USED AUTOMOBILES STUDY

Statistical research on the used automobiles market dynamic

Abstract

The present research on Used automobiles implements the different statistical tests learned in the INFO589_FA22 Application Stats for Business Analytics course.

By

Joanna Rendon Ospina

Table of Contents

Executive Summary

This analysis will be conducted on a random sample of 743 used cars that are listed in the used automobiles market.

Considering the variables that affect used cars market dynamics such as brand, price, mileage, age, origin, and type, the aim is to solve the following questions:
-Is there any relationship between price and origin?
-Is there any relationship between mileage and origin?
-To study the change in price, mileage, and age across different types of used cars

Hypothesis Testing between populations:
The car dealerships and used car sellers claim that the price, mileage, and age depend on the type of automobile. To evaluate the key effects in a large group of used cars that are listed in the market, we can check whether the type of car makes any difference in the selling price, mileage, and age.

Predicting the price of a used car:
Evaluate how the price of a used car is related to mileage, age, and categorical variables like origin, type, and brand. We have created different regression models to predict the price of a used car considering a set of independent numerical and dummy variables.

Predicting the probability of selecting a USA car:
The aim is to understand how the Origin code is linked to Price, Age, and Type code. A logistic regression model was built to predict the correct origin of a used car for a specific customer using both quantitative and dummy variables.

Major Findings of the study:

1. The proportion of used cars from a specific type depends on the origin. USA car proportion from a particular type differs from Asian or European car proportion.
2. The average mileage and age of a used car do not differ between USA cars and Asian or European cars.
3. The price range and type of a used car are dependent on the car's origin.
4. There is a statistically significant difference in selling price between the different types of used cars available in the market.
5. There is a statistically significant difference in mileage among the various types of used cars.
6. There is a statistically significant difference in age between the different types of used cars available in the market.
7. Mileage is a key variable to determine the selling price of a used car. Furthermore, categorical variables like origin code and brand code give a more accurate selling price due to their statistical significance.

Analysis of Model Estimates:

Linear regression model: Predicting the price of a used car.

Different regression models using quantitative and categorical variables were built to predict the price of a used car. The first proposed models only consider quantitative variables. Then, dummy variables were added representing categories to the existing models.

After comparing the results of the adjusted R squared and standard error, the model selected included mileage, and brand code variables to predict the price of a used car because it has the highest adjusted R squared and the lowest standard error.

Logistic regression model: Predicting the probability of selecting a USA car.

A model was created to help used car buyers to predict the correct origin of a used car. In this case, Origin Code is the response variable, so Origin was coded as a binary variable in two groups: Asia or Europe = 0 and USA = 1.

Furthermore, a logistic regression analysis was performed to predict the probability of success considering some independent variables that were significant. The variables found substantial in predicting the response variable are Price, Age, and type code.

INTRODUCTION AND THEORETICAL BACKGROUND

Introduction

History of the automobiles:

Even though the automobile was invented in Europe, it was also built in the United States shortly after its launch. Cars have been key in the development of the North American country facilitating chores, boosting tourism, and speeding up construction projects. According to the History Editorial report, in 1980 over ninety percent of Americans were auto-dependent (History.com Editors, 2018).

The motor-vehicles industry works with supply and demand parameters as most of the products in the market. However, its popularity and importance in people's lives have grown over time.

Definition of an automobile:

For some people, this object represents their reliable transportation method, their source of income, or a symbol of freedom and independence (United States: Used-Vehicle Supply Rises; Average Listing Price Dips from Record Levels, 2022).

Importance of used automobiles study:

Due to the reasons previously mentioned, used car dealership places, and used cars web pages came to life. The Manheim Market Report explained that new automobiles lose value over time very quickly. Older or used vehicles, on the other hand, show more stability in selling prices.

Nevertheless, because of the global economic deceleration in the past couple of years, price segments and inventory units for used cars have been very volatile, exhibiting high prices shortage of stock in some categories (United States: Used-Vehicle Supply Rises; Average Listing Price Dips from Record Levels, 2022).

Both buyers and sellers of used cars can overcome market volatility by addressing the effect of some quantitative and categorical explanatory variables.

Types of automobiles:

In the current market, there are ten types of automobiles: Sedan, Coupe, Sedan Wagon, Sports Car, Muscle Car, SUV, Minivan, Pickup Truck, Van, and Wagon.

Sedan: is a 4-door passenger car in a three-box configuration with separate compartments for an engine, passengers, and cargo (Kia Corporation, 2022).

Coupe: a passenger car with a sloping or truncated rear roofline and two doors (Corby, 2021).

Sedan Wagon: An automobile with one or more rows of folding or removable seats behind the driver (Merriam-Webster, n.d.).

Sports Car: usually 2-passenger automobile designed for quick response, easy maneuverability, and high-speed driving (Jocelyn & Biagi, 2021).

Muscle Car: Americans made two-door sports coupes with powerful engines designed for high-performance driving (Musclecarclub, 2022).

SUV: A rugged vehicle like a station wagon but built on a light-truck chassis. Sport utility vehicle (Hyundai, 2016).

Minivan: a small passenger van, somewhat larger than a station wagon, typically with side or rear windows and rear seats that can be removed for hauling small loads (Penguin Random House LLC and HarperCollins Publishers Ltd, 2019).

Pickup Truck: it is a small truck with an open cargo area (HarperCollins Publishers, 2022).

Van: a medium-sized motor vehicle with a boxy shape and high roof, used for transporting goods and/or passengers (Dictionary.com, 2022).

Wagon: usually a four-wheeled vehicle for transporting bulky commodities and drawn originally by animals (Merriam-Webster, n.d.).

At the same time, the automobile types previously mentioned can also be categorized as:

- Big car: very comfortable and spacious car where people usually transport more than two passengers such as SUVs, Minivan, Pickup Truck, Van, or Wagon.
- Small car: exhibits low cargo capacity and space, hence it is usually used to transport two passengers such as Sedan, Coupe, Sedan Wagon, Sports Car, and Muscle Car

Project Overview

An analysis of a random sample of 743 cars listed in a used car study.

The objective of the study:

The primary objective of this research is to analyze the used car market dynamics considering some stratification factors such as brand, price, mileage, age, origin, and type, with the aim of solving the following questions:

- Is there any relationship between price and origin?
- Is there any relationship between mileage and origin?
- Study the change in price, mileage, and age across different types of used cars
- Predict the price of a used automobile
- Predict the probability of choosing a USA-used automobile for a customer

Requirements:

- Use statistical techniques such as Z-test, T-Test, Chi-Square test, and ANOVA test to establish statistically significant differences between the numeric and categorical variables previously mentioned.
- Build the best-fit regression model contemplating quantitative and categorical explanatory variables to predict the price of a used automobile.
- Build a logistic regression model contemplating quantitative and categorical explanatory variables to predict the probability of choosing a USA-used automobile according to the customer's requirements.

