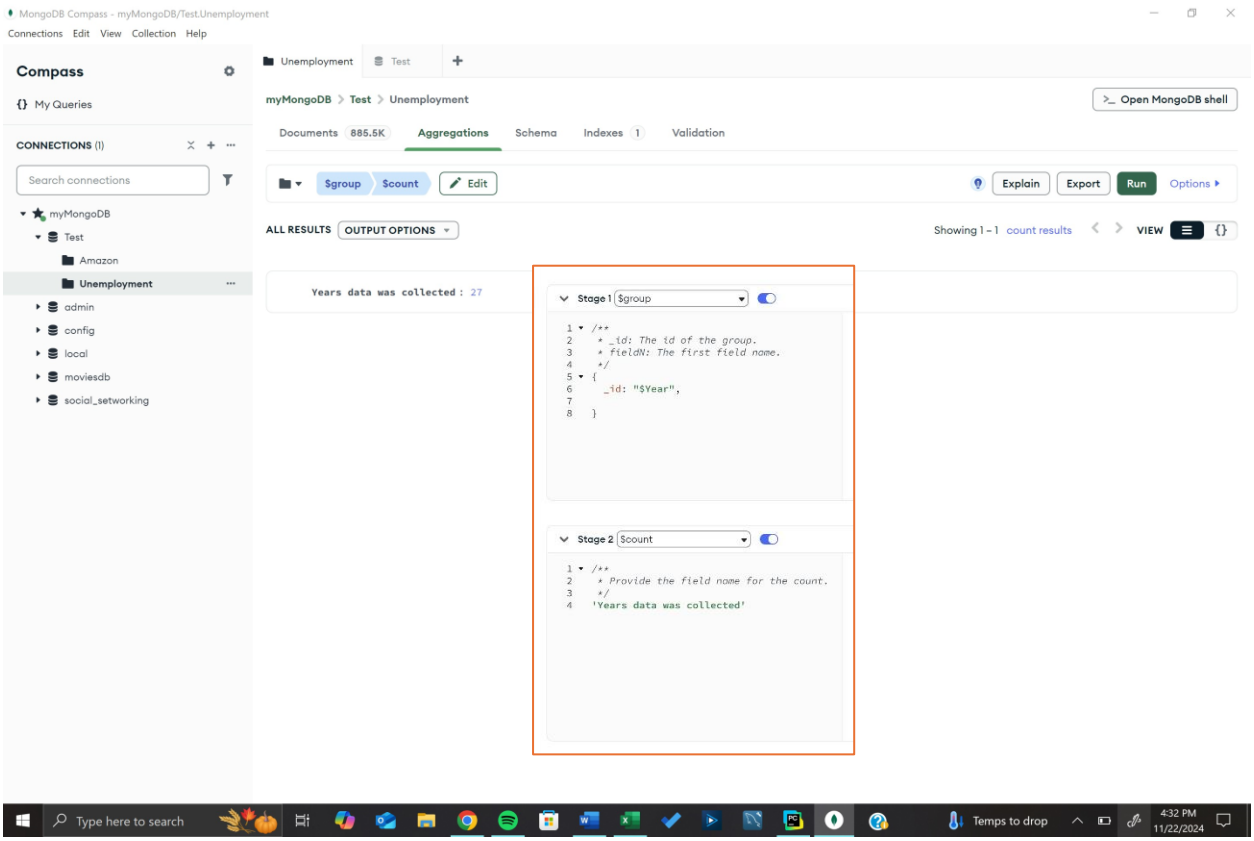
Christina DiMaggio

November 22, 2024

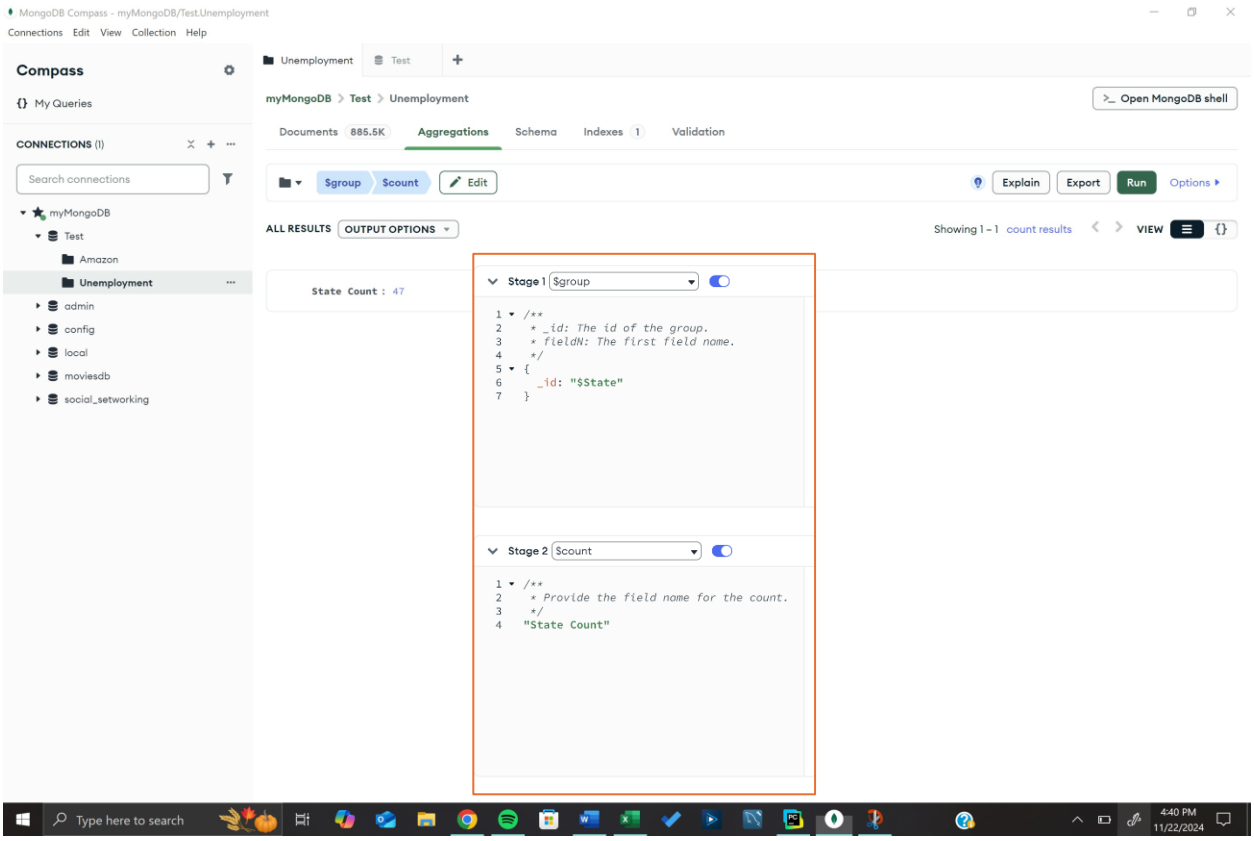
Assignment 5

// 1. Over how many years was the unemployment data collected?



