Predicting the Price of Used Cars Using Multiple Regression Modelling

Amelia Seemungal
Department of computing and information technology
University of the West Indies (St. Augustine)
816030432

Shainah Kalicharan
Department of computing and information technology
University of the West Indies (St. Augustine)
816030327

Maranda Alyssa Ragoobir
Department of computing and information technology
University of the West Indies (St. Augustine)
816034485

Anthony Jawahir
Department of computing and information technology
University of the West Indies (St. Augustine)

***Abstract*—In this paper, we investigate the application of machine learning techniques to predict the price of used cars. The predictions are based on a data set containing information on used cars sold in various states throughout the United States of America which was obtained from Kaggle. The techniques applied was the multiple linear regression model and an interactive dashboard was created for easy and efficient use by our target audience. The predictions are then evaluated to ensure that our model works as anticipated. A seemingly easy problem turned out to be indeed very difficult to resolve with high accuracy.**

I. Problem Statememnt

Prospective used car buyers with a lack of knowledge on the effect of a vehicle's mileage, make and body has on its value are highly likely to be overcharged at dealerships. Making this information available will help prospective buyers make more informed decisions and reduce the chances of buyers paying significantly more than market value for a particular used vehicle.

II. Introduction

The used vehicle market can be difficult to navigate for many individuals as they lack the proper knowledge of the market to make informed decisions. However, what most persons have in common is that they want the best value for their money. This often leads to individuals engaging in the typical "run around" of visiting various dealerships, viewing cars with similar attributes in an attempt to secure the best price. Bargaining and negotiating further add to the significant amount of effort and time required. Moreover, there's a constant race against time to secure the desired vehicle before it's claimed by another consumer. On the other hand, second-hand car dealerships aim to maximize their own profit and may perceive consumers' lack of knowledge of the used car market as an opportunity to exploit them, thereby maximizing their profit yield at the expense of the consumer. Such dynamics highlight the asymmetry of information and power between buyers and sellers in the used car market.

Several studies and articles have pointed out a notable increase in the prices of used cars in recent years, presenting various viewpoints on the underlying reasons. For example, one study highlights a significant rise in the average price of a three-year-old used car, which was just above 23,000 dollars prior to the COVID-19 pandemic in 2019, but has now surged to nearly 32,500 dollars, reflecting a 41 percent increase. Consequently, individuals with a comparable budget today would need to consider purchasing a six-year-old vehicle to maintain affordability at a similar level [1]. Some attribute this surge in prices directly to the onset of the COVID-19 pandemic in December 2019, which caused disruptions in the supply chain and a shortage of new cars, resulting in diminished supply and driving prices to unprecedented levels [2]. Conversely, alternative perspectives propose a potentially more nuanced explanation for the rise in second-hand vehicle prices. For instance, there is a suggestion that over the typical eight-year ownership period, women incur an additional 142 dollars per year in car ownership expenses compared to men, potentially resulting in an overall disparity of up to 7,800 dollars during the ownership period. Additionally, women are reported to pay approximately 117.12 dollars more than men when purchasing new cars [3]. Although there some articles that contradict this saying that used car prices have not gone up but instead have decreased "the gap between prices for new and used vehicles, which narrowed during the pandemic, has widened again" [4]. This source states that the global COVID-19 pandemic did not cause an increase in second hand vehicle prices but instead caused a decrease as there was a lack of demand from consumers due to the state of emergency or lockdown which would have cased prices to decrease.

These collected articles and studies have all identified differ-

ent reasons for the price increase, with some even disagreeing that prices have risen. However, one common theme prevails: the prices of second-hand vehicles remain in a constant state of fluctuation, with no fixed prices for any of them. Whether attributed to economic issues stemming from a global health crisis or discrimination based on gender, it has become increasingly challenging for the average person to purchase a second-hand car for their convenience without undertaking the exhausting task of finding a reliable vehicle that meets their expectations and requirements.

III. Related Works

There have been numerous studies conducted discussing topics similar in nature to ours, with some using similar approaches to this study and others using a completely different approach. In the publication by Monburinon et al. in the year 2018 [5], a comparative study on the performance of regression based on supervised machine learning models was conducted. Each model was trained using data from the used car market collected from a German e-commerce website. As a result, gradient boosted regression trees yielded the best performance with a mean absolute error of 0.28, followed by random forest regression with a mean squared error of 0.35, and multiple linear regression with a mean squared error of 0.55, respectively. This study applies similar concepts to our own with its use of multiple regression models, but it also applies a different approach with its use of the random forest regression model.

