

Ankita H. Savaliya

Professor Dr. Shilpa Balan

CIS 5250

1 Nov 2022

Used Car listing for the US and Canada

A. Data sets:

My project topic is Used Car listing for the US and Canada. I will be focusing on the only U.S used car market.

→**Dataset:** <https://www.kaggle.com/datasets/rupeshraundal/marketcheck-automotive-data-us-canada?select=us-dealers-used.csv> (Links to an external site.)

The project aims to analyze how used car feature impact market prices and use data from online listings to predict prices based on vehicle feature.

This proposal first presents the motivation and goals of the project. Then follows his dataset of over 1,000,000 observations on US used car listings. It also introduces first-level data **cleaning** and **analysis**. Finally, we **conclude** with a **static** summary and statistical tests.

B. Introduction

For buyers new to the used car market, there is general confusion about what to expect. Therefore, in this project, we will try to provide a used car analysis based on the dataset :

<https://www.kaggle.com/datasets/rupeshraundal/marketcheck-automotive-data-us-canada?select=us-dealers-used.csv> (Links to an external site.)

The purpose of this project is to discuss the characteristics of used cars in terms of their impact on market prices. Based on such analysis, we can build price prediction models and provide historical data. The final product takes as input the characteristics of a used car (make, model, body type, year, mileage, etc.) and outputs a price prediction.

The data used in the following analysis is from Kaggle and it includes his over 1,000,000 raw observations of 21 features collected from US used car listings).

→Following is some of the fields, I am using for the analysis.

Make	McLaren	Audi	Am	Lincoln	Isuzu	Mercury
	Ferrari	Land	General	Volkswagen	Subaru	FIAT
	Mercedes-Benz	Rover	Chevrolet	Eagle	GENESIS	Saturn
	Porsche	BMW	Nissan	Jeep	Acura	Scion
	Lamborghini	KARMA	Toyota	Hummer	Chrysler	Suzuki
	Rolls-Royce	Maserati	Honda	Hyundai	INFINITI	Oldsmobile
	Ford	Tesla	MINI	GMC	Mitsubishi	smart
	Bentley	Lexus	Kia	RAM	Sterling	Saab
	Aston Martin	Dodge	Volvo	Alfa Romeo	Mazda	Plymouth

	Jaguar	Cadillac Lotus	Maybach	Pontiac	Buick	Daewoo Geo Bugatti
Miles	Starting from 1 to 142820					
Year	2019-2022					
Price	Starting from \$1500 up to \$ 649000					

C. Data Description:

→The data set contains 21 variables, but I will be using only below fields for analysis.

Data Field	Description	Example Values
VIN	VIN of the Car Data type: Text	3N1AB7AP2KY253755, KNDPRCA69K7513914
Price	The car price as listed on the website Data type: Currency	Starting from \$1500 to \$649000
Miles	The car miles/odometer as listed on the website Data type: Number	Starting from 1 to 142820
Year	Model Year of the car Data type: Number	2019,2020,2021
Make	Manufacturer Data type: Text	Nissan, Toyota, Acura
Model	Model of the car Data type: Text	Sentra, Camry, TLX
Trim	Trim of the car Data type: Text	S, XSE, Premium
Body Type	The body type of the car Data type: Text	Sedan, Hatchback, SUV
State	USA state (short forms) Data type: Text	CA
City	Cities in California Data type: Text	West Covina, Bloomington

D. Data Cleaning:

Firstly, we load the raw data from Kaggle, and import common packages for data presentation and visualization. The raw data contains more than 1,000,000 observations and 21 values including the price and other information about the US used car listing. As we are only interested in some columns among 21 columns, the irrelevant columns are dropped.

Data Cleaning Steps:

Step 1:

Data Field: All fields

Category Name: Remove Duplicates

Pre-Cleaning:

Below screenshot is to provide number of rows, so can reference it after cleaning steps.

VIN	Price	Miles	Year	Make	Model	Trim	Body_Type	City	State
RSGWGF4CF3F284719	20998	115879	2015	Chevrolet	Express Cargo	Work Van	Cargo Van	Emmett City	MO
WHTT2BC50V1087514	279521	7339	2018	BMW	i3	s	Hatchback	Pompano Beach	FL
ML3JF4J13HJ010325	110555	39798	2018	Mitsubishi	Mirage G4	SE	Sedan	West Bend	WI
1GCGTE15K120191189	52997	28568	2019	Chevrolet	Colorado	ZR2	Pickup	Layton	UT
1G2ZU15K000000003	188485	2008	Pontiac	G5	Base	Coupe	Saint Peters	MO	
WHTT2AR13C9V200763	65000	74374	2010	Chevrolet	Giga World	RS	Hatchback	Kennesaw	NC
KGGBABD0KX5590012	105974	74374	2010	Chevrolet	HHR	LT	Mini MPV	Grand Blanc	MI
SYMCY008090040057	42000	2020	BMW	X6 M	Base	SUV	Miami	FL	
1GCGSC15E11950013	23024	131286	2016	Chevrolet	Colorado	LT	Pickup	Jefferson City	MO
WHDUD7U5BRA070249	169995	60150	2008	Mercedes-Benz	CL-Class	CL550	Coupe	Tustin	CA
ZMMAG45AS4K0037556	27999	60122	2008	Maserati	GranTurismo	Base	Coupe	Bethelville	AR
1G1WQ5EMX3A116446	111221	2010	Chevrolet	Impala	Police Sedan	Sedan	Somersett	KY	
JTHJ202408150484	17995	107146	2008	Lexus	GX	470	SUV	Charlottesville	VA
1GCGWQAFP000134969	30799	33015	2020	Chevrolet	Express Cargo	Work Van	Cargo Van	Burnsville	MN
1K0000000000000000	45999	143459	2019	Ford	F-250 2500 Pickup	Laramie	Pickup	DownersTown	MD
1HGAC6715RA232314	45999	143459	2008	Mercedes-Benz	GL450	SUV	SUV	Chicago	IL
1GCGTE1211136522	47848	7556	2020	Chevrolet	Colorado	ZR2	Pickup	Pleasanton	CA
WP1AD2A2JAE7A7213	40988	69610	2014	Porsche	Cayenne	GTS	SUV	Wilmington	NC
1C4AD1P13E036484	55595	28904	2018	Dodge	Durango	SRT	SUV	Englewood	NJ
1GCGWQAFP000134968	99999	60122	2008	Chevrolet	Express Cargo	Work Van	Cargo Van	Fair Wayne	IN
1MAGJPT18RC2A0389	15996	31371	2018	Ford	Ecosport	SE	Crossover	Naperville	IL
1GCGSHF95C8167729	16888	35001	2012	Chevrolet	Colorado	Work Truck	Pickup	Winiflow	AZ
25UXKUZ54KLN65802	72988	16579	2019	BMW	X5	50i	SUV	Bloomfield	MI
JTHJ2024081504848	20995	168460	2010	Ford	F-350 Super Duty	XL	Pickup	Paterson	NJ
1WV00000000000000	99999	143459	2019	Mercedes-Benz	SL-Class	S550	Sedan	Shreveboro	OH
1HMKL6538L669736	10500	141736	2008	Ford	Expedition	XLT	SUV	Hibbing	MN
1FDMK15F20C48510	60000	26114	2016	Ford	E-Series Cutaway	Base	Cutaway	Ennis	TX
3CGRURCBGG61689	36951	91682	2016	RAM	Ram 2500 Pickup	Tradesman	Pickup	Potsdam	NY
3C7CGRU1G1G16505	36997	0	2016	RAM	Ram 3500 Chassis Cab	Tradesman	Chassis Cab	Salina	KS
JTHJ2024081504849	219881	2008	Ford	Transit	430	SDV	Crossover	Cleveland	OH
SUXKUZ51015274	47000	51530	2018	BMW	328i	xDriveSi	SUV	Warren Robins	GA
1GCHK24G7E190586	53230	2007	Chevrolet	Silverado Classic 2500HD	Work	Pickup	Dubuque	Iowa	
JTHJ2049859000270	26446	82885	2006	Lexus	SC	430	Convertible	Durham	NC
WUWACAHF150900659	31995	99974	2013	Audi	RS 5	Base	Convertible	Lake Bluff	IL
1G4H4501000000009	95000	2010	Buick	Lucerne	CX	Sedan	Kirkwood	MO	

Average: 1.80869E+11 Count: 1016549 Sum: 5.42204E+17

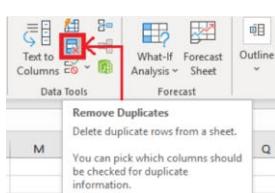
74°F Sunny 4:50 PM 10/13/2022

Post-Cleaning:

Average: 1.80869E+11 Count: 1016549 Sum: 5.42204E+17

74°F Sunny 4:51 PM 10/13/2022

-Before cleaning, dataset had more than 1,000,000 observations. I remove duplicate values from dataset.



-After cleaning, dataset reduced to around 860000 unique observations.

Step 2:

Data Field: VIN

Category Name: Find & Replace Missing Values

Pre-Cleaning:

	VIN Check	VIN	Price	Mile	Year	Make	Model	Trim	Body_type	City	Stat
1	Sort Smallest to Largest	20998 115879	2015 Chevrolet	Express Cargo	2018	BMW	Work Van		Cargo Van	Fayette City	MD
2	Sort Largest to Smallest	20921 7339					X	Hatchback		Pensacola Beach	FL
3		11093 23838					A	Sedan		Winnipeg Beach	WI
4		52997 28568	2019 Chevrolet	Colorado	2016		ZR2	Pickup		Layton	UT
5							LT	Coupe		Saint Peters	MO
6			188485	2008 Pontiac	GS			Base		Holbrook	Arizona
7			105972 100000	2018 Ford	Li			High-Wheel		Minneapolis	NE
8			9540 4424	2020 Chevrolet	MM		LT	Mini MPV		Grand Blanc	MI
9			105972 9549	2020 BMW				Base		Miami	FL
10			105972 100000	2018 Chevrolet	MM			SVT		Tustin	CA
11			42000	2008 Mercedes-Benz	CL-Class		CL550	Coupe		Bethesda	MD
12			106951 110615	2013 Mercedes-Benz	CLS-Class		CL550	Coupe		Wichita Falls	TX
13			27999 110615	2013 Chevrolet	Colorado		LS	Coupe		Orlando	FL
14			111223 100000	2010 Chevrolet	Impala		Police Sedan	Sedan		Somerset	KY
15			17995 101746	2008 Lexus	GS		470	SLV		Chantilly	VA
16			30799 101746	2013 Chevrolet	EQUINOX		LT	Van		Minneapolis	MN
17			45999 70954	2017 RAM	Express Cargo		WB7500	Pickup		Jessup	MD
18			143459	2018 Dodge	RAM 2500 Pickup		LT	Van		Chicago	IL
19			47845 100000	2018 Chevrolet	Colorado		LT	Pickup		Washington	CA
20			40988 69610	2014 Porsche	Cayenne		GTS	SLV		Wilmington	NC
21			55595 28904	2018 Dodge	Durango		SRT	SLV		Englewood	NJ
22			105972 100000	2018 Chevrolet	Colorado		LT	Van		Franklin	IN
23			125996 21271	2018 Ford	Expor		SL	Crossover		Nashville	TN
24			166888 35001	2012 Chevrolet	Colorado		LS	Pickup		Winnsboro	LA
25			27999 100000	2012 Ford	Expor		XLT	Van		Winnipeg	MB
26			29995 184490	2020 Ford	F-350 Super Duty		XL	Pickup		Paterson	NJ
27			39999 95746	2016 Mercedes-Benz	GL-Class		5550	Sedan		Shreveport	OH
28			100000 191746	2012 Ford	Expor		LT	SUV		Frisco	TX
29			60000 281514	2013 Ford	F Series Cutaway		Base	Cutaway		Portland	OR
30			30691 91682	2015 RAM	RAM 3500 Pickup		Tradesman	Pickup		Salina	KS
31			30691 91682	2015 RAM	RAM 3500 Chassis Cab		Chassis Cab	Pickup		Waukegan	IL
32			318095 211712	218095 0	2016 RAM		470	SLV		Warren Robins	GA
33			470000 51580	2018 BMW	XS		driveAss	SLV			
34	TRUE	DC1AVW1A1EG1049045	30691 91682	2015 RAM	RAM 3500 Chassis Cab		470	SLV			
35	TRUE	JTB870060511712	75120	2017 Chevrolet	Colorado		LS	Pickup		Dubuque	IA
36	TRUE	JTMN9K4C1D190546	20446	2018 Ford	Expor		XLT	Van		Montgomery	AL
37	TRUE	JWVAGGAHSDN990659	31995 99974	2013 Audi	RS 5		Base	Convertible		Lake Bluff	IL
38	TRUE	JG1HA1F14AU1024589	95000	2010 Buick	Lucerne		CX	Sedan		Kirkwood	MO

Post-Cleaning:

	VIN Check	VIN Replace With Unknown	VIN	Price	Mile	Year	Make	Model	Trim	Body_type	City	Stat
42238	FALSE	Unknown	3.20E+15 18988	130024	2009	Ford					Springfield	MO
42104	FALSE	Unknown	2.32E+16 18900	66900		Volkswagen					Van Nuys	CA
46590	FALSE	Unknown	1.23E+16 6995		2006	MINI	Truck				Richmond	VA
48146	FALSE	Unknown	2.02E+16 41995		2021	Mercury					Clearwater	MN
48406	FALSE	Unknown	2.02E+16 38500			Ford					Lillington	NC
53910	FALSE	Unknown	7.78E+16 26000	76104	2015	BMW	X5				Monroeville	PA
54800	FALSE	Unknown	1.00E+17	52798	2009	Chevrolet	Defender	SEDDAN	Pickup		Saddle Brook	NJ
54307	FALSE	Unknown	7.78E+16 0	0	2015	Jaguar	XF				Katy	TX
54535	FALSE	Unknown	3.33E+16		2015	Cadillac	DeVille				Bronx	NY
55961	FALSE	Unknown	1.00E+17	9000	2009	Chevrolet	Defender	SEDDAN	Pickup		Johnsbury	VT
56621	FALSE	Unknown	3.33E+16 9500		2015	Cadillac	DeVille				Jackson	TN
57010	FALSE	Unknown	1.20E+16 35977	62805	2008	BMW	3 Series	M3			Hatfield	PA
862635												
862636												
862637												
862638												
862639												
862640												
862641												
862642												
862643												
862644												
862645												
862646												
862647												
862648												
862649												
862650												
862651												
862652												
862653												
862654												
862655												
862656												
862657												
862658												

-By Using ISTEXT formula, if Blank or any other char have found, replace it with 'Unknown'.

