FURKAN DEMIRDOVEN

160828182

BAN100 Assignment 1- Data Audit on Cars_1993.xlsx dataset

The dataset cars_1993.xlsx contains 93 observations and 26 variables. What follows is an audit report consisting of metadata, descriptive statistics and some comments on the dataset.

Column1	Column2	Column3	Column4
Variables	**Type**	**Scale**	**Description**
Manufacturer	Character	Nominal	Company that produces that car. 30 Levels
Model	Character	Nominal	Model. 93 Levels
Category	Character	Ordinal	Category in terms of size. 6 Levels
Min_Price	Numeric	Continuous	Minimum Price
Mid_Price	Numeric	Continuous	Midrange Price
Max_Price	Numeric	Continuous	Maximum Price
City_Fuel	Numeric	Discrete	City Fuel Consumption
Hwy_Fuel	Numeric	Discrete	Highway Fuel Consumption
Air_Bags	Numeric	Integer	Air Bags Standard [1,2,3]
Drive_Train	Numeric	Integer	Drive Train [0,1,2]
Cylinders	Numeric	Discrete	Number of Cylinders
Engine_Size	Numeric	Continuous	Engine Size
Max_HP	Numeric	Discrete	Maximum Hourse Power
Max_HP_RPM	Numeric	Discrete	Maximum RPM
RPM_high	Numeric	Discrete	RPM High
Manual	Numeric	Binary	Manual Transmission [0,1]
Fuel_Tank	Numeric	Continuous	Fuel Tank Capacity
Passenger	Numeric	Discrete	Passenger Capacity
Length	Numeric	Discrete	Length
Wheel_Base	Numeric	Discrete	Wheel Base
Width	Numeric	Discrete	Width
U_Turn_Diam	Numeric	Discrete	U-turn diameter
Rear_Room	Numeric	Continuous	Rear Seat Room
Luggage	Numeric	Discrete	Luggage Capacity
Weight	Numeric	Discrete	Weight
Domestic	Numeric	Binary	Domestic [0,1]

Descriptive Statistics for Numeric Variables

	Column2	Mean	Standard Error	Median	Mode	Standard Deviation	Sample Variance	Kurtosis	Skewness	Range	Minimum	Maximum	Sum	Count	Missing
Min_Price		17.12580645	0.906920954	14.7	14.2	8.746028953	76.49302244	1.019418749	1.182989202	38.7	6.7	45.4	1592.7	93	0
Mid_Price		20.98172043	1.732143215	18.2	15.9	16.70418423	279.0297709	40.12227024	5.444590818	143.6	7.4	151	1951.3	93	0
Max_Price		21.89892473	1.143805087	19.6	21.7	11.03045679	121.6709771	7.4394262	2.03385866	72.1	7.9	80	2036.6	93	0
City_Fuel		22.3655914	0.582747307	21	18	5.619811513	31.58228144	4.00430589	1.704430146	31	15	46	2080	93	0
Hwy_Fuel		29.08602151	0.552874234	28	26	5.331726032	28.42730248	2.614200606	1.229896739	30	20	50	2705	93	0
Cylinders		4.967391304	0.136023563	4	4	1.304692183	1.702221691	-0.20435896	0.820190141	5	3	8	457	92	1
Engine_Size		2.667741935	0.107569533	2.4	3	1.037363012	1.07612202	0.381024905	0.859418415	4.7	1	5.7	248.1	93	0
Max_HP		143.827957	5.430973274	140	110	52.37440954	2743.078775	1.11088259	0.951728245	245	55	300	13376	93	0
Max_HP_RPM		5280.645161	61.87819371	5200	6000	596.7316899	356088.7097	-0.409478979	-0.258532686	2700	3800	6500	491100	93	0
RPM_high		2332.204301	51.48532827	2340	1690	496.5065252	246518.7295	0.220543861	0.281546015	2435	1320	3755	216895	93	0
Fuel_Tank		16.66451613	0.340054877	16.4	18	3.279370471	10.75427069	0.127206499	0.108146202	17.8	9.2	27	1549.8	93	0
Passenger		5.086021505	0.107737055	5	5	1.038978532	1.079476391	0.936192595	0.062516854	6	2	8	473	93	0
Length		183.2043011	1.514196428	183	184	14.60238154	213.2295465	0.449250413	-0.090094622	78	141	219	17038	93	0
Wheel_Base		103.9462366	0.707167207	103	103	6.819673573	46.50794764	-0.797455728	0.113726837	29	90	119	9667	93	0
Width		69.37634409	0.391862641	69	67	3.77898646	14.28073866	-0.246427704	0.264027381	18	60	78	6452	93	0
U_Turn_Diam		38.95698925	0.334236963	39	39	3.223264543	10.38943432	-0.732203957	-0.133568584	13	32	45	3623	93	0
Rear_Room		27.82967033	0.313339936	27.5	26.5	2.989072487	8.934554335	0.894682228	0.078259508	17	19	36	2532.5	91	2
Luggage		13.8902439	0.331070048	14	14	2.997966791	8.987804878	0.549687812	0.229565773	16	6	22	1139	82	11
Weight		3072.903226	61.16941859	3040	3470	589.8965102	347977.8927	-0.855115674	-0.143669038	2410	1695	4105	285780	93	0

