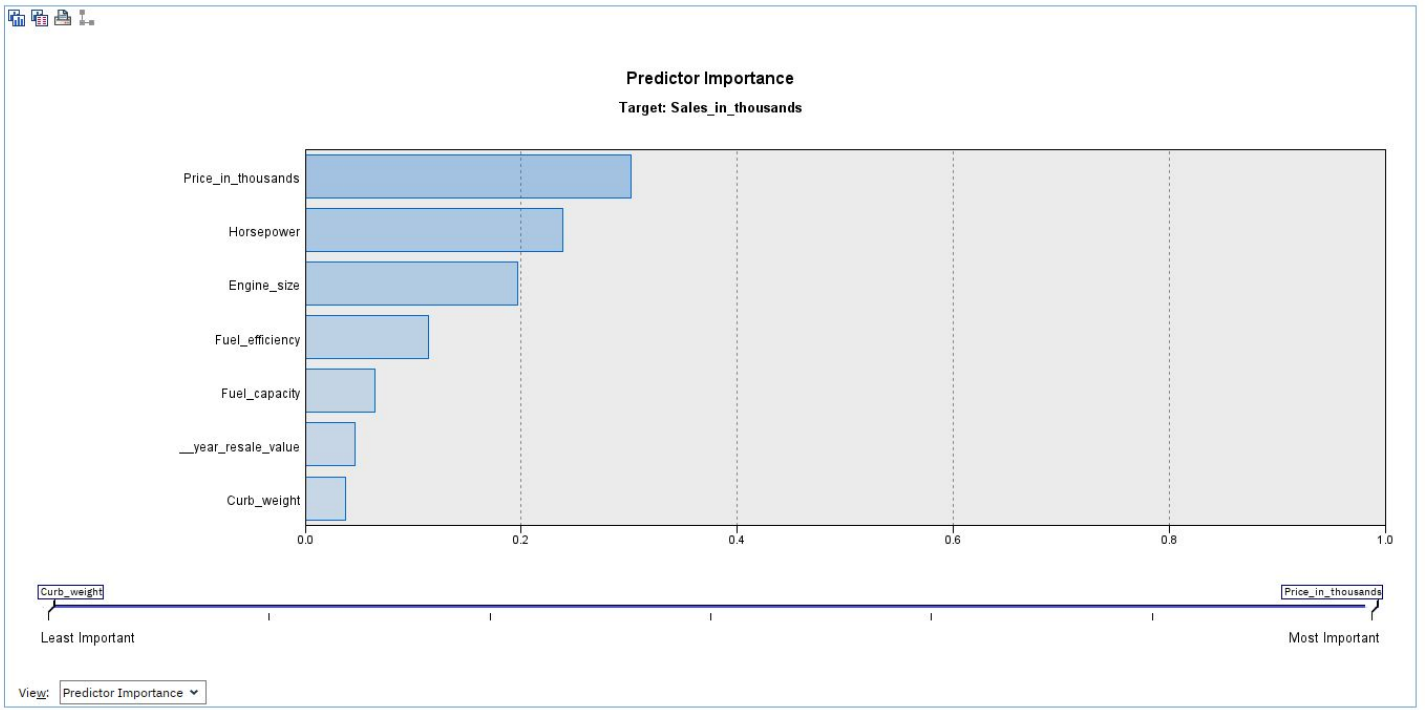
Collapse All | Expand All

- Analysis
  - __year_resale_value * 0.223 +
  - Price_in_thousands * -1570.6 +
  - Engine_size * 31656.5 +
  - Horsepower * -356.2 +
  - Curb_weight * -7442.0 +
  - Fuel_capacity * 4649.9 +
  - Fuel_efficiency * 1862.4 +
  - -42172.2
- Fields
- Build Settings
- Training Summary

Sales_in_thousands

File | Edit | Generate | View | Insert | Format | Preview

Model | Summary | Advanced | Settings | Annotations

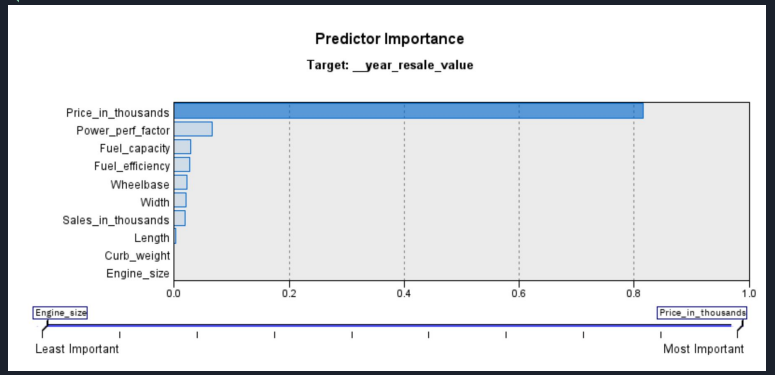


B. Data modeling (Conclusion)

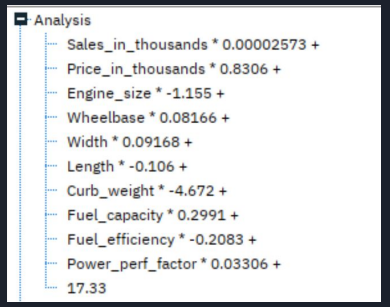
3. How to set resale prices?