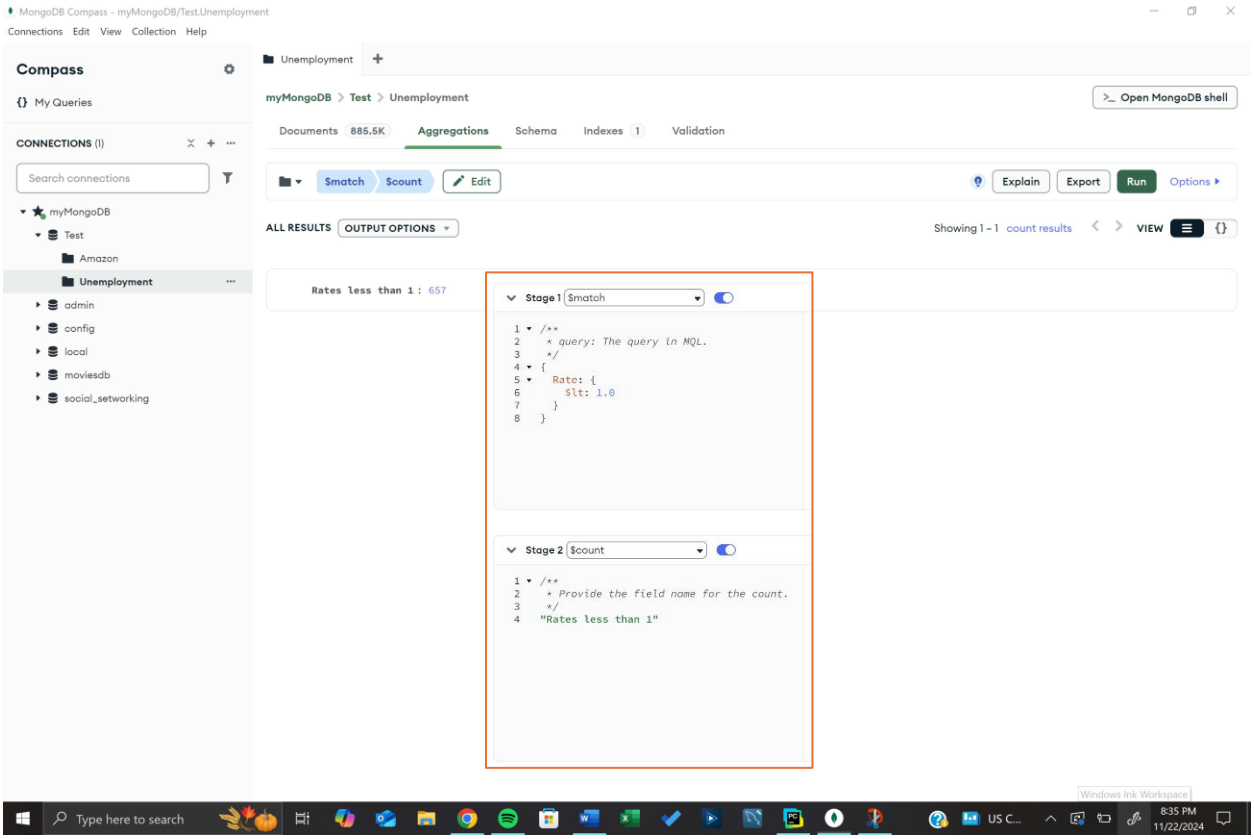
The unemployment data was collected over 27 years. The aggregation first groups data by the year field. In this stage, all the specific years in the dataset can be found and returned as documents, from 1990 to 2016. Then count is used to retrieve the number of documents in the collection that match the previous query. In this case, 27 years were returned from the group aggregate, so the count aggregate reported 27.

// 2. How many states were reported on in this dataset?



There are 47 states reported in this dataset. The aggregation first groups data by the state field. In this stage, all the specific states in the dataset can be found and returned as documents. Then count is used to retrieve the number of documents in the collection that match the previous query. In this case, 47 different states were returned from the group aggregate, so the count aggregate reported 47.

// 3. What does this query compute? db.unemployment.find({Rate : {$lt: 1.0}}).count()

