</talentlabs>

Assignment 3 - Data Wrangling with Google Sheets

Instructions

1. Download the Airbnb NYC.csv data file and import it into Google Sheets.
2. Complete the tasks and answer the question
3. Please submit the assignment through TalentLabs Learning System. You will need to submit a zip file which contains this word document (with answers) and the data wrangled submission.csv file.

Question 1:

Organise the dataset and paste a screenshot of the last 3 columns and the first 3 rows below:

	N	O	P
ew	reviews_per_month	calculated_host_listings_count	availability_365
19	0.21	6	365
21	0.38	2	355
		1	365
05	4.64	1	194
19	0.1	1	0
22	0.59	1	129
05	0.4	1	0
24	3.47	1	220
21	0.99	1	0
09	1.33	4	188
22	0.43	1	6
23	1.5	1	39
24	1.34	3	314
05	0.91	1	333
31	0.22	1	0
29	1.2	1	46
28	1.72	1	321
01	2.12	1	12
22	4.44	1	21
		2	249
28	0.07	1	0
01	1.09	6	347
01	0.37	6	364
02	0.61	6	304
19	0.73	2	233
23	1.37	2	85
		1	0
13	0.49	1	75
15	1.11	3	311
19	0.24	1	67
01	2.04	3	355
14	1.42	1	255
15	1.65	3	284
01	2.37	3	359

	A	B		C	D	
1	id	name		host_id	host_name	neighbo
2	2539	Clean & quiet apt home by the park		2787	John	Brookly
3	2595	Skylit Midtown Castle		2845	Jennifer	Manhat
4	3647	THE VILLAGE OF HARLEM....NEW YORK !		4632	Elisabeth	Manhat
5	3831	Cozy Entire Floor of Brownstone		4869	LisaRoxanne	Brookly
6	5022	Entire Apt: Spacious Studio/Loft by central park		7192	Laura	Manhat
7	5099	Large Cozy 1 BR Apartment In Midtown East		7322	Chris	Manhat

</talentlabs>

Question 2:

Inspect the dataset. What data cleaning operations needs to be carried out from the ones below? Select all that apply.

- ☑ A – Remove duplicate rows
- ☑ B – Handle missing values
- ☑ C – Correct data formats
- ☑ D – Drop irrelevant columns
- ☑ E – Fix inconsistent data entry
- ☑ F – Trim whitespaces
- ☑ G – Correct spelling errors
- ☑ H – Correct numerical errors

Question 3:

Columns M and N (last review and reviews per month) have missing values. Why do you think those values are missing? How would you suggest to handle those missing values?

> **Reason for Missing Values:**
> - The missing values in last review and reviews per month could be due to listings that have never been reviewed or newly added listings without any reviews yet.
>
> **Handling Missing Values:**
> - For last review, you can fill missing values with a placeholder date or leave them as NA if no review exists.
> - For reviews per month, you can fill missing values with 0, assuming no reviews have been made yet.

Question 4:

Clean the dataset and handle the missing values. Next, complete the following data enrichment operations:

1. Make all the ID column values 10 numbers in length (custom formatting – pad with 0s).
2. Make a new column. Name this column Clean. Search the Name column for the text "Clean". If the value is present, amend your new Clean column with the value 1. Else, 0.

</talentlabs>

3. Change the room type column, to only include one word - Private or Entire (instead of Private room and Entire home).
4. Filter the dataset for Brooklyn neighbourhood group only.

Once finished, export the dataset as a CSV file. Name the file submission.csv. Submit both this word document and the CSV file together in a zipped folder. Name the zip folder submission.zip.