TITLE

Manufacturer's Suggested Retail Price (MSRP) is associated with the Car weight independent of the region of where the car is manufactured

2023

Research Question

The purpose of this study was to identify the best predictors of Car Manufacturer's Suggested Retail Price (MSRP) from multiple related factors such as car weight, horsepower, number of cylinders and engine size. As a financial analyst, it is my responsibility to forecast car prices as they are introduced to the market. Having a better understanding of factors that are most likely to increase or decrease the retail selling price will allow me to identify which factors to focus on in order determine the total car finance price structure.

Research justification
If a model is implemented which can predict the retail price, the lending bank which I work for can easily determine how to finance cars before they are introduced to the market.

Sample

The sample included N=428 car models from the production batches of car manufacturing plants around the world from Jan 1,2014 to December 31, 2014. The respond dependent variable is the Manufacturer's suggested Retail Price measured in US dollars. The explanatory variables are in the table 2 below.

Table 1: Cars dataset, number of observations and variables. **Data Set Name**	WORK.CARS	**Observations**	428
Member Type	DATA	**Variables**	14

Table 2: Variables and variable types

Alphabetic List of Variables and Attributes					
#	**Variable**	**Type**			**Label**
8	Cylinders	Num			Number of Cylinders
4	DriveTrain	Char			Drive Train
7	EngineSize	Num			Engine Size (L)
9	Horsepower	Num			Car Horse power
14	Length	Num			Length (IN)
10	MPG_City	Num			MPG (City)
11	MPG_Highway	Num			MPG (Highway)
5	MSRP	Num			Manufacturer Retail Price
1	Make	Char			The Make of Car
3	Origin	Char			Origin of the vehicle
2	Type	Char			Type of the car
12	Weight	Num			Weight (LBS)
13	Wheelbase	Num			Wheelbase (IN)

Manufacturer?s Suggested Retail Price (MSRP) is associated with the Car weight independent of region of where the car is manufactured

Dependent Variable: MSRP

Table3 :The ANOVA Procedurelass Level Information		
Class	**Levels**	**Values**
Origin	3	Asia Europe USA