Step 3:

Data Field: Price

Category Name: Replacing Null Values

Pre-Cleaning:

AutoSave

UserCar.csv

File Home Insert Draw Page Layout Formulas Data Review View Help

Conditional Formatting Styles Cells Comments Share

General Number Styles Cell Styles

Autosum Sort & Find Filter Select Analyze Data

Savally, Ankita Hemkumhalkar

Price

B1	C1	D1	E1	F1	G1	H1	I1	J1	K1	L1	M1	N1	O1
VIN	Price	Miles	-Year	Make	Model	Trim	Body_type	City	State				
WV1P72KCF1B1284719	20998	115879	2015	Chevrolet	Express Cargo	Work Van	Cargo Van	West City	MO				
WPV172KCS5U073114	27921						Hatchback	Pompano Beach	FL				
ML3JF71ZJ2HF10325	11055						Sedan	West Bend	WI				
1G0PTE1E151219189	52997						Pickup	Layton	UT				
1GALL1F087312093							Coupe	Saint Peters	MO				
WB71PAC54SF500763							Hatchback	Kernersville	NC				
SYAHCN000400457	6500						Minivan	Grand Blanc	MI				
1G0GC15151G199513	105794						SUV	Man	FL				
WDOE0F1X0D0814550	23024						Pickup	Jefferson City	MO				
WDDOD1C9B8A170249	16995	110615	2011	Mercedes-Benz	CLS-Class	CL550	Coupe	Tustin	CA				
ZAMGUS5AS0037356	27995	60122	2008	Maserati	GranTurismo	CL550	Coupe	Bentonville	AR				
21GWD1A0T1A146446							Coupe	Portland	OR				
1KCNM1A146L177884	17995	11121	2010	Chevrolet	Impala	Police Sedan	Sedan	Somerset	KY				
1KCNM1A146L177884	30799	33800	2014	Chevrolet	Impala	4D	SUV	Charlottesville	VA				
30JUR5C1H065895	45999	70954	2017	Chevrolet	Express Cargo	Work Van	Cargo Van	Burnsville	MN				
418F71EFA2A73733	143459	2008	Mercedes-Benz	Colorado	Ram 2500 Pickup	Laramie	Pickup	Leonardtown	MD				
41GPTTE1D11163522	47848	7556	2000	Chevrolet	Colorado	GL-Class	SUV	Chicago	IL				
WP1AD2A81E8747813	40988	69610	2014	Porsche	Cayenne	GT2	Pickup	Pleasanton	CA				
1HPCF1C63E0613636	55595	28028	2008	Chevrolet	Durango	GTS	SUV	Wilmington	NC				
1HGWQAWP17H17884	35290	111589	2000	Chevrolet	Express Cargo	SRT	SUV	Englewood	NJ				
MA13P1TB8C140389	15996	31371	2018	Ford	EcoSport	Work Van	Cargo Van	Fort Wayne	IN				
1GHC5SP1R8C167724	16888	35001	2012	Chevrolet	Colorado	Work Truck	Crossover	Hauppauge	NY				
25UXJ254NL65802	72988	16578	2019	BMW	X5	50i	Pickup	Winslow	AZ				
1FTWW3B8L1E000000	20995	16840	2010	Ford	F-350 Super Duty	XL	SUV	Bloomfield	MI				
1HPCF1C63E0613636	39995	95746	2018	Mercedes-Benz	GLS-Class	Sedan	Pickup	Paterson	NJ				
1HPCF1C63E0613636	10590	58500	2018	Mercedes-Benz	GLS-Class	AMG	Sedan	Streetboro	OH				
1E0K4F53Z0C48610	60000	26114	2016	Ford	E-Series Cutaway	Base	Pickup	Wabash	MN				
3C0URSJCBG616189	36951	91682	2016	RAM	Ram 2500 Pickup	Tradesman	Pickup	Emmis	TX				
3CT7WRA1A1G40505	36997	0	2016	RAM	Ram 3500 Chassis Cab	Tradesman	Pickup	Potsdam	NY				
SUXKRN1Z0860117234		219883	2006	BMW	XS	470	Cabriolet	Salina	KS				
SUXKRN1Z0860117234	47000	51530	2018	BMW	XS	xDrive50i	SUV	Cleveland	OH				
1GHCQ497P190586		53230	2007	Chevrolet	Silverado Classic 2500HD	Work	Pickup	Warren Robins	GA				
1HPCF1C63E0613636	26446	99974	2013	Audi	RS 5	50i	Convertible	Dubuque	IA				
36WUAGFAHFS090659	31995	99974	2013	Audi	RS 5	50i	Convertible	Lake Bluff	IL				
1G4HAE11410489		95000	2010	Buick	Lucerne	CX	Sedan	Kirkwood	MO				

UserCar

Type here to search

Search (Alt+F)

Autosum

Sort & Find Filter Select

Analyze Data

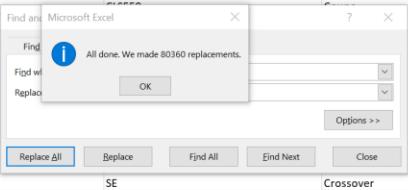
Comments

Share

Average: 25019.68409 Count: 783274 Sum: 19572223334

5:21 PM 10/13/2022

Post-Cleaning:



VIN	Price	Miles	Year	Make	Model	Trim	Body_type	City	State
1 GCGWGF3F1284719	20998	115879	2015	Chevrolet	Express Cargo	Work Van	Cargo Van	Ellicott City	MD
2 WBV72C59VBLB7514	27921	7339	2018	BMW	i3	s	Hatchback	Pompano Beach	FL
3 ML32F4J2H10325	11055	39798	2018	Mitsubishi	Mirage G4	SE	Sedan	West Bend	WI
4 1GCPTE15K1291189	52997	28568	2019	Chevrolet	Colorado	ZR2	Pickup	Layton	UT
5 1G2A1L8F087312093	0	188485	2008	Pontiac	G5	Base	Coupe	Saint Peters	MO
7 WBY124C54FV500763	0	54276	2015	BMW	i3	LT	Hatchback	Kernersville	NC
8 3GNBABDBXAS598032	6500	74274	2010	Chevrolet	HHR	X6 M	Mini Mpv	Grand Blanc	MI
9 5YMCYCOC08LD04057	105974	9589	2020	BMW	LT	Base	SUV	Miami	FL
10 1GCGSC15G1195013	23024	131286	2016	Chevrolet	Colorado	LT	Pickup	Jefferson City	MO
11 WDDDEJ71X0A014550	0	42000	2008	Mercedes-Benz	CL-Class	CLS50	Coupe	Tustin	CA
12 WDDDJ7CB9BA170249	16995	110615	2011	Mercedes-Benz	CLS-Class	GT	Coupe	Bentonville	AR
13 ZAMGJ45A58B0037556	27999	60122	2008	Maserati	GranTurismo	SE	Crossover	Portland	OR
14 2G1WD5EM3A1146446	0	111221	2010	Chevrolet	Impala	Work Truck	Pickup	Somerset	KY
15 JTJB120480150436	17995	107146	2008	Lexus	GX	50i	SUV	Chantilly	VA
16 1GCGWGAFT1134969	30799	33015	2020	Chevrolet	Express Cargo	F-350 Super Duty	Pickup	Burnsville	MN
17 3C6URSF12HG558595	45999	70954	2017	RAM	Ram 2500 Pickup	SE	Sedan	Leonardtown	MD
18 4JGBF71E78A327314	0	143459	2008	Mercedes-Benz	GL-Class	50i	Hatchback	Chicago	IL
19 1GCPTE12L1136522	47848	7556	2010	Chevrolet	Colorado	XL	Minivan	Pleasanton	CA
20 WP1AD2A28ELA78213	40988	69610	2014	Porsche	Cayenne	F-350 Super Duty	Pickup	Wilmette	NC
21 1C4SDIG12I366348	55595	28904	2018	Dodge	Durango	SE	Crossover	Englewood	NJ
22 1GCGWGAFT11379684	35290	12589	2020	Chevrolet	Express Cargo	50i	Pickup	Fort Wayne	IN
23 MAJ3P1TE8JC240389	15996	31371	2018	Ford	Ecosport	SE	Crossover	Naperville	IL
24 1GCCS9F95C8167274	16888	35001	2012	Chevrolet	Colorado	Work Truck	Pickup	Winslow	AZ
25 5UXU2C54KLN65802	72988	16578	2019	BMW	X5	50i	SUV	Bloomfield	MI
26 1FTFWV3BR3AE44859	20995	168460	2010	Ford	F-350 Super Duty	XL	Pickup	Paterson	NJ
27 WDDUG8FB1GA252600	39999	95746	2016	Mercedes-Benz	S-Class	SE	Crossover	Streetsboro	OH
28 1FMFK16538LA69736	10500	141736	2008	Ford	Expedition	5550	Sedan	Hibbing	MN
29 1FDXE4FS2GDC48510	60000	26114	2016	Ford	E-Series Cutaway	XLT	SUV	Ennis	TX
30 3C6URSF12BG361689	36951	91682	2016	RAM	Ram 2500 Pickup	Base	Cutaway	Potsdam	NY
31 3C7WR9A116G340505	36997	0	2016	RAM	Ram 3500 Chassis Cab	Tradesman	Tradesman	Salina	KS
32 JTJB120X860117128	0	218983	2006	Lexus	GX	xDrive50i	SUV	Cleveland	OH
33 5UXKR6C51J0U15224	47000	51530	2018	BMW	X5	Work	Pickup	Warren Robins	GA
34 1GCHK24G97E190586	0	53230	2007	Chevrolet	Silverado Classic 2500HD	430 Convertible	Pickup	Dubuque	IA
35 JTHFN48Y569000270	26446	82885	2006	Lexus	SC	RS 5	Convertible	Durham	NC
36 WU1ACRAFHSDN990659	31995	99974	2013	Audi	RS 5	Base	Convertible	Lake Bluff	IL

-I will check the price. If Price is blank, then replace with the 0. (Later will plan to neglect all 0

Values)

Step 4:

Data Field: Price

Category Name: Replace Price with \$

Pre-Cleaning:

The screenshot shows a Microsoft Excel spreadsheet titled "UsedCar.csv". The "Price" column is selected, and the "Format Cells" dialog box is open, specifically the "Currency" tab. The "Symbol" dropdown is set to "\$", and the "Decimal places" dropdown is set to 0. The preview area shows the value "\$1234" with a red outline. The main spreadsheet table has columns: VIN, Price, Miles, Year, Make, Model, Trim, Body_type, City, and State. The data includes various vehicle models like Ford Expedition, Ram 2500 Pickup, and Chevrolet Silverado, along with their respective trim levels, body types, cities, and states.

VIN	Price	Miles	Year	Make	Model	Trim	Body_type	City	State
1FMFK16538LA69736	10500	141736	2008	Ford	Expedition				
1F0XE6FS2GD048510	60000	26114	2016	Ford	E-Series Cutaway				
3C6UR5UJ8G361689	36951	91682	2016	RAM	Ram 2500 Pickup	Tradesman	Pickup	Potsdam	NY
3C7WR9A1J1G340505	36997	0	2016	RAM	Ram 3500 Chassis Cab	Tradesman	Chassis Cab	Salina	KS
32JTB120X86011728	0	21898	2006	Lexus	GX		SUV	Cleveland	OH
33SUXK86C510U15224	47000	51530	2018	BMW	xDrive50i		SUV	Warner Robins	GA
341GCHK24G97E190586	0	53230	2007	Chevrolet	Silverado Classic 2500HD	Work	Pickup	Dubuque	IA
35JTHFN48Y56900270	26446	82885	2006	Lexus	SC		430 Convertible	Durham	NC
36WUJAC6AFH5N999659	31995	99974	2013	Audi	RS 5	Base	Convertible	Jake Bluff	IL

Post- Cleaning:

	VIN	Price	Miles	Year	Make	Model	Trim	Body_type	City	State
2	1GCGWGF3FJ1284719	\$20,995	115,879	2015	Chevrolet	Express Cargo	Work Van	Cargo Van	Ellicott City	MD
3	WBV72BC59V887514	\$27,921	7339	2018	BMW	i3	s	Hatchback	Pompano Beach	FL
4	ML324F2J1HF10324	\$11,055	39,798	2018	Mitsubishi	Mirage G4	SE	Sedan	West Bend	WI
5	1GCGPFLK1Z191189	\$52,995	28,074	2019	Chevrolet	Colorado	ZR2	Pickup	Payson	UT
6	1GCGPFLK1Z191183	\$50,995	40,468	2019	Pontiac	G8	Base	Coupe	St. Peters	MO
7	WPW12AC54V1600763	\$0	54,276	2015	BMW	i8	Giga World	Hatchback	Kernersville	NC
8	3GNGB8D8XAS598032	\$6,500	74,774	2010	Chevrolet	HHR	LT	Mini Mpv	Grand Blanc	MI
9	3VMCYCOCB9U040057	\$10,974	9,589	2020	BMW	X6 M	Base	SUV	Miami	FL
10	1GCGSC1E151195013	\$21,302	131,286	2016	Chevrolet	Colorado	LT	Pickup	Jefferson City	MO
11	WDDEJ1X08A014550	\$0	42,000	2008	Mercedes-Benz	CL-Class	CL550	Coupe	Tustin	CA
12	WDODD7C9B8A170249	\$16,995	110,615	2011	Mercedes-Benz	CLS-Class	CL550	Coupe	Bentonville	AR
13	ZAMK02000168000556	\$27,799	60,122	2012	Mercedes-Benz	GranTurismo	Base	Coupe	Portland	OR
14	1GDN2EN1A1400066	\$0	59,101	2010	Chevrolet	Impala	Police Sedan	Sedan	West Chester	NY
15	1JRT2X0400150436	\$11,995	107,146	2008	Lexus	GS	470 SUV	Coupe	Chantilly	VA
16	1GCGWGF71134969	\$30,799	330,015	2012	Chevrolet	Express Cargo	Work Van	Cargo Van	Burnsville	MN
17	3CGRURF1ZHG58595	\$45,995	70,954	2017	RAM	Laramie	Pickup	Pickup	Leonardtown	MD
18	4JG871ET8AZ327314	\$0	143,459	2008	Mercedes-Benz	GL-Class	GL450	SUV	Chicago	IL
19	1GCGPTE1Z1136522	\$47,841	7556	2010	Chevrolet	Colorado	ZR2	Pickup	Pleasanton	CA
20	WPADAD2A28ELA78213	\$40,988	69,610	2014	Porsche	Cayenne	GTS	SUV	Wilmington	NC
21	1LCM828001000000	\$55,595	280,710	2010	Chevrolet	Durango	SRT	SUV	Englewood	NJ
22	1GCGWGF71137984	\$37,995	131,580	2010	Chevrolet	Express Cargo	Work Van	Cargo Van	Fort Wayne	IN
23	MA13P1TE1C140389	\$11,995	31,171	2018	Ford	Escape	SE	Crossover	Waukesha	WI
24	1GCSBF95C8167274	\$16,888	35,001	2012	Chevrolet	Colorado	Work Truck	Pickup	Winslow	AZ
25	SUXJU2C4KLH65802	\$72,995	165,78	2019	BMW	X5	50i	SUV	Bloomfield	MI
26	1FTWW3B83AC444858	\$30,995	168,460	2010	Ford	F-350 Super Duty	XL	Pickup	Paterson	NJ
27	WDUDG8F816A252608	\$30,995	95,746	2016	Mercedes-Benz	S-Class	SS550	Sedan	Streetsboro	OH
28	LFMPK16538LA69736	\$10,500	141,736	2008	Ford	Expedition	XLT	SUV	Hibbing	MN
29	1FTRU2C4K0361689	\$64,995	26,170	2011	Ford	E-Series Cutaway	Base	Cutaway	Ennis	TX
30	1GCGWGF816A2561689	\$36,951	91,882	2010	RAM	Ram 2500	Pickup	Pickup	South Jordan	UT
31	1CTHW041LG340505	\$36,997	0	2016	RAM	Ram 3500 Chassis Cab	Tradesman	Chassis Cab	Salina	KS
32	1JRT2X0800117128	\$0	218,983	2006	Lexus	GS	470 SUV	Pickup	Cleveland	OH
33	SUXKRNCS1J0U15224	\$47,000	51,530	2018	BMW	X5	xDrive50i	SUV	Warren Robins	GA
34	1GCKHZ4G97E190586	\$0	53,230	2007	Chevrolet	Silverado Classic 2500HD	Work	Pickup	Dubuque	IA
35	1THFN4BY569000270	\$26,446	82,885	2006	BMW	SC	430 Convertible	Convertible	Durham	NC
36	1M1L1AC1AH1YD9N9W59	\$31,995	99,974	2013	Audi	RS 5	Base	Convertible	Jake Bluff	IL

-Generally, we use \$ to represent price. Converting column to Currency with No Decimal Value.

Step 5:

Data Field: Trim, Body_Type

Category Name: Replace Blanks with Unknown

Pre-Cleaning:

The screenshot shows a Microsoft Excel spreadsheet titled "UsedCar.csv". The "Trim" column is currently populated with various car models like "Express C", "i3", "Mirage G4", etc. A dropdown menu is open over the cell containing "GranTurismo", showing a list of other trim options such as "Z4X S", "Z4X SE", "Z4X ST", "ZDX", "ZDX Comfort", "ZDX S", "ZDX SES", and "(Blanks)". The user has selected "Base" from this list.

The screenshot shows a Microsoft Excel spreadsheet titled "UsedCar.csv". The "Trim" column is currently populated with various car models like "Express C", "i3", "Mirage G4", etc. A dropdown menu is open over the cell containing "RS 5", showing a list of other trim options such as "Stripped Chassis Motorhome", "Accent", "F-350 Super Duty", "F-350 Super Duty", "Volt", "Impreza", "Frontier", "SS", "Truck", "Express", "Pacific", "Savana", "Low Cab Forward", "Model 3", "C-Class", "A6", "Transit", "All-New Wrangler Unlimited", "CTS Sport Sedan", "Ram 3500 Chassis Cab", and "(Blanks)". The user has selected "Base" from this list.

Post-Cleaning:

The screenshot shows a Microsoft Excel spreadsheet titled 'UsedCar'. A dropdown menu is open over the 'Trim' column, listing various vehicle models and their trim levels. A 'Find and Replace' dialog box is overlaid on the screen, with 'Unknown' selected in both the 'Find what:' and 'Replace with:' fields. The status bar at the bottom indicates that 2309 of 884493 records were found.

-Check the values in Trim field. If Trim is null, then replace with ‘Unknown’.

Pre-Cleaning:

The screenshot shows a Microsoft Excel spreadsheet titled 'UsedCar'. A dropdown menu is open over the 'Trim' column, listing various vehicle models and their trim levels. A 'Find and Replace' dialog box is overlaid on the screen, with the formula '#ISBLANK(Trim)' selected in the 'Find what:' field and 'Unknown' selected in the 'Replace with:' field. The status bar at the bottom indicates that 2309 of 884493 records were found.

