Final Project

Pratheek Sreerangam

2022-12-15

Problem

In this case study, tasks included assessing costs and figuring out the car characteristics the buyer will get. In this case study, we'll try to gather data on the least expensive, most affordable, average, and most expensive cars.

What are the specifications that clients will receive based on these four price ranges?

About dataset

For this project, the data came from Kaggle. The car dataset has 26 columns and 205 rows. All vehicle parameters, including body type, driving wheel position, engine position, wheelbase, length, breadth, and height, as well as a few more, are covered by the dataset.

Let’s look at dataset details

symboling	normalized	make	fuel-type	aspiration	num-of-d	body-sty	drive-wheels	engine-locati	wheel-base	length	width	height	curb-weigl	engine-type	num-of-cy	engine-siz	fuel-syst	bore	stroke	compres	horsepov	peak-rpn	city-mpg	highway-mpg	price
3	?	alfa-rom	gas	std	two	convertit	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	13495
3	?	alfa-rom	gas	std	two	convertit	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	16500
1	?	alfa-rom	gas	std	two	hatchbac	rwd	front	94.5	171.2	65.5	52.4	2823	ohcv	six	152	mpfi	2.68	3.47	9	154	5000	19	26	16500
2	164	audi	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337	ohc	four	109	mpfi	3.19	3.4	10	102	5500	24	30	13950
2	164	audi	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2824	ohc	five	136	mpfi	3.19	3.4	8	115	5500	18	22	17450
2	?	audi	gas	std	two	sedan	fwd	front	99.8	177.3	66.3	53.1	2507	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	15250
1	158	audi	gas	std	four	sedan	fwd	front	105.8	192.7	71.4	55.7	2844	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	17710
1	?	audi	gas	std	four	wagon	fwd	front	105.8	192.7	71.4	55.7	2954	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	18920
1	158	audi	gas	turbo	four	sedan	fwd	front	105.8	192.7	71.4	55.9	3086	ohc	five	131	mpfi	3.13	3.4	8.3	140	5500	17	20	23875
0	?	audi	gas	turbo	two	hatchbac	4wd	front	99.5	178.2	67.9	52	3053	ohc	five	131	mpfi	3.13	3.4	7	160	5500	16	22	?
2	192	bmw	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2395	ohc	four	108	mpfi	3.5	2.8	8.8	101	5800	23	29	16430
0	192	bmw	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2395	ohc	four	108	mpfi	3.5	2.8	8.8	101	5800	23	29	16925
0	188	bmw	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2710	ohc	six	164	mpfi	3.31	3.19	9	121	4250	21	28	20970
0	188	bmw	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2765	ohc	six	164	mpfi	3.31	3.19	9	121	4250	21	28	21105
1	?	bmw	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3055	ohc	six	164	mpfi	3.31	3.19	9	121	4250	20	25	24565
0	?	bmw	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3230	ohc	six	209	mpfi	3.62	3.39	8	182	5400	16	22	30760
0	?	bmw	gas	std	two	sedan	rwd	front	103.5	193.8	67.9	53.7	3380	ohc	six	209	mpfi	3.62	3.39	8	182	5400	16	22	41315
0	?	bmw	gas	std	four	sedan	rwd	front	110	197	70.9	56.3	3505	ohc	six	209	mpfi	3.62	3.39	8	182	5400	15	20	36880
2	121	chevrolet	gas	std	two	hatchbac	fwd	front	88.4	141.1	60.3	53.2	1488	l	three	61	2bbl	2.91	3.03	9.5	48	5100	47	53	5151
1	98	chevrolet	gas	std	two	hatchbac	fwd	front	94.5	155.9	63.6	52	1874	ohc	four	90	2bbl	3.03	3.11	9.6	70	5400	38	43	6295
0	81	chevrolet	gas	std	four	sedan	fwd	front	94.5	158.8	63.6	52	1909	ohc	four	90	2bbl	3.03	3.11	9.6	70	5400	38	43	6575
1	118	dodge	gas	std	two	hatchbac	fwd	front	93.7	157.3	63.8	50.8	1876	ohc	four	90	2bbl	2.97	3.23	9.41	68	5500	37	41	5572
1	118	dodge	gas	std	two	hatchbac	fwd	front	93.7	157.3	63.8	50.8	1876	ohc	four	90	2bbl	2.97	3.23	9.4	68	5500	31	38	6377
1	118	dodge	gas	turbo	two	hatchbac	fwd	front	93.7	157.3	63.8	50.8	2128	ohc	four	98	mpfi	3.03	3.39	7.6	102	5500	24	30	7957
1	148	dodge	gas	std	four	hatchbac	fwd	front	93.7	157.3	63.8	50.6	1967	ohc	four	90	2bbl	2.97	3.23	9.4	68	5500	31	38	6229
1	148	dodge	gas	std	four	sedan	fwd	front	93.7	157.3	63.8	50.6	1989	ohc	four	90	2bbl	2.97	3.23	9.4	68	5500	31	38	6692

https://www.kaggle.com/code/natigmamishov/eda-and-regression-on-automobile-dataset/data

