Assignment #9

👀 Click on this link: https://classroom.github.com/a/bxgyYosC (https://classroom.github.com/a/bxgyYosC) to accept this assignment in GitHub classroom. This will create your homework repository. Clone your new repository.

In this homework, you'll:

1. Install MongoDB
2. Run basic MongoDB queries
3. Use the aggregation pipeline

Setup

1. Install the MongoDB server for your OS (https://www.mongodb.com/docs/manual/tutorial/install-mongodb-on-os-x/) (also link may be this (https://www.mongodb.com/try/download/community))
2. (Optional) Install MongoDB Compass (https://www.mongodb.com/products/compass) if you'd like to use a graphical client
3. Download NYC Jobs data as a csv (https://data.cityofnewyork.us/City-Government/NYC-Jobs/kpav-sd4t)
4. Import data
   - use `mongoimport` (see the slides for an example (../slides/mongo/intro.html#31))
   - or import with MongoDB Compass (https://www.mongodb.com/blog/post/import-export-your-data-with-compass)
5. In `queries.js`, write the queries specified in the section below (you can run them in your client, then copy and paste int `queries.js` using your text editor of choice)
   - please prefix each query with the appropriate number in a comment
   - `// 1.`

Queries

1. Find all of the possible values for the `Career Level` field. Only show each value once. Hint: use `distinct`:

```
[
  'Entry-Level',
  'Executive',
  'Experienced (non-manager)',
  'Manager',
  'Student'
]
```

2. Show the Agency, Business Title, and Salary Range To of the top 3 hourly jobs based on salary, using the "high" end of the hourly rate. Only show results for jobs posted (Posting Type) externally. ⚠️ **The example output below will not match, as the jobs data changes over time; this is just what the output may *look* like →**

```
[
  {
 Agency: 'DEPARTMENT OF CORRECTION',
 'Business Title': 'Staff Physician  (Part-Time)',
 'Salary Range To': 98.84
  },
  {
 Agency: "ADMIN FOR CHILDREN'S SVCS",
 'Business Title': 'Psychologist',
 'Salary Range To': 95
  },
  {
 Agency: 'FIRE DEPARTMENT',
 'Business Title': 'Telemetry Physician',
 'Salary Range To': 84.86
  }
]
```

3. Show the number of external (Posting Type) and full time (Full-Time/Part-Time indicator) job postings per agency, along with the name of the agency. Sort in descending order of count. ⚠️ **The example output below will not match, as the jobs data changes over time; this is just what the output may *look* like →**

```
[
  { _id: 'DEPT OF ENVIRONMENT PROTECTION', count: 292 },
  { _id: 'NYC HOUSING AUTHORITY', count: 206 },
  { _id: 'HRA/DEPT OF SOCIAL SERVICES', count: 204 },
  { _id: 'DEPT OF HEALTH/MENTAL HYGIENE', count: 188 },
  { _id: 'DEPARTMENT OF CORRECTION', count: 150 },
  { _id: 'DEPARTMENT OF TRANSPORTATION', count: 147 },
  { _id: "ADMIN FOR CHILDREN'S SVCS", count: 116 },
  { _id: 'DEPT OF DESIGN & CONSTRUCTION', count: 77 },
  { _id: 'HOUSING PRESERVATION & DVLPMNT', count: 74 },
  { _id: 'DEPT OF YOUTH & COMM DEV SRVS', count: 64 },
  { _id: 'OFFICE OF THE COMPTROLLER', count: 63 },
  { _id: 'TAXI & LIMOUSINE COMMISSION', count: 62 },
  { _id: 'OFFICE OF MANAGEMENT & BUDGET', count: 62 },
  { _id: 'LAW DEPARTMENT', count: 60 },
  { _id: 'FINANCIAL INFO SVCS AGENCY', count: 59 },
  { _id: 'BRONX DISTRICT ATTORNEY', count: 58 },
  { _id: 'DEPT. OF HOMELESS SERVICES', count: 56 },
  { _id: 'FIRE DEPARTMENT', count: 53 },
  { _id: 'DEPT OF CITYWIDE ADMIN SVCS', count: 41 },
  { _id: 'DEPT OF PARKS & RECREATION', count: 37 }
]
```

4. Modify the previous query so that only agencies with more than 100 external full time job postings are shown. ⚠️ **The example output below will not match, as the jobs data changes over time; this is just what the output may *look* like →**

```
[
  { _id: 'DEPT OF ENVIRONMENT PROTECTION', count: 292 },
  { _id: 'NYC HOUSING AUTHORITY', count: 206 },
  { _id: 'HRA/DEPT OF SOCIAL SERVICES', count: 204 },
  { _id: 'DEPT OF HEALTH/MENTAL HYGIENE', count: 188 },
  { _id: 'DEPARTMENT OF CORRECTION', count: 150 },
  { _id: 'DEPARTMENT OF TRANSPORTATION', count: 147 },
  { _id: "ADMIN FOR CHILDREN'S SVCS", count: 116 },
]
```

5. Modify the previous query so that it outputs:
   - the name of the agency field is agency
   - does not include the _id field
   - the agency is lowercase (use `$toLower` (https://www.mongodb.com/docs/manual/reference/operator/aggregation/toLower/))
   - ⚠️ **The example output below will not match, as the jobs data changes over time; this is just what the output may *look* like →**

```
[
  { count: 292, agency: 'dept of environment protection' },
  { count: 206, agency: 'nyc housing authority' },
  { count: 204, agency: 'hra/dept of social services' },
  { count: 188, agency: 'dept of health/mental hygiene' },
  { count: 150, agency: 'department of correction' },
  { count: 147, agency: 'department of transportation' },
  { count: 116, agency: "admin for children's svcs" }
]
```

6. Find the average high end salary (`Salary Range To`) per year based on the `Posting Date`
   - to do this, use `$split` and `$arrayElemAt` for extract the date
   - use `$group` and `$avg` to find the average per year
   - ⚠️ **The example output below will not match, as the jobs data changes over time; this is just what the output may *look* like →**

```
[
  { _id: '2020', 'Average Salary': 173486 },
  { _id: '2021', 'Average Salary': 94162.5731707317 },
  { _id: '2022', 'Average Salary': 96516.19672943509 },
  { _id: '2023', 'Average Salary': 91135.94427291729 }
]
```