Elements of testing hypothesis:

- Null and alternative hypothesis
- Level of significance (5%) and confidence interval (95%)
- Test statistics
- P-values

<u>Hypothesis Testing between populations</u>

Z-Test:
Z-test is a statistical test for two population proportions. It finds the difference between two unknown proportions from independent samples. Origin and Type of car are the selected variables.

T-Test:
The T-test is a statistical test for the mean difference between two populations when the standard deviation is unknown. The distribution must be normal. Mileage and Origin, Age and Origin and Price and origin are the variables chosen.

Chi-Square Test:
The chi-square is a statistical test to determine whether there is a significant association between the two categorical variables. Price range and origin code, Age range and origin code and Type of car and origin code are the variables selected.

ANOVA Test:
An ANOVA test is a statistical test that measures the variance between the means of three or more independent populations. Price versus type of car, Mileage versus type of automobile and Age versus type of car are the chosen variables.

Predicting the price of a used car

The best regression model was developed to predict the price of a used car. Furthermore, numerical, and categorical variables to be included in the model were determined based on the R squared and standard error value.
Price is the response variable and age, mileage, type code, brand code and origin code are the explanatory variables to analyze.

Predicting the probability of selecting a USA car

A logistic regression analysis was performed to predict the probability of success considering some independent variables that were significant. The variables found substantial in predicting the response variable 'most suitable origin of a used car for a specific customer' are Price, Age, and type code.

Hypothesis Testing between populations

Research methodology:

Statement of the problem:
The car dealerships and used car sellers claim that the price, mileage, and age depend on the type of automobile. To evaluate the key effects in a large group of used cars that are listed in the market, we can check whether the type of car makes any difference in the selling price, mileage, and age.

Variables under investigation:

Variables Name	Description
Car	Brand of vehicle
Model	Model of vehicle
Price	Price in thousands of dollars (K) at sale of used vehicle
Mileage	Mileage at sale of used vehicle
Age	Age in years of used vehicle
Origin	Where the car is from: 1=USA 0=Asia or Europe
Type	The data is categorized using the type of vehicle as: Coupe - Minivan - Pickup T - Sedan - SUV - Wagon

Requirements:

Use statistical techniques such as Z-test, T-Test, Chi-Square test, and ANOVA test to establish statistically significant differences between the numeric and categorical variables previously mentioned.

Research Design:

Descriptive research techniques are used in this study.
A descriptive study is conducted to analyze the data available on the car dealership systems.

Statistical Design:

The data is analyzed with the help of the Data Analysis Tool pack in Excel.

Sample size:

Total 743 samples of various used cars.

Summary of the analysis:

1. Difference between two unknown proportions

The objective of the test:

Z-Test examines whether there is a significant proportion difference in the number of sedans, SUVs, Coupe, and minivans type of cars between two independent groups:

- Car from Asia or Europe
- Car from the USA

Count of Type	Origin	
Type	Asia or Europe	USA
Coupe	22	14
Minivan	**13**	**24**
Muscle Cr		1
PickupT	6	14
Sedan	**323**	**93**
SedanWg		6
Sports Cr	15	2
SUV	**90**	**95**
Van		6
Wagon	12	7
Grand Total	**481**	**262**

1)To test whether the proportion of imported sedan cars is greater than the proportion of USA sedan cars

Hypothesis:

- Null Hypothesis: The proportion of imported sedan cars is less than or equal to the proportion of domestic sedan cars.
- Alternative Hypothesis: The proportion of imported sedan cars is greater than the proportion of domestic sedan cars.

H0:P1-P2 ≤ 0
Ha: P1-P2 >0

Level of Significance:

5 % level of significance

Decision Rule:

If the P-value is less than (or equal to) alpha (α), then the null hypothesis (H0) is rejected in favor of the alternative hypothesis. If the P-value is greater than alpha, then the null hypothesis is not rejected.

Test in two populations whether there is a significant difference between the proportions of sedan cars:
- Number of sedan cars from Asia or Europe
- Number of sedan cars from the U.S.

Analysis:

Using the z-test to compare the number of sedan cars proportions of imported and domestic cars, the P-value found is less than 0.05. Thus, the null hypothesis (H0) will be rejected.

Conclusion:

Since the null hypothesis is rejected, the conclusion is that the imported sedan car proportion is greater than the proportion of domestic sedan cars.

2) To test whether the proportion of imported SUV cars is lower than the proportion of USA SUV cars

Hypothesis:

- Null Hypothesis: The proportion of imported SUV cars is greater than or equal to the proportion of USA SUV cars.
- Alternative Hypothesis: The proportion of imported SUV cars is lower than the proportion of USA SUV cars.

H0:P1-P2 ≥ 0
Ha: P1-P2 < 0

Level of Significance:

5 % level of significance

Decision Rule:

If the P-value is less than (or equal to) alpha (α), then the null hypothesis is rejected in favor of the alternative hypothesis. If the P-value is greater than alpha (α), then the null hypothesis is not rejected.

Test in two populations whether there is a significant difference between the proportions of SUV cars:

- Number of SUV cars from Asia or Europe
- Number of SUV cars from the U.S.

Analysis:

Using the z-test to compare the proportion of imported SUV and domestic SUV cars, the P-value found is zero at a 5% significance level. Thus, the H0 will be rejected.

Conclusion:

Since the null hypothesis was rejected, the conclusion is that the proportion of imported SUV cars is lower than the proportion of USA SUV cars.

3) To test whether the proportion of imported Minivan cars is greater than the proportion of USA Minivan cars

Hypothesis:

- Null Hypothesis: the proportion of imported minivan cars is less than or equal to the proportion of USA minivan cars.
- Alternative Hypothesis: the proportion of imported minivan cars is greater than the proportion of USA minivan cars.

$H0: P1-P2 \leq 0$

$Ha: P1-P2 > 0$

Level of Significance:

5 % level of significance

Decision Rule:

If the P-value is less than (or equal to) alpha (α), then the null hypothesis is rejected in favor of the alternative hypothesis. If the P-value is greater than alpha (α), then the null hypothesis is not rejected.

Test in two populations whether there is a significant difference between the proportions of Minivan cars:

- Number of Minivan cars from Asia or Europe
- Number of Minivan cars from the U.S.

Analysis:

Using the z-test to compare the proportion of imported minivan cars and domestic minivan cars, the P-value found is greater than 0.05. Thus, the null hypothesis cannot be rejected.

Conclusion:

Because the null hypothesis was not rejected, we cannot conclude that the proportion of imported minivan cars is less than the proportion of domestic minivan cars.

