Homework 3 – Mongo

First 6 questions deal with Drinkers and Bars collections. Last one with Penna.

SUBMIT: Mongo expressions, JSON object or code (last question)

1. Return the bar if either its phone or address is an empty string

2. Find the city that has more than 4 bars. Return the city name and the number of bars it has.

3. Return how many bars sell more than 5 kinds of beers.

4. Find the drinkers that have visited any bars either on Saturday or Sunday (or both)   [hint: go check out "$elemMatch" function]:

5. Find the drinker who has ordered "Blue Tattoo" beer more than once (in the drinker's history)

6. Insert Lucy to Drinker collection. Lucy is from Edison, lives at "433 river Road" with phone number 732-571-9871, she is 23 years old and her list of favorite bar foods consists of French fries, Onion rings, Nachos. and Wings

  7. Value Count (lecture on Nov 18th, we called it value space, but I prefer **value count)**

INTRODUCTION:

The value count of a JSON object J is the number of atomic values which appear in J.

For example value count of this JSON object is 6

- {
  "employees":[
    {"firstName":"John", "lastName":"Doe"},
    {"firstName":"Anna", "lastName":"Smith"},
    {"firstName":"Peter", "lastName":"Jones"}
  ]
  }

Value count of relation "Penna" is 10 (number of columns) * 486413 = 4.86M, while Value count of JSON Penna representation:

$3110 * (156 * 5) + 3110 * 5 = 2.45M$

With 156 (on average) timestamps per precinct.

Simple Example, Here is JSON representation of R with 7 tuples and value count of 7 x 4 =28

```
[{  "A": "1",  "B": "1",  "C": "1",  "D": "1" },
 {  "A": "1",  "B": "0",  "C": "1",  "D": "1" },
 {  "A": "1",  "B": "0",  "C": "0",  "D": "1" },
 {  "A": "1",  "B": "0",  "C": "1",  "D": "0" },
 {  "A": "0",  "B": "1",  "C": "1",  "D": "0" },
 {  "A": "0",  "B": "1",  "C": "1",  "D": "1" },
 {  "A": "1",  "B": "1",  "C": "0",  "D": "1" }]
```

But we can do better, since there are redundancies in R, 10 is repeated 3 times in AB, and 01 is repeated 2 times. This leads to

more compact representation of R with 3 JSON objects with two object "attributes": AB and CD

J=[{   "AB": 11,   "CD": [11]},

 {   "AB": 10,   "CD": [11, 01, 10]},

 {   "AB": 01,   "CD": [10,11]}]

With value count of J calculated as 4+8+6=18 (4 for the first object, 8 for second, and 6 for third)

JSON representation of a relation R will generally lead to more compact, reduced value count.

QUESTION 7

Given a relation R, represent R as JSON object J with the *smallest value count.*

R=

[{   "A": 1,   "B": 1,   "C": 0,   "D": 0 },

 {   "A": 1,   "B": 1,   "C": 0,   "D": 0 },

 {   "A": 1,   "B": 1,   "C": 0,   "D": 0 },

 {   "A": 1,   "B": 1,   "C": 1,   "D": 0 },

 {   "A": 0,   "B": 1,   "C": 1,   "D": 0 },

 {   "A": 0,   "B": 1,   "C": 0,   "D": 0 },

 {   "A": 0,   "B": 1,   "C": 1,   "D": 1 }]

8.   [Python/any other programming language plus Mongo] – Penna collection.

For each timestamp T, define TotIncrement as sum of totalvote increments over all precincts (totalvote increment, as defined in 2.1 of Election project (newPenna).

Finds timestamp(s) with largest value of TotIncrement along with this largest value.

Submit CODE + Result