# SQL Queries

## Query 1

### Question

A leading beverage company has announced a billion-dollar fund for removing debris from forests, rivers and mountains in the US. All states are interested. Which 2 states have the least chance to win a share of the fund?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Assumptions: It can be assumed the states with the most debris will receive the fund. Many fires are caused by debris and intuition leads us to believe the states with the most debris have the most fires caused by debris and the that the state with the least amount of fires caused by debris have the least amount of debris and are the least likely to receive the fund. Therefore, the state with the least amount of fires caused by debris will have the least chance to receive the fund. It wouldn’t make sense for a state with no debris to receive the money.

SOURCE\_REPORTING\_UNIT\_NAME column in fires table contained the most forests and was used in this query to represent forests.

The statistical cause code for fires caused by debris is 5. So, the count of statistical cause code where the Stat\_Cause\_Code is equal to 5, is displayed as fires caused by debris.

The states with least number of fires caused by debris are Washington DC and Puerto Rico.

### Translation

Select the Source reporting unit name as forest name, the number of fires caused due to debris as fires caused by debris and the state of the forest from the fires table. Group by the state. Order the number of fires caused due to debris in ascending order to display the forests with least number of fires. Limit the result to 2.

### Screen Shot of SQL Query and Results

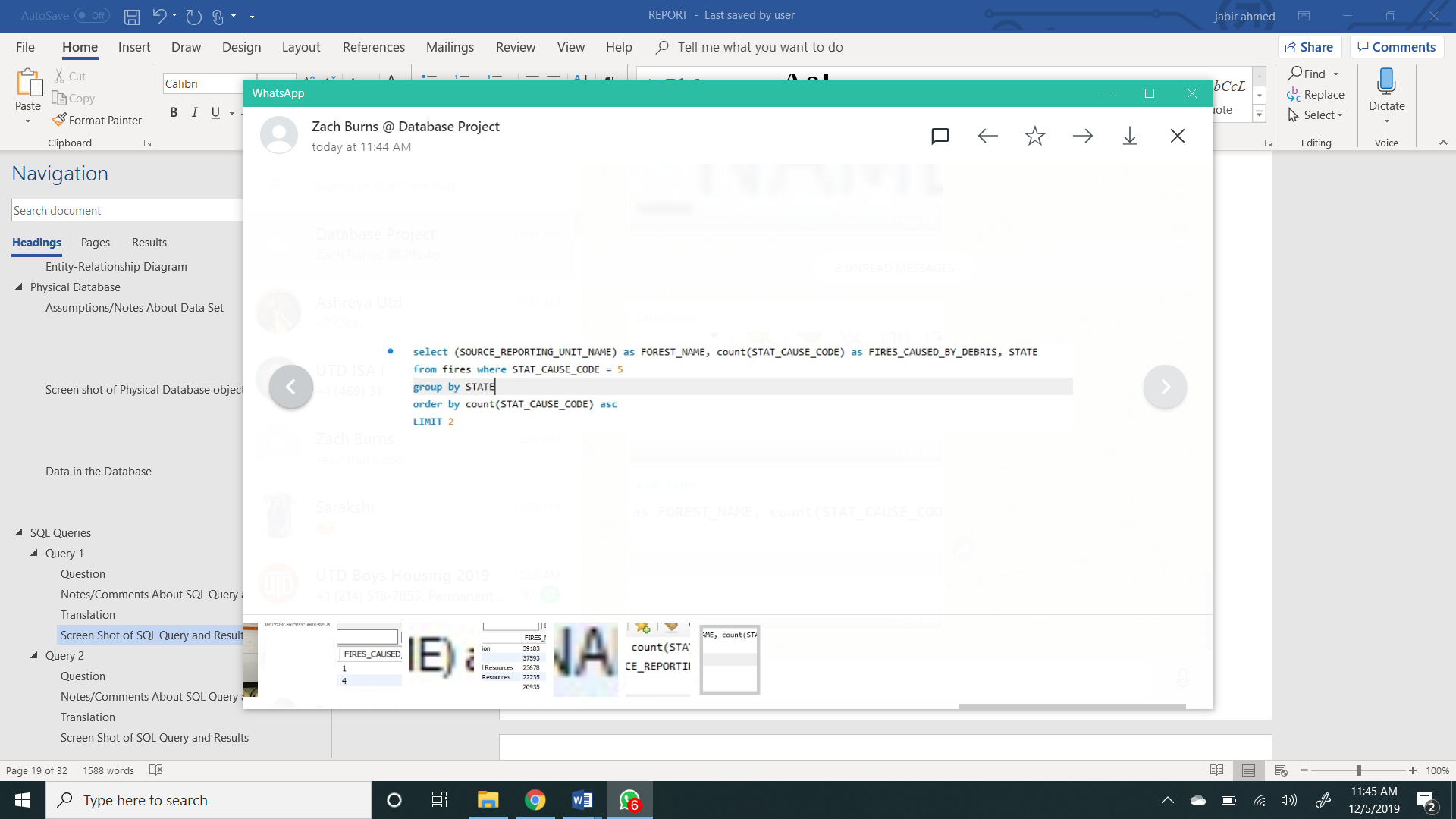
select (SOURCE\_REPORTING\_UNIT\_NAME) as FOREST\_NAME, count(STAT\_CAUSE\_CODE) as FIRES\_CAUSED\_BY\_DEBRIS, STATE

