Data Science Project

Kiva

* Waiting for zip file to download
* Got loan data from kivatools.com – doesn’t have a default variable for loans, has default variables for partners but not enough explanatory variables
* Someone who already tried predicting loan rates using decision trees:
  + <https://blog.bigml.com/2013/09/10/model-of-the-week-predicting-kiva-loan-defaults/>
  + Data vis: <https://bigml.com/user/ashikiar/gallery/model/522a3814035d0729c10069f9>

Cycling

* Strava: doesn’t seem to be publicly available data
* Sydney cycling survey data: <http://data.gov.au/dataset/cycling-data/resource/a08080f9-1ac2-46ca-8737-30850ecf7b12>
* Action:
  + Other cycling websites data

DataKind projects

* DataKind git has lots of information on projects from past datadives
* For some of them the data might not be available
* The datadives seem to have projects which use concepts that we have learnt in class
* E.g,
  + Random forest for predicting stem rust: <https://github.com/datakind/PlantDiseaseSpread/blob/master/Prediction_model_for_stem_rust.ipynb>
  + Using clustering on training images: <https://github.com/datakind/GiveDirectly/blob/master/ipython/notebook.ipynb>
* Action:
  + Get data
  + Run the code and see if I understand
  + See if there is anything else I might be able to add

Other lists (see notes from notebook)

* see the past projects
* see slides the Louis talked through
* Go through git and see past projects