Missing values were reported for Luggage, Rear_Room and Cylinders variables. As a result of an eyeball inspection, it is observed that Mid_Price variable has an outlier which is 151. This is not possible considering the fact that Max_Price variable has a maximum value of 80.

Following variables are categorical variables including the binary ones, and for this reason frequency distributions were deemed more proper for them.

Manufacturer
Model
Category
Air Bags
Drive_Train
Domestic
Manual

Row Labels	Count of Category
Compact	16
Large	11
Midsize	22
Small	21
Sporty	14
Van	9
Grand Total	**93**

Row Labels	Count of Air_Bags
0	34
1	43
2	16
Grand Total	**93**

Row Labels	Count of Drive_Train
0	16
1	67
2	10
Grand Total	**93**

Row Labels	Count of Domestic
0	45
1	48
Grand Total	**93**

Row Labels	Count of Manual
0	32
1	61
Grand Total	**93**

Row Labels	Count of Manufacturer
Acura	2
Audi	2
BMW	1
Buick	4
Cadillac	2
Chevrolet	8
Chrysler	3
Dodge	6
Eagle	2
Ford	8
Geo	2
Honda	3
Hyundai	4
Infiniti	1
Lexus	2
Lincoln	2
Mazda	5
Mercedes-Benz	2
Mercury	2
Mitsubishi	2
Nissan	4
Oldsmobile	4
Plymouth	1
Pontiac	5
Saab	1
Saturn	1
Subaru	3
Suzuki	1
Toyota	4
Volkswagen	4
Volvo	2
Grand Total	**93**

Row Labels	Count of Manufacturer
Acura	**2**
Integra	1
Legend	1
Audi	**2**
90	1
100	1
BMW	**1**
535i	1
Buick	**4**
Century	1
LeSabre	1
Riviera	1
Roadmaster	1
Cadillac	**2**
DeVille	1
Seville	1
Chevrolet	**8**
Astro	1
Camaro	1
Caprice	1
Cavalier	1
Corsica	1
Corvette	1
Lumina	1
Lumina APV	1
Chrysler	**3**
Concorde	1
Imperial	1
LeBaron	1
Dodge	**6**
Caravan	1
Colt	1
Dynasty	1
Shadow	1
Spirit	1
Stealth	1
Eagle	**2**
Summit	1
Vision	1
Ford	**8**
Aerostar	1
Crown Victoria	1
Escort	1