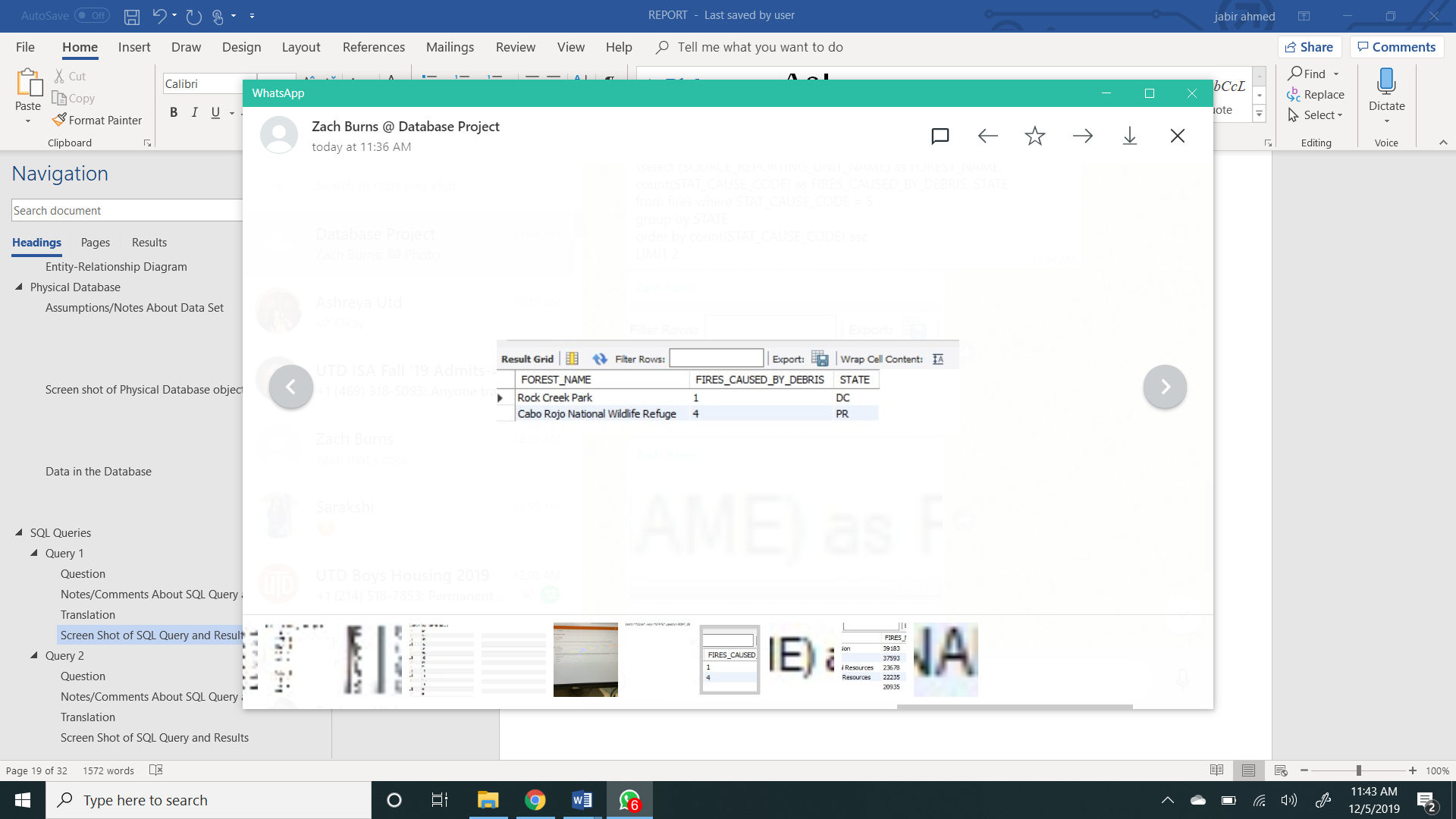
from fires where STAT\_CAUSE\_CODE = 5

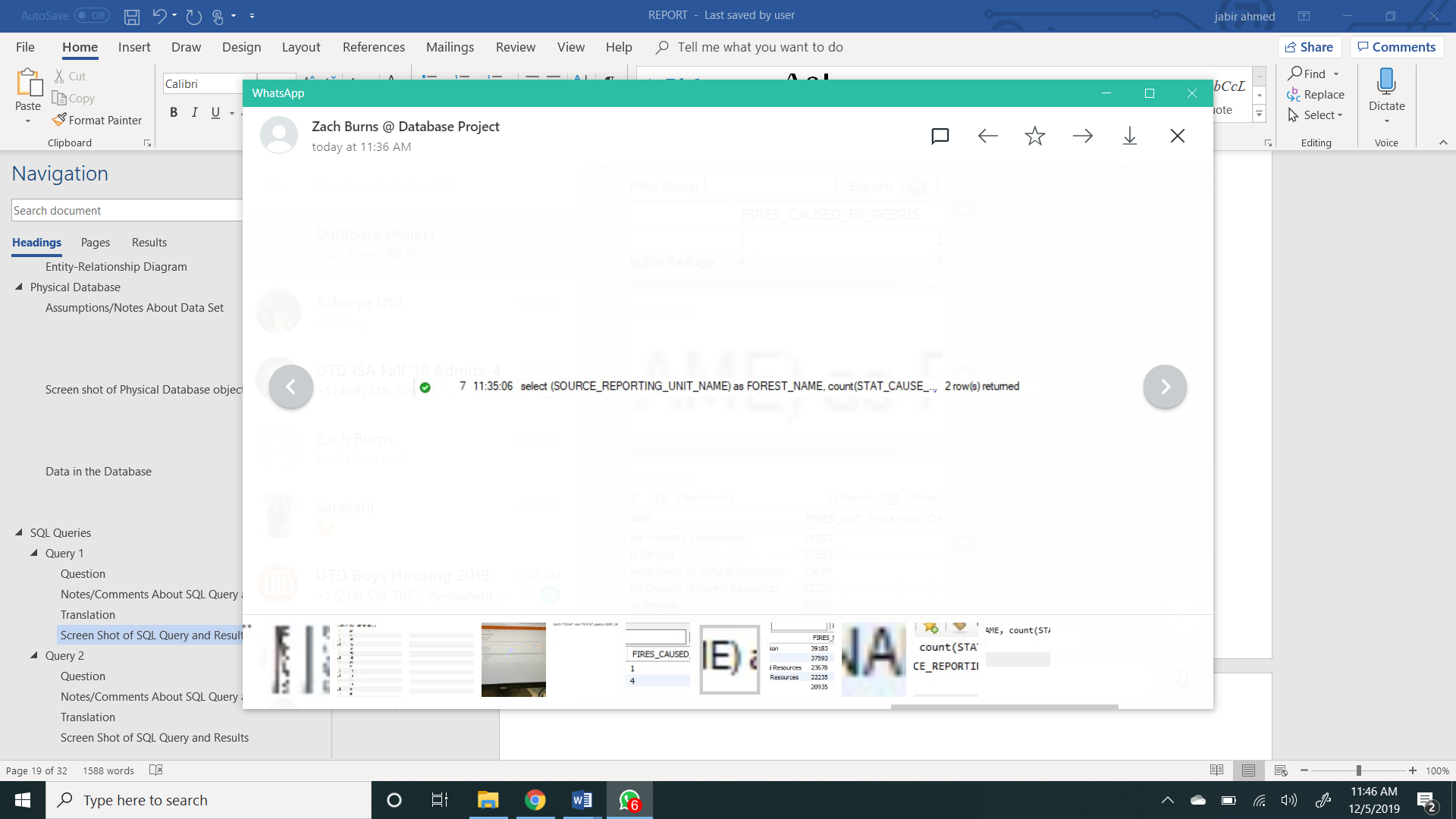
group by STATE

order by count(STAT\_CAUSE\_CODE) asc

LIMIT 2







## Query 2

### Question

One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 4 children in a group visiting a forest. Name top 5 forests where this would be the least appropriate.

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Assumptions: It is least appropriate to ban children from forests that have the highest number of fires that did not involve children.

SOURCE\_REPORTING\_UNIT\_NAME column in fires table contained the most forests and was used in this query to represent forests.

The statistical cause code for fires caused due to children is 8. So, the count of statistical cause code where the Stat\_Cause\_Code is not equal 8, is displayed as *FIRES\_NOT \_INVOLVING\_CHILDREN*.

Since the following five forests have the highest number of fires that did not involve any children, the

ban on children will be the least appropriate for these forests.

Forests with most fires not involving children: South Carolina Forestry Commission, Texas forest Service, Minnesota Department of Natural Resources, North Carolina Division of Forest Resources, Florida Forest Service.