Post-Cleaning:

The screenshot shows a Microsoft Excel spreadsheet titled 'UsedCar.xlsx' with data across multiple columns (A-J). Column A contains VIN numbers, and Column B contains Price. Column H contains the 'Body_type' field, which is being cleaned. A search dialog box is overlaid on the spreadsheet, specifically targeting the 'Body_type' column. The 'Find what:' field is set to 'Unknown' and the 'Replace with:' field is also set to 'Unknown'. Other buttons in the dialog include 'Replace All', 'Replace', 'Find All', and 'End Next'. The status bar at the bottom of the screen indicates '22:20 PM' and '89°F Sunny'.

VIN	Price	Miles	Year	Make	Model	Trim	Body_type	Check	Body_type	City
19653	\$19,990	83913	2009	Ford	LCF 450	Regular Cab DRW 2WD	TRUE	Unknown	Norco	
20698	\$64,990	12570	2009	Ford	Stripped Chassis Motorhome	Unknown	TRUE	Unknown	Murrysville	
42074	\$24,577	52121	2020	Chrysler	Voyager	LX	TRUE	Unknown	Ontario	
42075	\$35,877	16039	2020	RAM	Class	SLT	TRUE	Unknown	Orange	
42076	\$30,995	50	2015	Ford	Truck	XLT	TRUE	Unknown	Buda	
42077	\$11,118	80372	2012	Volkswagen	Rouanet	SE	TRUE	Unknown	Moon Township	
42078	\$21,999	46791	2009	GMC	Savana Cutaway	Work Van	TRUE	Unknown	Leominster	
42079	\$9,988	121579	2013	Chevrolet	Impala	SEDAN	TRUE	Unknown	Puyallup	
42080	\$44,900	105232	2007	Ford	Supreme	Unknown	TRUE	Unknown	Raleigh	
42081	\$34,394	22035	2020	RAM	Classic	SLT	TRUE	Unknown	Woodbury	
42082	\$34,995	29425	2008	Dodge	Truck	SLT	TRUE	Unknown	Elizabethtown	
42083	\$29,565	89083	2017	Mitsubishi	Truck	Commercial	TRUE	Unknown	Rocky Mount	
42084	\$9,500	188200	2007	Ford	E-450 Super Duty	Super	TRUE	Unknown	Sparta	
42102	\$50	45800	2008	Ford	Super Duty	SuperCab	TRUE	Unknown	Fort Wayne	
42103	\$1,900	8102	2012	Ford	Super Duty	SuperCab	TRUE	Unknown	Whitner	
42104	\$50	37171	2019	Isuzu	1	Find	TRUE	Unknown	Fort Worth	
42105	\$10,950	161772	2007	Ford	1	Find	TRUE	Unknown	Fort Worth	
42106	\$35,950	107898	2008	GMC	1	Find what:	Unknown	Unknown	Roanoke	
42107	\$54,995	13716	2020	Isuzu	1	Find what:	Unknown	Unknown	Franklin	
42108	\$0	23805	2009	Ford	1	Find what:	Unknown	Unknown	Acworth	
42109	\$21,900	8102	2013	Ford	1	Find what:	Unknown	Unknown	Manchester	
42110	\$53,000	19213	2020	RAM	1	Find what:	Unknown	Unknown	Taylorville	
42111	\$10,950	15250	2013	Ford	1	Find what:	Unknown	Unknown	Taylorsville	
42112	\$8,000	86872	2013	Chevrolet	1	Find what:	Unknown	Unknown	Harrison	
42113	\$36,600	30220	2020	RAM	1	Find what:	Unknown	Unknown	Gothenburg	
42114	\$16,995	89248	2017	Nissan	1	Find what:	Unknown	Unknown	Buford	
42115	\$47,595	1626	2019	Mitsubishi	1	Find what:	Unknown	Unknown	Marshall	
42116	\$9,525	123850	2016	Chevrolet	Impala Limited	SEDAN	TRUE	Unknown	Dallas	
42117	\$42,800	27800	2013	Chevrolet	Express	3LT	TRUE	Unknown	North Aurora	
42118	\$8,997	85431	2013	Chevrolet	Impala	LT	TRUE	Unknown	Temple	
42119	\$0	19343	2014	Mitsubishi	Lancer	Unknown	TRUE	Unknown	Collinsville	
42120	\$35,000	65684	2016	Ford	Truck	XL	TRUE	Unknown	Heath	
42121	\$32,590	21229	2020	RAM	Classic	SLT	TRUE	Unknown	Michigan City	
42122	\$35,990	23161	2020	RAM	Classic	Crew	TRUE	Unknown	Burlington	
42123	\$0	13334	2007	Isuzu	Truck	Unknown	TRUE	Unknown	Tama	
42124	\$32,998	135688	2016	RAM	Tradrman	TradeMaster	TRUE	Unknown		

-For body-type field, first check the filed if its blank or not. If it is blank replace with 'Unknown'.

Step 6:

Category Name: Formatting Data

Pre-Cleaning:

This screenshot shows the 'UsedCar' sheet in Microsoft Excel. The data is organized into columns:

- Column A:** VIN numbers.
- Column B:** Price.
- Column C:** Miles.
- Column D:** Year.
- Column E:** Make.
- Column F:** Model.
- Column G:** Trim.
- Column H:** Body_Type.
- Column I:** City.

The 'UsedCar' tab is highlighted at the bottom left of the window.

This screenshot shows the 'UsedCar' sheet in Microsoft Excel with the 'Clear Formats' context menu open over the first few rows of data. The menu includes options like Clear All, Clear Formats, Clear Contents, Clear Comments and Notes, and Clear Hyperlinks. The 'UsedCar' tab is selected at the bottom left.

Post-Cleaning:

VIN	Price	Miles	Year	Make	Model	Trim	Body_type	City	State
2 1GCWGFC3F1284719	20998	115879	2015	Chevrolet	Express Cargo	Work Van	Cargo Van	Elliott City	MD
3 WBV7ZBC59JVB87514	27921	7339	2018	BMW	i3	s	Hatchback	Pompano Beach	FL
4 ML324F12H1F0325	11055	39798	2018	Mitsubishi	Mirage G4	SE	Sedan	West Bend	WI
5 1GCPTEE15K1291189	52997	28568	2019	Chevrolet	Colorado	ZR2	Pickup	Layton	UT
6 3GNBABD0BXAS598032	6500	74274	2010	Chevrolet	HHR	LT	Mini Mpv	Grand Blanc	MI
7 5YMCNOCY08L04057	105974	9589	2020	BMW	X6 M	Base	SUV	Miami	FL
8 1GGC5CE15G1195013	23024	131286	2016	Chevrolet	Colorado	LT	Pickup	Jefferson City	MO
9 WDDDF7CB9BA170249	16995	110615	2011	Mercedes-Benz	CLS-Class	CL550	Coupe	Bentonville	AR
10 ZAMGJ45A580037556	27999	60122	2008	Maserati	GranTurismo	Base	Coupe	Portland	OR
11 JTBT20X480150438	17995	107146	2008	Lexus	GX	470 SUV	SUV	Chantilly	VA
12 1GCWGAFP7L1134969	30799	33015	2020	Chevrolet	Express Cargo	Work Van	Cargo Van	Burnsville	MN
13 3C6UR5F12H1G55895	45999	70954	2017	RAM	Ram 2500 Pickup	Laramie	Pickup	Leonardtown	MD
14 1GCPTEE12L1136522	47848	7556	2020	Chevrolet	Colorado	ZR2	Pickup	Pleasanton	CA
15 WP1AD2A28ELA78213	40988	69610	2014	Porsche	Cayenne	GTS	SUV	Wilmington	NC
16 1C45D1GJ2JC366348	55595	28904	2018	Dodge	Durango	SRT	SUV	Englewood	NJ
17 1GCWGAFP7L1179684	35290	12589	2020	Chevrolet	Express Cargo	Work Van	Cargo Van	Fort Wayne	IN
18 MAJ3P1TEB1C20389	15996	31371	2018	Ford	Ecosport	SE	Crossover	Naperville	IL
19 1GCCSBP95C8167274	16888	35001	2012	Chevrolet	Colorado	Work Truck	Pickup	Winslow	AZ
20 SUKXU2C54KLNG5802	72988	16578	2019	BMW	X5	50i	SUV	Bloomfield	MI
21 1FTWV3BR3AE444858	20995	168460	2010	Ford	F-350 Super Duty	XL	Pickup	Paterson	NJ
22 WDDUGFB1GA252600	39999	95746	2016	Mercedes-Benz	S-Class	5550	Sedan	Streetsboro	OH
23 1MF1K6538LA69736	10500	141736	2008	Ford	Expedition	XL7	SUV	Hibbing	MN
24 1FDXE4F52GD48510	60000	26114	2016	Ford	E-Series Cutaway	Base	Cutaway	Ennis	TX
25 3C6UR5CJBG361689	36951	91682	2016	RAM	Ram 2500 Pickup	Tradesman	Pickup	Potsdam	NY
26 SUKXR6C51J0U15224	47000	51530	2018	BMW	X5	xDrive50i	SUV	Warner Robins	GA
27 JTHFN48Y56900270	26446	82885	2006	Lexus	SC	430 Convertible	Convertible	Durham	NC
28 WIJAC6AFHSDN900659	31995	99974	2013	Audi	RS 5	Base	Convertible	Lake Bluff	IL
29 WDDA17F36M000101	298300	5534	2006	Mercedes-Benz	SLR McLaren	Base	Coupe	Great Neck	NY
30 WBY1Z4C53GV508046	24000	39145	2016	BMW	i3	Mega World	Hatchback	Charlotte	NC
31 ZHWUT4Z5LLA12830	175950	958	2020	Lamborghini	Huracan	EVO	Convertible	Denver	CO
32 5YMKT6C561DX21085	69999	34836	2018	BMW	X5 M	Base	SUV	Lincolnwood	IL
33 WBAYEBCSFD781585	23736	105846	2015	BMW	7 Series	750Li	Sedan	Bentonville	AR
34 1GCWGFCATE1210688	13995	142351	2014	Chevrolet	Express Cargo	Work Van	Cargo Van	Canal Fulton	OH
35 MAJ3P1TE6U2C23625	16999	42875	2018	Ford	Ecosport	SE	Crossover	Goose Creek	SC
36 MI32ASH11KH079298	9995	57876	2019	Mitsubishi	Mirage	IF	Hatchback	Herkimer	NY

In these steps, all data values are formatted. As shown in the above image, text data display on the left side and numeric data display on the right side.

Step 8:

Category Name: Spell Checking

Pre-Cleaning:

The screenshot shows the 'UsedCar.csv' spreadsheet in Microsoft Excel. The 'Proofing' tab is active in the ribbon. A search bar at the top says 'Search (Alt+Q)'. The spreadsheet has columns for VIN, Price, Miles, Year, Make, Model, Trim, Body_type, City, State, and other vehicle specifications. The 'Fis' row is highlighted in yellow.

The screenshot shows the 'UsedCar.csv' spreadsheet in Microsoft Excel. The 'Proofing' tab is active in the ribbon. A search bar at the top says 'Search (Alt+Q)'. A spell check dialog box is open over the spreadsheet, showing suggestions for the word 'Fisher'. The suggestions include Fisher, Fiske, Fish, and Rfish. The 'Ignore Once' button is highlighted.

Post-Cleaning:

The screenshot shows a Microsoft Excel spreadsheet titled "UsedCar.csv". The data consists of vehicle information such as VIN, Price, Miles, Year, Make, Model, Trim, Body_type, City, and State. A spell check dialog box from Microsoft Excel is overlaid on the screen, stating "Spell check complete. You're good to go!" with an "OK" button.

VIN	Price	Miles	Year	Make	Model	Trim	Body_type	City	State
1	20098	115879	2015	Chevrolet	Express Cargo	Work Van	Cargo Van	Elliott City	MD
2	29791	7339	2018	BMW	Mirage G4	i3	Hatchback	Pompano Beach	FL
3	11055	39798	2018	Mitsubishi	Ram 2500 Pickup	ZR2	Sedan	West Bend	WI
4	52997	28568	2019	Chevrolet	Colorado	A4	Pickup	Laton	UT
5	6500	74277	2010	Chevrolet	HHR	LT	Mini Mpv	Grand Blanc	MI
6	105974	9589	2020	BMW	X6 M	Base	SUV	Miami	FL
7	13024	119503	2016	Chevrolet	Colorado	LT	Pirkun	Jefferson City	MO
8	16995	110615	2011	Benz	CLS-Class				AR
9	27999	33015	2008	Maserati	GranTurismo				OR
10	17995	107146	2008	Lexus	GX				VA
11	30799	33015	2010	Chevrolet	Express Cargo				MN
12	45999	70594	2017	RAM	Ram 2500 Pickup				MD
13	47488	7556	2020	Chevrolet	Colorado				CA
14	40988	69610	2014	Porsche	Cayenne				NC
15	55959	28904	2018	Dodge	Durango				NJ
16	35290	12589	2020	Chevrolet	Express Cargo				IN
17	15996	31371	2018	Ford	Ecosport				IL
18	16888	35001	2012	Chevrolet	Colorado				AZ
19	72988	16578	2019	BMW	X5				MI
20	20995	168460	2010	Ford	F-350 Super Duty				NJ
21	30000	154176	2016	Benz	S-Class				OH
22	30000	154176	2016	Mercedes-Benz	Expedition				MN
23	10500	141736	2008	Ford	E-Series Cutaway				TX
24	60000	26114	2016	Ford	Ram 2500 Pickup				TX
25	36951	91682	2016	RAM	Tradesman				NY
26	47000	107146	2008	BMW	Pickup				WA
27	26446	80889	2006	Audi	xDrive50i				Robins
28	31995	99074	2013	Audi	430 Convertible				GA
29	298300	5534	2006	Mercedes-Benz	SLR McLaren				NC
30	24000	39145	2016	BMW	i3				Lake Bluff
31	175950	958	2020	Lamborghini	Mega World				NY
32	69999	34836	2018	BMW	EVO				Charlette
33	23736	105846	2015	BMW	X5 M				CO
34	13995	142351	2014	Chevrolet	Express Cargo				Lincolnwood
35	16999	42875	2018	Ford	Ecosport				Bentonville
36	9995	57876	2019	Mitsubishi	Mirage				AR

Pre-Cleaning:

The screenshot shows a Microsoft Excel spreadsheet titled "UsedCar.csv". The data consists of vehicle information such as VIN, Price, Miles, Year, Make, Model, Trim, Body_type, City, and State. A spell check dialog box from Microsoft Excel is overlaid on the screen, showing suggestions for "GranTurismo" including "GranTurismo", "Granter", "Granter's", "Grantur", "Grantur's", and "Grantis". The dictionary language is set to "English (United States)".

VIN	Price	Miles	Year	Make	Model	Trim	Body_type	City	State
1	20098	115879	2015	Chevrolet	Express Cargo	Work Van	Cargo Van	Elliott City	MD
2	29791	7339	2018	BMW	Mirage G4	i3	Hatchback	Pompano Beach	FL
3	11055	39798	2018	Mitsubishi	Ram 2500 Pickup	ZR2	Sedan	West Bend	WI
4	52997	28568	2019	Chevrolet	Colorado	A4	Pickup	Laton	UT
5	6500	74277	2010	Chevrolet	HHR	LT	Mini Mpv	Grand Blanc	MI
6	105974	9589	2020	BMW	X6 M	Base	SUV	Miami	FL
7	13024	119503	2016	Chevrolet	Colorado	LT	Pirkun	Jefferson City	MO
8	16995	110615	2011	Benz	CLS-Class				AR
9	27999	33015	2008	Maserati	GranTurismo				OR
10	17995	107146	2008	Lexus	GX				VA
11	30799	33015	2010	Chevrolet	Express Cargo				MN
12	45999	70594	2017	RAM	Ram 2500 Pickup				MD
13	47488	7556	2020	Chevrolet	Colorado				CA
14	40988	69610	2014	Porsche	Cayenne				NC
15	55959	28904	2018	Dodge	Durango				NJ
16	35290	12589	2020	Chevrolet	Express Cargo				IN
17	15996	31371	2018	Ford	Ecosport				IL
18	16888	35001	2012	Chevrolet	Colorado				AZ
19	72988	16578	2019	BMW	X5				MI
20	20995	168460	2010	Ford	F-350 Super Duty				NJ
21	30000	154176	2016	Benz	S-Class				OH
22	30000	154176	2016	Mercedes-Benz	Expedition				MN
23	10500	141736	2008	Ford	E-Series Cutaway				TX
24	60000	26114	2016	Ford	Ram 2500 Pickup				TX
25	36951	91682	2016	RAM	Tradesman				NY
26	47000	107146	2008	BMW	Pickup				WA
27	26446	80889	2006	Audi	xDrive50i				Robins
28	31995	99074	2013	Audi	430 Convertible				GA
29	298300	5534	2006	Mercedes-Benz	SLR McLaren				NC
30	24000	39145	2016	BMW	i3				Lake Bluff
31	175950	958	2020	Lamborghini	Mega World				NY
32	69999	34836	2018	BMW	EVO				Charlette
33	23736	105846	2015	BMW	X5 M				CO
34	13995	142351	2014	Chevrolet	Express Cargo				Lincolnwood
35	16999	42875	2018	Ford	Ecosport				Bentonville
36	9995	57876	2019	Mitsubishi	Mirage				AR

Post-Cleaning:

The screenshot shows a Microsoft Excel spreadsheet titled "UsedCar.csv". The spreadsheet contains data from columns A to M, listing various vehicles with their details such as ID, year, make, model, and location. A spelling check dialog box is overlaid on the screen, stating "Spell check complete. You're good to go!" with an "OK" button.