In another publication conducted by Chen et al. in the year 2017 [6], data was collected from over 100,000 used car dealing records throughout China to conduct empirical analysis on a thorough comparison of two algorithms: linear regression and random forest. They determined that random forest has a stable but not ideal effect on the price evaluation model for a certain car make, but it shows a great advantage in the universal model compared with linear regression. This study also uses a similar approach as that by Monburinon et al. (2018) [5], with its use of the random forest regression model.

Other works such as, Ozgur et al in the year 2016 [7], used a representative sample of 470 of all 2005 GM cars with the make of either Chevrolet or Pontiac. The purpose of this paper was to develop a relatively good regression equation for predicting the price of these cars. “It is known that there are many factors that influence the price of a car, but we do not know what factors will influence the price of the cars and how these factors influence the price”, this quoted line from the study shows how sensitive the market is when it comes to secondhand vehicles.

In this study conducted by Pudaruth in the year 2014 [8], investigated the application of supervised machine learning techniques to predict the price of used cars in Mauritius. The predictions were based on historical data collected from daily newspapers. Different techniques like multiple linear regression analysis, k nearest neighbor, naïve bayes and decision trees were used to make predictions.

A study conducted by Wu et al. in the year 2009 [9] proposed a system consisting of three parts: a data acquisition system, a price forecasting algorithm, and performance analysis. They utilized a neuro-fuzzy knowledge-based system to predict the price of used cars, using only three factors: the make of the car, the year of manufacture, and the engine style. This study produced similar results compared to simple regression methods. Car dealerships in the United States sell thousands of cars yearly via leasing. Most of these cars would be returned at the end of their leasing period and resold. Selling these cars at the proper price can have major economic implications. Due to this, the ODAV (Optimal Distribution of Auction Vehicles) system was developed by Du et al. [11], which predicts the best price for the vehicle and location to sell/buy. A k-nearest neighbor regression model was used to forecast the price.

In the study conducted by Catalina et al. in the year 2013 [10], a multiple regression prediction model was used to further develop their prediction model. The proposed model presented the following characteristics: three inputs and one output, simplicity, large applicability, good match with the simulations, and with the energy certification calculations, as well as human behavior correction. Despite this study and ours having differing topics, it uses a multiple regression model in a similar method as was done in our work and shows its practical application when developing an accurate, functional, and efficient prediction model, regardless of the topic being researched.

When it comes to the presentation of research, various methods can be used, such as heatmaps. In Bojko [12], heatmaps can be very effective in summarizing and communicating data, but it also states that they can often be used incorrectly and for the wrong reasons. This information is beneficial to this research project as it helps guide it through the proper usage of the heatmap.

Another form of data presentation is scatterplots. In Mayorga et al. [13], the use of scatter plots in research projects, along with their drawbacks and ways to address them, is discussed. This is beneficial to this project as it shows the proper way to handle a scatterplot.

Another form of data representation used in our research is the bar graph. In Fischer et al. [14], a series of experiments were conducted on the responses of participants who viewed bar graphs. The results show that the design of the graphs does help with the reader’s understanding of the information presented.

IV. Methodology

The data utilized in the research was collected from a Kaggle dataset as shown in figure 1 in the figures section, which presents information on various secondhand vehicles. This dataset was employed to construct a predictive model aimed at determining the optimal price for used cars based on specific consumer attributes. The researchers ensured that the data collected was recent, not older than 10 years, to maintain

accuracy regarding the price and availability of models, considering that time can influence car prices. Seasonal patterns were not taken into account during this study as it does not affect the selling or purchasing of used cars.

The following data was collected for each car: year, make, model, mileage, trim, body transmission, vehicle id, state, condition, colour, interior, seller, selling price, sale date and MMR . Among these collected data, only specific fields were utilized in the research, including make, model, body type, trim, condition, odometer reading, seller, MMR, and selling price. as the other fields were deemed unnecessary for this research, hence giving sound reasoning to the removal of the columns: transmission, vin, colour, interior and sale date this can be seen in figure 2 in the figures section.

Initially 550,000 + records were collected from the data set. However, after data cleaning, which involved the removal of rows with blank sections as well as the removal of any duplicate records,470,000 + records were left in the data set for use in our research. The values are then preprocessed in a form amenable to further processing using machine learning for the duration of our research project.The cleaning process can be seen in figures 3-6 in the figures section.

V. Solutions

In this research, several goals were established. One of these objectives was to utilize a multiple regression model to predict the selling price of a vehicle based on its make, body, and mileage.

Another goal was to show the relationship of condition, mileage, and market value by demonstrating the correlation between condition, mileage, and market value.