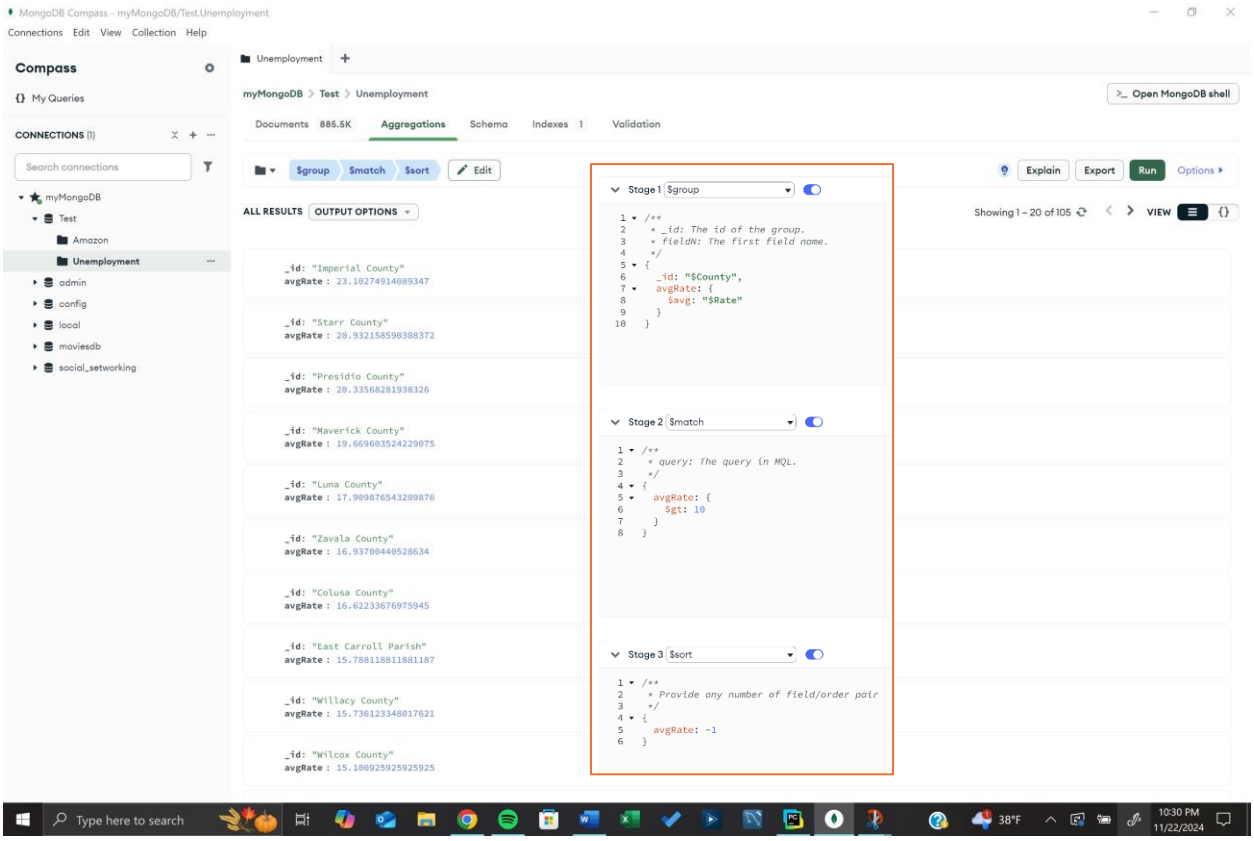


There are 657 documents in this dataset that have an unemployment rate of less than 1.0. The aggregation first matches data by the field rate if it is less than 1.0. In this stage, all the specific matches to this query in the dataset can be found and returned as documents. Then count is used to retrieve the number of documents in the collection that match the previous query. In this case, 657 entries of data have an unemployment rate of less than 1.0, so the count aggregate reported is 657.