Hypothesis	Claim	Z	p-value	Decision
H0:P1-P2 ≤ 0 Ha: P1-P2 > 0	The proportion of imported sedan cars is greater than the proportion of USA sedan cars	8.3051	0.0000	Reject H0
:P1-P2 ≥ 0 Ha: P1-P2 < 0	The proportion of imported SUV cars is lower than the proportion of USA SUV cars	-5.2851	0.0000	Reject H0
H0:P1-P2 ≤ 0 Ha: P1-P2 > 0	The proportion of imported Minivan cars is greater than the proportion of USA Minivan cars	-3.8662	0.9999	Do not reject H0

2. Mean difference between two populations with normal distribution

Using the t-test for two-sample assuming equal or unequal variances based on the f-test values we compared the amount of Mileage, Age, and Price, of two independent groups of used cars with similar Origins.

Hypothesis:

Populations / Variable: Mileage of US car, Mileage of Asian or Europe car

H0: μ mileage of US cars = μ mileage of Asian or Europe cars

HA: μ mileage of US cars ≠ μ mileage of Asian or Europe cars

Populations / Variable: Age of US car, Age of Asian or Europe car

H0: μ age of US cars sold = μ age of Asian or Europe cars sold
HA: μ age of US cars ≠ μ age of Asian or Europe cars sold

Populations / Variable: Price of US car, Price of Asian or Europe car

H0: μ price of US cars = μ price of Asian or Europe cars
HA: μ price of US cars ≠ μ price of Asian or Europe cars

Level of Significance:

5 % level of significance

Decision Rule:

If the P-value is less than (or equal to) alpha (α), then the null hypothesis is rejected in favor of the alternative hypothesis. If the P-value is greater than alpha (α), then the null hypothesis is not rejected.

Findings of the test:

S.no	Variable	P Value	Conclusion
1	Mileage	0.1174	The null hypothesis is not rejected.
2	Price	0	The null hypothesis is rejected.
3	Age	0.3998	The null hypothesis is not rejected

Analysis:

For Mileage, the hypothesis that the total miles traveled differs between cars from USA and Asia or Europe was tested. The p-value is greater than 0.05 which means at the 5% significance level, the null hypothesis cannot be rejected and cannot conclude that mileage for the USA cars and Asian or European cars differ.

For Price, the hypothesis that the average selling price differs between USA cars and Asian or European cars was tested. The p-value is less than 0.05, which means, at the 5% significance level, the null hypothesis was rejected hence the selling price of USA cars and Asian or European cars is not the same.

For Age, the hypothesis that the average age of a used car differs between USA cars and Asian or European cars was tested. The p-value is greater than 0.05 which means, at the 5% significance level, the null hypothesis cannot be rejected and cannot conclude that the average age for the USA and Asian or European cars is not the same.

Conclusion:

There is a statistically significant difference in price for used cars from the USA compared to those from Asia or Europe. Furthermore, there is no statistically significant difference in mileage and age between USA cars and

Asian or European cars.

3. Dependency between two categorical variables

Chi-Square is used to determine whether there is a significant association between the two categorical variables from a single population.

Hypothesis:

1) The Price range depends on the Origin of the car

Ho: Price range and Origin of the car are independent
Ha: Price range and Origin of the car are dependent

2) The Age Range of the cars and their Origin are dependent

Ho: The Age range of the cars and their origin are independent
Ha : The Age range of the cars and their origin are dependent

3) The type of car and Origin are dependent

Ho: The type of car and Origin are independent
Ha: The type of car and Origin are dependent

Level of significance:

5% level of significance

Decision Rule:

If the P-value is less than (or equal to) alpha (α), then the null hypothesis is rejected in favor of the alternative hypothesis. If the P-value is greater than alpha (α), then the null hypothesis is not rejected.

Findings of the test:

1) The Price range depends on the Origin of the car

Observed Frequencies:

Origin	High Price	Low Price	Medium Price	Grand Total
Asia or Europe	19	336	126	481
USA		238	24	262
Grand Total	19	574	150	743

Expected Frequencies:

Origin	High Price	Low Price	Medium Price	Grand Total
Asia or Europe	12.3001	371.5935	97.1063	481
USA	6.6999	202.4065	52.8937	262
Grand Total	19	574	150	743

P-value = 6.51398E-09 = 0

Analysis:

The Price variable was categorized into three groups:

- High Price (≥46)
- Medium Price (>23 < 46)
- Low Price (≤ 23)

Then, the three groups were split into Origin Asia or Europe and Origin USA. After conducting the calculations between the observed and expected variables, the p-value found was less than 0.05. Hence, at a 5% significance level, the null hypothesis that says price range and Origin are independent was rejected.

Conclusion:

The observed and expected values for the Price range are dependent on the origin of the car.

2) The Age Range of the cars and Origin are dependent

Observed Frequencies

Origin	Old Cars	Recent Cars	Grand Total
Asia or Europe	60	421	481
USA	32	230	262
Grand Total	92	651	743

Expected Frequencies

Origin	Old Cars	Recent Cars	Grand Total
Asia or Europe	59.5585	421.4415	481
USA	32.4415	229.5585	262
Grand Total	92	651	743

P value = 0.918033822

Analysis:

The Age attribute was classified into two groups:

- Old Cars (> 5 years)
- Recent Cars (≤ 5 years).

Then, the two groups were organized based on Asia or Europe origin and USA origin. After conducting the calculations between the observed and expected variables, the p-value found was greater than 0.05. Therefore, at a 5% significance level, the null hypothesis that says the Origin and Age Range of the cars are independent was not rejected.

Conclusion:

I cannot conclude that the Age Range of Cars and the Origin of the cars are dependent variables.

3) The type of car and Origin are dependent

Observed Frequencies

Origin	Coupe	Minivan	Muscle Cr	PickupT	Sedan	SedanWg	Sports Cr	SUV	Van	Wagon	Grand Total
Asia or Europe	22	13		6	323		15	90		12	481
USA	14	24	1	14	93	6	2	95	6	7	262
Grand Total	36	37	1	20	416	6	17	185	6	19	743

Expected Frequencies

Origin	Coupe	Minivan	Muscle Cr	PickupT	Sedan	SedanWg	Sports Cr	SUV	Van	Wagon	Grand Total
Asia or Europe	23.3055	23.9529	0.6474	12.9475	269.3082	3.8843	11.0054	119.7645	3.8843	12.3001	481
USA	12.6945	13.0471	0.3526	7.0525	146.6918	2.1157	5.9946	65.2355	2.1157	6.6999	262
Grand Total	36	37	1	20	416	6	17	185	6	19	743

P value = 1.05894E-16 = 0

Analysis:

The Types of cars were grouped into ten groups: Coupe, Minivan, Muscle Cars, Pickup trucks, Sedan, Sedan Wagon, Sports Car, SUVs, Van, and Wagon.