The final goal was to present the statistics for each state. This involved determining the standard deviation of the difference between market value and selling price for each state, as well as identifying which dealership offered the best prices in each state.

VI. Implementation, evaluation, and analysis

Multiple linear regression analysis was used to predict the prices of used vehicles in this project. The pair-wise correlation coefficient was computed between different pairs of attributes. The heatmap was generated, this can be seen in figure 7 of the figures section.

In this research, absolute error was used to evaluate the model . For the calculation of mean absolute error, the formula shown in figure 8 of the figures section below was used to calculate the mean absolute error of 1079.8497828258865. A MAE of 1079 indicates that, on average, the predictions deviate from the actual values by approximately 1079 units.

Given that the average selling price of a vehicle is 30,000 USD, we can infer that the MAE represents roughly 3.6 percent of the average value. This suggests that, on average, the predictions have an error of about 3.6 percent of the average value. In this model this level of error is acceptable since the goal of the model is to only give users a prediction of the value not an accurate value. It is just to give users a ballpark figure to guide decision making. In reality this actual value of the vehicle will depend on other factors such as market demand .

In the relation between condition and mileage there is a weak negative correlation, which means that the higher a car's mileage/odometer the lower the condition of the vehicle. This makes sense practically as cars with higher milage would have been driven more and would have begun to degrade.

Also, in the relation between condition and market value there is a weak positive correlation, which means that the higher the condition of a vehicle the higher the market value would be. This again makes sense logically as people would spend more money on higher quality vehicles. In the other relationship between mileage and market value there is a fairly strong negative correlation, which would indicate that the higher the mileage the lower the market value.

A scatterplot was generated (figure 9 of figures section)which shows the relationship between mileage and market value with condition. This scatterplot shows the concentration of different vehicles with different attributes as well as displaying the outliers of our dataset.

Another scatterplot (figure 10 in figures section)showing the relationship between the training data and the predicted values was generated. As can be seen that the majority of the data is concentrated in one area but there still are a few outliers seen.

Another analysis done on the data was the identifying if dealerships in various states with the lowest prices for vehicles with the consumer's desired attributes. The bar graph shown above shows this information with the y axis showing states and x axis showing prices this data can be seen in figure 11 of figures section.

VII. Challenges Faced

During the duration of the research multiple challenges were discovered one of which was a difficulty in processing non-numerical data in the regression model, the solution to this was to classify the columns as categorical data, hence making processing simpler.

Another challenge faced was abundance of categorical variables resulted in difficulty when dropping dummy variables. This issue proved itself to be verry difficult in nature so our solution to this was simply to use trial and error.

VIII. Conclusion

As evident from the preceding paragraphs, disparities in used car prices present an inconvenience to consumers, necessitating a solution to address the issue. In this study, a multiple linear regression model was employed to forecast the prices of vehicles based on various consumer preferences. Additionally, a user-friendly dashboard was developed for consumer convenience. Challenges encountered included handling non-numeric data in the regression model and dealing with numerous categorical variables, which made dropping dummy variables difficult. Nonetheless, these challenges were successfully addressed to complete the project. The group members extend their gratitude to Dr. Patrick Hosein and Mr. Sergio Maturin for the opportunity to conduct this research.

IX. References

[1] Gorzelany, J(2023), The Unfortunate Solution To Sky-High Used-Car Prices: Buy An Older Model, Forbes

[2] Delvillar, A (2023), The state of used car prices: why are certain car brands so high?, CBT news

[3]Tengler, S (2021), New "Pink Tax" Study Shows Women Pay Upwards Of 7,800 dollars More For Car Ownership, Forbes

[4] Haytt, D (2023), Used-Car Shoppers Are Getting a Break As Prices Fall—Unless They Need A Loan, Investopedia

[5] Monburinon et al (2018), Prediction of prices for used car by using regression models,IEEE

[6] Chen et al (2017), Comparative analysis of used car price evaluation models, AIP publishing

[7] Ozgur et al (2016), Multiple Linear Regression Applications Automobile Pricing, International Journal of Mathematics and Statistics Invention (IJMSI)

[8] Pudaruth,S (2014) Predicting the Price of Used Cars using Machine Learning Technique, International Journal of Information Computation Technology.

