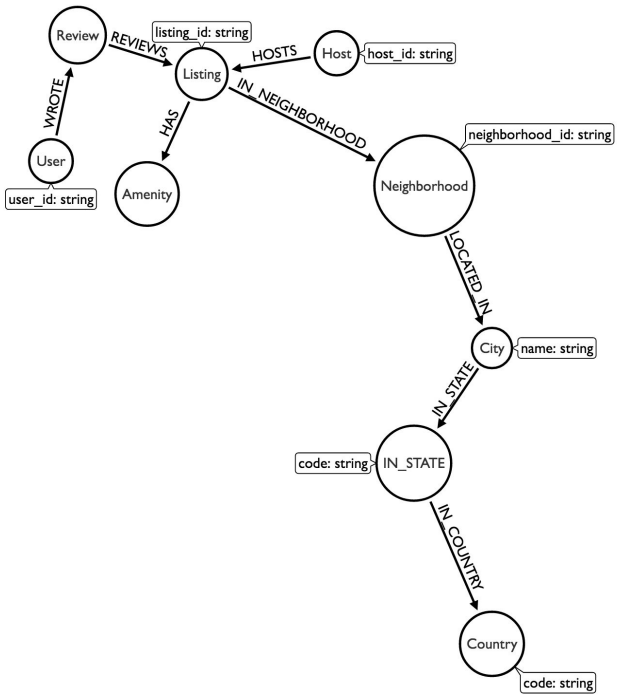
CS 327E Project 6, due Thursday, 10/29.

This project makes use of an Airbnb dataset which is modelled as follows:



Your job is to load the data into your Neoj database and then construct some queries on the Airbnb graph.

To start, open a terminal window in JupyterLab and download and extract the airbnb assets from Google Cloud Storage:

```
gsutil cp gs://cs327e-open-access/airbnb.zip .
unzip airbnb.zip
```

The extracted folder contains 3 files: `listings.csv`, `reviews.csv`, `data_load.cypher`.

Second, create a new Python Jupyter notebook and name it `project6.ipynb`.

Before loading any data, be sure your database is empty by running this command:

```
!$CONNECT "MATCH (n) DETACH DELETE n"
```

Of course, you'll need to set the `CONNECT` variable before running the above command!

Third, load the airbnb data into Neo4j as follows:

```
!cat /home/jupyter/airbnb/load_data.cypher | {CONNECT} --format plain
```

The script may take a few minutes to run. It outputs a count of the nodes which it has loaded.

Verify that all of the data has loaded correctly by retuning a total node count:

```
!{CONNECT} "MATCH (n) RETURN COUNT(n)"
```

You should get back 129,444 nodes.

Now, bring up a Neo4j Browser and explore the nodes and relationships visually.

Go back to your notebook and run a count for each node label in the graph.

You are now ready to construct some cypher queries. Translate the following questions into cypher and run them from your notebook.

Q1. How many hosts are located in "Austin, Texas, United States"?

Q2. Which listings does host_id = "4641823" have? Return the listing name, property_type, price, and availability_365 sorted by price. Limit the results to 10.

Q3. Which users wrote a review for listing_id = "5293632"? Return the user's id and name sorted alphabetically by name. Limit the results to 10.

Q4. Which users wrote a review for any listing which has the amenities "Washer" and "Dryer"? Return the user's id and name sorted alphabetically by name. Limit the results to 10.

Q5. Which listings have 3 bedrooms and are located in the Clarksville neighborhood? Return the listing name, property_type, price, and availability_365 sorted by price. Limit the results to 5.

Q6. Which amenities are the most common? Return the name of the amenity and its frequency. Sort the results by count in descending order. Limit the results to 5.

Q7. Which neighborhoods have the highest number of listings? Return the neighborhood's name and zip code (neighborhood_id) and number of listings they have sorted by number of listings in descending order. Limit the results to 5.

CS 327E Project 5 Rubric
Due Date: 10/29/20

Download and extract the airbnb assets to your jupyter notebook instance. **-5** no airbnb assets found in Jupyter instance	5
Create a new Python Jupyter notebook named `project6.ipynb`. **-3** incorrect file name	3
Run the data loader script (`load_data.cypher)` to populate the airbnb graph. **-10** script not run or run incorrectly	10
Run a query that returns a count for each node label. **-2** for each missing or incorrect count	12
Implement queries Q1 - Q7. **-10** for each missing or incorrect query **-7** for each missing output from query	70
`project6.ipynb` pushed to your group's private repo on GitHub. Your project **will not** be graded without this submission.	**Required**

`submission.json` submitted into Canvas. Your project **will not** be graded without this submission. The file should have the following schema:
 	
`{`
`  "commit-id": "your most recent commit ID from GitHub",`
`  "project-id": "your project ID from GCP"`
`}`
 	
Example:
 	
`{`
`  "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9",`
`  "project-id": "some-project-id"`
`}`
	Required

Total Credit:	**100**