Data Article

Used-car market dataset for Latvia 2018

Valerijs Skribans

Riga Technical University, Latvia

A R T I C L E   I N F O

Article history:
Received 15 November 2018
Received in revised form 5 December 2018
Accepted 21 December 2018
Available online 28 December 2018

A B S T R A C T

Used-car market monitoring was started in 2018, 03 January, when Federal Administrative Court in Leipzig (Germany) allowed cities and communes in Germany to impose bans for diesel cars in order to reduce the level of nitric oxide in the air. This decision can radically change used-car market in the EU, because Germany massive exports used cars to Eastern Europe. The database shows the influence of the Germany Court decision to the used-car market in Latvia, as part of the EU market. The database represents the used-car market monitoring in Latvia during 2018. In Latvia 20–22th. used-car are offered for sale each month. The database reflects the observation dynamics of 12 months. In the observations there are included main used-car characteristics such as price, first registration date, motor cubic capacity, motor type, gearbox type, mileage, color, producer (brand), model, body types, (following) technical inspection data, set (features) of cars and so on.

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Specifications table

| Subject area                        | Economics, Econometrics and Finance |
|-------------------------------------|-------------------------------------|
| More specific subject area          | Business and management, marketing, market study, econometrics |
| Type of data                        | Table                                |
| How data was acquired               | Market advertisements in internet monitoring with Data Mining technologies, for data collection used Rcrawler library |
| Data format                         | Raw                                  |

E-mail address: valerijs.skribans@rtu.lv

https://doi.org/10.1016/j.dib.2018.12.075
2352-2409/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Experimental factors
Data was grouped by months according to the collection process, no changes were made to the data.

Experimental features
Latvian used-car market not so big, so it is possible to observe market advertisements in internet with Data Mining technologies. There was collected average 21 thousands offers per month of used-car in Latvia in 2018 for 12 months.

Data source location
Latvia, European Union

Data accessibility
Data is attached with this article

Related research article
Skribans V., Hegerty, S.W. Data mining application for used-car market analysis (In press) [1]

Value of the data

- The data reflects the used-car market and preferences of cars buyers in Latvia.
- It is possible to analyze the effects of diesel-gate and German legal restrictions on the industry in Latvia, based on it, it is possible draw conclusions about the long-term sustainable development of the car manufacturing industry in EU (because Germany massive exports used cars to Eastern Europe, including Latvia). The data set confirms that in situation, when in Germany ratio of diesel cars decrease, in Latvia the ratio increase for used-car only. Based on data it possible forecast, when last diesel car in Latvia (and, probably, in Europa) will be utilized.
- Data can be used for illegal economy estimation in Latvia on used-car market (for it, it is needed to compare data with official institution data [2]).

1. Data

Database consists of Latvian used-car market advertisements monitoring, total 258.6 th. observations. The database fields are cars price, first registration date, motor cubic capacity, motor type, gearbox type, mileage, color, producer (brand), model, body types, (following) technical inspection date, set (features) of cars and so on.

2. Experimental design, materials, and methods

In Latvia, EU it is possible to observe market advertisements in internet with Data Mining technologies. There was collected average 21 thousands offers per month of used-car in Latvia, it is 3 times more offers then available in Latvian Vehicle Trade Register Information System [2], which is official state register, which was established on the basis of the LR Cabinet of Ministers Regulations No. 876 from 18 December 2007 [3]. This shows the advantage of the Data Science and Data Mining methods for collecting and providing information for different purposes.

The database was collected from biggest in the Latvia advertisement website www.ss.com [4] from its transport and cars section, by monthly repeated following R studio data scraping code:

```r
library(Rcrawler)
Rcrawler(Website = "https://www.ss.com/lv/transport/cars/", KeywordsFilter = c("auto"), no_cores = 4, no_conn = 4, 6, ExtractXpathPat = c("//[@id='tdo_8'"],
                      "/[@id='tdo_31']",
                      "/[@id='tdo_18']",
                      "/[@id='tdo_15']",
                      "/[@id='tdo_35']",
                      "/[@id='tdo_16']")
```
PatternsNames = c("Price","Model","First_registration_date","Motor", "Gearbox", "Mileage","Color","Body_types","Following_technical_inspection_date","Set_of_car")
Price < - unlist(lapply(DATA, `[`, 1))
Model < - unlist(lapply(DATA, `[`, 2))
First_registration_date < - unlist(lapply(DATA, `[`, 3))
Motor < - unlist(lapply(DATA, `[`,4))
Gearbox < - unlist(lapply(DATA, `[`,5))
Mileage < - unlist(lapply(DATA, `[`,6))
Color < - unlist(lapply(DATA, `[`,7))
Body_types < - unlist(lapply(DATA, `[`,8))
Following_technical_inspection_date < - unlist(lapply(DATA, `[`,9))
Set_of_car < - unlist(lapply(DATA, `[`,10))
mytable <- data.frame(Price, Model, First_registration_date, Motor, Gearbox, Mileage, Color, Body_types, Following_technical_inspection_date, Set_of_car)

completeFun <- function(data, desiredCols) {
  completeVec <- complete.cases(data[, desiredCols])
  return(data[completeVec, ])
}
write.csv(completeFun(mytable, "Price"), "auto.csv", row.names=F)

The data source represent biggest part of Latvian advertisements, the close competitor, reklama.lv [5], have has seven times less ads, so it is concluded, that the attached database fully represent Latvian used-car market in 2018. Collected data was compared with EU data [6].

In Riga Technical University monitoring process was started based on Latvian transport system leading experts [7–9] opinions about research topically for whole EU, Latvia and at regional level [10–12]. Parallel to data mining, RTU conduct different researches related both transport, logistic [13], international relations [14], petroleum products research [15] and several researches related with tax [16,17], education [18] and business [19]. For researches use different economical – mathematical methods, as optimization method [20], econometric method [21], ratio analyse method [22], as well as methods and approach development [23,24].

The relevance of the study is confirmed by the situation in the EU in November – December of 2018. In Belgium and France after diesel cars’ tax increase had start riots [25]. In Germany, the situation is not so tense, given that for Germany there is the possibility of exporting diesel cars to the Eastern European countries.

Transparency document. Supplementary material

Transparency document associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.12.075.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.12.075.