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

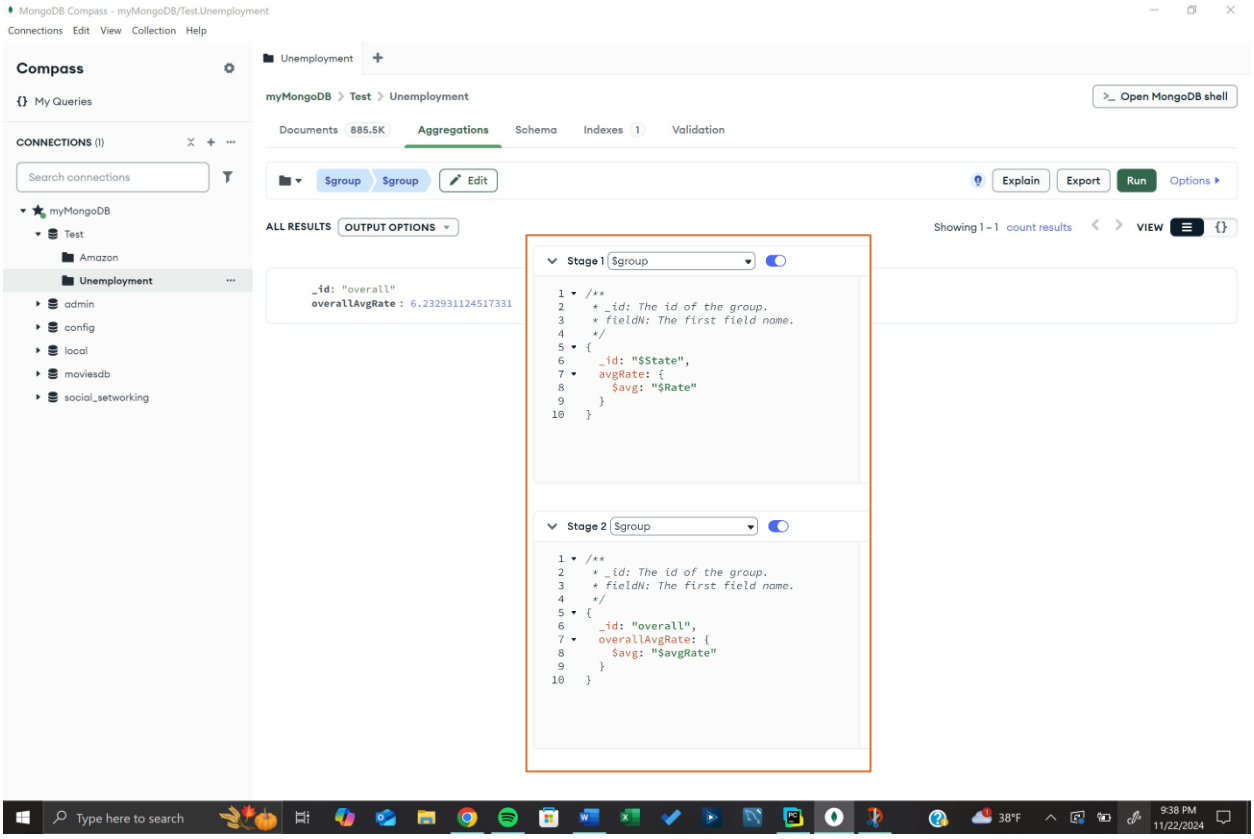
- MongoDB will search the unemployment collection for the documents that match the specific criteria:
    - Rate less than 1.0 is a query filter. Documents where the Rate field is less than 1.0 are returned
    - Count tells MongoDB to return the count of the number of documents to match the filter criteria instead of returning the documents themselves.

