Beepi – Data Scientist Test

The **`data.csv`** file contains 3 columns of values corresponding to keywords 1 (kw1), keywords 2, and frequency (count). This specifies the frequency of searches in which kw1 was followed by kw2. You may see that we have duplicated kw1 → kw2 as kw2 → kw1 so order does not matter when describing related searches.

The **`makes.csv`** file contains a unique MakeID for each car Make. The **models.csv** file contains the MakeID for each model Name. Note that it is possible for a given model Name to exist for multiple MakeID.

Please construct a keyword network for all unique make-model kw1 in the data.csv file. For each make-model pair, provide an ordered list of the most frequent search terms based on the count values. This keyword network should be only for make-model search terms. For example:

- "used honda accord" should be interpreted as "honda accord", a make-model pair.
- "2012 red camry" should be interpreted as "toyota camry", a make-model pair.
- "TOYOTA PRIUS C" should be interpreted as "toyota prius c", a make-model pair.
- "2012 200" should not be interpreted as a make-model pair because the "200" model cannot be uniquely assigned to Audi or Chrysler.
- "used Honda" should not be interpreted as a make-model pair because there is no model information.
- "less than $20000" should not be interpreted as a make-model pair.

The final deliverable should be one output file containing the unique make-model pairs and corresponding ordered lists of make-model values that contain the most frequent make-model search terms in descending order. Each list's first element should be the same as the search term's make-model and there should be no more than 20 elements in total. You also do not need to worry about order when there is counts tie in the data.

Feel free to use any reasonable software/programming language and output format. The example below is one possibility for two make-models but you may find that the real lists may differ.

```
{
    "acura ilx": [
        "acura ilx",
        "honda civic",
        "acura tsx",
        "volkswagen jetta"
     ] ,
    "mitsubishi outlander": [
        "mitsubishi outlander",
        "acura rdx",
        "honda pilot",
        "nissan altima"
        "hyundai genesis",
        "volkswagen gti"
     ]
}
```

Please include all files relevant to your solution in a zip archive with the output file clearly labeled.

Beepi – SQL Test

Assume a MySQL database where the server time zone is in UTC.

tblSales

Column_Name	Data_Type
sale_id	integer
buyer_id	integer d
seller_id	integer
region	text
device	enum('android', 'iphone', 'web')
status	enum('reserved', 'sale pending', 'delivered')
reserve_dt	timestamp with time zone
delivery_dt	timestamp with time zone

tblClients

Column_Name	Data_Type
user_id	integer
email	text
firstname	text
lastname	text
role	enum('buyer', 'seller')
signup_dt	timestamp with time zone

Write SQL queries that would answer the following questions:

1. In the "Los Angeles" and "San Diego" regions, what was the number of cars that were reserved for sale but not delivered in March 2015? What does it look like by device type?
2. In order to get customer reviews on the sell side in "San Francisco" we need to know all car sales last week (between 3/23/2014 and 3/30/2015 inclusive) by sale_id and the corresponding name and email of the seller.
3. For each region what is the percentage of buyers who reserved a car within 30 days of signing up for an account at Beepi.com?