Then, the groups' results were interpreted according to their Origin (Asia or Europe and USA). After conducting the calculations between the observed and expected variables, the p-value found was less than 0.05. Hence, at a 5% significance level, the null hypothesis that says the Origin and Type of car are independent variables was rejected.

Conclusion:

The Types of Cars and the Origin of the cars are dependent on each other.

4. Variance between the mean of six independent populations

One-Way ANOVA test is used to analyze how six different independent categorical types of used cars, can influence quantitative variables such as price, mileage, and age.

Level of significance:

5% significance level

Hypothesis:

1) Price Versus Type of Used car

H0: All types of used cars have the same price.

Ha: At least one type of used car has a different price.

2) Mileage Versus Type of Used car

H0: All types of used cars have the same mileage.
Ha: At least one type of used car has a different mileage

3) Age Versus Type of Used car

H0: All types of used cars have the same age
Ha: At least one type of used car has a different age

Decision rule:

If the p-value is greater than alpha, do not reject the null hypothesis. On the other hand, if the p-value is less than or equal to alpha, reject the null hypothesis.

Findings of the test:

1) Price vs. Type of used car

Groups	*Count*	*Sum*	*Average*	*Variance*		
Coupe	36	858.5	23.8472	229.4991		
Minivan	36	433.6	12.0444	17.4288		
PickupT	20	336	16.8000	44.1011		
Sedan	36	721.5	20.0417	83.9196		
SUV	36	655.1	18.1972	46.3666		
Wagon	19	378.9	19.9421	314.1526		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2,715.7681	5	543.1536	4.8813	**0.0003**	2.2652
Within Groups	19,695.1621	177	111.2721			
Total	22,410.9303	182				

2) Mileage vs. Type of used car

Groups	*Count*	*Sum*	*Average*	*Variance*		
Coupe	36	1242.2	34.5056	418.4028		
Minivan	36	1735.1	48.1972	675.5706		
PickupT	20	912.9	45.6450	1151.3479		
Sedan	36	1437.5	39.9306	460.2148		
SUV	36	1109.7	30.8250	231.5162		
Wagon	19	697.5	36.7105	537.5032		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7,192.17	5	1438.4332	2.7071	**0.0220**	2.2652
Within Groups	94,050.32	177	531.3577			
Total	101,242.49	182				

3) Age vs. Type of used car

Groups	*Count*	*Sum*	*Average*	*Variance*		
Coupe	36	163	4.5278	4.9421		
Minivan	36	140	3.8889	5.5302		
PickupT	20	70	3.5000	3.2105		
Sedan	36	137	3.8056	3.9325		
SUV	36	106	2.9444	1.8825		
Wagon	19	56	2.9474	3.9415		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	58.3249	5	11.6650	2.9412	**0.0141**	2.2652
Within Groups	702.0029	177	3.9661			
Total	760.3279	182				

Analysis and inference:

A statistical inference test was used to assess the hypothesis that the quantitative means of price, mileage and age are equal across the six different independent types of used cars. The One-Way ANOVA test result allows to conclude that there is a statistically significant difference between price and type of used car, mileage, and type of used car, and between age and type of used car because their p-values are less than 0.05 (0.0003, 0.0220, and 0.0141 respectively).

To determine which type of used car price means were different, Fisher's LSD method and Tukey's HSD method were conducted and found that Coupe and Minivan means are different same as Minivan and Sedan means.

To establish which type of used car mileage means were different, Fisher's LSD method and Tukey's HSD method were performed and found that Minivan and SUV means are different.

To find out which type of used car age means were different, Fisher's LSD method and Tukey's HSD method were conducted and discovered that Coupe and SUV means are different.

Predicting the price of a used automobile

Research methodology:

Statement of the problem:

The goal was to analyze how the price of a used car is related to mileage, age, and categorical variables like origin, type, and brand. Different regression models were built to predict the price of a used car considering a set of independent numerical and dummy variables.

Variables under investigation:

Variables Name	Description
Car	Brand of vehicle
Model	Model of vehicle
Price	Price in thousands of dollars (K) at sale of used vehicle
Mileage	Mileage at sale of used vehicle
Age	Age in years of used vehicle
Origin	Where the car is from
Origin Code	Where the car is from: 1=USA 0=Asia or Europe
Type	The data is categorized using the type of vehicle as: Coupe - Minivan - Pickup T - Sedan - SUV - Wagon
Type Code	The data is categorized using the Type of car as -Small cars (0)= Sedan, Coupe, Sedan Wagon, Sports Car and Muscle Car -Big Cars (1) = SUV, Minivan, PickupT, Van, Wagon
Brand Code	The data is categorized using the Brand of a car as: Luxury car brands (0) = BMW, Buick, Cadillac, Chrysler, Hummer, Jaguar, Land Rover, Lexus, Lincoln, MercBenz, MiniCoo, Porsche, Saab Commercial car brands (1) = Acura, Chevrolet, Dodge, Ford, GMC, Honda, Hyundai, Infiniti, Isuzu, Jeep, Kia, Mazda, Mercury, Mitsubishi, Nissan, Oldsmble, Pontiac, Saturn, Scion, Smart, Subaru, Suzuki, Toyota and Volkswagen

Objectives:

1) Build the best-fit regression model with meaningful quantitative independent variables. Use the correlation matrix, the R squared and the standard error value to determine which variables must be included in the model.

2) Build the best fit multiple regression model with a set of categorical and numerical independent variables. Use the R squared and the standard error value to determine which variables must be included in the model.

Statistical Design:

The data is analyzed with the help of the Data Analysis Tool pack in Excel.

Sample size:

Total 743 samples of various used cars.

Level of significance:

5% significance level

Correlation between variables:

A correlation matrix was used to verify the correlation between the response variable Price and the quantitative explanatory variables Mileage and Age. Furthermore, the matrix also included the correlation between the two explanatory variables, and it was found that Mileage and Age are highly correlated with a coefficient of 0.7321.

	Price	*Mileage*	*Age*
Price	1		
Mileage	-0.4662	1	
Age	-0.3671	**0.7321**	1

Due to this correlation between Mileage and Age, a Multicollinearity situation was identified. For that reason, the decision was not to include both variables because their presence together might result in inaccurate slope coefficients.

Mileage is the best explanatory variable to use in the regression to predict the price of a used car because its slope coefficient of -0.4662 is higher than the estimate for Age.

Summary of the analysis and findings:

1) Simple regression model based on quantitative independent variables

In this section, two different numerical variables were selected to build two models. The quantitative variables used are Mileage and Age to predict the selling price.