A	B	C	D	E	F	G	H	I	J	K	L	M
477843	2GNAXEV161352343	21929	29184	2018 Chevrolet	Equinox	LT	SUV	Lockport	NY			
477844	2GNAXEV0K16186124	23888	26934	2019 Chevrolet	Equinox	LT	SUV	Blue Springs	MO			
477845	2GNAXEV916245699	20886	28891	2018 Chevrolet	Equinox	LT	SUV	Tilton	NH			
477846	3GNAXEV7KS540416	19645	24033	2019 Chevrolet	Equinox	LS	SUV	Brunswick	GA			
477847	3GNAXEV16L5006553	25991	12207	2020 Chevrolet	Equinox	LT	SUV	Benton Harbor	MI			
477848	3GNAXEV5M105736	25998	4827	2021 Chevrolet	Equinox	LT	SUV	Atlanta	GA			
477849	2GNAXEV1K16198914	20982	39960	2019 Chevrolet	Equinox	LT	SUV	Scottsboro	AL			
477850	3GNAXEV1S1624886	22995	31366	2018 Chevrolet	Equinox	1LT	SUV	Noblesville	IN			
477851	3GNAXREV7JS564129	19714	32535	2018 Chevrolet	Equinox	LS	SUV	Red Wing	MN			
477852	2GNAXEV0J6334691	20995	18650	2018 Chevrolet	Equinox	LT	SUV	Scarsdale	NY			
477853	2GNAXEV916291033	18984	75507	2018 Chevrolet	Equinox	LT	SUV	Grand Rapids	MI			
477854	2GNAXEV6L126498	27995	25334	2020 Chevrolet	Equinox	LT	SUV	Holdredge	NE			
477855	3GNAXEV41569810	21981	35004	2018 Chevrolet	Equinox	LT	SUV	Sault Sainte Marie	MI			
477856	3GNAXREV6S16567260	19933	25667	2018 Chevrolet	Equinox	LT	SUV	Stafford	VA			
477857	3GNAXEV915616106	22995	26941	2018 Chevrolet	Equinox	LT	SUV	Daleville	IN			
477858	2GNAXEV916148213	21988	49868	2018 Chevrolet	Equinox	LT	SUV	Lowell	MA			
477859	3GNAXMEV0I131149	23500	31538	2018 Chevrolet	Equinox	LT	SUV	Westbury	NY			
477860	2GNAXEV161228464	18988	23110	2018 Chevrolet	Equinox	LT	SUV	Paw Paw	MI			
477861	2GNAXEV916260734	20877	28727	2019 Chevrolet	Equinox	LT	SUV	Conway	AR			
477862	2GNAXEV7J6285621	22142	35226	2018 Chevrolet	Equinox	LT	SUV	Nelliston	NY			
477863	2GNAXEVX616126439	23471	2512	2018 Chevrolet	Equinox	LT	SUV	Sunrise	FL			
477864	2GNAXEV7J6248878	24995	19378	2018 Chevrolet	Equinox	LT	SUV	Manheim	PA			
477865	2GNAXEV916290421	17945	45679	2018 Chevrolet	Equinox	LS	SUV	Grayslake	IL			
477866	2GNAXEV8J6304387	20211	27406	2018 Chevrolet	Equinox	LS	SUV	Texarkana	TX			
477867	2GNAXEV216156292	20984	49000	2018 Chevrolet	Equinox	LT	SUV	Orleans	IN			
477868	3GNAXEV615150705	22179	24622	2018 Chevrolet	Equinox	LT	SUV	Fenton	MO			
477869	3GNAXEV315161560	21495	18230	2018 Chevrolet	Equinox	LT	SUV	Fridley	MN			
477870	2GNAXEV12U177678	25599	21844	2020 Chevrolet	Equinox	LT	SUV	Limerick	PA			
477871	3GNAXEV15578683	26988	37869	2018 Chevrolet	Equinox	LT	SUV	Santa Fe	NM			
477872	3GNAXEV6LL138867	22199	47264	2020 Chevrolet	Equinox	1LT	SUV	Lombard	IL			
477873	3GNAXEV41124889	19766	64322	2018 Chevrolet	Equinox	LT	SUV	Wood River	IL			
477874	3GNAXEV4K5561372	19852	46689	2019 Chevrolet	Equinox	LS	SUV	Plainwell	MI			
477875	3GNAXEV415647643	19790	23444	2018 Chevrolet	Equinox	LT	SUV	Bloomington	IL			
477876	3GNAXEV41398733	19350	34392	2018 Chevrolet	Equinox	LS	SUV	Fremont	NE			
477877	2GNAXMEV0I16115032	21995	68282	2018 Chevrolet	Equinox	Premier	SUV	Irving	TX			
477878	2GNAXEV8J6200609	22395	30926	2018 Chevrolet	Equinox	LT	SUV	Franklin	TN			

-In these steps, I am checking the spelling errors. If there is any mistake in spelling, then these are very important step to find out the mistake.

Step 9:

Category Name: Format Cells

Pre-Cleaning:

The screenshot shows a Microsoft Excel spreadsheet titled "UsedCar.csv". The data consists of approximately 36 rows of vehicle information across columns A through M. The columns include VIN, Price, Miles, Year, Trim, Body Type, City, State, and other descriptive fields. A context menu is open over the data, with the "Format Cells..." option highlighted under the "Cells" tab. The "Format Cells" dialog box is visible, showing the "Number" category and "General" format.

This screenshot shows the same Excel spreadsheet and data structure as the previous one. However, the "Format Cells" dialog box is now open, and the "Currency" category is selected. The "General" format is still chosen within the currency settings. The main Excel window remains the same, displaying the vehicle data table.

Post-Cleaning:

VIN	Price	Miles	Year	Make	Model	Trim	Body_type	City	State
16CWGFCF3F1284719	\$20,998	115879	2015	Chevrolet	Express Cargo	Work Van	Cargo Van	Ellicott City	MD
WBVY7B8C59JVB87514	\$27,921	7339	2018	BMW	i3	s	Hatchback	Pompano Beach	FL
M1L3ZT4F12JHF10325	\$11,055	39798	2018	Mitsubishi	Mirage G4	SE	Sedan	West Bend	WI
16CPCTEE15K1291189	\$52,997	28568	2019	Chevrolet	Colorado	ZR2	Pickup	Layton	UT
3GNBABDXA5S98032	\$6,500	74274	2010	Chevrolet	HHR	LT	Mini Mpv	Grand Blanc	MI
5VMCYOCOB9J04057	\$105,974	9589	2020	BMW	X6 M	Base	SUV	Miami	FL
16CGSC1E1G1195013	\$23,024	131286	2016	Chevrolet	Colorado	LT	Pickup	Jefferson City	MO
WDDDJ7C98A170249	\$16,995	110615	2011	Mercedes-Benz	CLS-Class	CLS550	Coupe	Bentonville	AR
ZAMG145A58037556	\$27,995	60122	2008	Maserati	GranTurismo	Base	Coupe	Portland	OR
JTB720X480150436	\$17,995	107146	2008	Lexus	GX	470 SUV	Chantilly	VA	
16CWGFAFP7L1134969	\$30,799	33015	2020	Chevrolet	Express Cargo	Work Van	Cargo Van	Burnsville	MN
3C6URSFJ2HG558595	\$45,999	70954	2017	RAM	Ram 2500 Pickup	Laramie	Pickup	Leonardtown	MD
16CPCTEE1211136522	\$47,848	7556	2020	Chevrolet	Colorado	ZR2	Pickup	Pleasanton	CA
WP1AD2A28ELA78213	\$40,988	69610	2014	Porsche	Cayenne	GTS	SUV	Wilmington	NC
1C45D1G2J1C366348	\$55,595	28904	2018	Dodge	Durango	SRT	SUV	Englewood	NJ
16CWGFAFP7L1179684	\$35,290	12589	2020	Chevrolet	Express Cargo	Work Van	Cargo Van	Fort Wayne	IN
MAJ3P1TE8IC240389	\$15,996	31371	2018	Ford	Ecosport	SE	Crossover	Naperville	IL
16CCSBF95C8167274	\$16,888	35001	2012	Chevrolet	Colorado	Work Truck	Pickup	Winslow	AZ
5UXUJ2C54KLN65802	\$72,988	16578	2019	BMW	X5	50i	SUV	Bloomfield	MI
1FTWW3B83EA44858	\$20,995	168460	2010	Ford	F-350 Super Duty	XL	Pickup	Paterson	NJ
WDDUG8F816A252600	\$39,995	95746	2016	Mercedes-Benz	S-Class	S550	Sedan	Streetsboro	OH
1MPFK16338L469736	\$10,500	141736	2006	Ford	Expedition	XLT	SUV	Hilbing	MN
1FDXF4F52GD48510	\$60,000	26114	2016	Ford	E-Series Cutaway	Base	Cutaway	Ennis	TX
3C6URSCJ8GG361689	\$36,951	91682	2016	RAM	Ram 2500 Pickup	Tradesman	Pickup	Potsdam	NY
5UXKRG6C51J0U15224	\$47,000	51530	2018	BMW	X5	xDrive50i	SUV	Warner Robins	GA
JTHFN48Y569000270	\$26,446	82885	2006	Lexus	SC	430 Convertible	Convertible	Durham	NC
WIAUC6AFH5D0900659	\$31,995	99974	2013	Audi	RS 5	Base	Convertible	Lake Bluff	IL
WDDAJ76F36M001012	\$298,300	5534	2006	Mercedes-Benz	SLR McLaren	Base	Coupe	Great Neck	NY
WBVY124C53GV508046	\$24,000	39145	2016	BMW	i3	Mega World	Hatchback	Charlotte	NC
ZHWU742F5LLA12830	\$175,950	958	2020	Lamborghini	Huracan	EVO	Convertible	Denver	CO
5YMK76C56J0X21085	\$69,995	34836	2018	BMW	X5 M	Base	SUV	Lincolnwood	IL
WBAYEB8CSXP781585	\$23,736	105846	2015	BMW	7 Series	750Li	Sedan	Bentonville	AR
16CWGFCATE71210688	\$13,995	142351	2014	Chevrolet	Express Cargo	Work Van	Cargo Van	Canal Fulton	OH
MAJ3P1TE8IC236325	\$16,995	42875	2012	Ford	Ecosport	SE	Crossover	Goose Creek	SC
M132A5H1KH00928	\$9,995	57876	2019	Mitsubishi	Mirage	1 F	Hatchback	Herkimer	NY

-In these steps, we can format the data in particular field. We represent year and miles data in numeric field only. To assign specific data type to field is very important step in data cleaning process.

Pre-Cleaning:

Average: 848.845597 Count: 462/9648 Sum: 8143824.3

10:14 PM 10/18/2022

Average: 848.845597 Count: 462/9648 Sum: 8143824.3

10:14 PM 10/18/2022

Post-Cleaning:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	VIN	Price	Miles	Year	Make	Model	Trim	Body_type	City	State			
2	1GCGWGFCCF3F1284719	\$20,998	115879	2015	Chevrolet	Express Cargo	Work Van	Cargo Van	Ellicott City	MD			
3	WBV72C59VBB75714	\$27,921	7339	2018	BMW	i3	s	Hatchback	Pompano Beach	FL			
4	ML32F4FJ2JH10325	\$11,055	39798	2018	Mitsubishi	Mirage G4	SE	Sedan	West Bend	WI			
5	1GCPTE15K1291189	\$52,997	28568	2019	Chevrolet	Colorado	ZR2	Pickup	Payton	UT			
6	3GNGABBDVXAS598032	\$6,500	74274	2010	Chevrolet	HHR	LT	Mini Mpv	Grand Blanc	MI			
7	SYMCYOC08L9D04507	\$105,974	9589	2020	BMW	X6 M	Base	SUV	Miami	FL			
8	1GCGSC1E15G1195013	\$23,024	131286	2016	Chevrolet	Colorado	LT	Pickup	Jefferson City	MO			
9	WDDDJ7CB9BA170249	\$16,995	110615	2011	Mercedes-Benz	CLS-Class	CLS550	Coupe	Bentonville	AR			
10	ZAMGI45A5B0037556	\$27,999	60122	2008	Maserati	GranTurismo	Base	Coupe	Portland	OR			
11	JTB720X480150436	\$17,995	107146	2008	Lexus	GX	470	SUV	Chantilly	VA			
12	1GCGWGAFT7L1134969	\$30,799	33015	2020	Chevrolet	Express Cargo	Work Van	Cargo Van	Burnsville	MN			
13	3GCUURSF12H6558595	\$45,999	70954	2017	RAM	Ram 2500 Pickup	Laramie	Pickup	Leonardtown	MD			
14	1GCPTEE12L1136522	\$47,848	7556	2020	Chevrolet	Colorado	ZR2	Pickup	Pleasanton	CA			
15	WP1AD2A28ELA78213	\$40,998	69610	2014	Porsche	Cayenne	GTS	SUV	Wilmington	NC			
16	1C4SDJG1ZJC366348	\$55,595	28904	2018	Dodge	Durango	SRT	SUV	Englewood	NJ			
17	1GCGWGAFT7L1179684	\$35,290	12589	2020	Chevrolet	Express Cargo	Work Van	Cargo Van	Fort Wayne	IN			
18	MAJ3P1TE8C240389	\$15,996	31371	2018	Ford	Ecosport	SE	Crossover	Naperville	IL			
19	1GCCS9F95C8167274	\$16,888	35001	2012	Chevrolet	Colorado	Work Truck	Pickup	Winslow	AZ			
20	SUXU2U2C54KLNL65802	\$72,988	16578	2019	BMW	X5	50i	SUV	Bloomfield	MI			
21	1FTVWW3BR3AE44858	\$20,995	168460	2010	Ford	F-350 Super Duty	XL	Pickup	Paterson	NJ			
22	WDDUG8FB1GA252600	\$39,999	95746	2016	Mercedes-Benz	S-Class	S550	Sedan	Streetsboro	OH			
23	1FMFKL6538LA69736	\$10,500	141736	2008	Ford	Expedition	XLT	SUV	Hibbing	MN			
24	1FDXE4F52GDC48510	\$60,000	26114	2016	Ford	E-Series Cutaway	Base	Cutaway	Ennis	TX			
25	3GCUURSF12H6561689	\$36,951	91682	2016	RAM	Ram 2500 Pickup	Tradesman	Pickup	Potsdam	NY			
26	SUXKRC51J0U15224	\$47,000	51530	2018	BMW	X5	xDrive50i	SUV	Warner Robins	GA			
27	JTHFN48Y569000270	\$26,446	82885	2006	Lexus	SC	430	Convertible	Durham	NC			
28	WUAC6AHF5DN900659	\$31,995	99974	2013	Audi	RS 5	Base	Convertible	Lake Bluff	IL			
29	WDDAJ76F36M001012	\$298,300	5534	2006	Mercedes-Benz	SLR McLaren	Base	Coupe	Great Neck	NY			
30	WBV124CS36VG508046	\$24,000	39145	2016	BMW	i3	Mega World	Hatchback	Charlotte	NC			
31	ZHWUT4Z2F5LA12830	\$175,950	958	2020	Lamborghini	Huracan	EVO	Convertible	Denver	CO			
32	5YMTK6C56D0X1085	\$69,999	34836	2018	BMW	X5 M	Base	SUV	Lincolnwood	IL			
33	WBAYE8C5XF7D781582	\$23,736	105846	2015	BMW	7 Series	750Li	Sedan	Bentonville	AR			
34	1GCGWGFCA7E1210688	\$13,995	142351	2014	Chevrolet	Express Cargo	Work Van	Cargo Van	Canal Fulton	OH			
35	MAJ3P1TE6C23625	\$16,999	42875	2018	Ford	Ecosport	SE	Crossover	Goose Creek	SC			
36	1M132A5H11K400392R	\$9,995	57876	2019	Mitsubishi	Mirage	IE	Hatchback	Herkimer	NY			

-Same as above, we use text to represent make, model, trim, body type, city, and state. Text data type is used to assign text as well as numeric combinations.

