← Data leakages Graded Quiz - 8 min Exploratory data analysis
EDA examples
USA examples
USA examples
Used because
Used because
Used because
Used because of the color base
dense
Used because of the color base
texts
Used because of the color
Used because of the color
Used because
Used beca Data leakages Data leakages Suppose that you have a credit scoring task, where you have to create a ML model that approximates expert evaluation of an individual's credit-worthiness. Which of the following can potentially be a data leakage? Select all that apply. First half of the data points in the train set has a score of 0, while the second half has scores > 0. Grade
100%
We keep your highest score Among the features you have a company id, an identifier of a company where this person works. It turns out that this feature is very important and adding it to the model significantly improves your score. 6 P P What is the most fedgroof way to set up a time series competition?
 Solt train public and private parts of data by time. Remove all features except libs (e.g. timestamp) from test sets on that persocipants will generate all the features based on part and join them stematics. momentees.

Make a time based split for traininest and a zandom split for publiciprivate.

Split train, public and private parts of data by time. Remove time variable from test set, keep the features. Suppose that you have a binary destification task being evaluated by logloss metric. You know that
there are 3600 rows in profit chools of test set and that constant 3.2 prediction gives the public
there are 3600 rows in profit chools of 0.44 What is the mean of target verticals in profit
part of test data (up to 4 decimal places)? 4. Suppose that you are solving image classification task. What is the label of this picture?

