

[Case Study]

Assignment Fakes

We create the next.

Don't buy a fake!



Summary

Ordering goods at online marketplaces is the go-to choice for many consumers. If it is easier to buy and sell on a market platform, then more individuals and businesses offer their products, prices are lower, diversity of goods is higher – and the market platform makes more money.

However, with more potential customers, the prevalence of fake goods increases as well. Receiving a less-than-genuine product may just be annoying for the customer, but it has high costs for the platform. Identifying fake offerings, and removing them from the central inventory, is therefore a crucial task.

Challenges

The online platform *Amazing* challenges you to build a machine learning model, which can identify fakes with data they have collected. It is your task to build the model(s) for the customer. Keep in mind that it is important, that you provide the customer with the whole data science workflow rather than for instance a thorough data exploration. Note, you do not need to provide the customer with a saved model, an expressive code suffices. Also, do not get hung up on trivialities. If there is a column within the data that you do not understand – just leave it out.

Submission

Please take 10 minutes of your preparation time to write down what you would have done differently / additionally if you were given more time - expressive bullet points at the end of the code suffice. Please provide us your zipped code by attaching it as an email response to your interview invitation.

Data overview

| VARIABLE | DESCRIPTION | ROLE |
|---------------------|--|---------|
| Fake | Identifies if the good is a fake, i.e. a product of inferior quality (=1) for the observed transaction | Target |
| RefId | Unique (sequential) number assigned to transactions | Feature |
| PurchDate | The date the good was purchased at the auction | Feature |
| FulfillmentType | Whether good was sold from the “Marketplace“, via “Auction“ or via the “Business“ platform | Feature |
| MarketDate | The year this good appeared on the platform | Feature |
| ProductAge | The years elapsed since the good was first sold on the platform | Feature |
| AveragePrice | The average price of the good on the platform | Feature |
| ConfirmedPrice | The price of the good for a sample that was confirmed as genuine by the platform | Feature |
| ProductName | Name of the good on the platform | Feature |
| Category | A broad product category | Feature |
| Brand | Given brand of the product | Feature |
| VNST | City from which the seller originated | Feature |
| TransactionPrice | Price for the observed transaction | Feature |
| VNZIP | Zipcode from which the seller originated (pseudonymized) | Feature |
| product_description | Excerpt from the product description posted on the platform | Feature |
| top_review | Top rated review or latest review (if no review is rated) if available | Feature |

> good.luck()