OpenRefine Airbnb Submission and Autograding

What to submit

You need to submit two files:

1. ***airbnb_clean.csv***: Export your clean file from OpenRefine by clicking the *Export* button on the top right, then choose *comma-separated value*. Name the file *airbnb_clean.csv.*

2. ***airbnb_recipe.json:*** Click the *Extract* button from the *Undo/Redo* tab and copy the content on the right side into a preferred text editor. Name this file *airbnb_recipe.json.*

Note that the file names must be **exactly** the same as mentioned above in order for the autograder to recognize them! **Do not** add any extensions (or your name) to the file names!

How to submit

Submit the above two files on Coursera:

- Week 3 → OpenRefine Homework → My submission → Upload.

Autograding

Our autograder will evaluate the two files of your submission:

Clean file: We will compare your *airbnb_clean.csv* with our expected solution 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 the difficulty of the tasks. The table below lists the maximum score for each column.

In addition, for your recipe file:

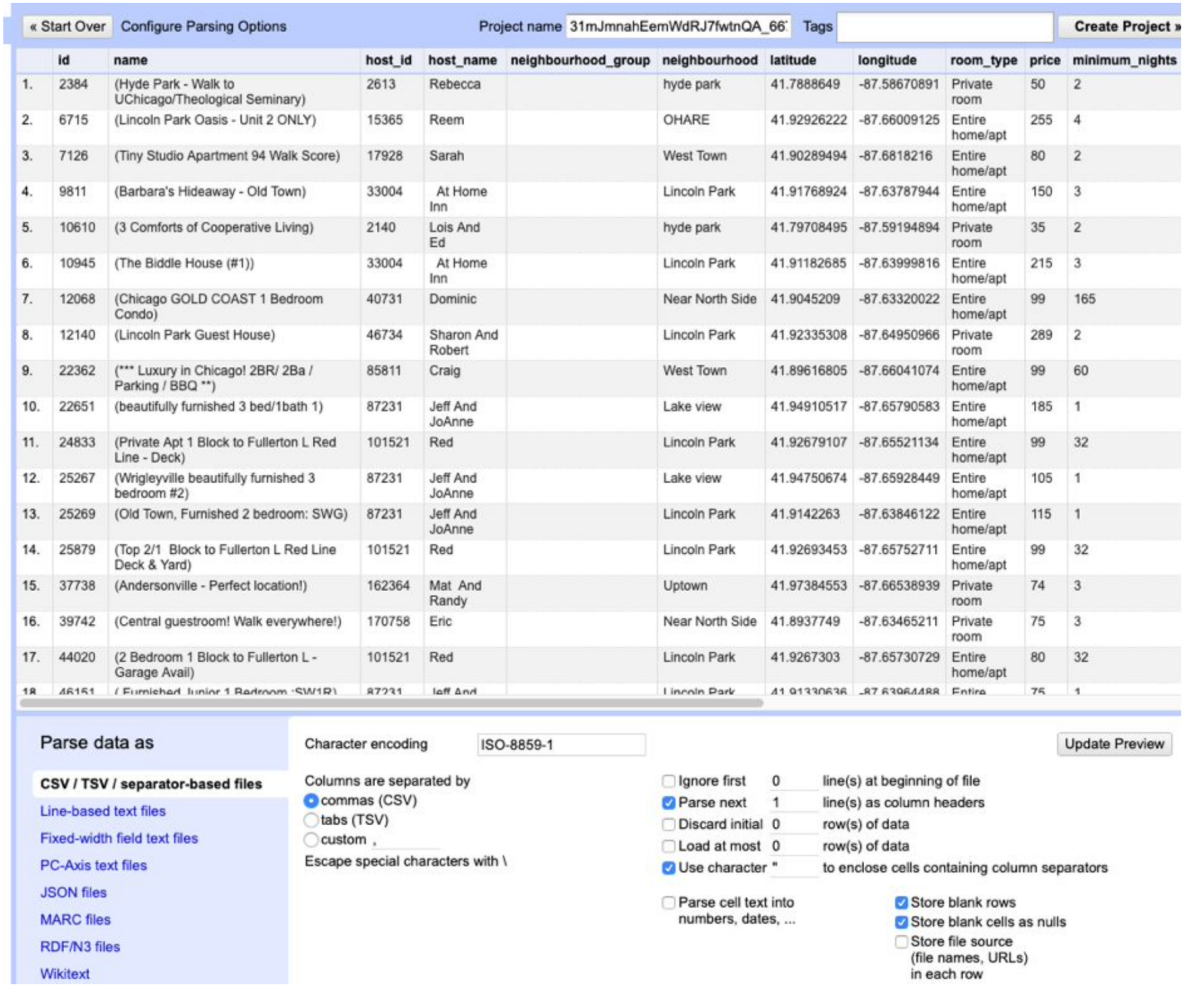
Recipe file: the recipe file will *not* be graded. However, the TAs will randomly select some of your recipe files for a "sanity check" of your submissions. It will still show on Coursera that it's weighted as 1 point (you can disregard this).