[9] Wu et al (2009), An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference, Science Direct

[10] Catalina et al (2013), Multiple regression model for fast prediction of the heating energy demand, Science Direct

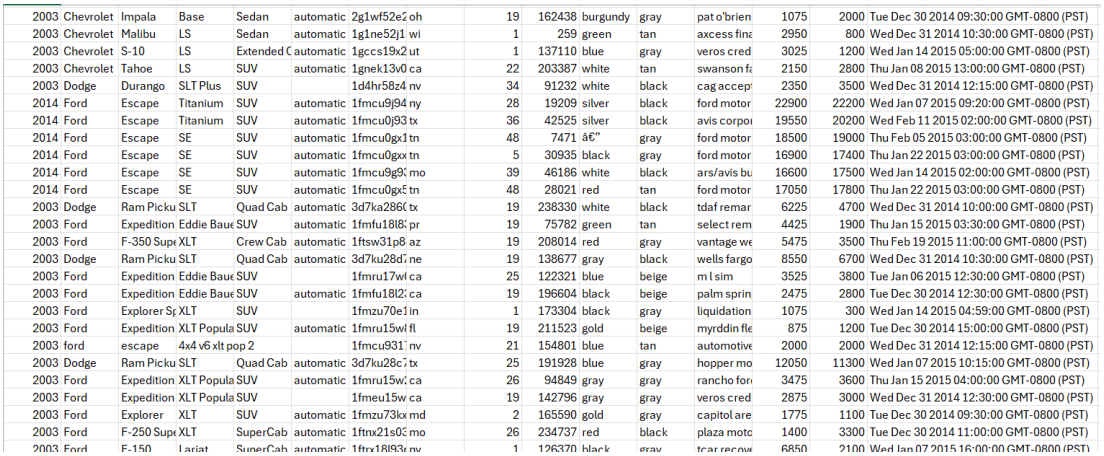
[11] Du et al (2009), PIN Optimal Distribution of Auction Vehicles System: Applying Price Forecasting, Elasticity Estimation, and

[12] Bojko, A (2009), Informative or Misleading? Heatmaps Deconstructed, Springer link

[13] Mayorga et al,(2013), Splatterplots: Overcoming Overdraw in Scatter Plots,IEEE Xplore.

[14]Fischer,(2015), Designing bar graphs: orientation matters, Wiley online library