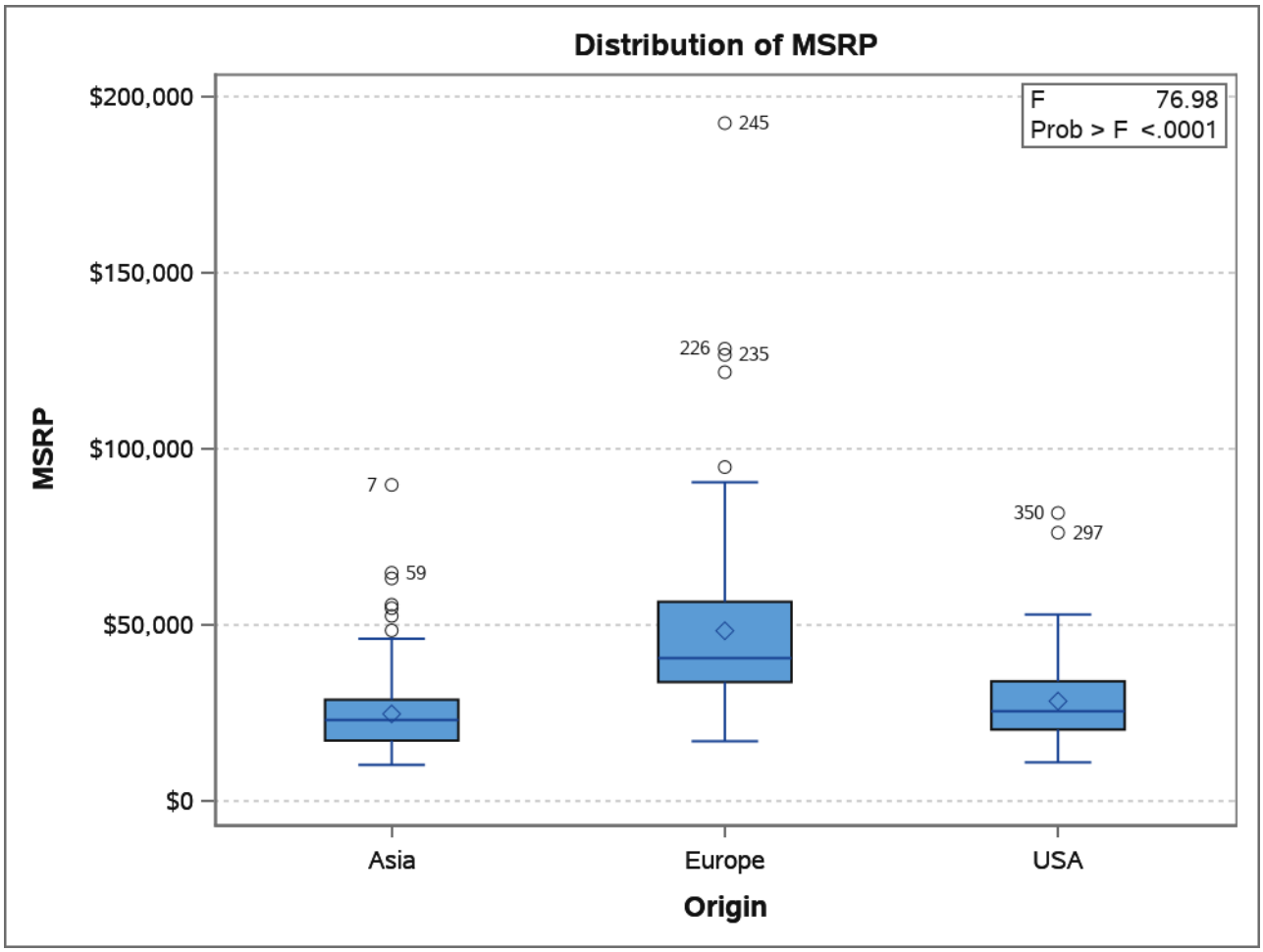
The p = 0.0001 < 0.05, we deduced that there is no difference between the means of the 3 regions USA, Europe and Asia. The region is not a contributing confronting factor the regression of MSRP and the explanatory variables. However, R-Square = 0.26 meaning only 26.5% of the MSRP is explained by the region.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	42876714994	21438357497	76.98	<.0001
Error	425	118354903709	278482126.37		
Corrected Total	427	161231618703			

R-Square	Coeff Var	Root MSE	MSRP Mean
0.265932	50.91642	16687.78	32774.86

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Origin	2	42876714994	21438357497	76.98	<.0001

The boxplot below visually shows that there is no different between the regional means of the MSRP.



Manufacturer?s Suggested Retail Price (MSRP) is associated with the Car weight independent of Drive Train of where the car is manufactured