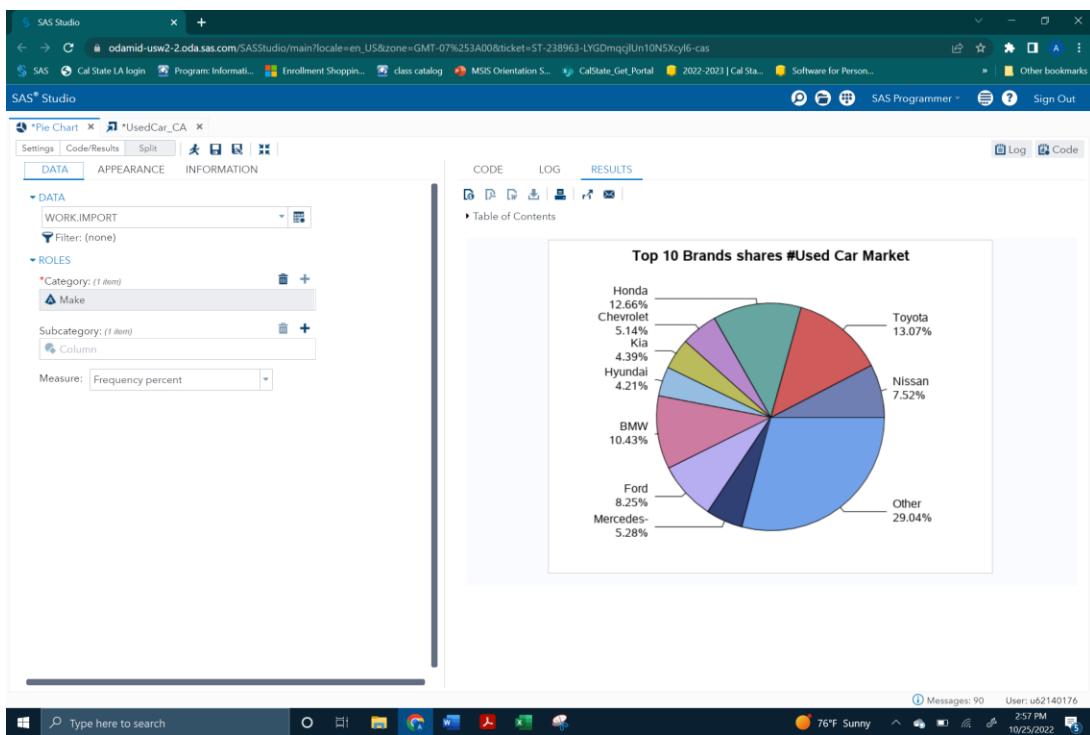
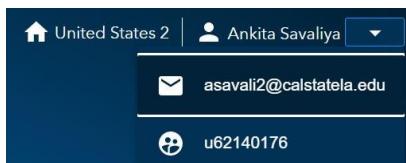
E. Analysis & Visualizations

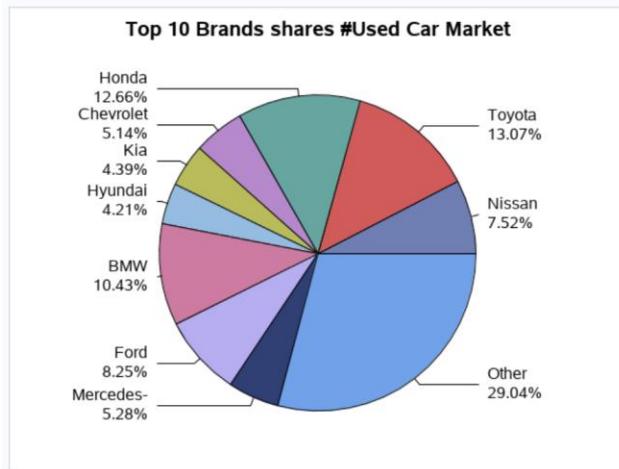
Data visualization is method for analyzing data. Data visualization is a graphical representation of the relationships between data, ideas, information, and concepts that is easy to understand and one of the most important learning strategies.

The main purpose of the analysis is to find out the following questions.

Q- Which top 10 cars maker was in the market for sales in CA between 2019 to 2021?

For reference, my login detail for SAS is:



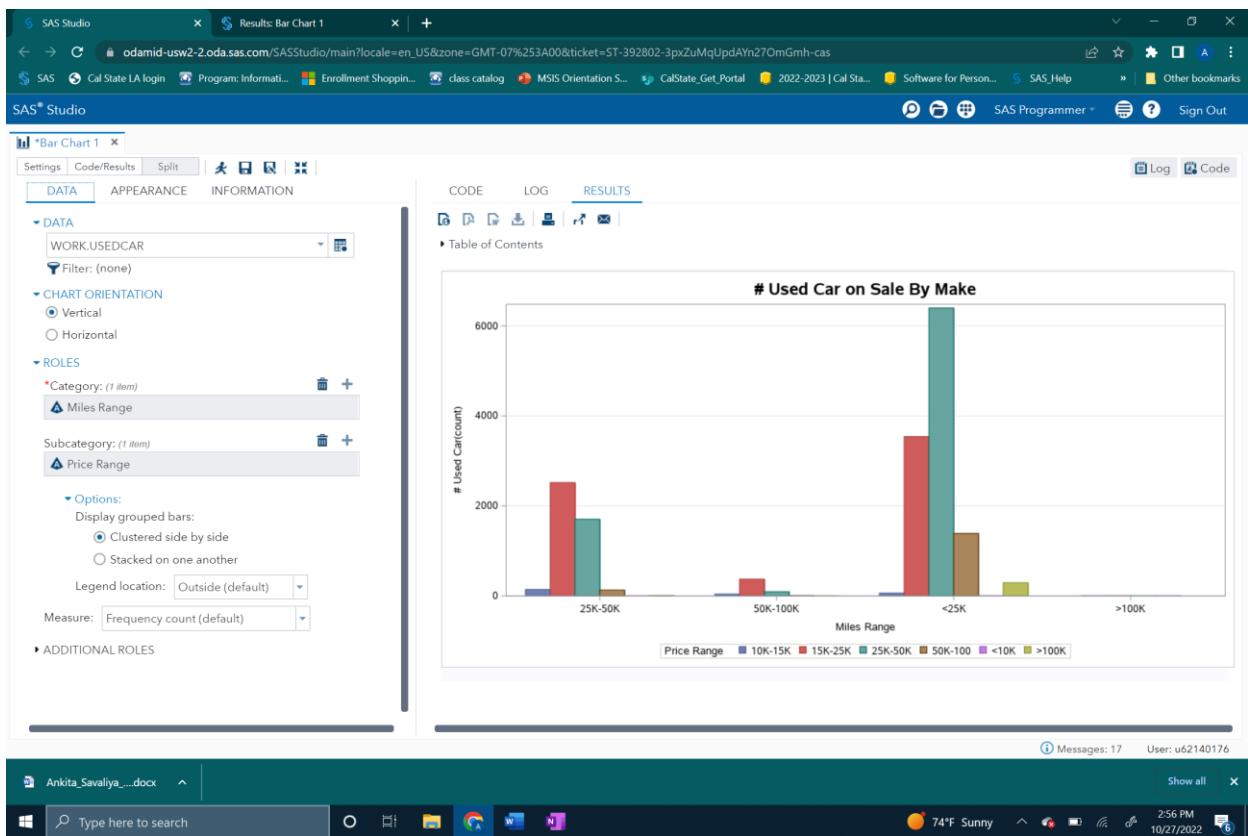


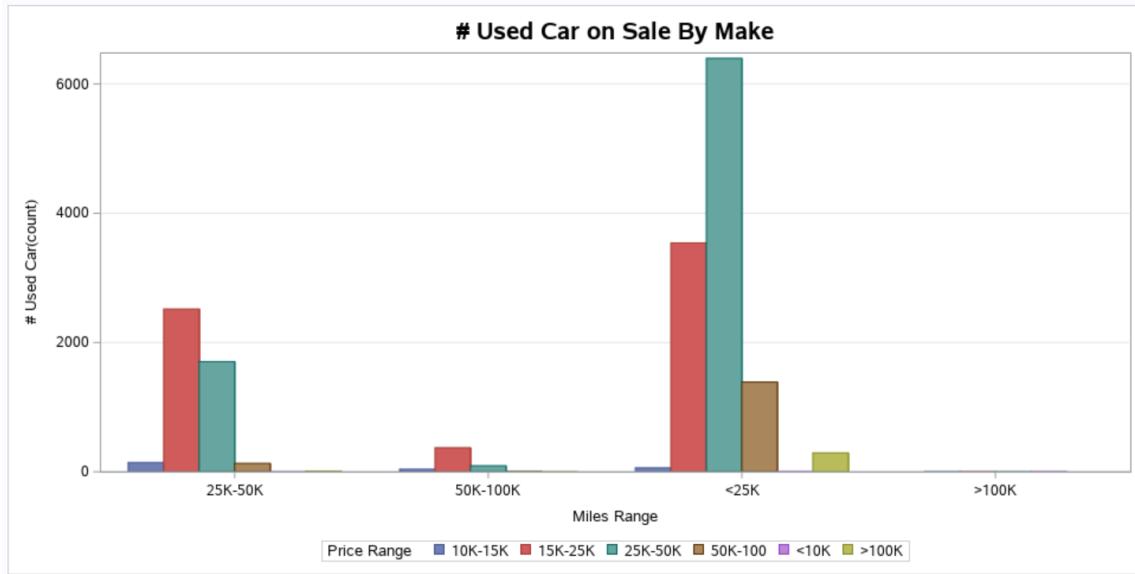
‘Other’ (29.04%) Category has below distribution of used Car.

Make	Percentage
KARMA	0.01%
Rolls-Royce	0.01%
Isuzu	0.02%
FIAT	0.02%
McLaren	0.04%
Aston Martin	0.07%
Ferrari	0.08%
Bentley	0.11%
Alfa Romeo	0.11%
Chrysler	0.12%
Lamborghini	0.15%
Maserati	0.16%
GENESIS	0.27%
Tesla	0.36%
Cadillac	0.48%
Buick	0.62%
Lincoln	0.62%
Jaguar	0.62%
Mitsubishi	0.62%
Dodge	0.67%
RAM	0.67%
MINI	0.77%
INFINITI	0.85%
GMC	0.96%
Volvo	1.07%
Land Rover	1.08%
Acura	1.23%
Porsche	2.02%
Subaru	2.14%
Audi	2.35%
Lexus	2.53%
Mazda	2.56%
Jeep	2.69%
Volkswagen	2.97%

The top 10 automakers in the used car sales market in California from 2019 to 2021 were Honda, Chevrolet, Kia, Hyundai, BMW, Ford, Toyota, Nissan, Mercedes, and others. According to the graph, we can see that Toyota is the highest. Used cars sold in CA are 13.07%. Second place is Honda with a used car share of 12.66%, and third place is BMW with 10.43%. Others include Karma, Rolls-Royce, and many others, as listed below. Other stocks account for 29.04% of the stock market.

Q-Which used car are in more demand by Price and Miles in California?





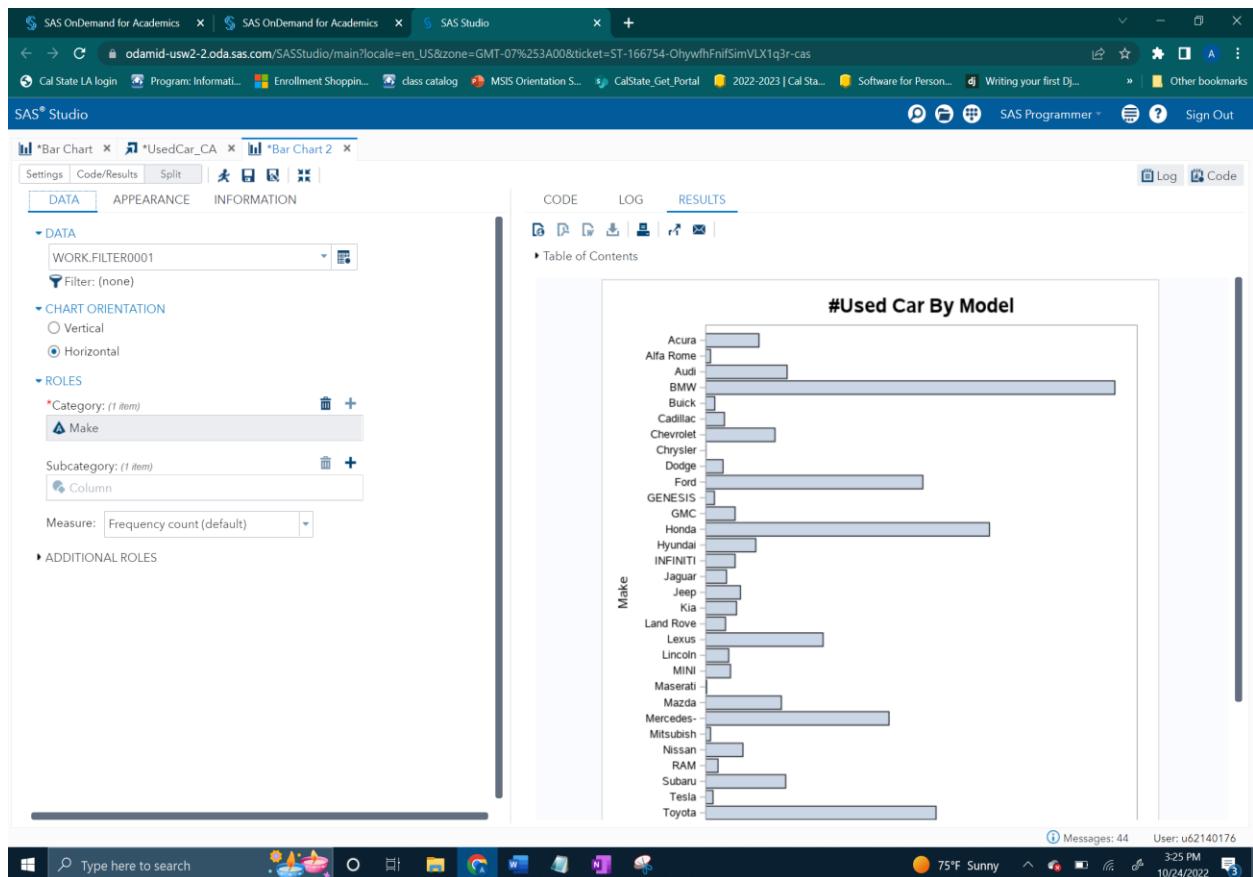
To answer to above question, Dataset has been further categorized with range of Price and Range of Miles. Mile Range has been grouped as “<25K, 25K-50K, 50K-100K, >100K” and Price Range defined as “<10K, 10K-15K,15K-25K,25K-50K,50K-100k,>100K”.