After analyzing the numerical independent variables, the best-fit simple regression model to predict the selling price of a used car is: $\hat{y} = \beta 0 + \beta 1 x 1$

Used Car Price= 26.1863 - 0.2115 (Mileage)

Summary Output - Model 1 (Mileage):

Regression Statistics						
Multiple R	0.4662					
R Square	0.2173					
Adjusted R Square	**0.2163**					
Standard Error	**8.6277**					
Observations	743					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	15,317.80	15,317.80	205.78	0.00	
Residual	741	55,158.17	74.44			
Total	742	70,475.96				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	26.1863	0.6226	42.0598	0.0000	24.9641	27.4086
Mileage	-0.2115	0.0147	-14.3451	0.0000	-0.2404	-0.1825

Coefficient of Slope	**Relationship**	**Interpretation:**
Mileage (0.2115)	Negative	The selling price of a used car is predicted to decrease by $0.21 as the Mileage increases by 1 mile

Comparison of the two models:

The following variables were under analysis:

Mileage	Age

	Model 1	Model 2
Intercept	26.1863	25.0305
	p-value = 0	pvalue = 0
Mileage	-0.2115	N/A
	p-value = 0	
Age	N/A	-1.943
		p-value = 0
R Square	0.2173	0.1348
Adjusted R Square	**0.2163**	**0.1336**
Standard Error	**8.6277**	**9.0715**

Based on the results of the above table, Model 1 has the highest Adjusted R Square of 0.2163 and the lowest Standard Error of 8.6277. Hence, Model 1 is the best fit to predict the selling price of a used car.

This confirms what I thought after analyzing the correlation matrix and guarantees the best explanatory variable to predict the price of a used car is Mileage.

Predict Y (selling price):

A numerical value is assumed and by putting this value in the final regression model, now it is possible to predict the value of y (Selling Price).

Let's suppose the quantitative value is the following:

- Mileage = 45

So, we can predict the selling price by putting the value in the model:

Price= 26.1863 - 0.2115 (Mileage)
Price= 26.1863 - 0.2115 (45) = 16.6699

The selling price of a used car with 45 miles is $16.7K

Confidence and predicted interval for a used car:

Using the Rstudio program, the confidence interval and predicted interval were calculated:

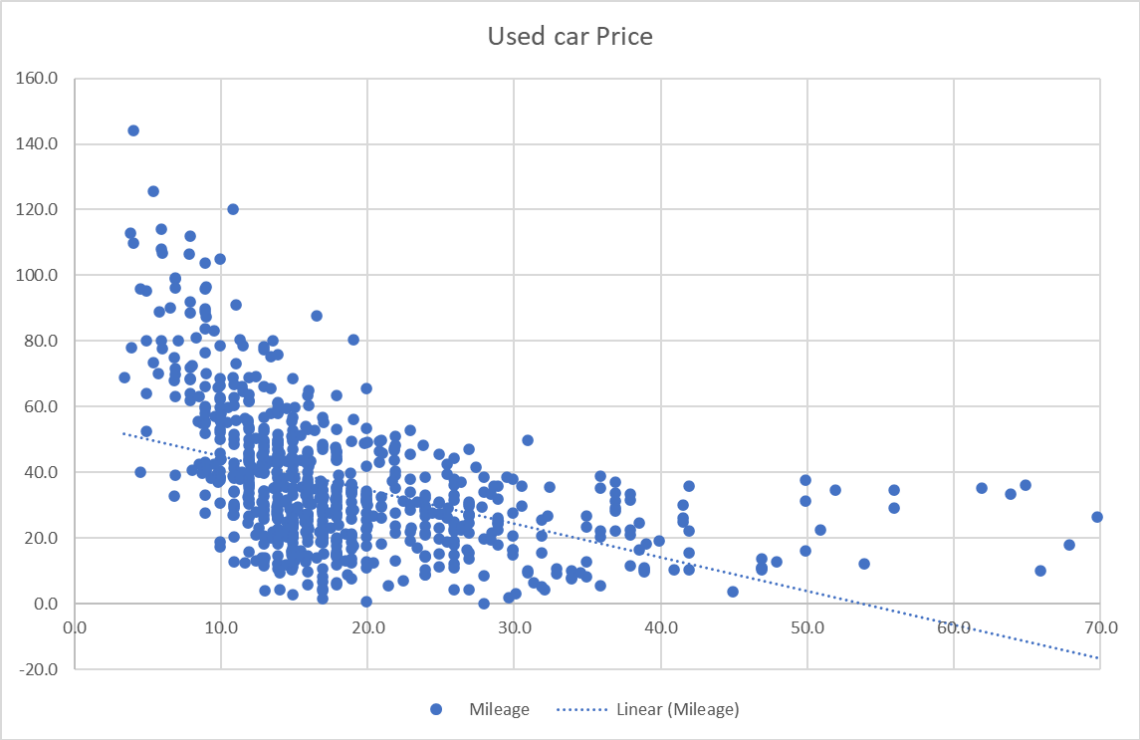
- 95% confidence interval estimate

Confidence Interval	
Coefficient	16.6699
Lower limit	16.0001
Upper limit	17.3396

- 95% predicted interval estimate

Predicted Interval	
Coefficient	16.6699
Lower limit	-0.2811
Upper limit	33.6208

Scatterplot:



The scatterplot graph shows a negative relationship between price and mileage.

2) Multiple regression model based on quantitative and categorical independent variables

For this section, different variables were selected, and three models were proposed using numerical variables like mileage and age to predict the selling price, and at the same time dummy variables such as origin code, type code, and brand code were added.

1. Does the origin of a car affect the selling price?

The aim of this test is to predict a used car's selling price, based on its mileage and origin code. Used car market dynamics consider mileage as a key determinant of selling price and in this case also the origin of the automobile. The expectation was that imported cars (Asia or Europe) show higher prices than the selling prices of USA cars.

Asia or Europe =	0
USA =	1

Estimated model:

Price = 27.6264 – 0.2055 (Mileage) – 4.7047 (Origin code)

Summary of models estimated:

	Numeric Model	**Numeric + categorical model (Origin code)**
Intercept	26.1863	27.6264
Mileage	-0.2115	-0.2055
Origin Code	N/A	-4.7047
R Square	0.2173	0.2704
Adjusted R Square	0.2163	**0.2685**
Standard Error	8.6277	**8.3356**

Analysis of model estimates:

The numeric model only uses one quantitative independent variable. The second model uses a numerical variable and a dummy variable which is the origin code. The coefficients, adjusted R square and standard error value were compared, and the numeric plus categorical model was selected as the best fit because it has a higher R square and a lower standard error value.

