Case study on regression

Predicting price of pre-owned cars

Problem statement

Storm Motors is an e-commerce company who act as mediators between parties interested in selling and buying pre-owned cars.

For the year 2015-2016, they have recorded data about the seller and car including-

> *Specification details*
>
> *Condition of car*
>
> *Seller details*
>
> *Registration details*
>
> *Web advertisement details*
>
> *Make and model information*
>
> *Price*

Storm Motors wishes to develop an algorithm to predict the price of the cars based on various attributes associated with the car.

Variable description

Total size :50000 x 19

Data file : cars_sample.csv

Variables	Data Type	Description	Categories
dateCrawled	date	date when the ad first crawled, all field values are taken from this date	--
name	string	string consisting of car name, brand, model etc	combination of strings
seller	string	nature of seller	private, commercial
offerType	string	whether the car is on offer or has the buyer requested for an offer	offer, request
price	integer	price on the ad to sell the car ($)	--
abtest	string	two versions of ad	test, control
vehicleType	string	types of cars	cabrio, suv, coupe and 5 more
yearOfRegistration	integer	year in which was first registered	--

Data Science   ‹#›

Variable description

Total size :50000 x 19

Variables	Data Type	Description	Categories
gearbox	string	type of gearbox	manual or automatic
powerPS	integer	power of the car (HP)	--
model	string	model type of the car	3er,xc_reihe and 248 more
kilometer	integer	number of kilometres the car has travelled	--
monthOfRegistration	integer (qualitative)	month of registration	1,2,3…,12
fuelType	string	types of fuel	petrol, diesel and 5 more
brand	string	make of car	bmw, mercedes and 38 others
notRepairedDamage	string	status of repair for damages if yes damages have not been rectified; if no damages were taken care of	yes, no

Variable description

Variables	Data Type	Description	Categories
dateCreated	date	date at which the ad at storm motor was created	--
postalCode	integer	postal code of seller	--
lastSeen	date	when the crawler saw this ad last online	--

Variable description

The variables can be grouped in to different buckets based on the information

Details	Variables
Specification details	*gearbox, power, fuelType*
Condition of car	*notRepairedDamaged, kilometer*
Seller details	*seller, postalCode*
Registration details	*yearOfRegistration, monthOfRegistration*
Make and model	*brand, model, vehicleType*
Advertisement details	*dateCrawled, name, abtest, dateCreated, lastSeen, offerType*

THANK YOU