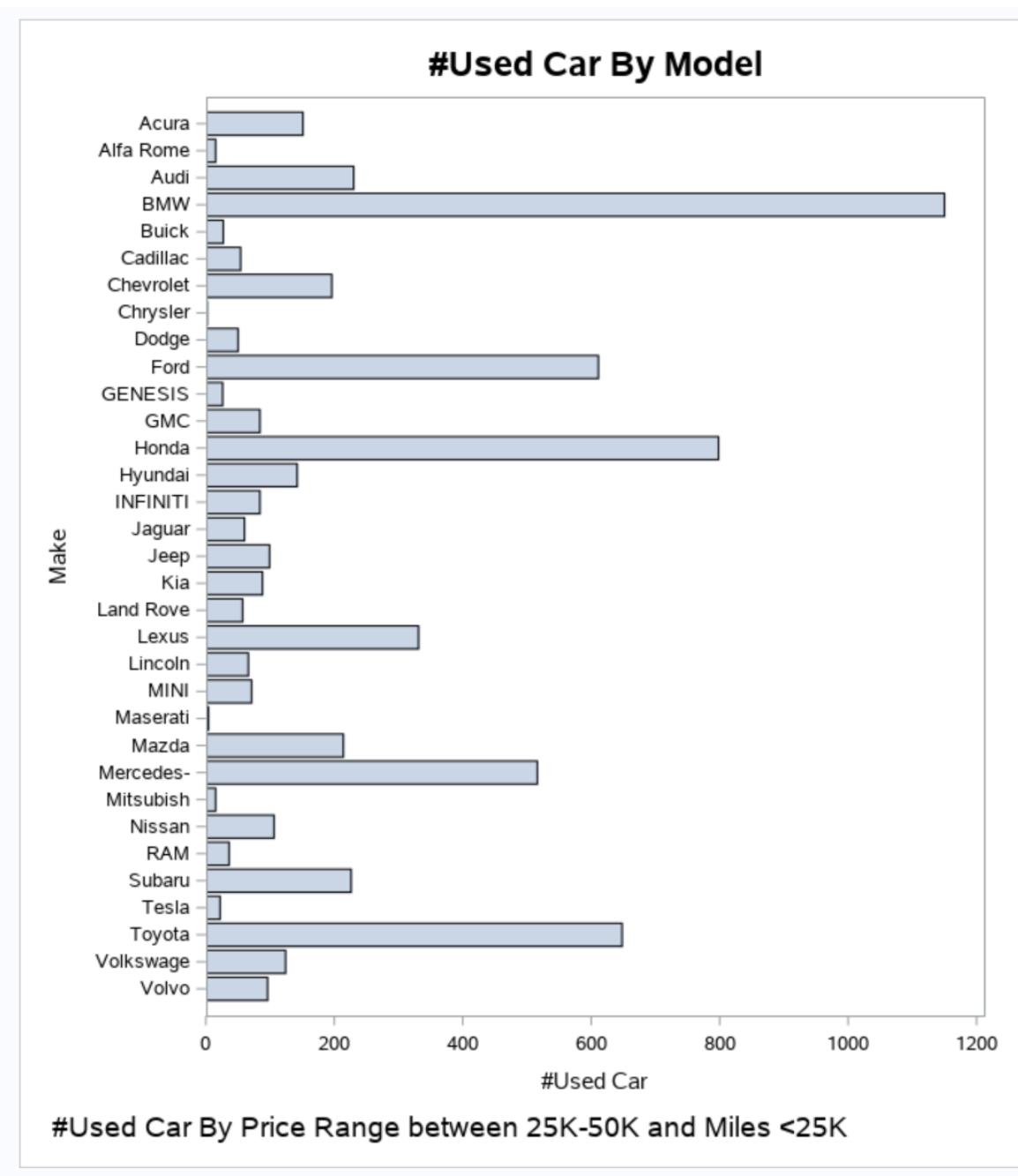
The first (25K-50K) mile group represents 15K-25K Price range has more car listed and likely to sale base on the demand. Sellers are not selling significant cars with Mile range of 10K-15K and 50K-100K price. So, 25K-50K will join the adjacent group to make it more sensible.

The 50K-100K mile range car has less combined cars for sale than 25K-50K price range for (25K-50K) Miles.

When it comes to less driven cars (<25K Miles), Again, 15K-50K Price tops the category with 25K-50K more than 6000 Cars and 15K-15K Little less than 4000 cars, make it most valuable cars on the listing.

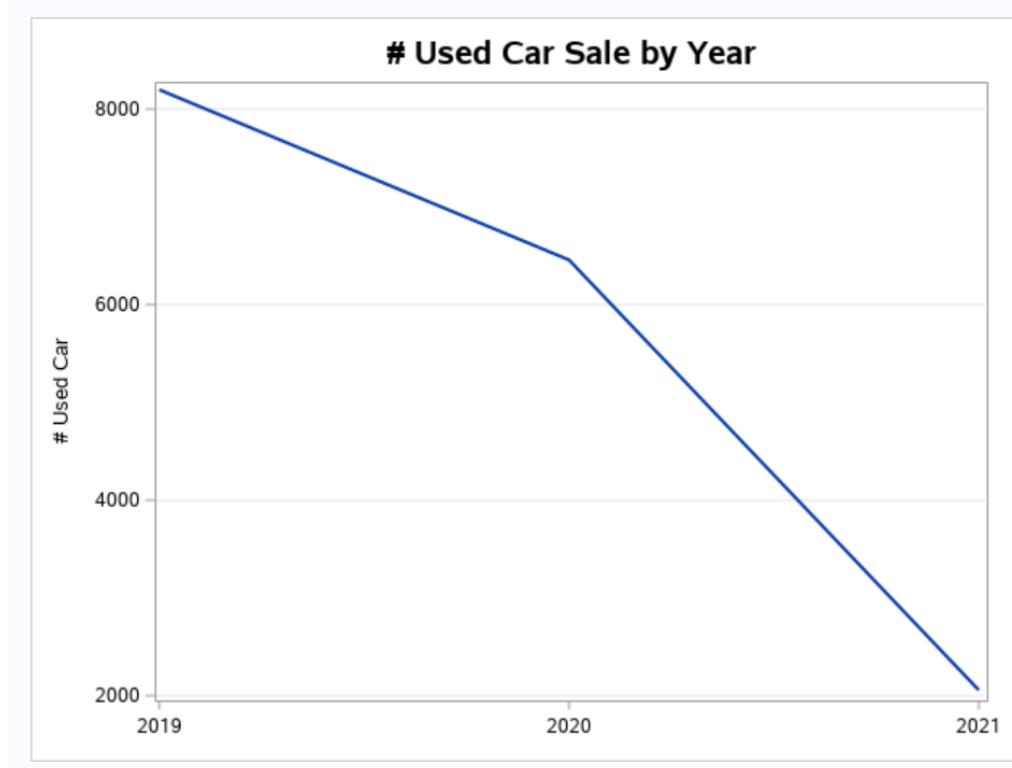
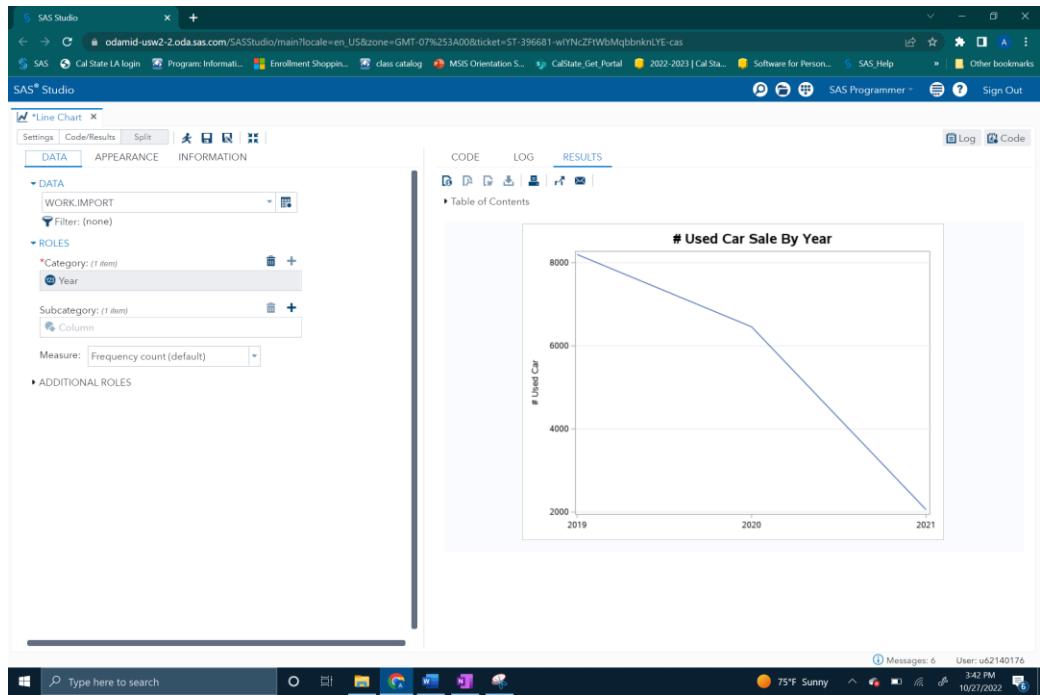
For the further analysis, it would be interesting to know that which model is more popular which has less than 25K mileage and price range between 25K-50K.

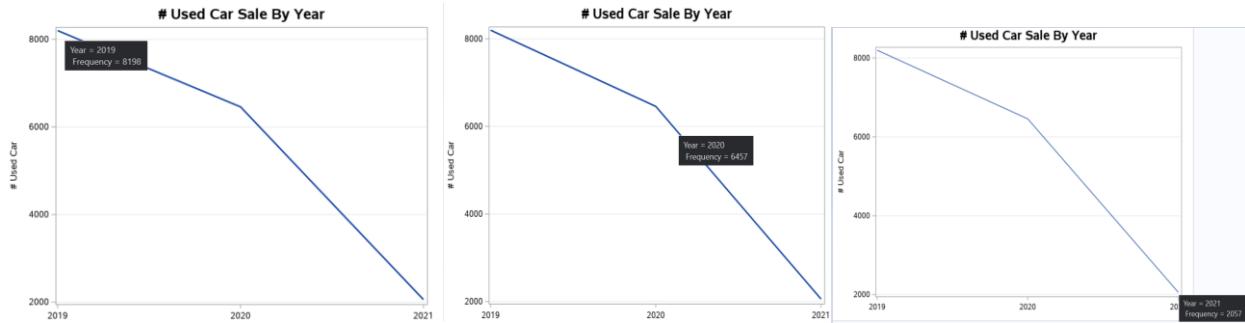




Above graph concludes that BMW, Ford, Honda, Mercedes, and Toyota listed as more used car in the market for sale in California. Among them, BMW was more popular car in the market. Besides, Maserati and Chrysler were least listed car in the market.

Q-How many used cars were on sale between 2019 to 2021 in California (By Year)?





I will only focus used car for sale in market between 2019 to 2021 in California. It shows that used car in the market is rapidly declining. The used car listed for sale in the market in 2019 was 8198. In 2020, the number of cars was less than year 2019 which was 6457. In 2021, less cars that is only 2057 were available for sale in California.

F. Statistical Summary:

It provides descriptive statistics for variables across all observations and within groups of observations.

The screenshot shows the SAS Studio interface with the 'Summary Statistics' window open. On the left, under 'DATA', there is a dropdown menu set to 'WORK.IMPORT'. Under 'ROLES', 'Analysis variables' include Miles and Price. Under 'Classification variables', there is a single entry for Year. On the right, the 'RESULTS' tab is selected, displaying a table of contents and a detailed summary statistics table for three years: 2019, 2020, and 2021. The table includes columns for Year, N Obs, Variable, Mean, Std Dev, Minimum, Maximum, Median, and N Miss. The data shows that Miles and Price are the variables being analyzed, with various statistical measures provided for each year.

Year	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	Median	N Miss
2019	8198	Miles	25061.98	14872.98	4.0000000	142820.00	23075.00	0
		Price	30430.45	23359.98	7490.00	599900.00	25078.00	0
2020	6457	Miles	14189.19	11749.19	2.0000000	100000.00	10366.00	0
		Price	36561.29	34454.85	1500.00	1495000.00	28500.00	0
2021	2057	Miles	3207.90	4213.82	1.0000000	100000.00	2842.00	0
		Price	43975.83	26681.39	15999.00	649000.00	41661.00	0

Year	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	Median	N Miss
2019	8198	Miles	25061.98	14872.98	4.0000000	142820.00	23075.00	0
		Price	30430.45	23359.98	7490.00	599900.00	25078.00	0
2020	6457	Miles	14189.19	11749.19	2.0000000	100000.00	10366.00	0
		Price	36561.29	34454.85	1500.00	1495000.00	28500.00	0
2021	2057	Miles	3207.90	4213.82	1.0000000	100000.00	2842.00	0
		Price	43975.83	26681.39	15999.00	649000.00	41661.00	0

The statistical summary includes year, number of observations, variables, mean, standard deviation, minimum, maximum, median, and number of missing values.

Year: It represents the time span starting from 2019 to 2021 as year.

N Obs: It represents total number of used cars were listed in the market for sale in California.

Variable: Considering **Miles** and **Price** as two variables which can be comparable to each other in reference to year.

Mean: The average of car mileage was 25061.98 and price was 30430.45 among 8198 cars was in the market for sale in 2019. In 2020, number of used cars for sale was 6457, and the average mileage was 14189.19, but the price is higher i.e., 36561.29 than the average price in 2019. In 2021, the number of cars was only 2057, and average mileage was only 3207.90, but the average price is higher than the year 2019 and 2020. *In 2021, number of cars for sale was less but the average selling price was higher with less mileages on odometer.*

Std Deviation: In 2019, the minimum mileage starting from 4 to maximum mileage 42820, the standard deviation wasn't nearer to the mean. But for the year 2020, the minimum and maximum mileage was less than the previous year, but standard deviation is close to the mean. For the year 2021, minimum mileage was less than the previous year, but the maximum mileage was equal to the previous year, *the standard deviation for mileage was higher than the mean, but it is opposite for price.*

Minimum: It shows minimum mileage and price in the market for used car in California. In 2019, minimum mileage was 4 (assuming new car with title change) and price was \$7490. For 2020, the minimum milage was 2 and price was \$1500. Similarly for 2021, the minimum milage was 1 and price was \$15999.

Maximum: It represents maximum mileage and price in the market for used car in California. In 2019, maximum mileage was more than 142K and price was \$5,99,900. For 2020, the maximum milage was 100K and price was \$14,95,000. Similarly for 2021, the number of cars for sell was

less, but the maximum milage was same as in the year 2020, but the maximum price was less than the year 2020.

Median: The median gives the middlemost value for Miles and Price across number of used car in the for sale in the California between the year 2019 to 2021. The highest median value was in the year 2021, because we had only 2057 observations and median mileage was 2842 and price was \$41661. We had the least median value in the year 2019, the median for miles was 23075 which is higher than in the year 2021, but the price was \$25078 which is less than in the year 2021.

N Miss: It shows number of missing values. Summary table shows that report doesn't have any missing values (It all handled in the data cleaning steps).

G. Statistical Tests

A test used to determine the statistical significance of an observation.

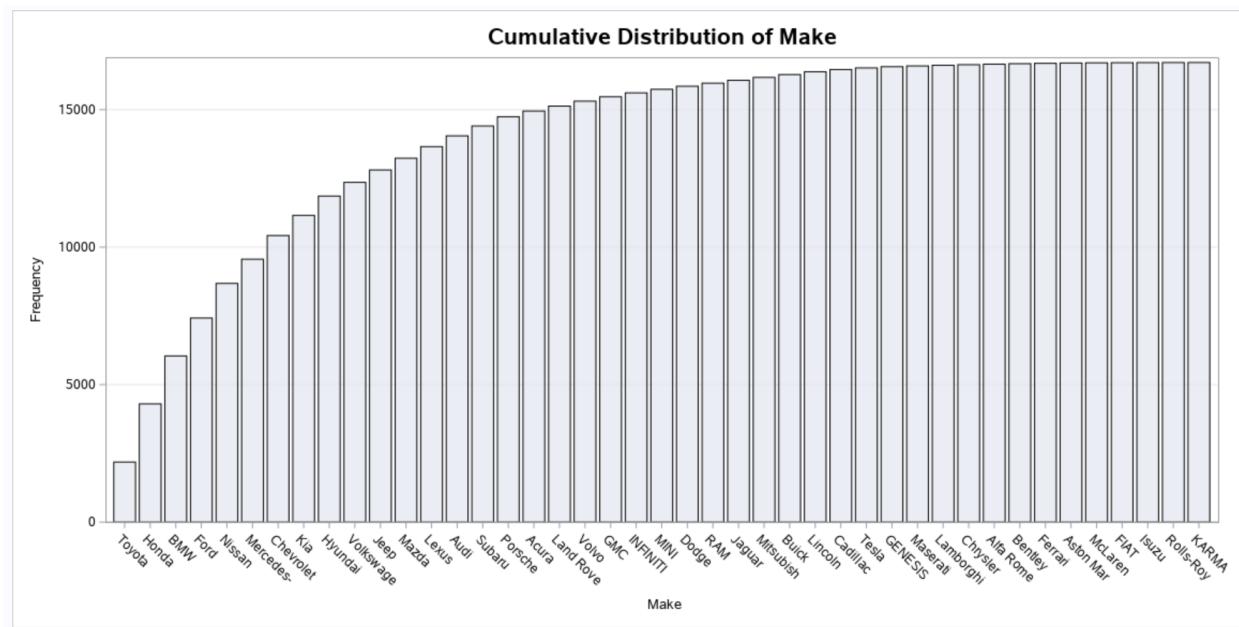
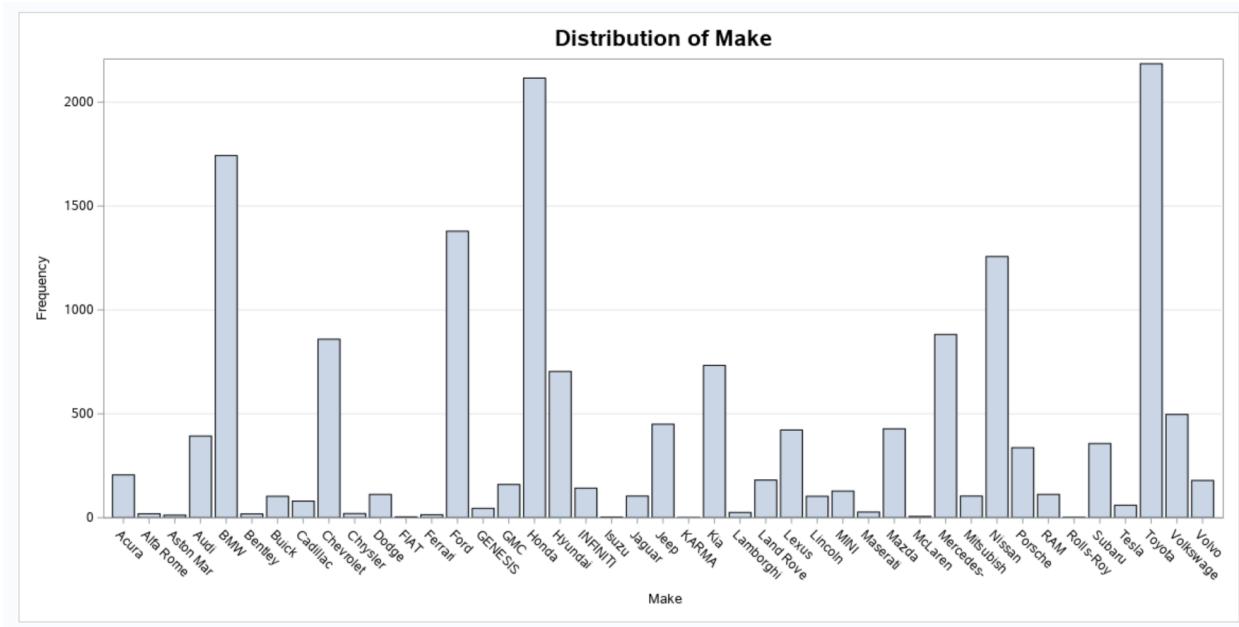
- **One-way Frequency:**

One-way frequency represents the data in a tabular form which only examines one categorical variable at a time.

