DS4300 HW3 - Mongo!

Members: *Danish Farooq*

Database:

New York Restaurants from Professor Rachlin, source otherwise unknown

Questions, Queries, and Answers

1. *How Many McDonalds are there in New York?*

   Query

```
db["New York"].aggregate(
  [
    {
      $match: {
        name: "Mcdonald'S"
      }
    },
    {
      $count: "totalMcDonalds"
    }
  ]
)
```

   Results:

```
[
  {
    "totalMcDonalds": 208
  }
]
```

   So there are 208 McDonalds in New York City

2. *How many reviews does each restaurant have on average?*

   Query

```
const aggregate = [
  {
    $project: {
      _id:0,
      numReviews: {$size:"$grades"},
```

   Results:

```
[
  {
    "_id": 0,
    "avgReviews": 3.6855948578413975
  }
]
```

```
    }
  },
  {
    $group: {
      _id:0,
      avgReviews: {$avg:"$numReview
s"},
    }
  }
]
db["New York"].aggregate(aggregate)
```

So there are 3.68 reviews for each restaurant on average

3. *What are the most critically acclaimed restaurants (highest average scores) with more than the average number of reviews?*

   Query

```
const aggregate = [
  {
    $match: {
      "grades.3": {$exists: true}
    }
  },
  {
    $project: {
      _id: "$restaurant_id",
      totalScore:  {
        $reduce: {
          input: "$grades",
          initialValue: 0,
          in: { $add : ["$$value",
 "$$this.score"] }
        }
      },
      numReviews: {$size: "$grade
s"},
      name: true
    }
  },
  {
    $project: {
      averageScore: { $divide: ["$to
talScore", "$numReviews"]},
      numReviews: true,
      name: true
    }
  },
```

Result

```
[
  {
    "name": "Bella Napoli",
    "_id": "40393488",
    "numReviews": 5,
    "averageScore": 38.6
  },
  {
    "name": "Tenda Asian Fusion",
    "_id": "41267350",
    "numReviews": 5,
    "averageScore": 37.4
  },
  {
    "name": "Red Chopstick",
    "_id": "41602559",
    "numReviews": 7,
    "averageScore": 36.285714285714285
  },
  {
    "name": "El Mixteco",
    "_id": "41550400",
    "numReviews": 5,
    "averageScore": 34.8
  },
  {
    "name": "Concrete Restaurant",
    "_id": "41363541",
    "numReviews": 4,
    "averageScore": 34.75
  }, ...
]
```

```
  {
    $sort: { "averageScore": -1 }
  }
];

db["New York"].aggregate(aggregate)
```

4. *What is the frequency of McDonalds across each borough?*

   Query

```
const aggregate = [
    {
        $match: {name: "Mcdonald'S"}
    },
    {
        $group: {
            _id: "$borough",
            count: { $sum: 1}
        }
    }
]

db["New York"].aggregate(aggregate)
```

Oh lucky Staten Island

5. *What ZIP Code in Brooklyn has the highest reviewed restaurants on average (with more than 4 reviews)?*

   Query

```
const aggregate = [
  {
    $match: {
      "grades.3": {$exists: true},
      "borough": "Brooklyn",
      "cuisine": { $regex: "Italian"
```

So the best restaurant to go to is Bella Napoli

Result


```
[
  {
    "_id": "Queens",
    "count": 54
  },
  {
    "_id": "Bronx",
    "count": 42
  },
  {
    "_id": "Brooklyn",
    "count": 48
  },
  {
    "_id": "Staten Island",
    "count": 8
  },
  {
    "_id": "Manhattan",
    "count": 56
  }
]
```


Result


```
[
  {
    "_id": "11249",
    "zipcode": "11249",
    "avgReviewScore": 14.25
  },
  {
```

```
}
    }
  },
  {
    $project: {
      _id: "$restaurant_id",
      totalScore:  {
        $reduce: {
          input: "$grades",
          initialValue: 0,
          in: { $add : ["$$value",
 "$$this.score"] }
        }
      },
      numReviews: {$size: "$grade
s"},
      name: true,
      zipcode: "$address.zipcode"
    }
  },
  {
    $group: {
      _id: "$zipcode",
      totalScore: { $sum: "$totalSco
re" },
      totalReviews: { $sum: "$numRev
iews" }
    }
  },
  {
    $project: {
      zipcode: "$_id",
      avgReviewScore: { $divide:
 ["$totalScore", "$totalReviews"]}
    }
  },
  {
    $sort: { "avgReviewScore": -1 }
  }

];

db["New York"].aggregate(aggregate)
```

6. *Which Borough has the highest rated restaurants with more than the average number of reviews?*

   Query

```
    "_id": "11222",
    "zipcode": "11222",
    "avgReviewScore": 13.833333333333334
  },
  {
    "_id": "11207",
    "zipcode": "11207",
    "avgReviewScore": 13.3
  },
  {
    "_id": "11233",
    "zipcode": "11233",
    "avgReviewScore": 13
  },
  {
    "_id": "11225",
    "zipcode": "11225",
    "avgReviewScore": 13
  },
 ...
]
```

So the best Italian restaurants in Brooklyn are in the ZIP Code 11249, which is in Williamsburg by the East River.

