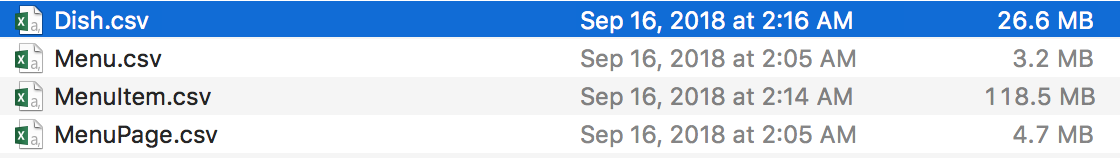
Run Time Models Experiment

1. Dataset:

Menu.csv from New York Public Library. [[Menu]](http://menus.nypl.org/data)

Data size: 3.2 MB

Data rows: 17545



1. Run Time Model Methodology
2. Sequential/Linear Run-time Model:

Raw Dataset A Operation B. ……. Clean Dataset X

1. Parallel Run-time Model:

Raw Dataset A

Operation B …… …..

Clean Dataset B

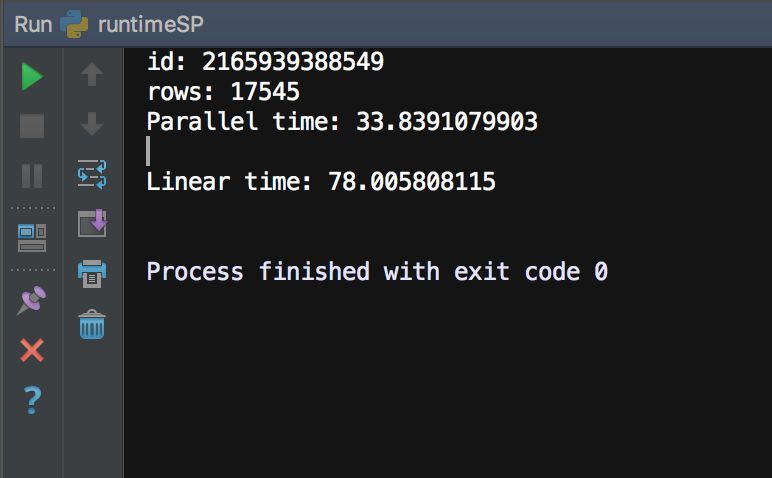
1. Implementation [[Github](https://github.com/LanLi2017/RuntimeCompare/blob/master/openrefine-client-master/runtime_compare.py) ]

*Linear Run-time Model:*

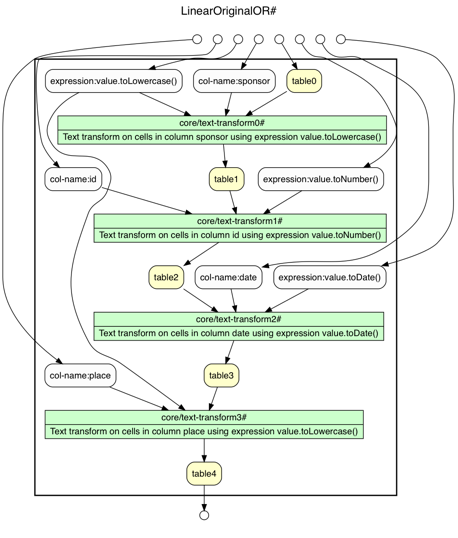
1. Read in csv file and set the project name
2. Create project with [OpenRefine-Client Library](https://github.com/opencultureconsulting/openre%20ne-%20client).
3. ***Set start time timer here***
4. Traverse the dictionaries in the JSON format csv file. And pass parameters to operations in OpenRefine-Client Library.
5. ***Set End time timer here***
6. Repeat step 4 for 1000 times. And get the difference of the End time and Start time.

*Parallel Run-time Model:*

1. Read in csv file and set the project name
2. Create project with [OpenRefine-Client Library](https://github.com/opencultureconsulting/openre%20ne-%20client).
3. Use [pool](https://docs.python.org/3.4/library/multiprocessing.html?highlight=process) to distribute processes (data parallelism)
4. ***Set start time timer here***
5. Do the operations (B,C,…) with the processes.
6. ***Set End time timer here***
7. Repeat step 5 for 1000 times. And get the difference of the End time and Start time.
8. Results.



1. Conceptual Model with Yesworkflow



Parallel model:

