Codebook for Cars Dataset

Data Overview

Credentials

This data set can be downloaded from Kaggle data sets.

The direct link to data is: https://www.kaggle.com/orgesleka/used-cars-database/downloads/autos.csv

Business goal

This data was collected to answer the following questions:

What are the main factors influencing second-hand cars price and selling time?

What are the factors influencing car sales in different postal codes/areas?

Data description

This data set is a data frame of 20 variables over 371540 rows. Each row represents a car selling deal in ebay.

Variables description

Variable Name	Description	Type	Possible values
dateCrawled	when this ad was first crawled, all field-values are taken from this date	character, can be date	dates, "%m/%d/%y" format
name	"name" of the car	character	the cars names, separated by _
seller	private or dealer	character	privat and gewerblich
offerType	whether an offer or petition	character	Angebot (offer) or Gesuch (petition)
price	the price on the ad to sell the car	Integer	0-2147483647
abtest	A/B test for the ad? Cars?	character	"test", "control" or "benzin"
vehicleType	Vehicle Type	character	9 types

yearOfRegistration	at which year the car was first registered	Integer	1000-9999
gearbox	The gearbox Type	character	manuell (manual) or automatik (automatic)
powerPS	power of the car in PS	Integer	0-20000
model	the cars model	character	name
kilometer	how many kilometers the car has driven	character, can be numeric	5000-150000
monthOfRegistration:	at which month the car was first registered	Integer	1 to 12
fuelType	the car's fuel type	character	one of 7 types
brand	the car's brand	character	one of 40 types
notRepairedDamage	if the car has a damage which is not repaired yet	character	ja (yes) or nein (no)
dateCreated	the date for which the ad at ebay was created	character, can be date	dates, "%m/%d/%y" format
nrOfPictures	number of pictures in the ad	Integer	all 0
postalCode	Postal code	Integer	1067-71546
lastSeen	when the crawler saw this ad last online	character, can be date	dates, "%m/%d/%y" format

https://www.kaggle.com/timucinanuslu/data-crunchers