Result

```
const aggregate = [
  {
    $match: {
      "grades.3": {$exists: true},
      "borough": {$not: {$regex: "Mi
ssing"}}
    }
  },
  {
    $project: {
      _id: "$restaurant_id",
      totalScore:  {
        $reduce: {
          input: "$grades",
          initialValue: 0,
          in: { $add : ["$$value",
 "$$this.score"] }
        }
      },
      numReviews: {$size: "$grade
s"},
      name: true,
      borough: true
    }
  },
  {
    $group: {
      _id: "$borough",
      totalScore: { $sum: "$totalSco
re" },
      totalReviews: { $sum: "$numRev
iews" }
    }
  },
  {
    $project: {
      avgReviewScore: { $divide:
 ["$totalScore", "$totalReviews"]}
    }
  },
  {
    $sort: { "avgReviewScore": -1 }
  }

];

db["New York"].aggregate(aggregate)
```

```
[
  {
    "_id": "Queens",
    "avgReviewScore": 11.885062265387594
  },
  {
    "_id": "Brooklyn",
    "avgReviewScore": 11.734872657655044
  },
  {
    "_id": "Manhattan",
    "avgReviewScore": 11.705527127101675
  },
  {
    "_id": "Staten Island",
    "avgReviewScore": 11.67306052855925
  },
  {
    "_id": "Bronx",
    "avgReviewScore": 11.134045243782493
  }
]
```

The results are pretty even, but I guess the best place to go is Queens!

7. *What are the restaurants around 2 kms of Jackson Heights, NY?*

Query

```
db["New York"].find(
    {
        "address.coord" : {
            $near : {
                $geometry : {
                    coordinates : [
 -73.8831, 40.7557 ] },
                    $maxDistance : 2
000
                    }
                }
    },
    {
        "name": 1,
        "cuisine": 1,
        "_id": 0,
        "zipcode": "$address.zipcod
e",
        "coords": "$address.coord"
    }
)
```

We first need to create a geospatial index on address.coords, specifying that it is a 2D sphere

Results


```
[{
    "cuisine": "Chinese",
    "name": "Ming Wok Kitchen",
    "zipcode": "11372",
    "coords": [
      -73.88329340000001,
      40.7557171
    ]
  },
  {
    "cuisine": "Latin (Cuban, Dominican,
Puerto Rican, South & Central America
n)",
    "name": "D'Antigua Restaurant",
    "zipcode": "11372",
    "coords": [
      -73.8828671,
      40.7555714
    ]
  },
  {
    "cuisine": "Latin (Cuban, Dominican,
Puerto Rican, South & Central America
n)",
    "name": "Pio-Pio",
    "zipcode": "11372",
    "coords": [
      -73.8834287,
      40.7557017
    ]
  },
  {
    "cuisine": "Latin (Cuban, Dominican,
Puerto Rican, South & Central America
n)",
    "name": "Rubi Rosa Restaurant & Ba
r",
    "zipcode": "11372",
    "coords": [
      -73.8831979,
      40.756027
    ]
  },
  {
    "cuisine": "Peruvian",
    "name": "Amaru",
    "zipcode": "11372",
    "coords": [
      -73.8830342,
      40.7560464
```

```
    ]
  }, ...]
```

8. *What are the frequencies of cuisine types around Jackson Heights, NY?*

   Query

```
const aggregate = [
    {
        $geoNear: {
            near: { type: "Point", coordin
ates: [ -73.8831, 40.7557 ] },
            maxDistance: 2000,
            query: { "grades.3": {$exists:
true} },
            distanceField: "calcDistance"
        }
    },
    {
        $group: {
          _id: "$cuisine",
          count: { $sum: 1},
        }
    },
    {
        $sort: {
            "count": -1
        }
    },
    {
        $limit: 5
    }
]

db["New York"].aggregate(aggregate)
```

   Result

```
[
  {
    "_id": "Latin (Cuban, Dominica
n, Puerto Rican, South & Central A
merican)",
    "count": 105
  },
  {
    "_id": "Spanish",
    "count": 47
  },
  {
    "_id": "Chinese",
    "count": 45
  },
  {
    "_id": "Mexican",
    "count": 41
  },
  {
    "_id": "American",
    "count": 40
  }
]
```

   There are a lot of restaurants from Spanish Speaking countries!