### Translation

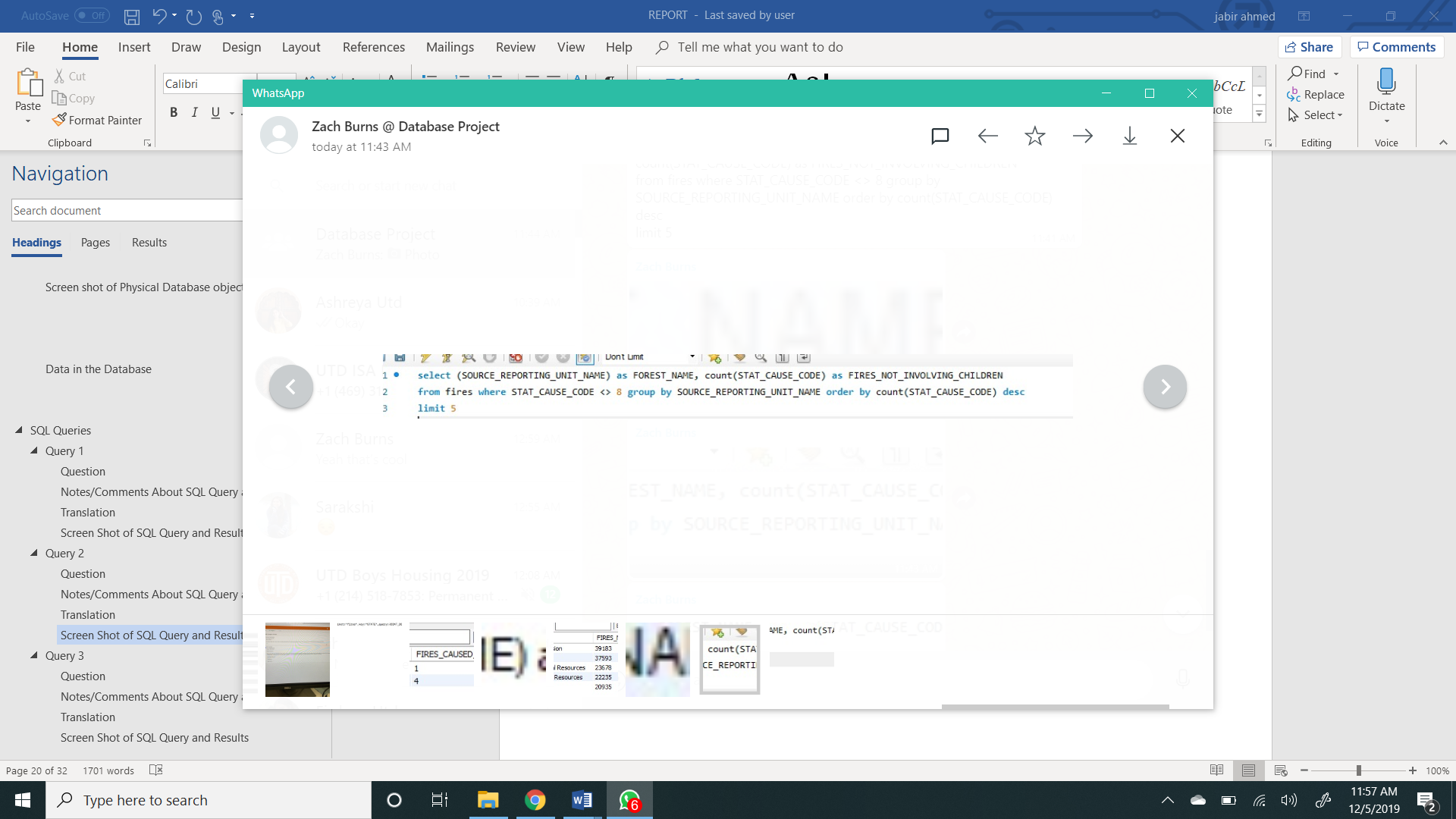
Select the Source reporting unit name as forest name and the number of fires that were not caused by children as fires not involving children from the fires table. Group by the source reporting unit name. Order the number of fires that were not caused by children in descending order. Limit the result to 5.

