## Kaggle

### Competitions: https://www.kaggle.com/competitions?search=tag:%27finance%27

Featured: these are public competitions with significant prize money meant to solve commercial problems. Prize winners grant the sponsor a non-exclusive license to their work, and will present their results via a detailed write-up and possible screencast (see below).

Playground: are public competitions set up to be fun, quirky and idea-driven, rather than to solve a specific business or research problem. These are for fun and not for any prize.

#### Playground

* DonorsChoose.Org Application Screening: https://www.kaggle.com/c/donorschoose-application-screening
  + Predict whether teacher’s proposals are accepted
  + The goal of the competition is to predict whether or not a DonorsChoose.org project proposal submitted by a teacher will be approved, using the text of project descriptions as well as additional metadata about the project, teacher, and school.
  + Comes with a tutorial that provides an overview of the dataset as well as a quick-start guide to building my first model in TensorFlow and submitting my entry to Kaggle
  + Is a Machine Learning course
  + Competition ended 6 months ago
  + Competition sponsored by Donorschoose.org and Google Machine Learning Crash Course

#### Featured

* Home Credit Default Risk: <https://www.kaggle.com/c/home-credit-default-risk>
  + Finished 2 months ago, ran for 1 week with 70k of prize money
  + Predict clients’ repayment abilities
  + Sponsored by Home Credit International
* Santander Product Recommendation : <https://www.kaggle.com/c/santander-product-recommendation>
  + predict which products their existing customers will use in the next month based on their past behavior and that of similar customers
  + $60k of prize money
  + Ran for about 1 week in 2016

### Datasets:

* Credit Card Fraud Detection: https://www.kaggle.com/mlg-ulb/creditcardfraud/home
  + The datasets contains transactions made by credit cards in September 2013 by european cardholders.
  + we have 492 frauds out of 284,807 transactions (unbalanced dataset)
  + features not original: transformed via PCA to preserve confidentiality; only time, amount, and Class (fraud or not) are preserved
* Daily News for Stock Market Prediction: <https://www.kaggle.com/aaron7sun/stocknews/home>
  + Data is news data (from reddit) and stock data from DJIA
  + Binary classification task: DJIA rose or stayed the same
  + Guy said for his NLP and Deep Learning course, so maybe not good
* Lending Club Loan Data: <https://www.kaggle.com/wendykan/lending-club-loan-data/home>
  + Data from 2007-2015
  + Predict loan acceptance
* Default of Credit Card Clients Dataset: https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset/home
  + Default Payments of Credit Card Clients in Taiwan from 2005
  + Payment default prediction
  + 25 variables (including default payment (1=yes, 0=no)
  + Some ideas for exploration:
    - How does the probability of default payment vary by categories of different demographic variables?
    - Which variables are the strongest predictors of default payment?
  + US Consumer Finance Complaints: <https://www.kaggle.com/cfpb/us-consumer-finance-complaints>
    - Predict something about the company response from the complaint?