a. *Created two linear regression models to decide resale price.*

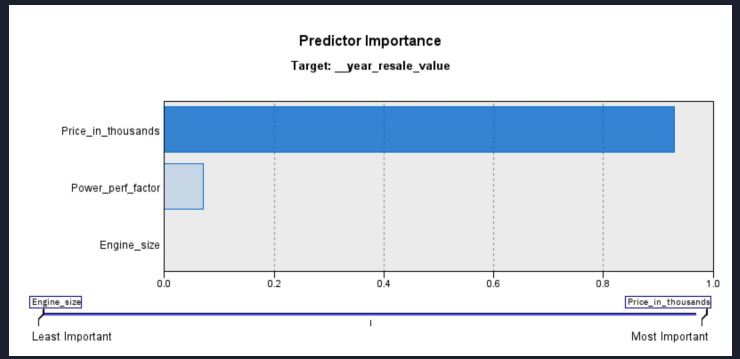
Used All variables



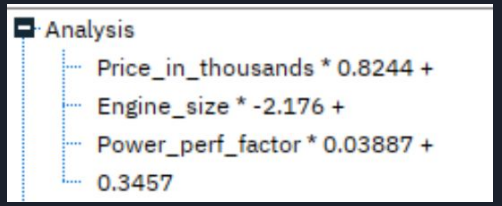
Regression Model



Used 3 high correlated variables



Regression Model



B. Data modeling (Conclusion)

3. How to set resale prices?

b. Compared between Actual Resale Value and Linear Regression Value

Used All variables

Results for output field __year_resale_value

Comparing $E-__year_resale_value with __year_resale_value

Minimum Error	-8.924
Maximum Error	7.16
Mean Error	-0.0
Mean Absolute Error	2.052
Standard Deviation	2.671
Linear Correlation	0.973
Occurrences	117

The linear Regression used all variables has a slightly better performance

Used 3 high correlated variables

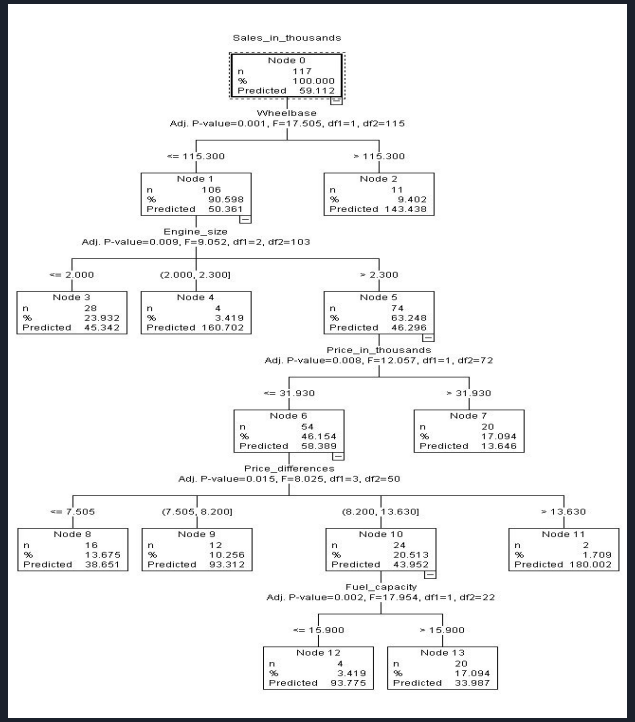
Results for output field __year_resale_value

Comparing $E-__year_resale_value with __year_resale_value

Minimum Error	-8.55
Maximum Error	10.743
Mean Error	-0.0
Mean Absolute Error	2.391
Standard Deviation	3.138
Linear Correlation	0.963
Occurrences	117

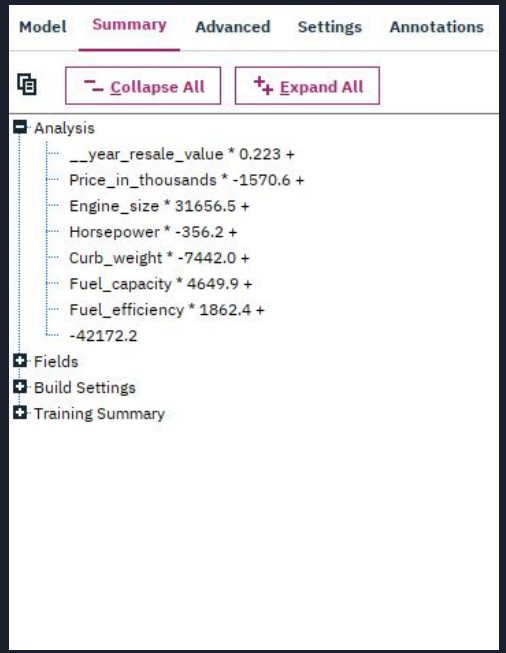
Conclusion

1. What are the key factors influencing the volume of sales?



Conclusion

2. What are the key factors predicting the volume of sales?



Target: Volume of sales

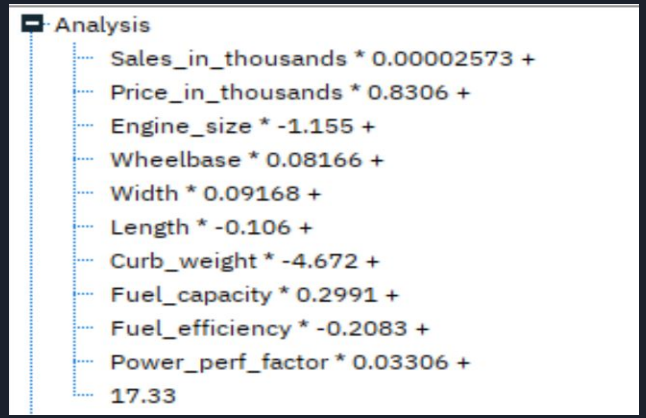
Top factors: Price, Resales value, engine size, horsepower, curb_weight, fuel_capacity, fuel_efficiency

Conclusion

3. How to set resale prices?

In our case study, we determine the key factors for future price prediction

Regression Model



Thank you !