Summary of the analysis and findings:

Two used automobiles with the same criteria with the exception that the first one is from Asia and the second one is from the USA were assessed. The two cars had 35 mileages (in thousands). The selling price for the Asian car was $20.4K and for the USA car was $15.7K.

There is a significant difference in price based on the origin of the used car. Hence, the hypothesis that Asian cars selling price is higher than American cars selling price was confirmed.

Predict Y (selling price):

Let's suppose the numerical and categorical values are the following:

- Mileage = 35
- Asia or Europe = 0
- USA = 1

Predict the selling price of an Asian or European car by putting the above values in the model:

Price = 27.6264 – 0.2055 (Mileage) – 4.7047 (Origin code)
Price = 27.6264 – 0.2055 (35) – 4.7047 (0) = 20.4354

Confidence and predicted interval for an Asian or European car:

Using the Rstudio program the confidence interval and predicted interval were calculated:

- 95% confidence interval estimate (0)

Confidence Interval	
Coefficient	20.4354
Lower limit	19.6891
Upper limit	21.1816

- 95% predicted interval estimate (0)

Predicted Interval	
Coefficient	20.4354
Lower limit	4.0542
Upper limit	36.8166

Predict the selling price of a USA car by putting the above values in the model:

Price = 27.6264 – 0.2055 (Mileage) – 4.7047 (Origin code)
Price = 27.6264 – 0.2055 (35) – 4.7047 (1) = 15.7307

Confidence and predicted interval for a USA car:

Using the Rstudio program the confidence interval and predicted interval were calculated:

- 95% confidence interval estimate (1)

Confidence Interval	
Coefficient	15.7307
Lower limit	14.7162
Upper limit	16.7453

- 95% predicted interval estimate (1)

Predicted Interval	
Coefficient	15.7307
Lower limit	-0.6649
Upper limit	32.1263

Summary Output:

Regression Statistics						
Multiple R	0.5200					
R Square	0.2704					
Adjusted R Square	0.2685					
Standard Error	8.3356					
Observations	743					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	2	19,059.57	9,529.79	137.16	0.00	
Residual	740	51,416.39	69.48			
Total	742	70,475.96				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	27.6264	0.6327	43.6632	0.0000	26.3843	28.8685
Mileage	-0.2055	0.0143	-14.4014	0.0000	-0.2335	-0.1775
OriginCODE	-4.7047	0.6411	-7.3384	0.0000	-5.9633	-3.4461

2. Does the type of car affect the selling price?

The purpose of this test is to predict a used car's selling price, based on its mileage, age, and type code. Used car market dynamics consider mileage as a key determinant

of selling price and in this case also the type of automobile. The expectation was that the type of car affects the selling cost.

Small cars =	0
Big cars =	1

Estimated model:

Price = 26.2539 – 0.2111 (Mileage) – 0.2258 (Type code)

Summary of models estimated:

	Numeric Model	Numeric + categorical model (Type code)
Intercept	26.1863	26.2539
Mileage	-0.2115	-0.2111
Type Code	N/A	-0.2258
R Square	0.2173	0.2175
Adjusted R Square	**0.2163**	0.2154
Standard Error	**8.6277**	8.6329

Analysis of model estimates:

The numeric model only uses one quantitative independent variable. The second model uses a numerical variable and a dummy variable which is the Type Code. The coefficients, adjusted R square, and standard error value were compared, and the numeric model was selected as the best fit because it has a higher adjusted R square and a lower standard error value.

Summary of the analysis and findings:

Two used automobiles with the same criteria with the exception that the first one is a small car and the second one is a big car were assessed. The two cars had 35 miles (in thousands). The selling price for the small car was $18.9k and for the big car was $18.6K.

There is no significant difference in selling price between a small used car and a big used car.

Predict Y (selling price):

Let's suppose the numerical and categorical values are the following:

- Mileage = 35

- Small car = 0
- Big car = 1

Predict the selling price of a small car by putting the above values in the model:

Price = 26.2539 – 0.2111 (Mileage) – 0.2258 (Type code)
Price = 26.2539 – 0.2111 (35) – 0.2258 (0) = 18.8653

Confidence and predicted interval for a small car:

Using the Rstudio program the confidence interval and predicted interval were calculated:

- 95% confidence interval estimate (0)

Confidence Interval	
Coefficient	18.8653
Lower limit	18.0884
Upper limit	19.6421

- 95% predicted interval estimate (0)

Predicted Interval	
Coefficient	18.8653
Lower limit	1.8996
Upper limit	35.8309

Predict the selling price of a big car by putting the above values in the model:

Price = 26.2539 – 0.2111 (Mileage) – 0.2258 (Type code)
Price = 26.2539 – 0.2111 (35) – 0.2258 (1) = 18.6395

Confidence and predicted interval for a big car:

Using the Rstudio program the confidence interval and predicted interval were calculated:

- 95% confidence interval estimate (1)

Confidence Interval	
Coefficient	18.6395
Lower limit	17.5974
Upper limit	19.6816

- 95% predicted interval estimate (1)

Predicted Interval	
Coefficient	18.6395
Lower limit	1.6597
Upper limit	35.6193

Summary Output:

Regression Statistics						
Multiple R	0.4663					
R Square	0.2175					
Adjusted R Square	0.2154					
Standard Error	8.6329					
Observations	743					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	2	15,326.47	7,663.23	102.83	0.00	
Residual	740	55,149.50	74.53			
Total	742	70,475.96				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	26.2539	0.6537	40.1625	0.0000	24.9706	27.5372
Mileage	-0.2111	0.0148	-14.2720	0.0000	-0.2401	-0.1821
Type Code	-0.2258	0.6619	-0.3411	0.7331	-1.5252	1.0736

3. Does the brand of the car affect the selling price?

The objective of this test is to predict a used car's selling price, based on its mileage and brand code. Used car market dynamics consider mileage as a key determinant of selling price and in this case also the brand of the automobile. The belief was that the brand of a car affects the selling cost.

Luxury brand cars =	0
Commercial brand cars =	1

Estimated model:

Price = 33.0953 – 0.1870 (Mileage) – 10.4417 (Brand code)

Summary of models estimated:

	Numeric Model	Numeric + categorical model (Brand code)
Intercept	26.1863	33.0953
Mileage	-0.2115	-0.1870
Brand Code	N/A	-10.4417
R Square	0.2173	0.4317
Adjusted R Square	0.2163	**0.4301**
Standard Error	8.6277	**7.3570**

Analysis of model estimates:

The numeric model only uses one quantitative independent variable. The second model uses a numerical variable and a dummy variable which is the brand code. The coefficients, adjusted R square and standard error value were compared, and the numeric plus categorical model was selected as the best fit because it had a higher R square and a lower standard error value.

