

## Title Page

- **Title:** DB Assignment 5
- **Your Name:** Kenny Chau
- **Date:** 11/22/2024

Query 1: Over how many years was the unemployment data collected?

Group the data by the years and then count each distinct year.

---

```
Distinct_Year : 27
```

Query 2: How many states were reported on in this dataset?

Group the data by states and then count each distinct state.

---

```
Unique_States : 47
```

Query 3: What does this query compute?

```
db.unemployment.find({Rate : {$lt: 1.0}}).count()
```

The query outputs the count of counties that had a unemployment rate less than 1%

```
> db.unemployment.find({Rate : {$lt: 1.0}}).count()
< 657
```

Query 4: Find all counties with unemployment rate higher than 10%

Project just the county and rate and use the \$match to find counties that had a rate higher than 10%.

Output after `$match` stage (Sample of 10 documents)

```
County : "Kemper County"
Rate : 10.6
```

```
County : "Jefferson County"
Rate : 14.3
```

Query 5: Calculate the average unemployment rate across all states.

Group the data by the average rate and use `project` to remove `_id`.

```
averageUnemployment : 5.577907
```

Query 6: Find all counties with an unemployment rate between 5% and 8%.

Project the counties and rates and then use `$match` to find counties that had an unemployment rate between 5% and 8%. I assumed that it did not include the percentages 5 and 8.

Output after `$match` stage (Sample of 10 documents)

```
County : "Newton County"
Rate : 6.1
```

```
County : "Monroe County"
Rate : 7.9
```

Query 7: Find the state with the highest unemployment rate.

Group the states by the average rate of unemployment in each state. Then sort from descending order and limit the result to only 1.

Output after `$limit` stage (Sample of 1 document)

```
_id: "Arizona"
averageUnemployment : 8.719809523809523
```

Query 8: Count how many counties have an unemployment rate above 5%.

Project only the counties and their rates. Then use `$match` to find counties whose rate was above 5%. Then use `$count` to count these counties.

```
Counties : 510173
```

Query 9: Calculate the average unemployment rate per state by year.

Create a projection that ignores `_id` and month. Then group the state and years together by the average rate of unemployment at that state during that year. Project again to show the year, state, and unemployment rate.

Output after `$project` stage (Sample of 10 documents)

```
rate : 6.155800653594771
year : 2015
state : "Illinois"
```

```
rate : 4.677827380952381
year : 2014
state : "Montana"
```

Query 10) For each state, calculate the total unemployment rate across all counties (sum of all county rates)

Group the states by the sum of every county in that state.

```
_id: "New York"  
rate : 12273.2
```

```
_id: "Michigan"  
rate : 13005.4
```

Query 11) The same as Query 10 but for states with data from 2015 onward.

Group the states and years by the sum of every counties' rate in that state during that year. Then use \$match to find data that came from 2015 and onwards. Then project the year, state, and rate.

Output after `$project` stage (Sample of 10 documents)

```
rate : 5.528333333333333  
year : 2015  
state : "Rhode Island"
```

```
rate : 4.746536796536796  
year : 2015  
state : "Oklahoma"
```