// 4. Find all counties with unemployment rate higher than 10%



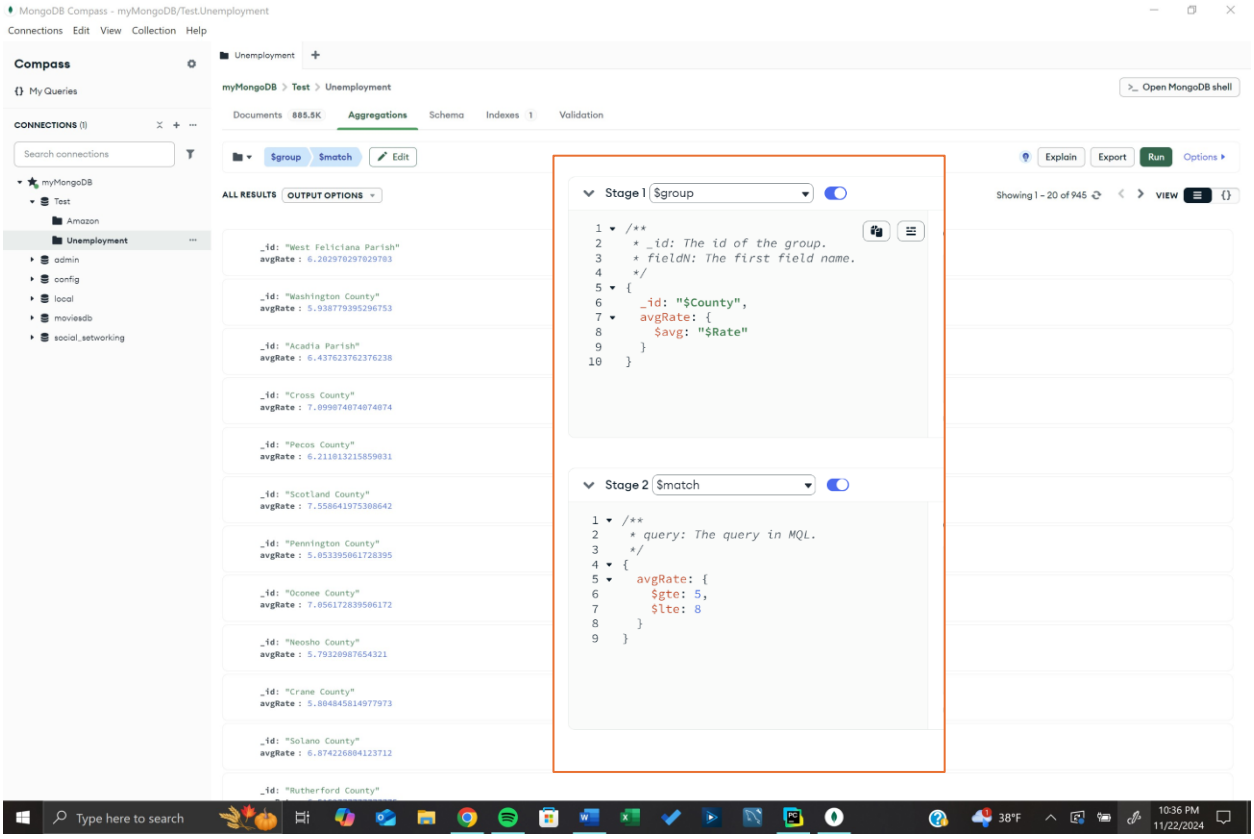
There are 105 counties with unemployment rates higher than 10%. The output shows a sample of the documents (image above). The aggregation groups data by the country and its average rate of unemployment over years data was collected. This signifies they have an overall unemployment rate higher than 10%. The aggregation then matches data by the field rate if it is greater than 10%. In this stage, all the specific matches to this query in the dataset can be found and returned as documents. In this stage, all the counties that match the filter of the average rate above 10% in the dataset can be found and returned as documents. Sort -1 is used to order the results from the county with the highest unemployment rate to the lowest.