X. figures



2003	Chevrolet	Impala	Base	Sedan	automatic	2g1wf52e2	oh	19	162438	burgundy	gray	pat o'brien	1075	2000	Tue Dec 30 2014 09:30:00 GMT-0800 (PST)
2003	Chevrolet	Malibu	LS	Sedan	automatic	1g1ne52j1	wi	1	259	green	tan	axcess fina	2950	800	Wed Dec 31 2014 10:30:00 GMT-0800 (PST)
2003	Chevrolet	S-10	LS	Extended C	automatic	1gccs19x2	ut	1	137110	blue	gray	veros cred	3025	1200	Wed Jan 14 2015 05:00:00 GMT-0800 (PST)
2003	Chevrolet	Tahoe	LS	SUV	automatic	1gnek13v0	ca	22	203387	white	tan	swanson fa	2150	2800	Thu Jan 08 2015 13:00:00 GMT-0800 (PST)
2003	Dodge	Durango	SLT Plus	SUV		1d4hr58z4	nv	34	91232	white	black	cag accep	2350	3500	Wed Dec 31 2014 12:15:00 GMT-0800 (PST)
2014	Ford	Escape	Titanium	SUV	automatic	1fmcu9j94	ny	28	19209	silver	black	ford motor	22900	22200	Wed Jan 07 2015 09:20:00 GMT-0800 (PST)
2014	Ford	Escape	Titanium	SUV	automatic	1fmcu0j93	tx	36	42525	silver	black	avis corpo	19550	20200	Wed Feb 11 2015 02:00:00 GMT-0800 (PST)
2014	Ford	Escape	SE	SUV	automatic	1fmcu0gx1	tn	48	7471	â€"	gray	ford motor	18500	19000	Thu Feb 05 2015 03:00:00 GMT-0800 (PST)
2014	Ford	Escape	SE	SUV	automatic	1fmcu0gxx	tn	5	30935	black	gray	ford motor	16900	17400	Thu Jan 22 2015 03:00:00 GMT-0800 (PST)
2014	Ford	Escape	SE	SUV	automatic	1fmcu9g9x	mo	39	46186	white	black	ars/avis bu	16600	17500	Wed Jan 14 2015 02:00:00 GMT-0800 (PST)
2014	Ford	Escape	SE	SUV	automatic	1fmcu0gx5	tn	48	28021	red	tan	ford motor	17050	17800	Thu Jan 22 2015 03:00:00 GMT-0800 (PST)
2003	Dodge	Ram Picku	SLT	Quad Cab	automatic	3d7ka286(	tx	19	238330	white	black	tdaf remar	6225	4700	Wed Dec 31 2014 10:00:00 GMT-0800 (PST)
2003	Ford	Expedition	Eddie Baue	SUV	automatic	1fmfu18l8	pr	19	75782	green	tan	select rem	4425	1900	Thu Jan 15 2015 03:30:00 GMT-0800 (PST)
2003	Ford	F-350 Supe	XLT	Crew Cab	automatic	1ftsw31p8	az	19	208014	red	gray	vantage we	5475	3500	Thu Feb 19 2015 11:00:00 GMT-0800 (PST)
2003	Dodge	Ram Picku	SLT	Quad Cab	automatic	3d7ku28d7	ne	19	138677	gray	black	wells fargo	8550	6700	Wed Dec 31 2014 10:30:00 GMT-0800 (PST)
2003	Ford	Expedition	Eddie Baue	SUV		1fmru17w(	ca	25	122321	blue	beige	m l sim	3525	3800	Tue Jan 06 2015 12:30:00 GMT-0800 (PST)
2003	Ford	Expedition	Eddie Baue	SUV	automatic	1fmfu18l2	ca	19	196604	black	beige	palm sprin	2475	2800	Tue Dec 30 2014 12:30:00 GMT-0800 (PST)
2003	Ford	Explorer Sp	XLT	SUV		1fmzu70e1	in	1	173304	black	gray	liquidation	1075	300	Wed Jan 14 2015 04:59:00 GMT-0800 (PST)
2003	Ford	Expedition	XLT Popula	SUV	automatic	1fmru15w(	fl	19	211523	gold	beige	myrddin fle	875	1200	Tue Dec 30 2014 15:00:00 GMT-0800 (PST)
2003	ford	escape	4x4 v6 xlt pop 2			1fmcu931'	nv	21	154801	blue	tan	automotive	2000	2000	Wed Dec 31 2014 12:15:00 GMT-0800 (PST)
2003	Dodge	Ram Picku	SLT	Quad Cab	automatic	3d7ku28c7	tx	25	191928	blue	gray	hopper mo	12050	11300	Wed Jan 07 2015 10:15:00 GMT-0800 (PST)
2003	Ford	Expedition	XLT Popula	SUV	automatic	1fmru15w(	ca	26	94849	gray	gray	rancho for	3475	3600	Thu Jan 15 2015 04:00:00 GMT-0800 (PST)
2003	Ford	Expedition	XLT Popula	SUV		1fmeu15w	ca	19	142796	gray	gray	veros cred	2875	3000	Wed Dec 31 2014 12:30:00 GMT-0800 (PST)
2003	Ford	Explorer	XLT	SUV	automatic	1fmzu73kx	md	2	165590	gold	gray	capitol are	1775	1100	Tue Dec 30 2014 09:30:00 GMT-0800 (PST)
2003	Ford	F-250 Supe	XLT	SuperCab	automatic	1ftnx21s0(	mo	26	234737	red	black	plaza moto	1400	3300	Tue Dec 30 2014 11:00:00 GMT-0800 (PST)
2003	Ford	F-150	Lariat	SuperCab	automatic	1ftrx18l83	nv	1	126370	black	gray	tcar recov	6850	2100	Wed Jan 07 2015 16:00:00 GMT-0800 (PST)

Fig. 1. kaggle dataset used for study

	year	make	model	body	state	odometer	seller	mmr	sellingprice
0	2015	Kia	Sorento	SUV	ca	16639.0	kia motors america inc	20500.0	21500.0
1	2015	Kia	Sorento	SUV	ca	9393.0	kia motors america inc	20800.0	21500.0
2	2014	BMW	3 Series	Sedan	ca	1331.0	financial services remarketing (lease)	31900.0	30000.0
3	2015	Volvo	S60	Sedan	ca	14282.0	volvo na rep/world omni	27500.0	27750.0
4	2014	BMW	6 Series Gran Coupe	Sedan	ca	2641.0	financial services remarketing (lease)	66000.0	67000.0
...	...	...	...	...	...	...	...	...	...
558831	2011	BMW	5 Series	Sedan	fl	66403.0	lauderdale imports ltd bmw pembrok pines	20300.0	22800.0
558833	2012	Ram	2500	Crew Cab	wa	54393.0	i -5 uhlmann rv	30200.0	30800.0
558834	2012	BMW	X5	SUV	ca	50561.0	financial services remarketing (lease)	29800.0	34000.0
558835	2015	Nissan	Altima	sedan	ga	16658.0	enterprise vehicle exchange / tra / rental / t...	15100.0	11100.0
558836	2014	Ford	F-150	SuperCrew	ca	15008.0	ford motor credit company llc pd	29600.0	26700.0

