* Table of content
* Hotel review problem
  + Cleaning
  + Exploration
  + Feature engineering
  + Data
  + Sentiment
  + Hotel review data
    - Shape
    - Format
    - # feature
    - # observation
    - Integer
* Score distribution
  + Round
  + Bar
  + Imbalance
  + Sample
* Data cleaning
* Feature engineering
  + Tokenizer function
    - Identify
    - Punctuation
    - Split
      * Space
    - Stop word
    - Stem
  + Text vectorization
    - Identify
      * Train
    - Quantify
      * Train
      * Test
  + Tokenizer validation
  + Positive repetition
  + Concatenation of vectoirzed review
    - Replace
* Logistic regression
  + Feature
  + Accuracy