// 5. Calculate the average unemployment rate across all states.



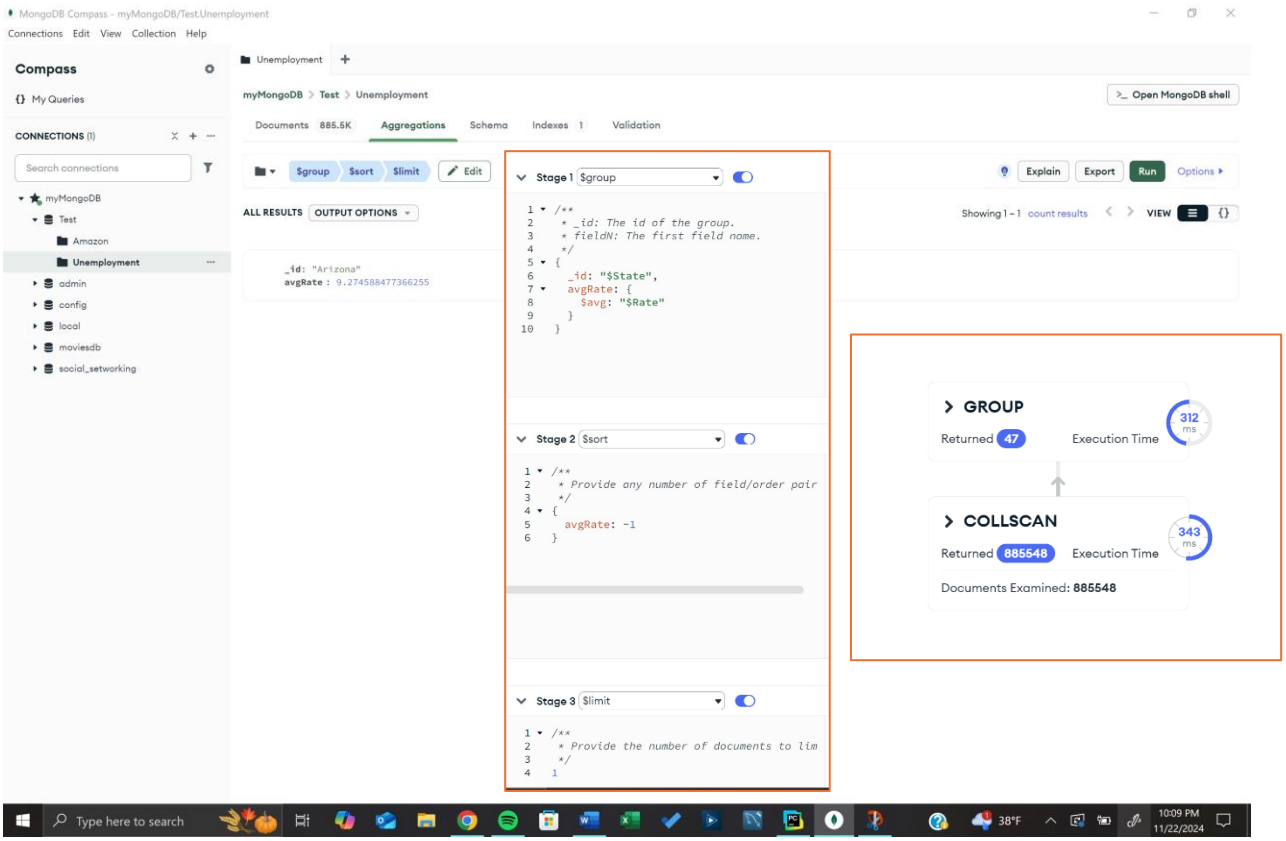
The average overall unemployment rate across all states in the data set is 6.23%. The aggregation first groups data by the state field and creates the average rate for each state in the dataset. In this stage, all the specific states and their respective average rates from the dataset can be found and returned as documents. These documents are then averaged together to create the total overall unemployment rate found across all states from the data set, 6.23%.