Summary of the analysis and findings:

Two used automobiles with the same criteria with the exception that the first one is from a luxury brand and the second one is a commercial brand car were assessed. The two cars had 45 miles (in thousands). The selling price for the luxury brand car was $24.7K and for the commercial brand car was $14.2K.

There is a significant difference in the selling price of a used car based on which brand the car is. Luxury car brands are more expensive than commercial brand cars.

Predict Y (selling price):

Let’s suppose the numerical and categorical values are the following:

- Mileage = 45
- Luxury brand = 0
- Commercial brand = 1

Predict the selling price of a luxury brand car by putting the above values in the model:

Price = 33.0953 – 0.1870 (Mileage) – 10.4417 (Brand code)

Price = 33.0953 – 0.1870 (45) – 10.4417 (0) = 24.6810

Confidence and predicted interval for a luxury brand car:

Using the Rstudio program the confidence interval and predicted interval were calculated:

- 95% confidence interval estimate (0)

Confidence Interval	
Coefficient	24.6810
Lower limit	23.5799
Upper limit	25.7821

- 95% predicted interval estimate (0)

Predicted Interval	
Coefficient	24.6810
Lower limit	10.1960
Upper limit	39.1659

Predict the selling price of a commercial brand car by putting the above values in the model:

Price = 33.0953 – 0.1870 (Mileage) – 10.4417 (Brand code)
Price = 33.0953 – 0.1870 (45) – 10.4417 (1) = 14.2392

Confidence and predicted interval for a commercial brand car:

Using the Rstudio program the confidence interval and predicted interval were calculated:

- 95% confidence interval estimate (1)

Confidence Interval	
Coefficient	14.23924
Lower limit	13.6007
Upper limit	14.8778

- 95% predicted interval estimate (1)

Predicted Interval	
Coefficient	14.23924
Lower limit	-0.2179
Upper limit	28.6964

Summary Output:

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.6570					
R Square	0.4317					
Adjusted R Square	0.4301					
Standard Error	7.3570					
Observations	743					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	2	30,423.39	15,211.70	281.05	0.00	
Residual	740	40,052.57	54.13			
Total	742	70,475.96				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	33.0953	0.6730	49.1781	0.0000	31.7741	34.4164
Mileage	-0.1870	0.0127	-14.7743	0.0000	-0.2118	-0.1621
Brand Code	-10.4417	0.6250	-16.7059	0.0000	-11.6688	-9.2147

Summary and Conclusion:

After evaluating all the proposed models, the conclusion was that the best-fit regression model to predict the selling price of a used car has the following quantitative variable: mileage. Furthermore, it is recommended to include a categorical variable like the origin code or brand code to get a more accurate price due to its higher adjusted R square and lower standard error value.

Predicting the probability of selecting a USA car

Research Methodology:

The most common types of cars on the market are sedans and SUVs. Hence, these types of cars were used as categorical variables. Moreover, quantitative variables like price and age also contributed to having a significant logistic regression to define which is the most suitable origin of a car for a particular client.

In this case, the aim is to understand how the Origin code is linked to Price, Age, and Type code. A logistic regression model was built to predict the correct origin of a used car for a specific customer using both quantitative and dummy variables.

Variables under investigation:

Variables Name	Description
Origin Code	Where the car is from: 1=USA 0=Asia or Europe
Price	Price in thousands of dollars (K) at sale of used vehicle
Age	Age in years of used vehicle
Type Code	The data is categorized using the Type of car as -Sedan = 0 -SUV = 1

Objectives:

1. Build the logistic regression equation.
2. Analyze the statistical significance of the variables in the model.
3. Predict the natural log of the estimated odds ratio.
4. Estimate the odds ratio of success to failure.
5. Calculate the estimated probability of success.

Statistical Design:

The data is analyzed with the help of PHSTAT in Excel and RStudio.

Sample Size:

Total 601 samples of different cars.

Logistic Regression Equation:

There were four variables analyzed during this study:
1. Y, whether the most suited origin of a car is Asia or Europe = 0 or USA = 1
2. X1, selling price of a used car in thousands of dollars
3. X2, age of a used car in years
4. X3, whether the car type preferred is a Sedan = 0 or a SUV = 1

Predictor	Coefficients	SE Coef	*Z*	p-Value
Intercept	1.8147	0.4231	4.2889	**0.0000**
Price	-0.1254	0.0180	-6.9587	**0.0000**
Age	-0.3100	0.0627	-4.9435	**0.0000**
TypeCode:1	1.6186	0.2108	7.6766	**0.0000**
Deviance	626.4242			

$$Y = \frac{\text{Exp}(1.8147 - 0.1254\,X1 - 0.3100\,X2 + 1.6186X3)}{1 + \text{Exp}(1.8147 - 0.1254\,X1 - 0.3100\,X2 + 1.6186X3)}$$

Analysis of findings:

Determine whether X1, X2, and X3 are individually significant at a 5% significance level. The hypotheses are the following:

H0: β1=0	H0: β2=0	H0: β3=0
Ha: β1≠0	Ha: β2≠0	Ha: β3≠0

The logistic regression calculated the p-values for Price (0.0000), Age (0.000) and Type code (0.5966). The null hypothesis was rejected for all the variables mentioned before because the p-value was less than 0.05 in all those cases.

The logistic regression had a 75.71% accuracy rate, so the model does a good job of predicting probabilities. At a 5% significance level, concluded that Price, Age, and Type Code are statistically significant variables to predict the most suitable origin of a used car based on customer preferences.

-b0: 1.8147. This means that for a car with 0 age, price of $0, and type of car equal to SUV, the estimated natural logarithm of the odds ratio of finding the most suitable origin of a car for a client is 1.8147.

-b1: -0.1254. This means that holding constant the effect of age and whether the type of car is an SUV or Sedan, for an increase of $1,000 dollars in price, the estimated natural logarithm of the odds ratio of finding the most suitable origin of a car for a client decreases by 0.1254. Hence, cars with high prices are less likely to be from the USA.

-b2: -0.3100. This means that holding constant the effect of price and whether the type of car is an SUV or Sedan, for an increase of 1 year in age, the estimated natural logarithm of the odds ratio of finding the most suitable origin of a car for a client decreases by 0.3100. Hence, cars with many years are less likely to be from the USA.

-b3: 1.6186. This means that holding constant the effect of price and age, the estimated natural logarithm of the odds ratio of finding the most suitable origin of a car for a client increases by 1.6186 for SUV cars compared with Sedan cars. Hence, SUV cars are much more likely to be better from a USA origin.

Predicting the natural log of the estimated odds ratio:

Only significant variables were included to predict the probability of selecting a USA car.

