

Featured Dataset

102

Used cars database

Over 370,000 used cars scraped from Ebay Kleinanzeigen



Orges Leka • last updated a year ago

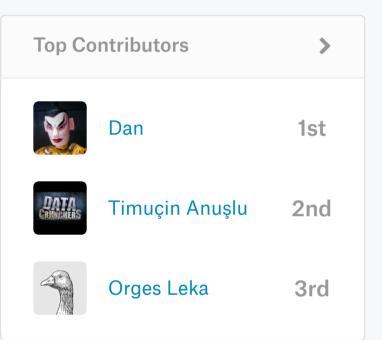
Overview Data Kernels Discussion Activity Download (18 MB) New Kernel

Tags

automobiles

medium

featured





Discussion	>
Earthquake Data Visual 3 days ago	5 replies
Problem converting the 7 days ago	1 reply
Get the clean Dataset i a month ago	4 replies

Description

Over 370000 used cars scraped with Scrapy from Ebay-Kleinanzeigen. The content of the data is in german, so one has to translate it first if one can not speak german. Those fields are included: **autos.csv**:

- dateCrawled: when this ad was first crawled, all field-values are taken from this date
- name: "name" of the car
- seller : private or dealer
- offerType
- price: the price on the ad to sell the car
- abtest
- vehicleType
- yearOfRegistration: at which year the car was first registered
- gearbox
- powerPS: power of the car in PS

- model
- kilometer: how many kilometers the car has driven
- monthOfRegistration: at which month the car was first registered
- fuelType
- brand
- notRepairedDamage: if the car has a damage which is not repaired yet
- dateCreated: the date for which the ad at ebay was created
- nrOfPictures: number of pictures in the ad (unfortunately this field contains everywhere a 0 and is thus useless (bug in crawler!))
- postalCode
- lastSeenOnline: when the crawler saw this ad last online

The fields lastSeen and dateCreated could be used to estimate how long a car will be at least online before it is sold.

brought to you by Orges Leka

Regression on average Price per Year based on this dataset

Table of value loss of an average used car per year

The second file is produced in MySQL from the first one through the query:

```
select
count(*) as count,
 kilometer,
yearOfRegistration,
20*round(powerPS/20) as powerPS,
min(price) as minprice,
max(price) as maxPrice,
avg(price) as avgPreis,
sqrt(variance(price)) as sdPreis from items where
    yearOfRegistration > 1990 and yearOfRegistration < 2016
   and price > 100 and price < 100000
    and powerPS < 600 and powerPS > 0
group by yearOfRegistration, round(powerPS/20),kilometer
having count > 10
into outfile '/tmp/cnt_km_year_powerPS_minPrice_maxPrice_avgPrice_sdPrice.csv'
fields terminated by ',' lines terminated by '\n';
```

Happy Coding!

Released Under CCO: Public Domain 1

Did you find this Dataset useful?
Show your appreciation with an upvote























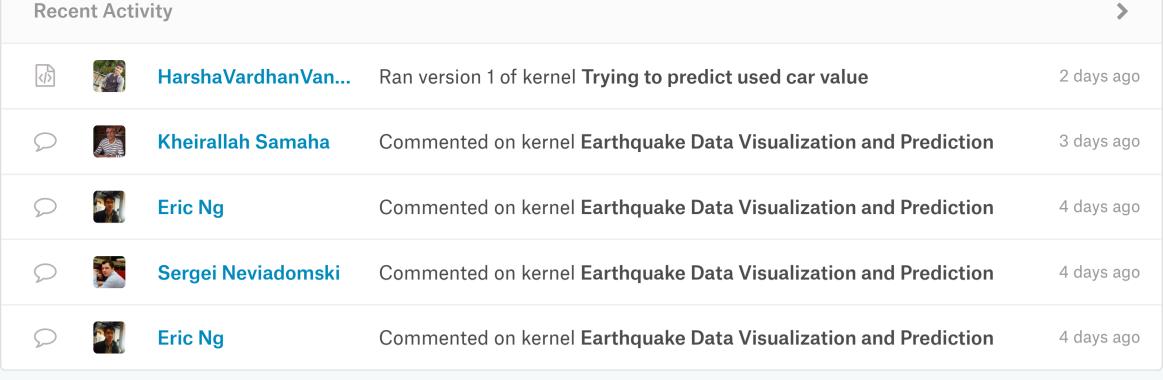












Similar Datasets



1.6M accidents & traffic flow over 16 years



Caravan Insurance Challenge



Car Sale Advertisements



New Car Sales in Norway



Parking Violations, December 2015

62,384 views - 8,682 downloads - 21 topics

© 2017 Kaggle Inc

Our Team Terms Privacy Contact/Support



Loading [Contrib]/a11y/accessibility-menu.js