The screenshot shows the SAS Studio interface with a "One-Way Frequencies" report open. The left sidebar contains sections for DATA, ROLES, and ADDITIONAL ROLES, with "WORK.SORTDS" selected under DATA and "Make" listed under ROLES. The main area displays a table titled "Table of Contents" with data for various car makes. The table has columns: Make, Frequency, Percent, Cumulative Frequency, and Cumulative Percent. The data is as follows:

Make	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Toyota	2185	13.07	2185	13.07
Honda	2116	12.66	4301	25.74
BMW	1743	10.43	6044	36.17
Ford	1379	8.25	7423	44.42
Nissan	1257	7.52	8680	51.94
Mercedes-	882	5.28	9562	57.22
Chevrolet	859	5.14	10421	62.36
Kia	733	4.39	11154	66.74
Hyundai	704	4.21	11858	70.96
Volkswage	497	2.97	12355	73.93
Jeep	450	2.69	12805	76.62
Mazda	428	2.56	13233	79.18
Lexus	422	2.53	13655	81.71
Audi	393	2.35	14048	84.06
Subaru	357	2.14	14405	86.20
Porsche	337	2.02	14742	88.21
Acura	206	1.23	14948	89.44
Land Rover	181	1.08	15129	90.53
Volvo	179	1.07	15308	91.60
GMC	160	0.96	15468	92.56
INFINITI	142	0.85	15610	93.41
MINI	128	0.77	15738	94.17
Dodge	112	0.67	15850	94.84
RAM	112	0.67	15962	95.51
Jaguar	104	0.62	16066	96.13
Mitsubishi	104	0.62	16170	96.76
Buick	103	0.62	16273	97.37
Lincoln	103	0.62	16376	97.99
Cadillac	80	0.48	16456	98.47
Tesla	60	0.36	16516	98.83

Make	Frequency	Percent	Frequency	Percent
Toyota	2185	13.07	2185	13.07
Honda	2116	12.66	4301	25.74
BMW	1743	10.43	6044	36.17
Ford	1379	8.25	7423	44.42
Nissan	1257	7.52	8680	51.94
Mercedes-	882	5.28	9562	57.22
Chevrolet	859	5.14	10421	62.36
Kia	733	4.39	11154	66.74
Hyundai	704	4.21	11858	70.96
Volkswage	497	2.97	12355	73.93
Jeep	450	2.69	12805	76.62
Mazda	428	2.56	13233	79.18
Lexus	422	2.53	13655	81.71
Audi	393	2.35	14048	84.06
Subaru	357	2.14	14405	86.20
Porsche	337	2.02	14742	88.21
Acura	206	1.23	14948	89.44
Land Rove	181	1.08	15129	90.53
Volvo	179	1.07	15308	91.60
GMC	160	0.96	15468	92.56
INFINITI	142	0.85	15610	93.41
MINI	128	0.77	15738	94.17
Dodge	112	0.67	15850	94.84
RAM	112	0.67	15962	95.51
Jaguar	104	0.62	16066	96.13
Mitsubish	104	0.62	16170	96.76
Buick	103	0.62	16273	97.37
Lincoln	103	0.62	16376	97.99
Cadillac	80	0.48	16456	98.47
Tesla	60	0.36	16516	98.83
GENESIS	45	0.27	16561	99.10
Maserati	27	0.16	16588	99.26
Lamborghi	25	0.15	16613	99.41
Chrysler	20	0.12	16633	99.53
Alfa Rome	19	0.11	16652	99.64
Bentley	18	0.11	16670	99.75
Ferrari	14	0.08	16684	99.83
Aston Mar	12	0.07	16696	99.90
McLaren	6	0.04	16702	99.94
FIAT	4	0.02	16706	99.96
Isuzu	3	0.02	16709	99.98
Rolls-Roy	2	0.01	16711	99.99
KARMA	1	0.01	16712	100.00

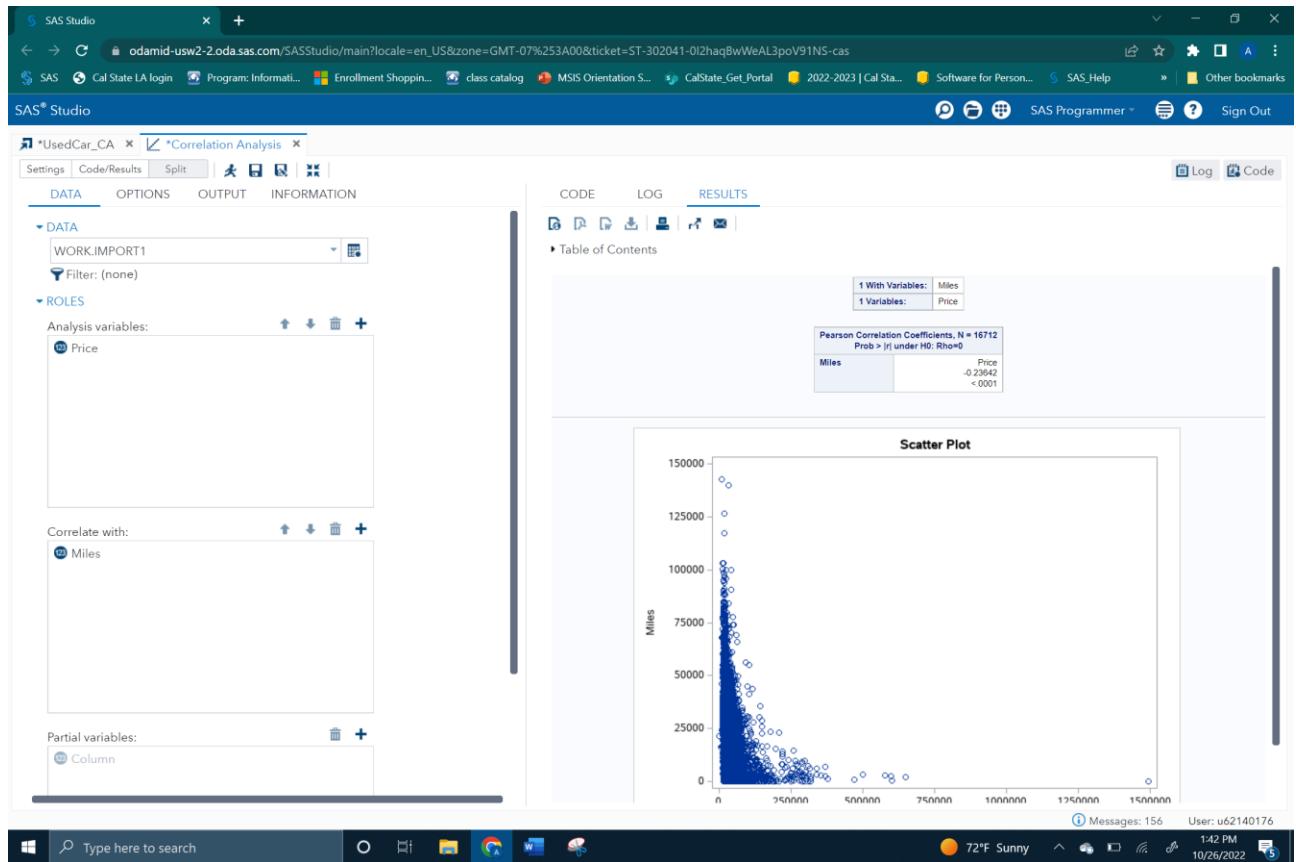


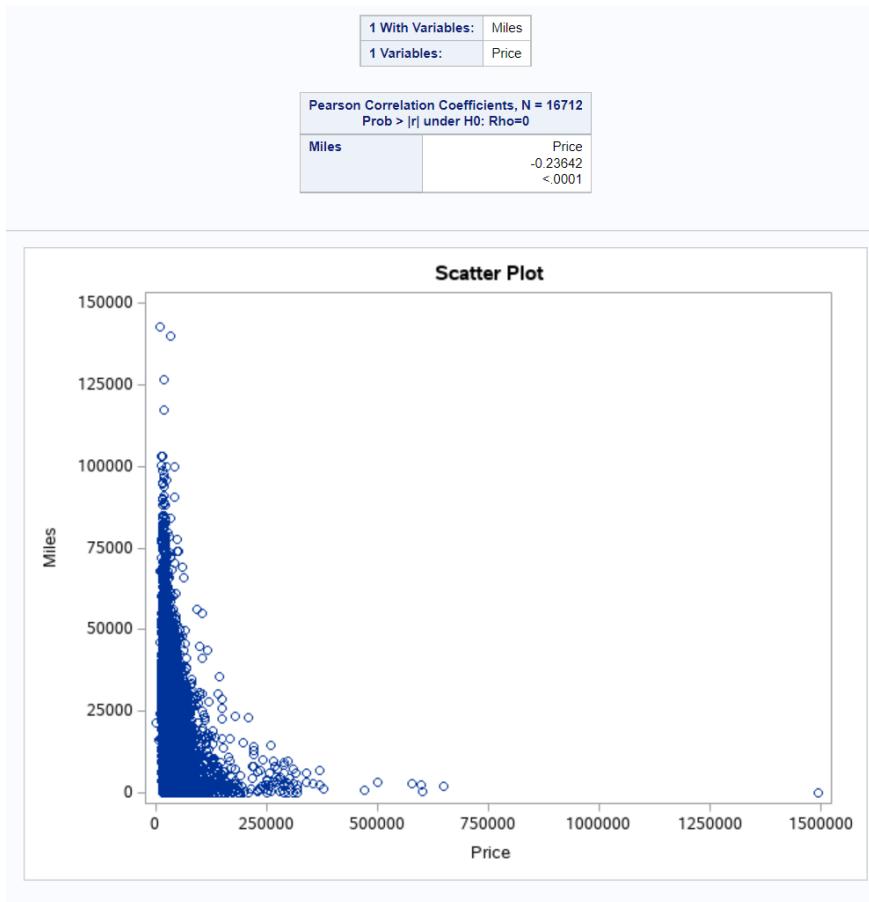
The following one-way frequency table reports the frequency, percent, cumulative frequency, and cumulative percent of the categorial variable ‘Make’. It represents number of used cars in the market for the sale for all Cars Brand Name (Make) between 2019 to 2021 in CA. By observing this table one can determine that the 2 brands, KARMA and Rolls-Roy has least amount of contribution in this dataset (Both has only 1 car in the market for sale in all respective years) while

Toyota and Honda contribute almost 25% to used car market (13.07% and 12.66% respectively) with 2185 and 2116 used car. In detail, out of 43 car brands in Market, 5 Car makers (Toyota, Honda, BMW, Ford, Nissan) has more than 50% cumulative cars in market for sales which defines the popularity of these car brands.

- **Correlation analysis:**

It is a method of visualizing relationships between pairs of variables.





The Correlation table displays the Pearson correlation coefficient between the two variables, including a p-value that tells us if the correlation is statistically significant.

From the output we can see:

- N (Number of observations) :16712
- Pearson correlation coefficient(ρ): -0.2364
- P-value: <.0001

A negative (inverse) correlation occurs when the correlation coefficient(ρ) is less than 0. This is an indication that both variables (i.e., Miles and Price) move in the opposite direction. The above output tells us that there is a weak negative correlation between Price and Miles. The correlation is statistically significant since the p-value (<.0001) is less than $\alpha = .05$.

- **Linear Regression:**

Linear Regression is the best way to identify the relationship between one or more independent variables or a dependent variable. The model of relationship is first proposed, and then the estimation of the parameter values is made to develop a regression equation (estimated).

The screenshot shows the SAS Studio interface with a Linear Regression project open. The left sidebar displays the DATA tab with 'WORKIMPORT3' selected, showing a list of variables including 'Miles' (dependent variable) and 'Price' (continuous variable). The MODEL tab shows the specification of the model as 'Dependent Variable: Miles'. The RESULTS tab is active, displaying the following output:

Model: MODEL1
Dependent Variable: Miles

Number of Observations Read	16712
Number of Observations Used	16712

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.084234E11	2.084234E11	989.28	<.0001
Error	16710	3.498727E12	208661080		
Corrected Total	16711	3.69315E12			

Root MSE 14445 **R-Square** 0.059
Dependent Mean 18171 **Adj R-Sq** 0.0558
Coeff Var 79.49471

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	22364	173.94322	128.57	<.0001
Price	1	-0.12165	0.00387	-31.45	<.0001

Observed by Predicted for Miles

A scatter plot titled "Observed by Predicted for Miles" showing the relationship between the observed value (Y-axis, 0 to 100,000) and the predicted value (X-axis). The data points show a strong negative linear trend, indicating a good fit of the regression model.

Model: MODEL1
Dependent Variable: Miles

Number of Observations Read	16712
Number of Observations Used	16712

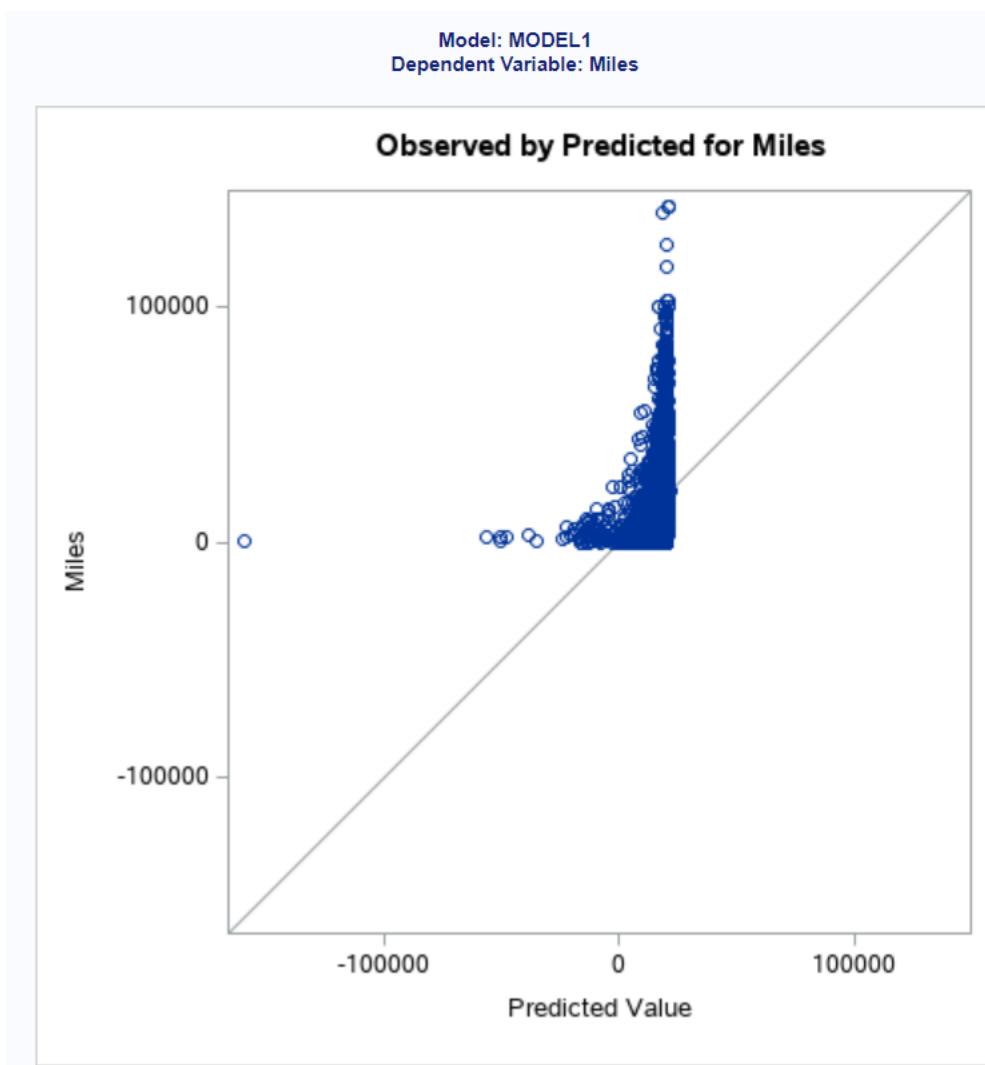
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.064234E11	2.064234E11	989.28	<.0001
Error	16710	3.486727E12	208661080		
Corrected Total	16711	3.69315E12			