Festiva	1
Mustang	1
Probe	1
Taurus	1
Tempo	1
Geo	**2**
Metro	1
Storm	1
Honda	**3**
Accord	1
Civic	1
Prelude	1
Hyundai	**4**
Elantra	1
Excel	1
Scoupe	1
Sonata	1
Infiniti	**1**
Q45	1
Lexus	**2**
ES300	1
SC300	1
Lincoln	**2**
Continental	1
Town Car	1
Mazda	**5**
323	1
626	1
MPV	1
Protege	1
RX-7	1
Mercedes-Benz	**2**
190E	1
300E	1
Mercury	**2**
Capri	1
Cougar	1
Mitsubishi	**2**
Diamante	1
Mirage	1
Nissan	**4**
Altima	1
Maxima	1
Quest	1
Sentra	1

Oldsmobile	**4**
Achieva	1
Cutlass Ciera	1
Eighty-Eight	1
Silhouette	1
Plymouth	**1**
Laser	1
Pontiac	**5**
Bonneville	1
Firebird	1
Grand Prix	1
LeMans	1
Sunbird	1
Saab	**1**
900	1
Saturn	**1**
SL	1
Subaru	**3**
Justy	1
Legacy	1
Loyale	1
Suzuki	**1**
Swift	1
Toyota	**4**
Camry	1
Celica	1
Previa	1
Tercel	1
Volkswagen	**4**
Corrado	1
Eurovan	1
Fox	1
Passat	1
Volvo	**2**
240	1
850	1
Grand Total	**93**

Row Labels	Count of Manufacturer
Acura	**2**
Integra	1
Legend	1
Audi	**2**
90	1
100	1
BMW	**1**
535i	1
Buick	**4**
Century	1
LeSabre	1
Riviera	1
Roadmaster	1
Cadillac	**2**
DeVille	1
Seville	1
Chevrolet	**8**
Astro	1
Camaro	1
Caprice	1
Cavalier	1
Corsica	1
Corvette	1
Lumina	1
Lumina APV	1
Chrysler	**3**
Concorde	1
Imperial	1
LeBaron	1
Dodge	**6**
Caravan	1
Colt	1
Dynasty	1
Shadow	1
Spirit	1
Stealth	1
Eagle	**2**
Summit	1
Vision	1
Ford	**8**
Aerostar	1
Crown Victoria	1
Escort	1

Festiva	1
Mustang	1
Probe	1
Taurus	1
Tempo	1
Geo	**2**
Metro	1
Storm	1
Honda	**3**
Accord	1
Civic	1
Prelude	1
Hyundai	**4**
Elantra	1
Excel	1
Scoupe	1
Sonata	1
Infiniti	**1**
Q45	1
Lexus	**2**
ES300	1
SC300	1
Lincoln	**2**
Continental	1
Town Car	1
Mazda	**5**
323	1
626	1
MPV	1
Protege	1
RX-7	1
Mercedes-Benz	**2**
190E	1
300E	1
Mercury	**2**
Capri	1
Cougar	1
Mitsubishi	**2**
Diamante	1
Mirage	1
Nissan	**4**
Altima	1
Maxima	1
Quest	1
Sentra	1

Oldsmobile	**4**
Achieva	1
Cutlass Ciera	1
Eighty-Eight	1
Silhouette	1
Plymouth	**1**
Laser	1
Pontiac	**5**
Bonneville	1
Firebird	1
Grand Prix	1
LeMans	1
Sunbird	1
Saab	**1**
900	1
Saturn	**1**
SL	1
Subaru	**3**
Justy	1
Legacy	1
Loyale	1
Suzuki	**1**
Swift	1
Toyota	**4**
Camry	1
Celica	1
Previa	1
Tercel	1
Volkswagen	**4**
Corrado	1
Eurovan	1
Fox	1
Passat	1
Volvo	**2**
240	1
850	1
Grand Total	**93**

There are no missing values for any of the categorical variables.