Fig. 2. image of dataset after removing unnecessary data

	year	make	model	trim	body	transmission	vin	state	condition	odometer	color	interior
0	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg566472	ca	5.0	16639.0	white	black
1	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg561319	ca	5.0	9393.0	white	beige
2	2014	BMW	3 Series	328i SULEV	Sedan	automatic	wba3c1c51ek116351	ca	45.0	1331.0	gray	black
3	2015	Volvo	S60	T5	Sedan	automatic	yv1612tb4f1310987	ca	41.0	14282.0	white	black
4	2014	BMW	6 Series Gran Coupe	650i	Sedan	automatic	wba6b2c57ed129731	ca	43.0	2641.0	gray	black

Fig. 3. image of dataset after removing null values (a)

body	transmission	vin	state	condition	odometer	color	interior	seller	mmr	sellingprice	saledate
SUV	automatic	5xyktca69fg566472	ca	5.0	16639.0	white	black	kia motors america inc	20500.0	21500.0	Tue Dec 16 2014 12:30:00 GMT-0800 (PST)
SUV	automatic	5xyktca69fg561319	ca	5.0	9393.0	white	beige	kia motors america inc	20800.0	21500.0	Tue Dec 16 2014 12:30:00 GMT-0800 (PST)
Sedan	automatic	wba3c1c51ek116351	ca	45.0	1331.0	gray	black	financial services remarketing (lease)	31900.0	30000.0	Thu Jan 15 2015 04:30:00 GMT-0800 (PST)
Sedan	automatic	yv1612tb4f1310987	ca	41.0	14282.0	white	black	volvo na rep/world omni	27500.0	27750.0	Thu Jan 29 2015 04:30:00 GMT-0800 (PST)
Sedan	automatic	wba6b2c57ed129731	ca	43.0	2641.0	gray	black	financial services remarketing (lease)	66000.0	67000.0	Thu Dec 18 2014 12:30:00 GMT-0800 (PST)

Fig. 4. image of dataset after removing null values (b)

	year	make	model	trim	body	transmission	vin	state	condition	odometer	color	interior
0	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg566472	ca	5.0	16639.0	white	black
1	2015	Kia	Sorento	LX	SUV	automatic	5xyktca69fg561319	ca	5.0	9393.0	white	beige
2	2014	BMW	3 Series	328i SULEV	Sedan	automatic	wba3c1c51ek116351	ca	45.0	1331.0	gray	black
3	2015	Volvo	S60	T5	Sedan	automatic	yv1612tb4f1310987	ca	41.0	14282.0	white	black

Fig. 5. image of dataset after removing duplicated values (a)

n	body	transmission	vin	state	condition	odometer	color	interior	seller	mmr	sellingprice	saledate
X	SUV	automatic	5xyktca69fg566472	ca	5.0	16639.0	white	black	kia motors america inc	20500.0	21500.0	Tue Dec 16 2014 12:30:00 GMT-0800 (PST)
X	SUV	automatic	5xyktca69fg561319	ca	5.0	9393.0	white	beige	kia motors america inc	20800.0	21500.0	Tue Dec 16 2014 12:30:00 GMT-0800 (PST)
3i V	Sedan	automatic	wba3c1c51ek116351	ca	45.0	1331.0	gray	black	financial services remarketing (lease)	31900.0	30000.0	Thu Jan 15 2015 04:30:00 GMT-0800 (PST)
5	Sedan	automatic	yv1612tb4f1310987	ca	41.0	14282.0	white	black	volvo na rep/world omni	27500.0	27750.0	Thu Jan 29 2015 04:30:00 GMT-0800 (PST)

Fig. 6. image of dataset after removing duplicated values (b)

