# **Folder Structure**

# **Main structure**

**project/**

**├── input/**

**├── ref/**

**├── result/**

**├── work/**

**└── script.ipynb**

# **Explanation**

1. Have the python script files that do things in the project root.
2. **[input/]**: Contains the original, or raw, data files. Contents in this folder should be treated as read-only.
3. **[ref/]**: Contains reference files, i.e. from research.
4. **[result/]**: Contains the final results and explanatory markdown files.
5. **[work/]**: The working directory, from my SAS days, was used to store temporary data files. I like having the main console code directly in the project folder outside of the sub-folders, and storing intermediate data that I may want to reference later in the work sub-folder.

# **Rules for naming script files**

1. 1\_dataPrepare\_[what]\_[when]\_[other comment].ipynb

example: 1\_dataPrepare\_review\_2018\_test.ipynb

1. 2\_featureExtract\_[what]\_[when]\_[other comment].ipynb

example: 2\_featureExtract\_user\_2018\_test.ipynb

1. 3\_model\_[what]\_[when]\_[other comment].ipynb

example: 3\_model\_xgBoost\_2018.ipynb

1. 4\_interprete\_[what]\_[when]\_[other comment].ipynb

example: 4\_interprete\_accuracy.ipynb

# **Details**

1. Root
   1. [2\_featureExtract\_user\_1819\_v1.ipynb]
2. [input/]
3. [ref/]
4. [result/]
5. [work/]
   1. [./user.csv]: users after 2018 with extracted features