

Analyzing Used Car Listings on eBay Kleinanzeigen

This presentation will explore a dataset of used car listings from eBay Kleinanzeigen, a classifieds section of the German eBay website. We will analyze the data to gain insights into the used car market in Germany.

 by Siya Beniwal





Data Overview

Data Source

The dataset was originally scraped and uploaded to Kaggle. We are working with a sample of 50,000 data points prepared by Dataquest.

Data Dictionary

The dataset includes 20 columns, such as dateCrawled, name, seller, offerType, price, abtest, vehicleType, yearOfRegistration, gearbox, powerPS, model, odometer, monthOfRegistration, fuelType, brand, notRepairedDamage, dateCreated, nrOfPictures, postalCode, and lastSeen.

Data Cleaning

1 Column Names

We will change the column names from camelcase to snakecase and update some wordings for clarity.

2 Dropping Unnecessary Columns

We will drop columns like num_photos, seller, and offer_type, as they contain mostly one value.

3 Cleaning Numeric Columns

We will clean and convert the price and odometer columns, which are numeric values stored as text.

colived Laimis
le
gleshs:
akatloy:
wis erlets:
t:
vooctomcy deigg:
olle staper:
olfe ctange:
lcrating:
olle hesw
virs'er Istens:
nst bla:
Whiceler vasicctple:
ouper fale
tangs
vailons
or hoves casper
age
erlotieu:
vriertellopd:
die Lipo
ti canges
lip tale
enate weigst
coume:
reilting
loer tupont
licreutsc-chalg eavals:
ne ipo
encrises lear clons:
titles
urhons naies
of the wouialle spele:
albesonley taldo:
ne taje
esriogs
rtaisceat vaicelopdi:
clogs
l lo logp:
v cath vactee
vati:
e tange:
e pareinstare:
ealuge:
ieflaetr scnehing:
achill vacke:
eilectlog:
eology
e rrtogy:
aleconaup:

} clean column nome,
past, chilst, tape.



Exploring Odometer and Price

Odometer

The odometer values are rounded, suggesting pre-set options. There are more high mileage vehicles than low mileage vehicles.

Price

Prices seem rounded, with a wide range of values. There are 1,421 cars listed with \$0 price, and a few listings with very high values.

Price		\$9470	\$91319	\$9210
\$1050	3200	9.190	3.960	3,670
\$.2.50	\$000	1.50	1,160	1.110
\$.3.50	3000	1,60	1,300	4,10



Cleaning Price Data

We will remove listings with prices above \$350,000, as they seem unrealistic. We will also keep listings with \$1 price, as they could be legitimate opening bids on an auction site.

30

25



Exploring Date Columns

date_crawled

The site was crawled daily over a one month period in March and April 2016.

ad_created

Most ad created dates fall within 1-2 months of the listing date, but a few are quite old.