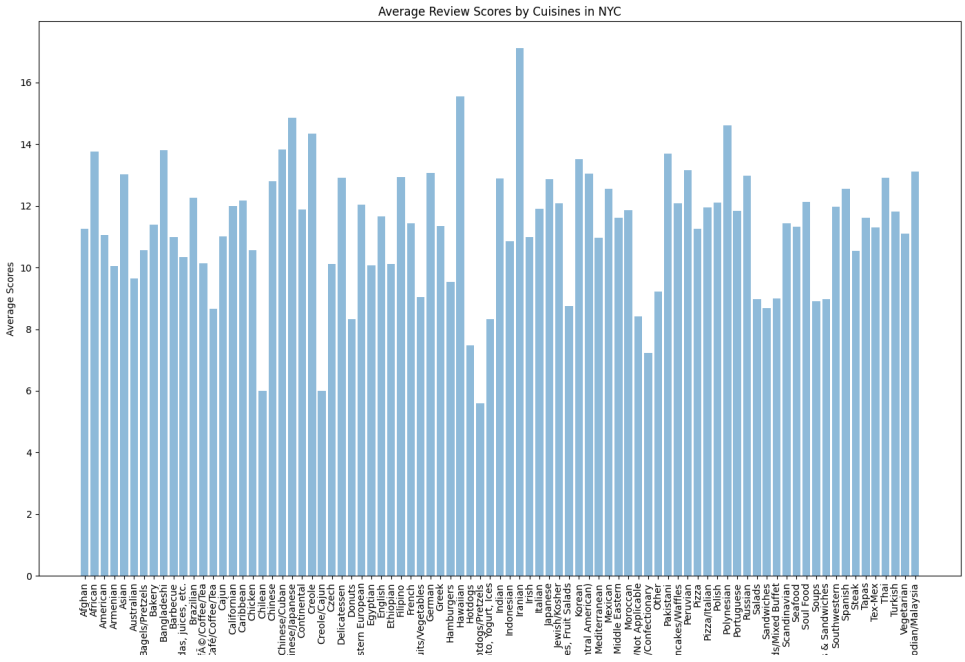
9. *How are different cuisines reviewed across NYC?* ***(VISUALIZATION)***

   We aggregate the average scores by grouping each cuisine, and graph it! Done in python using pymongo and visualized using matplotlib; code + bigger version of the image attached.

   But this is the result! The lowest rated cuisine is Hotdogs/Pretzels with an average review score of about 5.8. The maximal average score is Iranian food,

with an average review score of about 17. Most other cuisines hover around a score of 12. After some investigation into the data, it is apparent that not every reviewer grades on the same scale, which would skew scores for certain cuisines.

Also, I apologize for the letters getting cut off at the bottom, this is the first time I have used matplotlib!



1. *What are the best rated restaurants on average in the worst reviewed cuisine?*

Query

```
const aggregate = [
  {
    $match: {
      "grades.0": {$exists: true},
      "cuisine": "Hotdogs/Pretzels"
    }
  },
```

Result

```
[
  {
    "name": "Auntie Anne'S",
    "_id": "50006096",
    "numReviews": 2,
    "averageScore": 21.5
  },
  {
    "name": "Circle Line Xvii",
    "_id": "41560843",
```

```
  {
    $project: {
      _id: "$restaurant_id",
      totalScore:  {
        $reduce: {
          input: "$grades",
          initialValue: 0,
          in: { $add : ["$$value",
 "$$this.score"] }
        }
      },
      numReviews: {$size: "$grade
s"},
      name: true
    }
  },
  {
    $project: {
      averageScore: { $divide: ["$to
talScore", "$numReviews"]},
      numReviews: true,
      name: true
    }
  },
  {
    $sort: { "averageScore": -1 }
  },
  {
    $limit: 5
  }
];

db["New York"].aggregate(aggregate)
```

```
    "numReviews": 4,
    "averageScore": 13
  },
  {
    "name": "Island Soft Pretzel Stop",
    "_id": "41681666",
    "numReviews": 3,
    "averageScore": 10.333333333333334
  },
  {
    "name": "Jamaica Multi Plex Cinema
s",
    "_id": "41595437",
    "numReviews": 5,
    "averageScore": 8.8
  },
  {
    "name": "Ny Skyride",
    "_id": "41266025",
    "numReviews": 4,
    "averageScore": 6.75
  }
]
```

If there is something I've learned while working with this dataset, the reviewers do not have a consistent grading scale. I can personally attest that Auntie Anne's is probably not the best Hotdog/Pretzel in NYC.