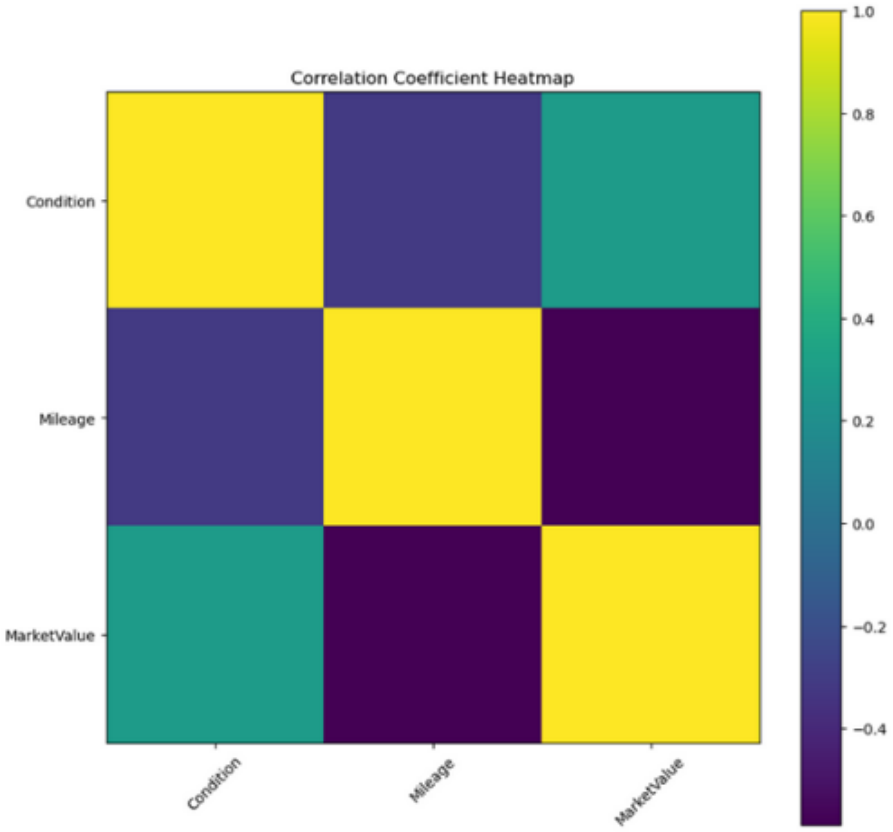


Fig. 7. heatmap generated to show the pairwise correlation of various vehicle attributes