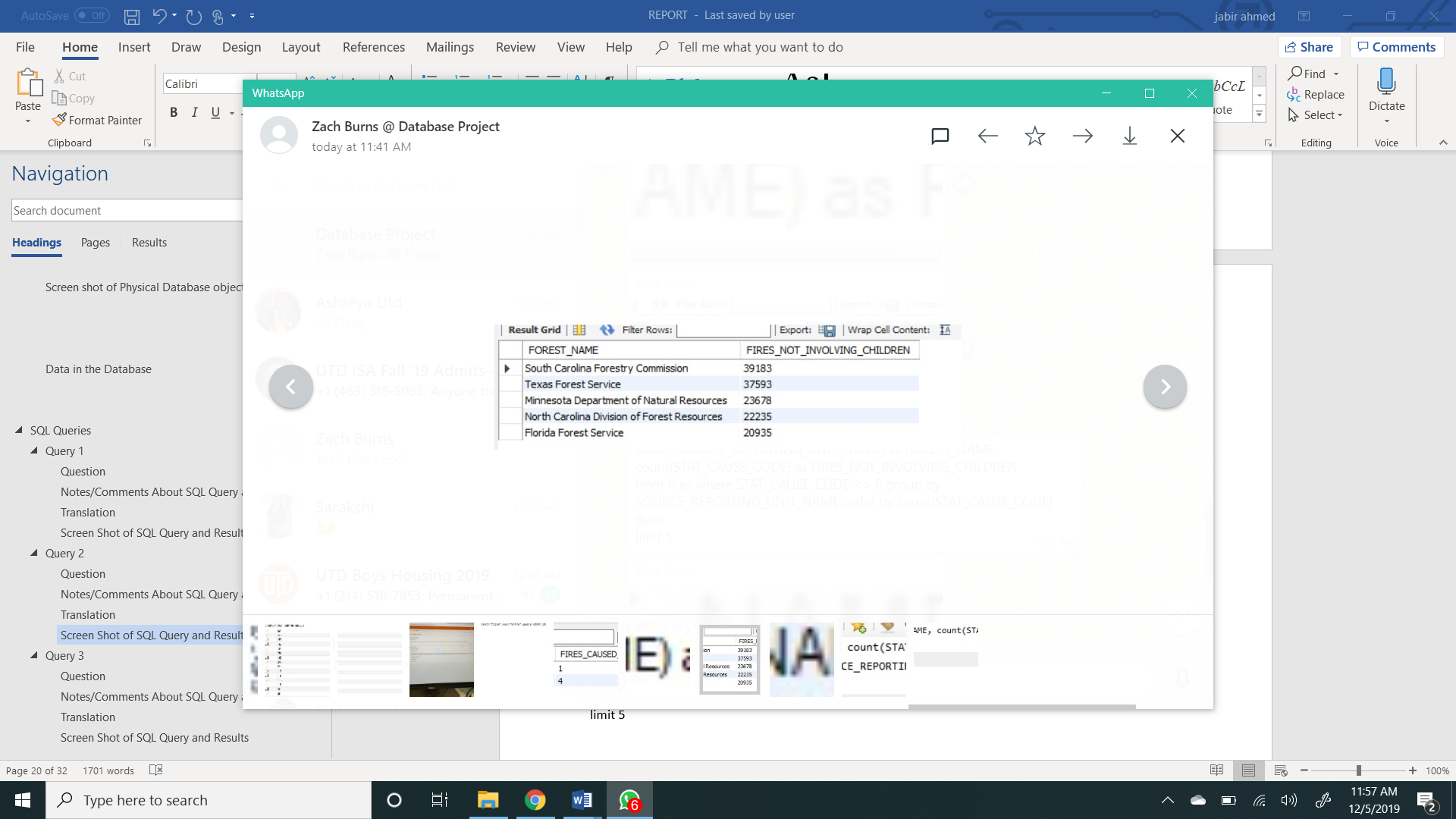
### Screen Shot of SQL Query and Results

