

Business Data Analytics and Prediction – Exercise 1

Group 4: Rotem Zarur, Dor Heldstein, Lior Rosen ("Cars" Database)

Business model goal:

Investigate used car market and its dynamics to understand what motivates purchase decisions in the used car market.

Business questions that can be answered with this data:

1. What are the main factors influencing on second-hand cars price?
2. What are the main factors influencing on second-hand cars selling time?
3. What age of used-cars are the "hottest" in the market?
4. Which vehicle type are sold the most and which models are sold the quickest?
5. What are the factors influencing car sales in different postal codes/areas?

Who needs to review the business questions before you analyze?

The senior management of the business organization financing the analysis should review the business questions before they are being analyzed; that's because these business questions will derive the costs associated in the efforts of building the relevant models and investigating them, as well as the potential benefit to the organization once these insights have been presented.

Obviously, once reviewed and approved by senior management, it should be presented and approved also by the organization's Board of Directors.

Find academic article (patent or blog) that relates to similar data. What was the main conclusion in it? Can you look for similar conclusions in your data?

Here is an article in Forbes which deals with used car data:

<https://www.forbes.com/sites/gurufocus/2016/07/06/the-used-car-market-whats-driving-growth/#691c32e240f0>

The main conclusion is that the used car market is driven by two key factors:

1. the average price per used is in continuous increase in the last few years, as people are more interested in younger used cars which are more expensive.
2. Despite the increasing average used car value, sales volume isn't slowing down. The numbers of cars sold have been growing on a yearly basis.

As for conclusion #1 - We can indeed look at price trends over the years in our database, as well as whether such trend is related to the car year model.

As for conclusion #2 – as our database covers only about a year time, we will not be able to comment about growth in sales over yearly basis. We may be able to investigate trends in sales on a monthly basis though.

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>