Let's suppose the explanatory variables have the following values:

- Price = 25
- Age = 4
- Type Code = 1 = SUV

The logistic regression equation is (1.8147 – 0.1254 X1 - 0.3100 X2 + 1.6186X3). Using the values mentioned above (1.8147 – 0.1254 (25) - 0.3100 (4) + 1.6186 (1)), the response variable equals -0.9422.

Estimated Odds Ratio of "Success" to "Failure":

The odds ratio is a way to interpret the logistic model finding the ratio of the probability of success to the probability of failure.

To obtain the odd ratio of success, use the natural log of the odds ratio and find its exponential value.

Exp (-0.9422), which is equal to (0.3898)

Estimate the probability of Success:

To find the probability of success, use the estimated odds ratio/ 1+ estimated odds ratio.

The predicted probability of selecting a USA car for a specific customer with the values mentioned above is:

$$Y = \frac{\text{Exp } (1.8147 – 0.1254\ X1 - 0.3100\ X2 + 1.6186X3)}{1+ \text{Exp } (1.8147 – 0.1254\ X1 - 0.3100\ X2 + 1.6186X3)}$$

$$Y = \frac{\text{Exp } (1.8147 – 0.1254\ (25) - 0.3100\ (4) + 1.6186\ (1))}{1+ \text{Exp } (1.8147 – 0.1254\ (25) - 0.3100\ (4) + 1.6186\ (1))} = \frac{0.3898}{1.3898} = 0.2805$$

Calculate the odds ratio using the following formula:

$$\frac{\text{Probability of success (selecting a USA car)}}{\text{1- probability of success (selecting a USA car)}} = \frac{0.2805}{1-0.2805} = 0.3898$$

Using the values previously mentioned, the odds ratio is again equal to 0.3898.

Testing of the logistic regression:

The above logistic equation was used to predict the probability of selecting a USA car with 4 years of age, selling price of 25K and SUV type (1) of automobile is 0.2805. This means that for this case an Asian car will be a best suited origin.

The logistic regression was also used to recalculate the probability using the same age, price but for a Sedan type of car (0) and the result was 0.0717.

These outcomes predict the probability of selecting a USA car based on the type of car, price, and age.

Summary

After finishing different statistical tests and running logistic and linear regression models, the findings were:

1. The proportion of used cars from a specific type depends on the origin. USA car proportion from a particular type differs from Asian or European car proportion.

2. The average mileage and age of a used car do not differ between USA cars and Asian or European cars.

3. The price range and type of a used car are dependent on the car's origin.

4. There is a statistically significant difference in selling price between the different types of used cars available on the market.

5. There is a statistically significant difference in mileage among the various types of used cars.

6. There is a statistically significant difference in age between the different types of used cars available on the market.

7. Mileage is a key variable to determine the selling price of a used car. However, categorical variables like origin code and brand code give a more accurate selling price due to their statistical significance.

8. Mileage and Age are correlated variables; hence, it is better to remove age to avoid multicollinearity in the linear regression model.

9. To predict the probability of selecting a USA car, the dependent variables are price, age, and type of car. Sedan types of cars had lesser odds in choosing a USA car compared to a SUV type of car.

Conclusion

To analyze the used car market dynamics, some stratification factors such as brand, price, mileage, age, origin, and type of car were considered. Each of these categories were divided into various groups to conduct the following test: Z-test, T-test, Chi-Square, and ANOVA. A set of hypotheses were tested to determine the statistical significance and dependency between these independent variables based on every type of car in the study and the origin code.

To predict the selling price of a used automobile it is better to use a multiple regression model. It tested the effect of independent variables like mileage, age and origin code or brand code on determining the selling price. Numeric models and numeric plus categorical variables were built and found that the best model to predict the selling price of a used automobile consists of mileage and brand code because it had the highest adjusted R squared and the lowest standard error.

The probability of selecting a USA car based on specific preferences of a customer was predicted. A logistic regression model tested if the most suitable origin of used car is related to Price, Age, and Type (Sedans and SUV's). In summary, selecting the most suitable origin of a used car is dependent on the variables previously mentioned and rejected the null hypothesis. All the independent variables were statically significant.

Overall, the stratification factors that influence the used car market dynamics can be analyzed from different angles based on what hypothesis want to probe and the selection of the appropriate statistical test. Different tests were performed and run a couple of regression models having a clear view of the role that each variable plays.

This used car study will help sellers and buyers to keep the market constantly moving. For that reason, it is important to continue to analyze the variables and compare results with future datasets collected.

References

History.com Editors. (2018, August 21). Automobile History. HISTORY. Retrieved September 9, 2022, from https://www.history.com/topics/inventions/automobiles

"United States: Used-Vehicle Supply Rises; Average Listing Price Dips From Record Levels." TendersInfo News 21 Feb. 2022: NA. Business Insights: Global. Web. 9 Sept. 2022.

What is a sedan? Kia Corporation. (2022). Retrieved September 24, 2022, from https://www.kia.com/dm/discover-kia/ask/what-is-a-sedan.html

Corby, S. (2021, April 26). Coupé vs sedan: what's the difference? Mercedes-Benz New Zealand. Retrieved September 24, 2022, from https://www.mercedes-benz.co.nz/passengercars/experience/mercedes-me-magazine/performance/articles/coupe-vs-sedan-difference/stage.module.html

Merriam-Webster. (n.d.). Station wagon definition & meaning. Merriam-Webster. Retrieved September 24, 2022, from https://www.merriam-webster.com/dictionary/station%20wagon

Jocelyn, V., & Biagi, L. (2021, December). Sports cars report 2021. Statista. Retrieved September 24, 2022, from https://www.statista.com/study/49991/sports-cars-report/

Musclecarclub. (2022, August 31). Muscle car definition. Muscle Car Club. Retrieved September 24, 2022, from https://musclecarclub.com/muscle-car-definition/

Hall Hyundai Newport. (2016). Difference Between CUV and SUV? Hall Hyundai Newport News. Retrieved October 14, 2022, from https://www.hallhyundai.com/difference-between-a-cuv-and-suv.htm#:~:text=An%20SUV%20is%20a%20rugged,a%20minivan%20or%20large%20sedan.

Penguin Random House LLC and HarperCollins Publishers Ltd. (2019). Minivan . Collins English Dictionary. Retrieved September 24, 2022, from https://www.collinsdictionary.com/us/dictionary/english/minivan

HarperCollins Publishers Ltd. (2022). Pickup truck . Collins English Dictionary. Retrieved October 14, 2022, from https://www.collinsdictionary.com/us/dictionary/english/pickup

Dictionary.com. (2022). Van definition & meaning. Dictionary.com. Retrieved September 24, 2022, from https://www.dictionary.com/browse/van

Merriam-Webster. (n.d.). Wagon definition & meaning. Merriam-Webster. Retrieved September 24, 2022, from https://www.merriam-webster.com/dictionary/wagon