$$MAE = 1/n \sum_{i=1}^{n} |y_i - x_i|$$

Fig. 8. equation of MAE

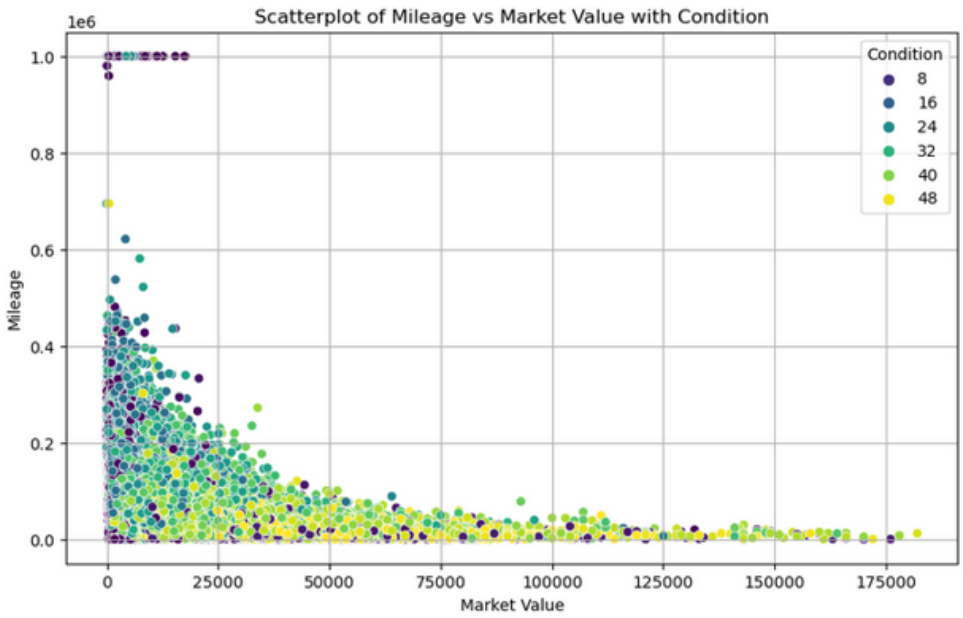


Fig. 9. scatterplot showing relationship between mileage and market price

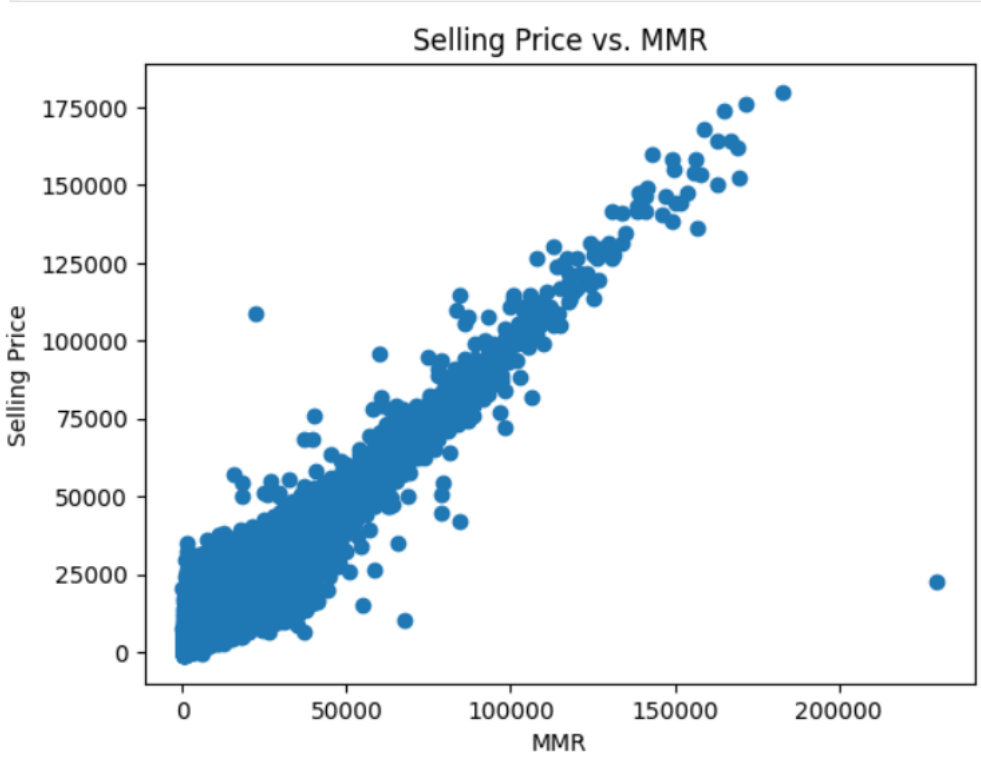


Fig. 10. scatterplot showing relationship between selling price and market price

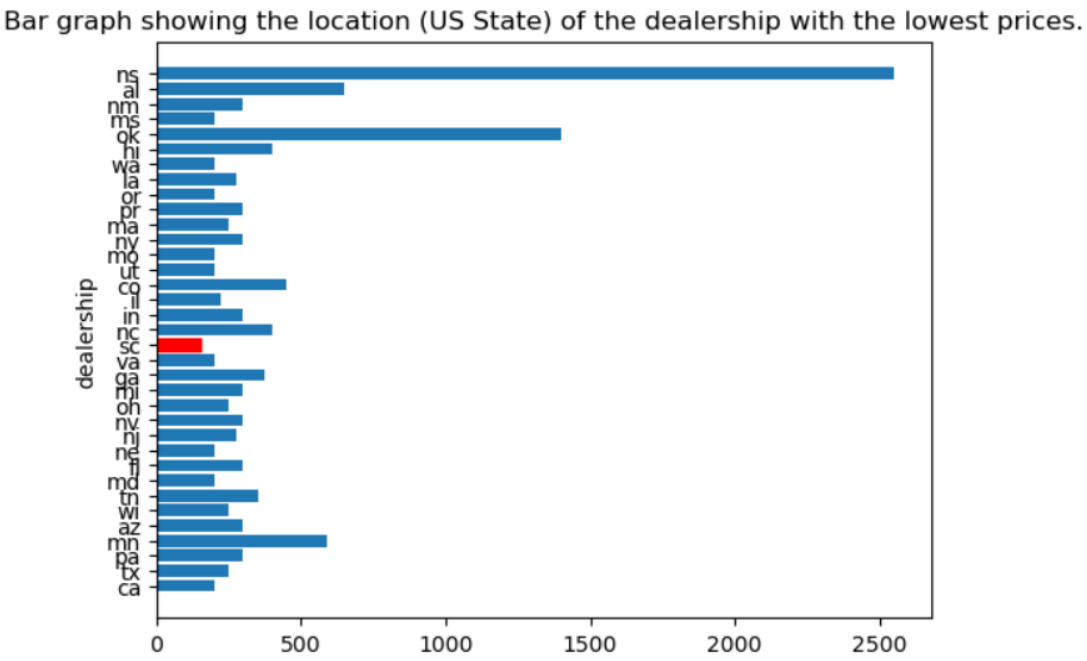


Fig. 11. bar graph showing the location (US state) of the dealership with the lowest prices