1

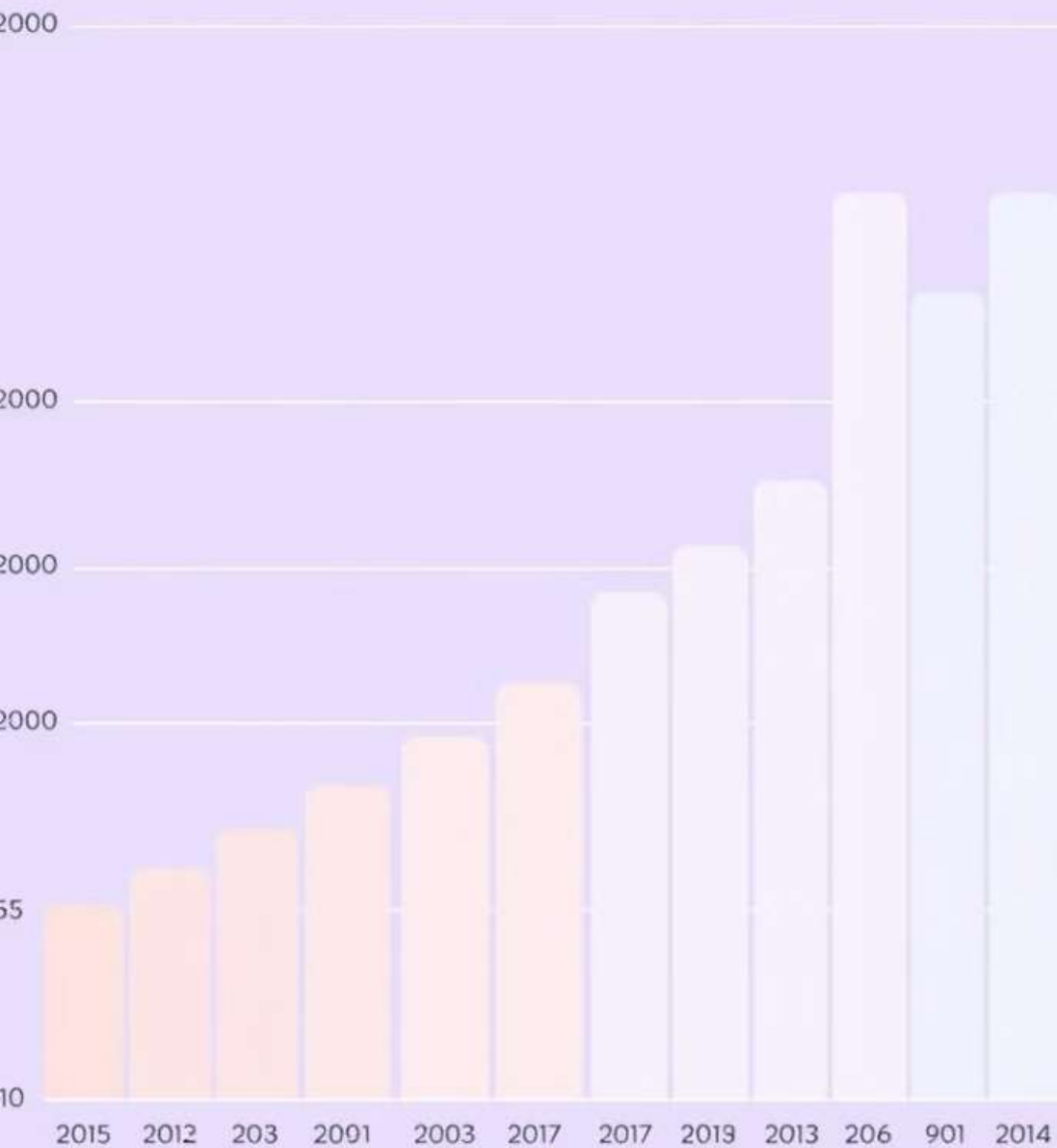
2

3

last_seen

The crawler recorded the date it last saw any listing, allowing us to determine when a listing was removed.

