How Much Did It Rain? II - Kaggle Competition

Wiki page. All important documents related to things like- **domain** of this problem, **important links**, **how to run code**,etc. go here.

Stuff to read before starting (and for future refenrence)

- Data Description and Download page.
- Forums!!, should read the pinned post.
- GOOD article describing the domain of the problem Understanding polarimetric radar measurements .

All About Data!! <- Put all data related details in this page

- train.csv(~1.16 GB):
  - Before Removing Id's with all nulls in Ref column
    - Rows- 13,765,201
    - Unique ID's- 1,180,945
  - After Removing Id's with all nulls in Ref column
    - Rows- 9,125,329
    - Unique ID's- 731,556
- test.csv(~618 MB):
  - Before Removing Id's with all nulls in Ref column
    - Rows- 8,022,756
    - Unique ID's- 717,625
  - After Removing Id's with all nulls in Ref column
    - Rows- 5,161,185
    - Unique ID's- 426,094

Pre-Post Process Related

- **Atleast one ref value non-null**: "MAE is now being computed only on Ids where at least one of the Ref values is non-null". Someone also suggested to **remove rows from train that have ONLY ref column** value as 'null'. link-ignored ids, original thread, permalink.New Post- Ref Column Only

Important comments from Forums (to help in algo design)

- Some of the extremely high Expected values may be due to the melting of ice precipitation that has collected in the rain gauge, which would release a flood of water in a relatively short period of time. If the rain gauges are heated, then snow filling the gauge and then melting later should not be as much of an issue, but hail or graupel filling the top of the gauge could cause underestimation initially by blocking rain from entering the gauge, then overestimation due to melting and draining into the gauge. The real trick will be to train an algorithm to recognize these instances. link

R - important tutorials

- https://github.com/dmlc/xgboost/blob/master/R-package/vignettes/discoverYourData.Rmd

Software Download

- Python 3.5.0- make sure to download 64bit(file name:python-3.5.0-amd64-webinstall)
  - Python for Windows Guide
- R-Windows Binaries
- R-Studio Download
- Tortoise Git
- Git Binaries

Updated 2015-10-30