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



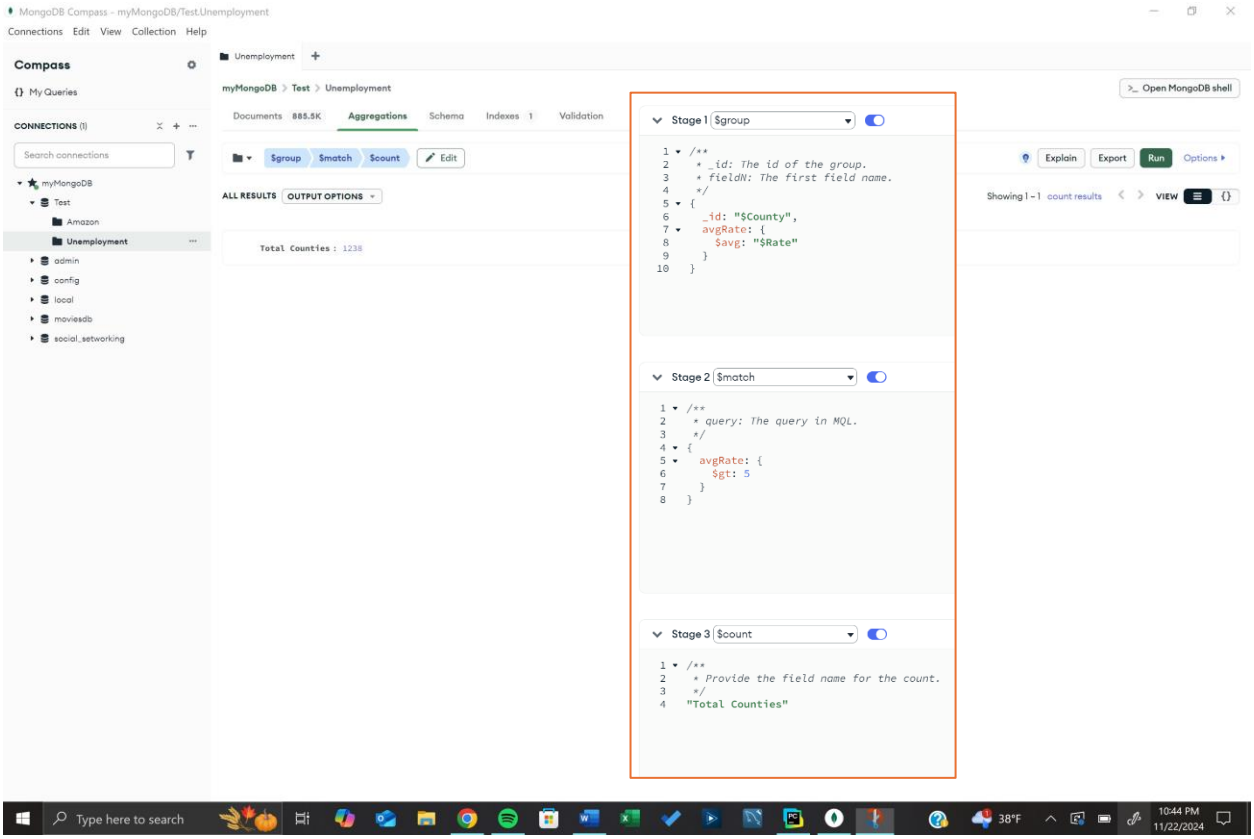
There are 945 counties with an unemployment rate between 5% and 8%. The output shows a sample of documents (image above). The aggregation groups data by the country and its average rate of unemployment over years data was collected. The aggregation then matches data by the field rate if it is greater than or equal to 5% or less than or equal to 8%. In this stage, all the specific matches to this query in the dataset can be found and returned as documents. This signifies they have an overall unemployment rate between 5% and 8%. In this stage, all the counties that match the filter in the dataset can be found and returned as documents.