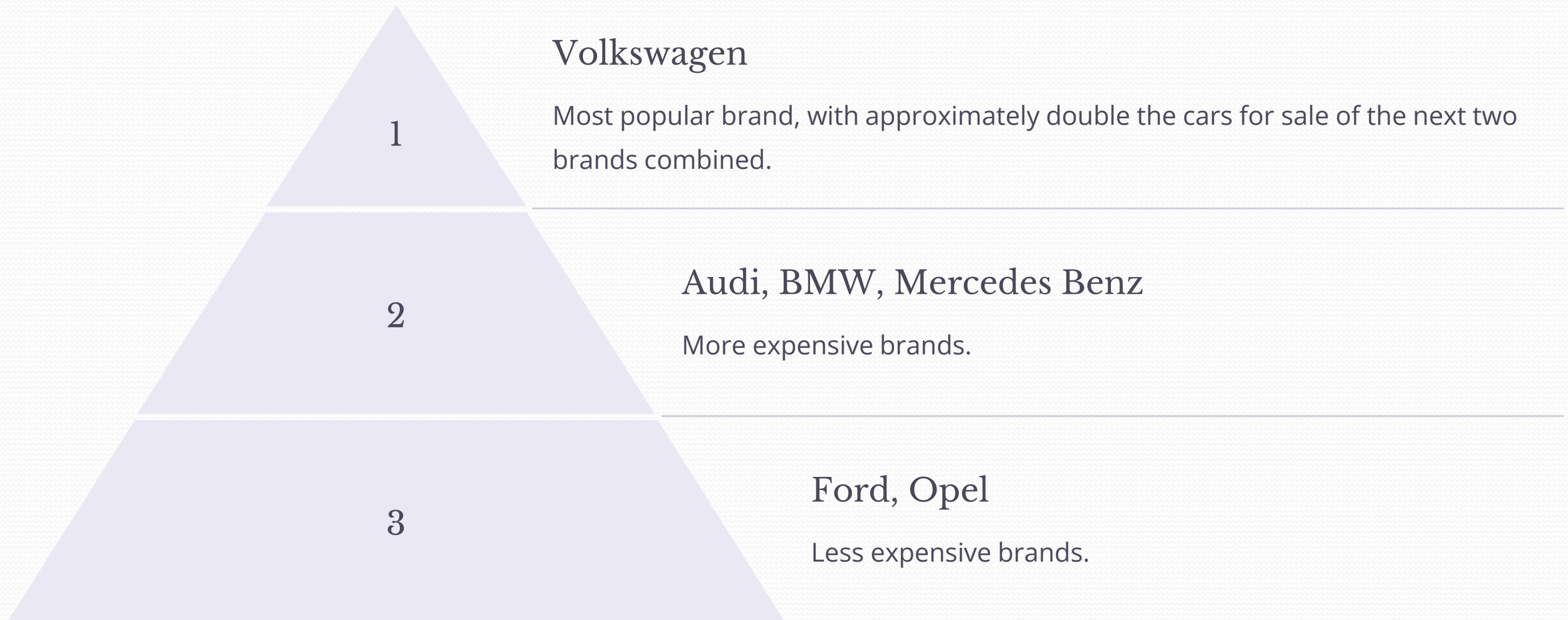
Registration, mad;
waldiny . yest



Dealing with Incorrect Registration Year Data

The yearOfRegistration column contains some odd values, including years before cars were invented and years in the future. We will remove listings with registration years above 2016, as a car cannot be registered after it was last seen.

Exploring Price by Brand





Exploring Mileage

The range of car mileages does not vary as much as the prices do by brand. There is a slight trend to the more expensive vehicles having higher mileage, with the less expensive vehicles having lower mileage.

Next Steps

Further Data Cleaning

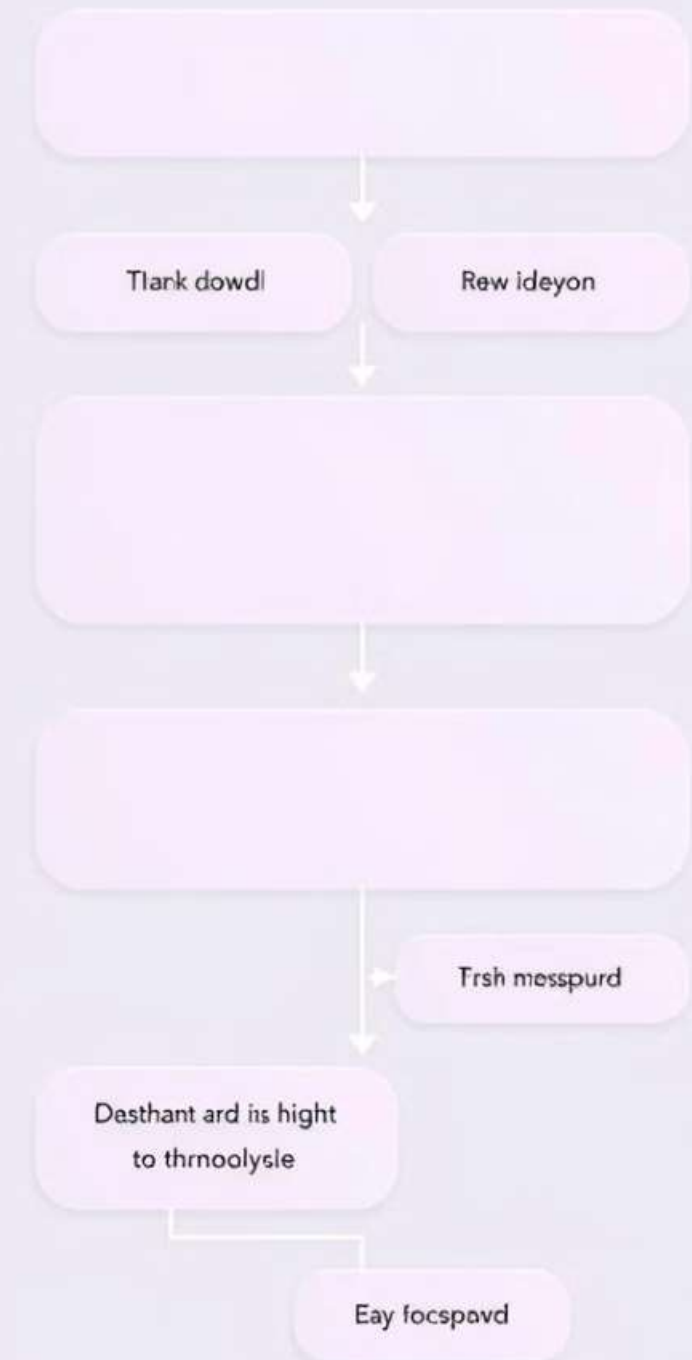
We will continue to clean the data, addressing any remaining inconsistencies or errors.

Feature Engineering

We will create new features from existing data, such as age of the car and price per kilometer.

Model Building

We will build predictive models to analyze the relationship between car features and price.





Conclusion

Price Trends

Used car prices have been steadily increasing in recent years. This trend is likely driven by factors such as supply chain disruptions and rising demand.

Popular Models

Certain models are more popular than others, with some experiencing higher demand than others. The data can be used to identify the most sought-after cars and their pricing trends.

Regional Variations

There are significant regional variations in car prices. Understanding these variations can be helpful for identifying potential opportunities or challenges.



Thank You

Do you have any Questions?