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 the autograder will also look at the special characters and count any differences as a mismatch!

The server machine is using an “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:



	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights
1.	2384	(Hyde Park - Walk to UChicago/Theological Seminary)	2613	Rebecca		hyde park	41.7888649	-87.58670891	Private room	50	2
2.	6715	(Lincoln Park Oasis - Unit 2 ONLY)	15365	Reem		OHARE	41.92926222	-87.66009125	Entire home/apt	255	4
3.	7126	(Tiny Studio Apartment 94 Walk Score)	17928	Sarah		West Town	41.90289494	-87.6818216	Entire home/apt	80	2
4.	9811	(Barbara's Hideaway - Old Town)	33004	At Home Inn		Lincoln Park	41.91768924	-87.63787944	Entire home/apt	150	3
5.	10610	(3 Comforts of Cooperative Living)	2140	Lois And Ed		hyde park	41.79708495	-87.59194894	Private room	35	2
6.	10945	(The Biddle House (#1))	33004	At Home Inn		Lincoln Park	41.91182685	-87.63999816	Entire home/apt	215	3
7.	12068	(Chicago GOLD COAST 1 Bedroom Condo)	40731	Dominic		Near North Side	41.9045209	-87.63320022	Entire home/apt	99	165
8.	12140	(Lincoln Park Guest House)	46734	Sharon And Robert		Lincoln Park	41.92335308	-87.64950966	Private room	289	2
9.	22362	(*** Luxury in Chicago! 2BR/ 2Ba / Parking / BBQ **)	85811	Craig		West Town	41.89616805	-87.66041074	Entire home/apt	99	60
10.	22651	(beautifully furnished 3 bed/1bath 1)	87231	Jeff And JoAnne		Lake view	41.94910517	-87.65790583	Entire home/apt	185	1
11.	24833	(Private Apt 1 Block to Fullerton L Red Line - Deck)	101521	Red		Lincoln Park	41.92679107	-87.65521134	Entire home/apt	99	32
12.	25267	(Wrigleyville beautifully furnished 3 bedroom #2)	87231	Jeff And JoAnne		Lake view	41.94750674	-87.65928449	Entire home/apt	105	1
13.	25269	(Old Town, Furnished 2 bedroom: SWG)	87231	Jeff And JoAnne		Lincoln Park	41.9142263	-87.63846122	Entire home/apt	115	1
14.	25879	(Top 2/1 Block to Fullerton L Red Line Deck & Yard)	101521	Red		Lincoln Park	41.92693453	-87.65752711	Entire home/apt	99	32
15.	37738	(Andersonville - Perfect location!)	162364	Mat And Randy		Uptown	41.97384553	-87.66538939	Private room	74	3
16.	39742	(Central guestroom! Walk everywhere!)	170758	Eric		Near North Side	41.8937749	-87.63465211	Private room	75	3
17.	44020	(2 Bedroom 1 Block to Fullerton L - Garage Avail)	101521	Red		Lincoln Park	41.9267303	-87.65730729	Entire home/apt	80	32
18.	46151	(Furnished Junior 1 Bedroom :SW1B)	87231	Jeff And		Lincoln Park	41.91330636	-87.63964488	Entire	75	1

Immediate feedback from autograder

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**	