OpenRefine Airbnb Submission and Autograding

What to submit

You need to submit two files:

- 1. airbnb_clean.csv: Export your clean file from OpenRefine and name it airbnb_clean.csv.
- 2. **airbnb_recipe.json:** Export your recipe. In OpenRefine operation history \rightarrow copy and paste the cleaning steps (your recipe) to a text editor \rightarrow save the text file with a json extension and name it **airbnb_recipe.json**

Note that the file names MUST BE exactly the same as mentioned above in order for the autograder to recognize them. Please DO NOT add any extensions nor your names to the file names.

How to submit

Submit the above two files on Coursera: Week 3 \rightarrow OpenRefine Homework \rightarrow My submission \rightarrow Upload the files accordingly on Coursera.

Autograding

Our autograder will evaluate your clean file by:

Clean file: We will compare your *airbnb_clean.csv* with our expected solution clean file and count the number of correct cells for each column. Your score for each column will be weighted by the fraction of correct cells in that particular column. Scores for each column are different based on difficulty level of the tasks. The table below lists the maximum score for each column.

In addition, for your recipe files:

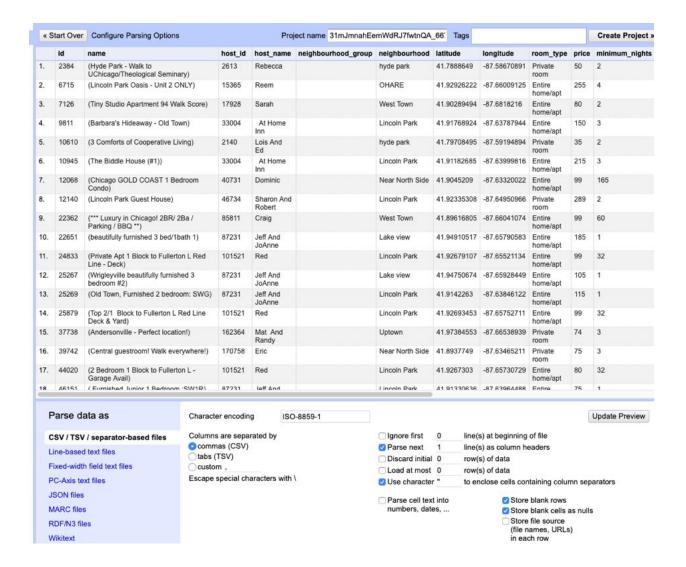
Recipe file: the Recipe file will not be graded. However, the TAs will randomly select some of your recipe files for a sanity check of your submissions. It will still show on Coursera that it's weighted as 1 point -- you can disregard this.

Grading will be based on your overall score (up to 100) of the clean file. If your overall score is 99.4%~99.7%, it might be due to different encoding schemes or special characters mismatches issues (See notes about the Encoding section below). If this is the case, you can try to perfect the score by changing to the ISO encoding scheme, but it is also okay if you do not do that. This is a known OpenRefine encoding difference, and Coursera will round up the scores to 100 if it's 99.5 and above.

Openrefine Encoding

As mentioned in the assignment instructions, do not remove other special characters that are not related to questions/instructions because grader will also look at the special characters and count any differences as a mismatch.

The server machine is using "ISO-8859-1" encoding, if you want to switch the encoding in OpenRefine, you can do so when you create new projects and import the file, as shown below:



Immediate feedback from autograder for each submission

During the assignment period, you can submit the files (clean file or recipe file, or both) as many times as you want and our autograder will give you feedback for every new submission. The feedback looks like the following table:

Column name	Score	Associated task
name	2	Trim spaces
name_grel	10	GREL: remove outer parentheses
name_grel_star	10	GREL: remove exclamation marks and asterisks
host_id	1	To Number
host_name	2	Trim spaces
host_name 1	6	Split column using regex And
host_name 2	6	Split column using regex And
neighbourhood	2	Trim spaces
neighbourhood_case	5	To titlecase
neighbourhood_loop	6	edit 'Loop'
neighbourhood_cluster	10	cluster 'O'Hare' and 'West Garfield Park'
latitude	1	To Number
longitude	1	To Number

room_type	2	Trim spaces
price	1	To number
price_crazy	10	Numeric facets >\$5000
minimum_nights	1	To Number
minimum_nights_long	10	Numeric facets > 300 nights
number_of_reviews	1	To Number
last_review	4	To Date
last_review_timeless	6	GREL: edit time
reviews_per_month	1	To Number
calculated_host_listings	1	To Number
availability_365	1	To Number
TOTAL	100	