The ANOVA Procedure

Dependent Variable: MSRP

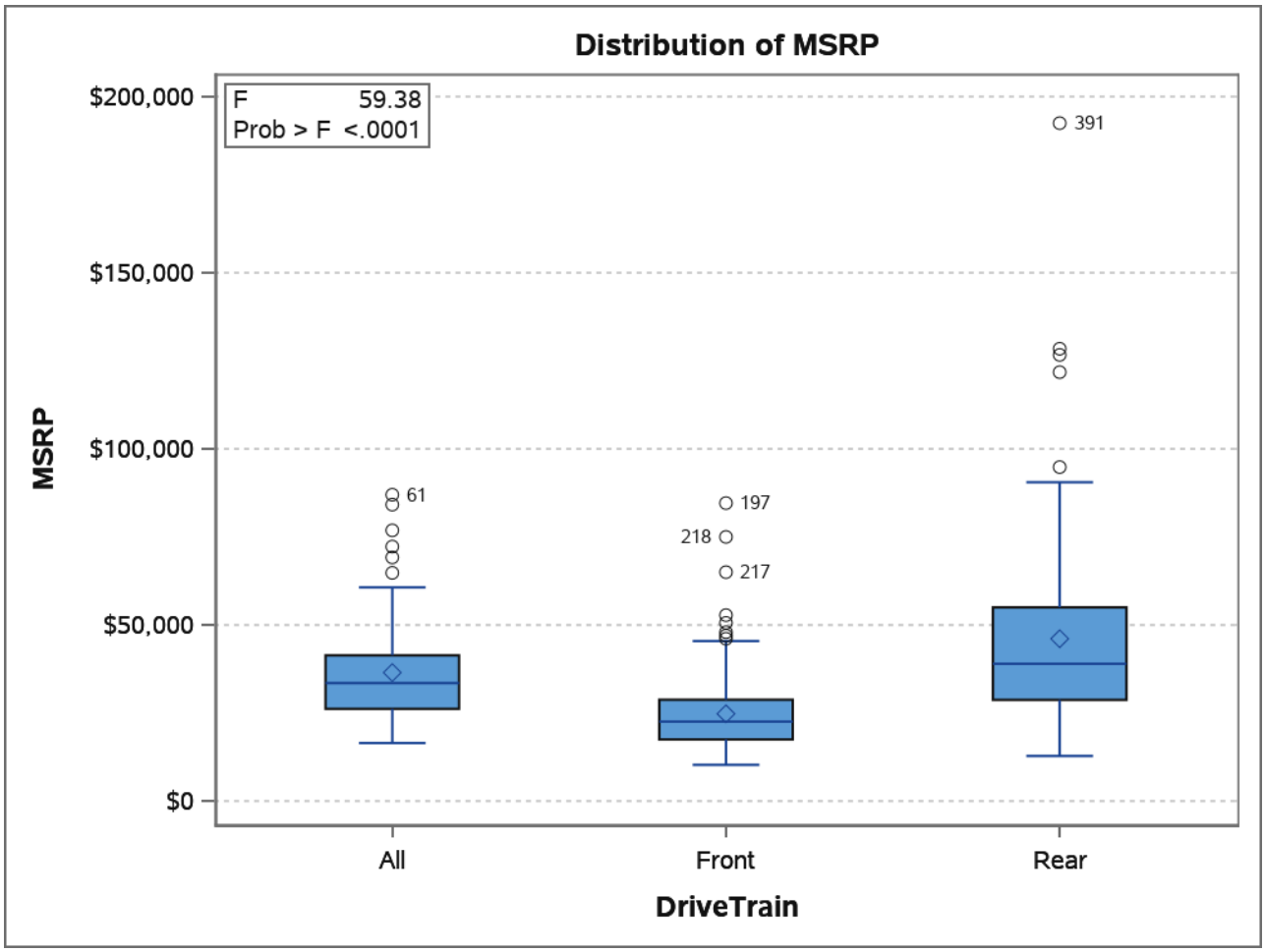
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	35214392497	17607196248	59.38	<.0001
Error	425	126017226206	296511120.49		
Corrected Total	427	161231618703			

R-Square	Coeff Var	Root MSE	MSRP Mean
0.218409	52.53875	17219.50	32774.86

The p = 0.0001 < 0.05, we deduced that there is no difference between the means of the drive trains. The region is not a contributing confronting factor the regression of MSRP and the explanatory variables. However, R-Square = 0.22 meaning only 22% of the MSRP is explained by the Drive Train.

Source	DF	Anova SS	Mean Square	F Value	Pr > F
DriveTrain	2	35214392497	17607196248	59.38	<.0001

The boxplot below visually shows that there is no different between the Drive Train means of the MSRP.



Manufacturer?s Suggested Retail Price (MSRP) is associated with the Car weight independent of region of where the car is manufactured

The MEANS Procedure

Analysis Variable : MSRP						
Origin	N Obs	N	Mean	Std Dev	Minimum	Maximum
Asia	**158**	158	24741.32	11321.07	10280.00	89765.00
Europe	**123**	123	48349.80	25318.60	16999.00	192465.00
USA	**147**	147	28377.44	11711.98	10995.00	81795.00

The means of the three origins are not significantly different.

7 Variables:	Cylinders EngineSize Horsepower MPG_City MPG_Highway MSRP Weight

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
Cylinders	426	5.80751	1.55844	2474	3.00000	12.00000	
EngineSize	428	3.19673	1.10859	1368	1.30000	8.30000	Engine Size (L)
Horsepower	428	215.88551	71.83603	92399	73.00000	500.00000	
MPG_City	428	20.06075	5.23822	8586	10.00000	60.00000	MPG (City)
MPG_Highway	428	26.84346	5.74120	11489	12.00000	66.00000	MPG (Highway)
MSRP	428	32775	19432	14027638	10280	192465	
Weight	428	3578	758.98321	1531364	1850	7190	Weight (LBS)

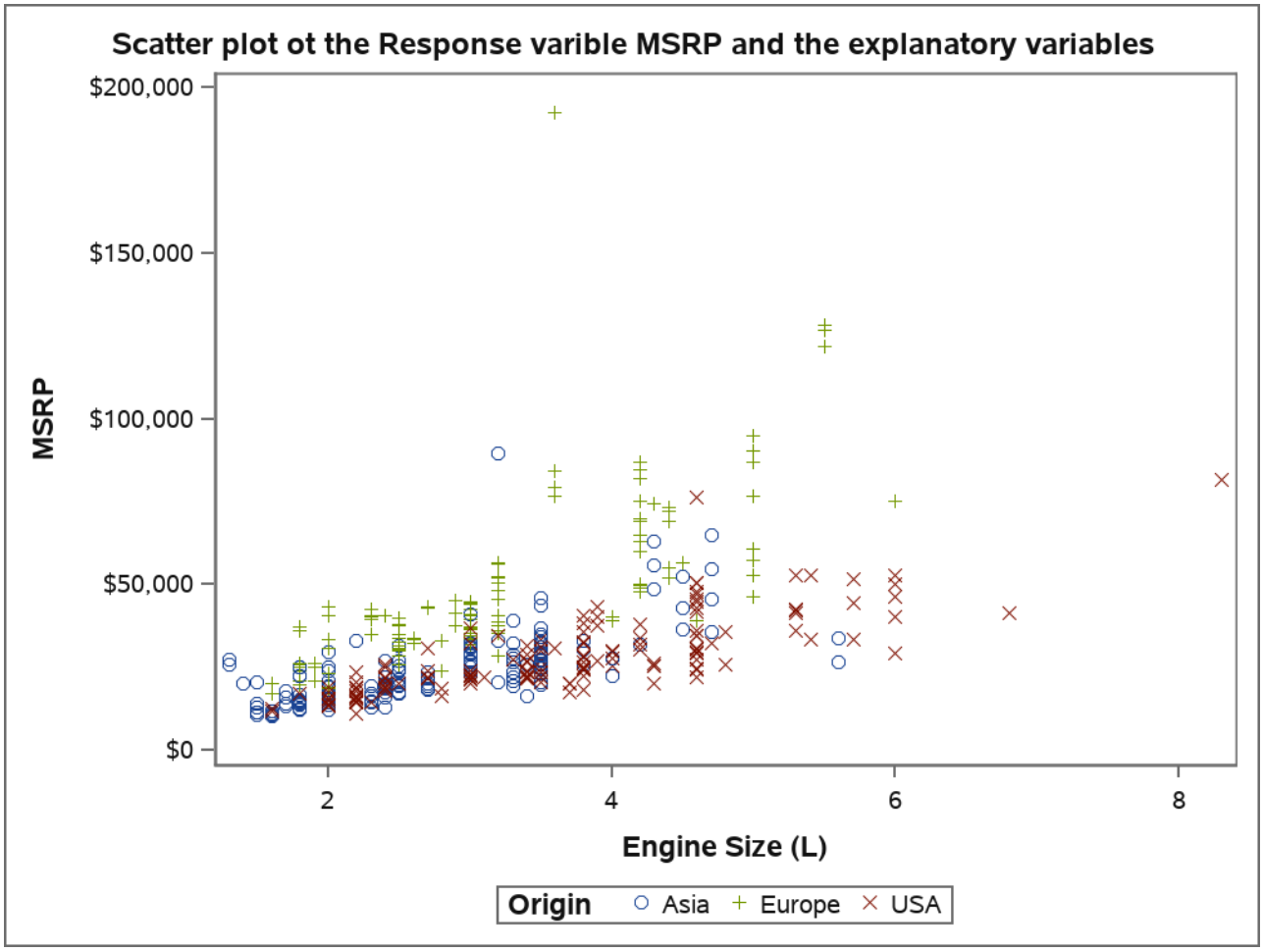
Correlation Coefficients

The p-value = 0.0001 < 0.05 for all the combinations of the variables. All the correlations are significant, however some have a negative correlation.