// 7. Find the state with the highest unemployment rate. Hint. Use { $limit: 1 }



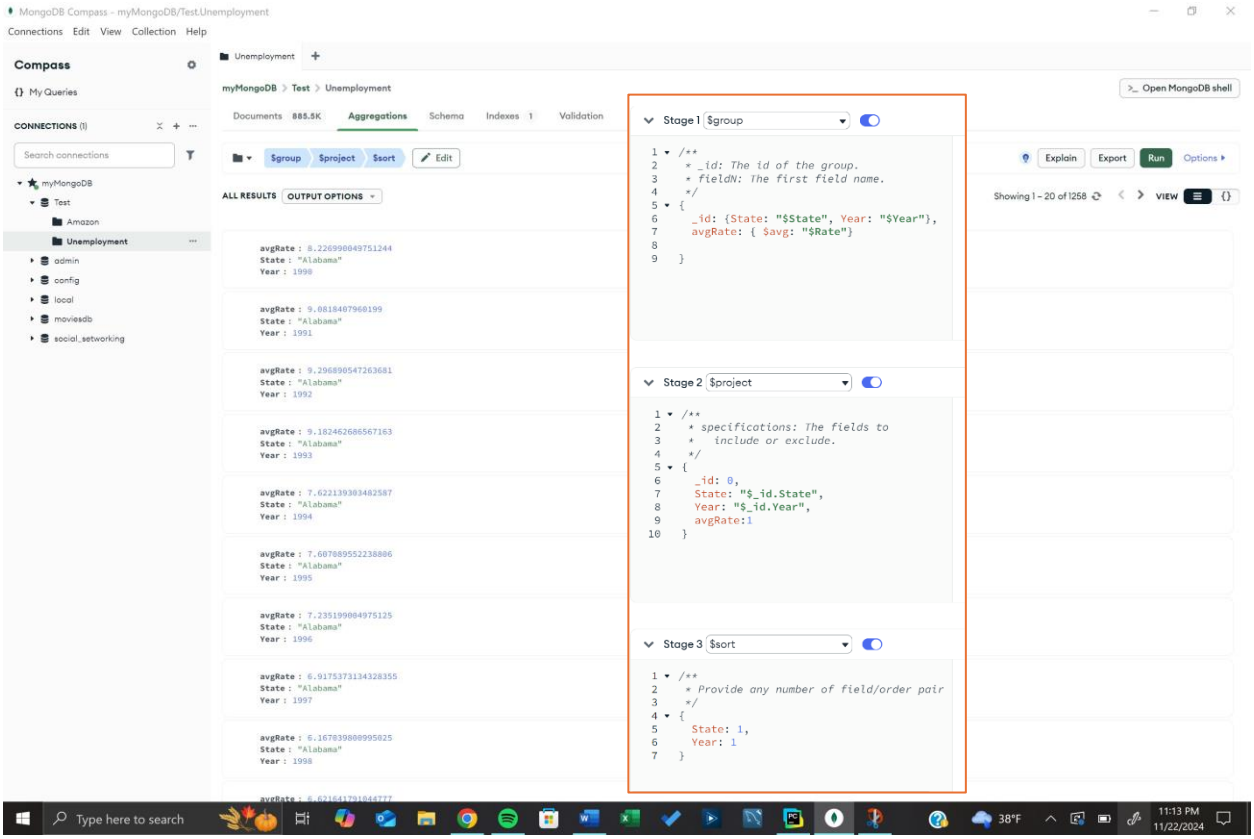
The state with the highest unemployment rate is Arizona with an average rate of 9.27%. The aggregation first groups data by the state field and creates the average rate for each state in the dataset. In this stage, all the specific states and their respective average rates from the dataset can be found and returned as documents. Sort -1 is used to order the results from the State with the highest unemployment rate to the lowest. Finally, the aggregation of limit 1 only returns the one document with the highest unemployment rate, which is Arizona at 9.27%.

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



There are 1238 counties with an unemployment rate above 5%. The aggregation groups data by the country and its average rate of unemployment over years data was collected. The aggregation then matches data by the field rate if it is greater than 5%. In this stage, all the specific matches to this query in the dataset can be found and returned as documents. This signifies they have an overall unemployment rate above 5%. In this stage, all the counties that match the filter in the dataset can be found and returned as documents. The final stage is count to determine the number of documents found in the previous stage. This is where it is determined that 1238 counties have an unemployment rate above 5%.

// 9. Calculate the average unemployment rate per state by year.



There are 1258 documents stating the average unemployment rate per state by year. The aggregation groups data by state and year to calculate its average rate of unemployment. The aggregation then projects to show in the document the state and the given year's average rate of unemployment. If this was not added, then _id would read as object which can become confusing for the organization. To organize the documents, sort 1 was used for states to be in alphabetical order and their respective years in numerical order as well.