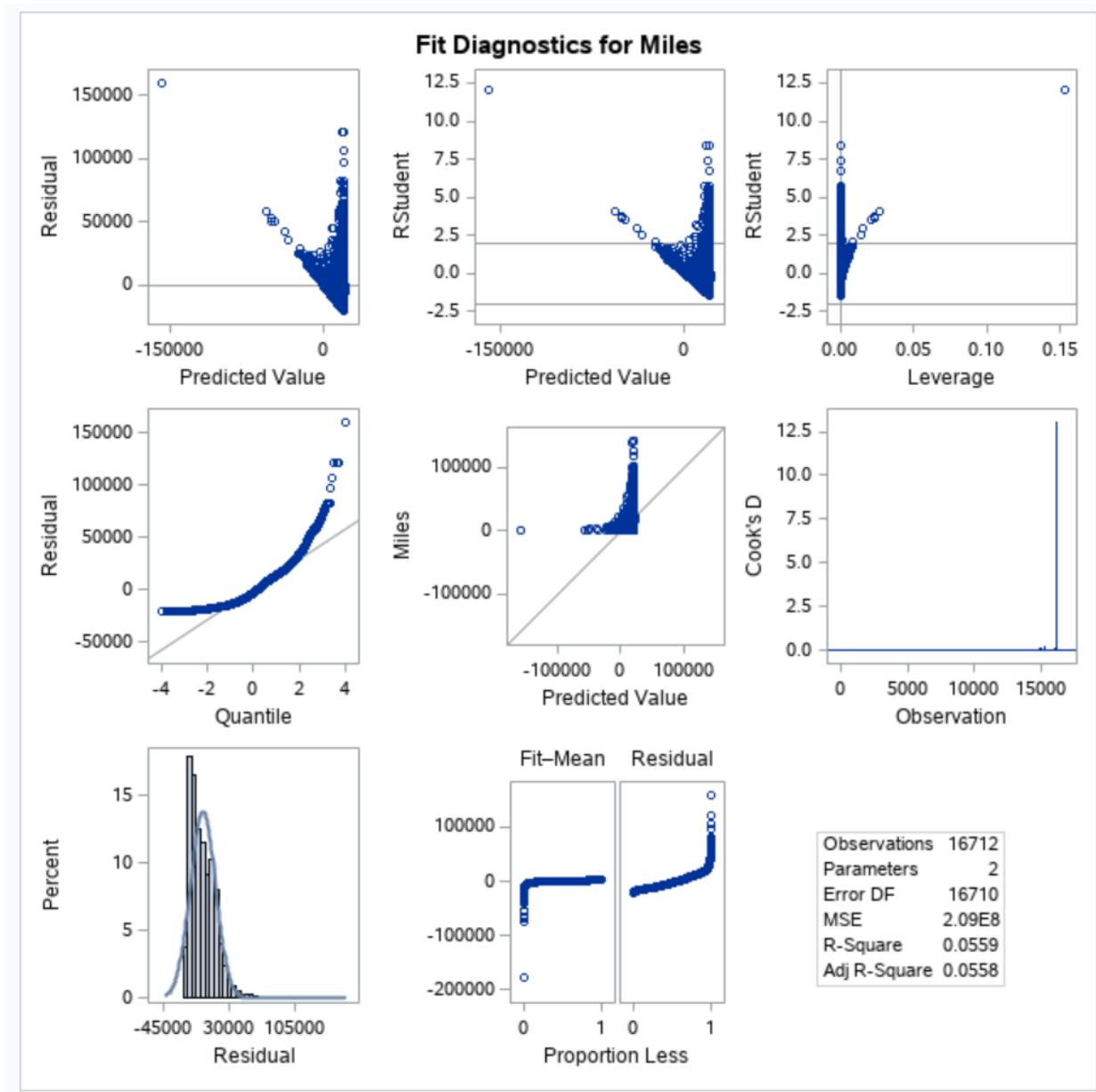
Root MSE	14445	R-Square	0.0559
Dependent Mean	18171	Adj R-Sq	0.0558
Coeff Var	79.49471		

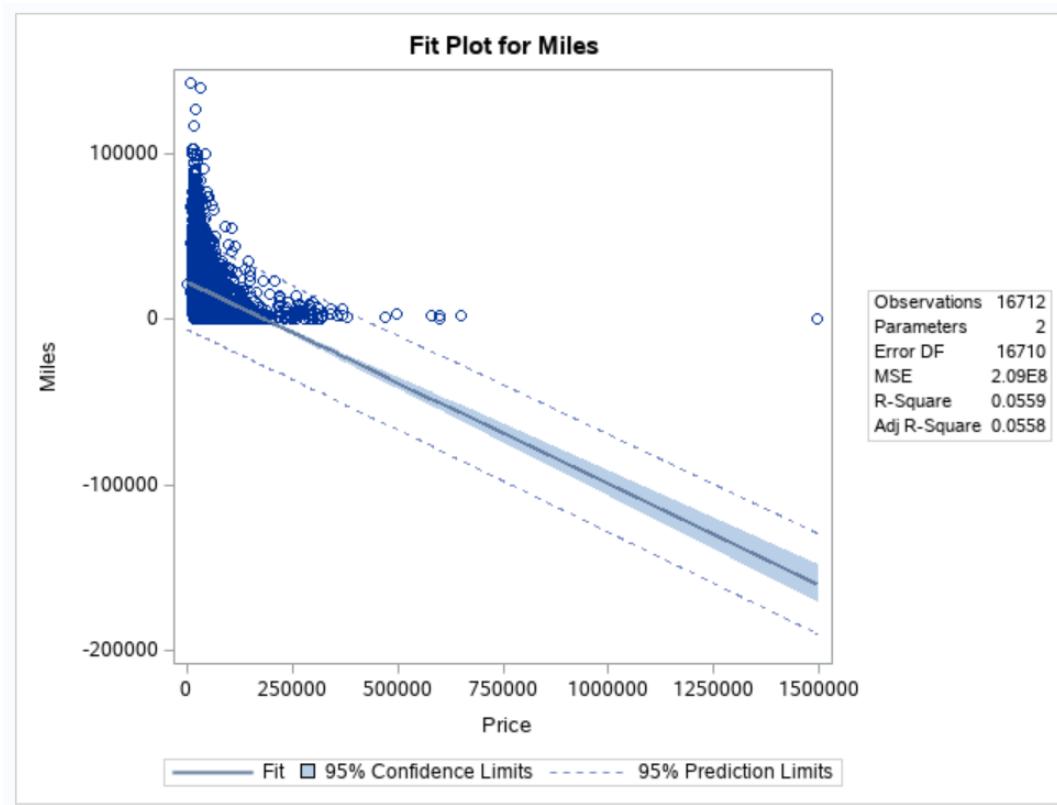
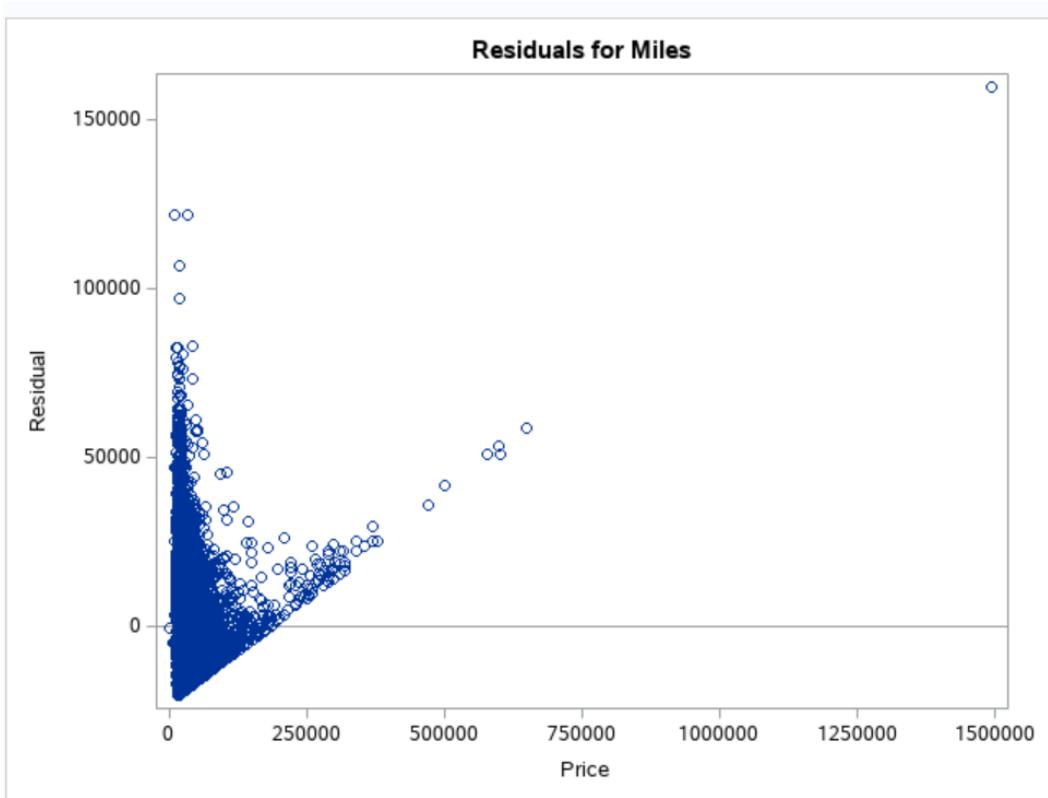
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	22364	173.94322	128.57	<.0001
Price	1	-0.12165	0.00387	-31.45	<.0001

Linear regression assumes that the dependent variable (e.g., Y) is linearly depending on the independent variable (X), i.e., $Y = mx + b$, where b is the intercept and m is the slope. The output shows the parameters of m and b respectively, i.e., Miles = -0.12 * Price + 22364. Therefore, one can predict miles given price using this linear function between the two variables. There were 16712 observations. No missing values are reported.

Furthermore, under the heading "Parameter Estimates" are columns labeled "standard error value," and "Pr > |t|". The T values and the associated p-value test the hypothesis that the parameter is zero, or in other words, whether the (linear) effect of price on miles is zero. Assume a value of P=0.05, it is statistically significant because it is greater than 0.0001(which is the probability value shown in the Table). This means that any variation in the miles can be explained by price variable. R squared value (0.06) is 6% which tends to correlation between price and miles in this example.







- **T-test:**

A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment influences the population of interest, or whether two groups are different from one another.

There are three types of T-tests:

1. One sample t-test
2. Independent two-sample t-test
3. Paired sample t-test

For analysis, One-sample test has been chosen.

The screenshot shows the SAS Studio interface with a tab titled "Tests 1". The left pane displays the "DATA" tab, which includes a "WORKIMPORT" dataset and a "t test" section set to "One-sample test" for the "Miles" variable. The right pane shows the "RESULTS" tab with three tables. The first table, "Variable: Miles", provides summary statistics: N=181712, Mean=181712.2, Std Dev=14866.1, Std Err=115.0, Minimum=1.0000, and Maximum=142820. The second table, "Mean", shows the mean and 95% confidence interval: Mean=181712.2, 95% CL Mean=17945.7 to 18396.6, Std Dev=14866.1, and 95% CL Std Dev=14708.4 to 15027.2. The third table, "DF | t Value | Pr > |t|", shows the degrees of freedom (18711) and a p-value of 158.02 < 0.0001. The bottom of the screen shows the Windows taskbar with a search bar, icons for File Explorer, Word, and Edge, and a system tray with weather, battery, and date/time information.

Variable: Miles

N	Mean	Std Dev	Std Err	Minimum	Maximum
16712	18171.2	14866.1	115.0	1.0000	142820

Mean	95% CL Mean	Std Dev	95% CL Std Dev
18171.2	17945.7	18396.6	14866.1

DF	t Value	Pr > t
16711	158.02	<.0001

In a one-sample t-test, we compare the average (or mean) of miles against the set average (or mean). This set average can be any theoretical value.

The first table shows descriptive statistics for the variable miles, including the N (Number of observations, mean, Std Dev (Standard Deviation), Std Err (Standard Error), Minimum, and Maximum. The second table shows 95% confidence interval mean and Std Dev for miles. The third and the last table shows the t-test value and its corresponding p-value.

Performing the Hypothesis Test:

1. Null Hypothesis: $H_0: \rho = 0$. There is no significance in miles from the mean.
2. Alternate Hypothesis $H_a: \rho \neq 0$. There is a significance in miles from the mean.

Using a p-value to decide:

- If the p-value is less than the significance level ($\alpha=0.05$):

Decision: Reject the null hypothesis.

- If the p -value is NOT less than the significance level ($\alpha = 0.05$)

Decision: Do not reject the null hypothesis.

From the table the output is:

- t-value: 158.02
- p-value :<.0001

Here, $\alpha = 0.05$, and p-value is 0.0001.

So, $p\text{-value} < \alpha$

$0.0001 < 0.05$. This means the p-value is less than the significance level, so we reject the null hypothesis(H_0). Thus, there is significance in miles from the mean.

Citations:

-www.ibisworld.com. *Used Car Dealers in the US - Market Size 2003–2028*(Updated: July 29, 2022): <https://www.ibisworld.com/industry-statistics/market-size/used-car-dealers-united-states/>

-www.prnewswire.com. *Used Car Market size in the US to increase by 3.91 million units / High growth expected in mid-size segment / Technavio* (NEWS PROVIDED BY Technavio Mar 25, 2022, 02:15 ET): <https://www.prnewswire.com/news-releases/used-car-market-size-in-the-us-to-increase-by-3-91-million-units--high-growth-expected-in-mid-size-segment--technavio-301510065.html>

-www.grandviewresearch.com. *Used Car Market Size, Share & Trends Analysis Report By Vehicle Type (Hybrid, Conventional, Electric), By Vendor Type, By Fuel Type, By Size, By Sales Channel, By Region, And Segment Forecasts, 2022 - 2030*: <https://www.grandviewresearch.com/industry-analysis/used-car-market>

-www.statology.org. *How to Calculate Correlation in SAS (With Examples)* By ZACH (updated on DECEMBER 17, 2021): <https://www.statology.org/correlation-in-sas/>

-mathworld.wolfram.com. *Statistical Test*: <https://mathworld.wolfram.com/StatisticalTest.html>

- www.simplilearn.com. *SAS Linear Regression and Other Statistical Procedure* By Simplilearn (Last updated on Sep 28, 2022): <https://www.simplilearn.com/tutorials/sas-tutorial/sas-linear-regression>

- towardsdatascience.com. *Everything you need to know about interpreting correlations* By Zakaria Jaadi (Last updated on Oct 15, 2019) : <https://towardsdatascience.com/eveything-you-need-to-know-about-interpreting-correlations-2c485841c0b8#:~:text=If%20the%20test%20shows%20that,not%20present%20in%20the%20en>
- www.investopedia.com. *What Do Correlation Coefficients Positive, Negative, and Zero Mean?* By STEVEN NICKOLAS, Reviewed by JULIUS MANSA, and Fact checked by KATRINA MUNICHIELLO(Updated on May 31, 2021) : <https://www.investopedia.com/ask/answers/032515/what-does-it-mean-if-correlation-coefficient-positive-negative-or-zero.asp>
- www.stat.purdue.edu. *Linear regression with SAS* https://www.stat.purdue.edu/~tqin/system101/method/method_linear_sas.htm#regression-report
- www.scribbr.com. *An Introduction to T-Tests / Definitions, Formula and Examples.* Published on January 31, 2020, by Rebecca Bevans. Revised on July 9, 2022. <https://www.scribbr.com/statistics/t-test/>
- www.courses.lumenlearning.com. *Testing the Significance of the Correlation Coefficient:* <https://courses.lumenlearning.com/introstats1/chapter/testing-the-significance-of-the-correlation-coefficient/>