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	Cylinders	EngineSize	Horsepower	MPG_City	MPG_Highway	MSRP	Weight
Cylinders	1.00000 426	0.90800 <.0001 426	0.81034 <.0001 426	-0.68440 <.0001 426	-0.67610 <.0001 426	0.64974 <.0001 426	0.74221 <.0001 426
EngineSize Engine Size (L)	0.90800 <.0001 426	1.00000 428	0.78743 <.0001 428	-0.70947 <.0001 428	-0.71730 <.0001 428	0.57175 <.0001 428	0.80787 <.0001 428
Horsepower	0.81034 <.0001 426	0.78743 <.0001 428	1.00000 428	-0.67670 <.0001 428	-0.64720 <.0001 428	0.82695 <.0001 428	0.63080 <.0001 428
MPG_City MPG (City)	-0.68440 <.0001 426	-0.70947 <.0001 428	-0.67670 <.0001 428	1.00000 428	0.94102 <.0001 428	-0.47502 <.0001 428	-0.73797 <.0001 428
MPG_Highway MPG (Highway)	-0.67610 <.0001 426	-0.71730 <.0001 428	-0.64720 <.0001 428	0.94102 <.0001 428	1.00000 428	-0.43962 <.0001 428	-0.79099 <.0001 428
MSRP	0.64974 <.0001 426	0.57175 <.0001 428	0.82695 <.0001 428	-0.47502 <.0001 428	-0.43962 <.0001 428	1.00000 428	0.44843 <.0001 428
Weight Weight (LBS)	0.74221 <.0001 426	0.80787 <.0001 428	0.63080 <.0001 428	-0.73797 <.0001 428	-0.79099 <.0001 428	0.44843 <.0001 428	1.00000 428

Scatter plot

Bivariate Analyses Scatter plots for the association between the MSRP response variable and quantitative predictors.



Scatter plot ot the Response varible MSRP and the explanatory variables

The GLM Procedure

Dependent Variable: MSRP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	115778611545	19296435258	178.20	<.0001
Error	419	45371498950	108285200.36		
Corrected Total	425	161150110495			

R-Square	Coeff Var	Root MSE	MSRP Mean
0.718452	31.72126	10406.02	32804.55

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Cylinders	1	68031867386	68031867386	628.27	<.0001
EngineSize	1	256930182	256930182	2.37	0.1242
Horsepower	1	46573841887	46573841887	430.10	<.0001
MPG_City	1	730802424	730802424	6.75	0.0097
MPG_Highway	1	84515939	84515939	0.78	0.3775
Weight	1	100653727	100653727	0.93	0.3355

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Cylinders	1	1533702734	1533702734	14.16	0.0002
EngineSize	1	2852355345	2852355345	26.34	<.0001
Horsepower	1	45871680351	45871680351	423.62	<.0001
MPG_City	1	14827785	14827785	0.14	0.7115
MPG_Highway	1	150590670	150590670	1.39	0.2390
Weight	1	100653727	100653727	0.93	0.3355

Scatter plot ot the Response varible MSRP and the explanatory variables

The GLM Procedure

Dependent Variable: MSRP

The p -values < 0.05 of the intercept , cylinders, engine size, and horse power are significant. Therefore, we include them in the regression.

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-37048.26353	7815.582166	-4.74	<.0001
Cylinders	3119.66587	828.937980	3.76	0.0002
EngineSize	-6569.03305	1279.924006	-5.13	<.0001
Horsepower	263.33319	12.794338	20.58	<.0001
MPG_City	109.90212	296.997293	0.37	0.7115
MPG_Highway	339.83126	288.170048	1.18	0.2390
Weight	1.28856	1.336512	0.96	0.3355

Scatter plot ot the Response varible MSRP and the explanatory variables

Cylinders	EngineSize	Horsepower	MPG_City	MPG_Highway	MSRP	Weight
426.000	428.000	428.000	428.000	428.000	$428	428.00
3.000	1.300	73.000	10.000	12.000	$10,280	1850.00
12.000	8.300	500.000	60.000	66.000	$192,465	7190.00
5.808	3.197	215.886	20.061	26.843	$32,775	3577.95
1.558	1.109	71.836	5.238	5.741	$19,432	758.98

Conclusion

Correlation coefficients describe the strength and direction of an association between variables. A Pearson correlation is a measure of a linear association between 2 normally distributed random variables. A Spearman rank correlation describes the monotonic relationship.

The regression equation only include the intercept , cylinders, engine size, and horse power