Approach

I'll try grouping the automobiles based on their specifications. I'll remove the normalized-losses and symbolling variables from the data set and only use the specification variables, which are categorical and numerical, for this.

The K-means method is one of the most frequently used clustering techniques. Simply described, K is a collection of clusters into which the data can be divided according to the similarities and differences in their attribute values. The center of each cluster is more like adjacent observatories. In K-means clustering, the number of similarities that will be utilized to form a group can be stated as the distance between two observations. This distance is then used in the calculation to establish which cluster each observational member belongs to. With each new observation, a new cluster center is found, and the correct cluster is then given to the new observations.

Why K-means

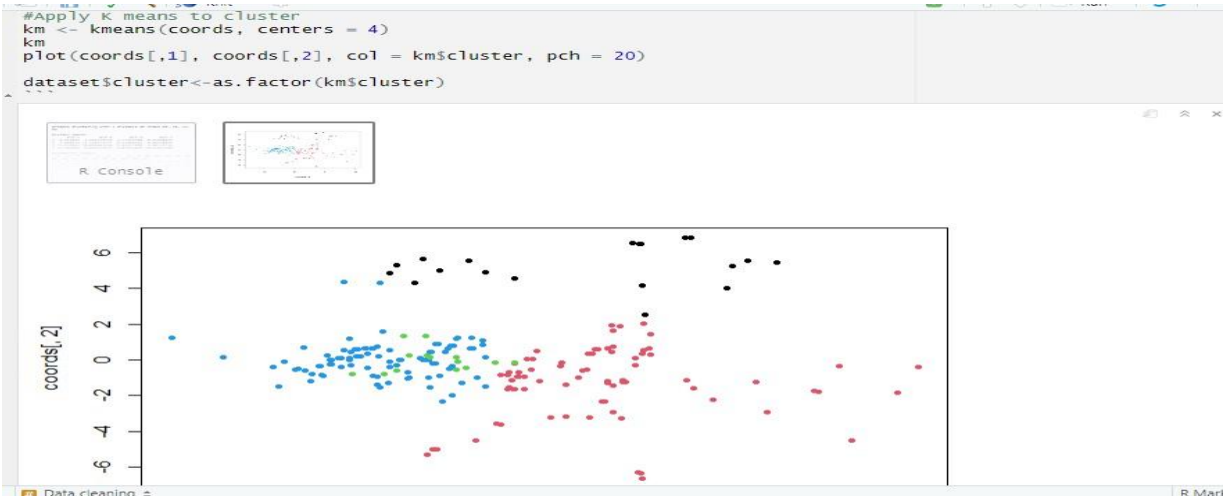
K-means is the simplest. to erect and maintain. Simply select "k" and repeat the process numerous times. Most clever algorithms require a lot more parameters to be configured and are far more complicated to design. Additionally, most people don't need high-quality clusters. Anything that can be done for them remotely makes them happy. Additionally, they are unsure of what to do when they have more complicated clusters. They need K-means, which reduces a vast amount of data to centroids to build clusters using the most basic model ever.

I chose to apply the techniques listed below to obtain the desired outcomes.

- **Launch the necessary libraries**
- **Importing the data**
- **Cleaning data**
- **Factorial evaluation**
- **Cluster Analysis**

The conclusions I came to because of my investigation are as follows.

I selected 4 centers to apply the K-means algorithm by utilizing Factorial analysis of mixed data (PCAmix) to analyze a data table in which observations are described using both quantitative and qualitative factors. I found the results accordingly



The four clusters' results are in line with these analyses.

Cluster 1: Cheapest, most efficient, compact, small-engine vehicles Cluster

Cluster 2: standard vehicles with standard specifications and standard prices

Cluster3: Diesel, four-door sedans, or wagons that are reasonably expensive, large, and heavy, and have low RPM

Cluster 4: Most expensive, powerful, and powerful engines Cluster

Conclusion

Based on the results and considering their desired features and spending limits, people can make direct car choices.

Companies frequently provide their customers with advice on which qualities, based on their budget, they should purchase.