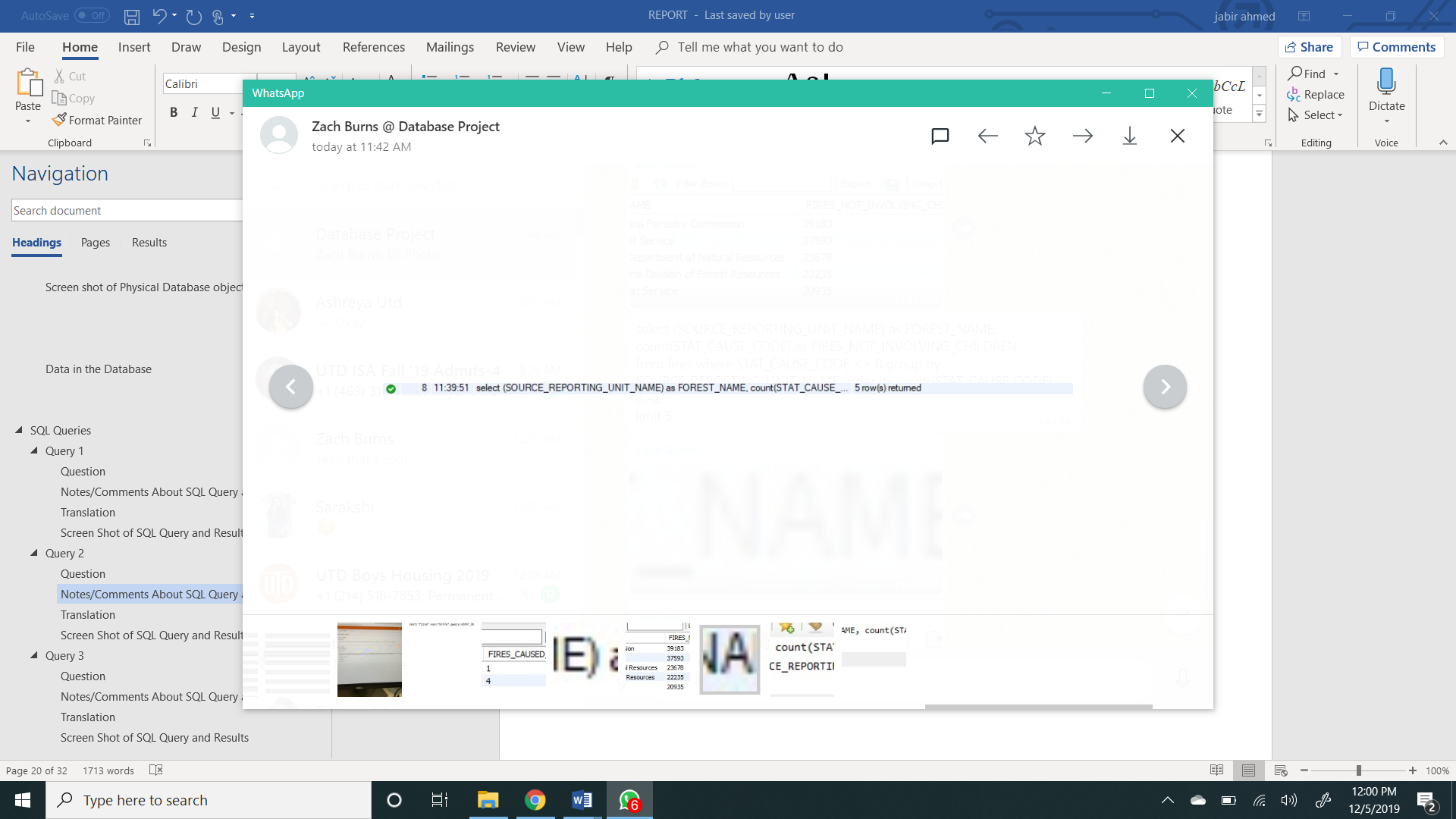
select (SOURCE\_REPORTING\_UNIT\_NAME) as FOREST\_NAME, count(STAT\_CAUSE\_CODE) as FIRES\_NOT\_INVOLVING\_CHILDREN

from fires where STAT\_CAUSE\_CODE <> 8 group by SOURCE\_REPORTING\_UNIT\_NAME order by count(STAT\_CAUSE\_CODE) desc

limit 5







## Query 3

### Question

One advocacy group says human actions and nature are equally to blame for most wildfires. Write a query that can help determine the truth of this statement.

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Assumptions: In the database the only natural cause of fire is lightning. All the other fires were caused due to human actions.

The statistical cause code for fires caused due to lightning is 1. So, the count of statistical cause code where the Stat\_Cause\_Code is not equal to 1, is displayed as *HUMAN\_CAUSE*, and the count of statistical cause code where the Stat\_Cause\_Code is equal to 1, is displayed as *NATURAL\_CAUSE*.

Since the number of fires caused by human actions is significantly greater than natural causes, human actions and nature are not equally to blame for most wildfires.

### Translation

Select the number of fires caused due to human actions and the number of fires caused due to natural causes from the fires table.

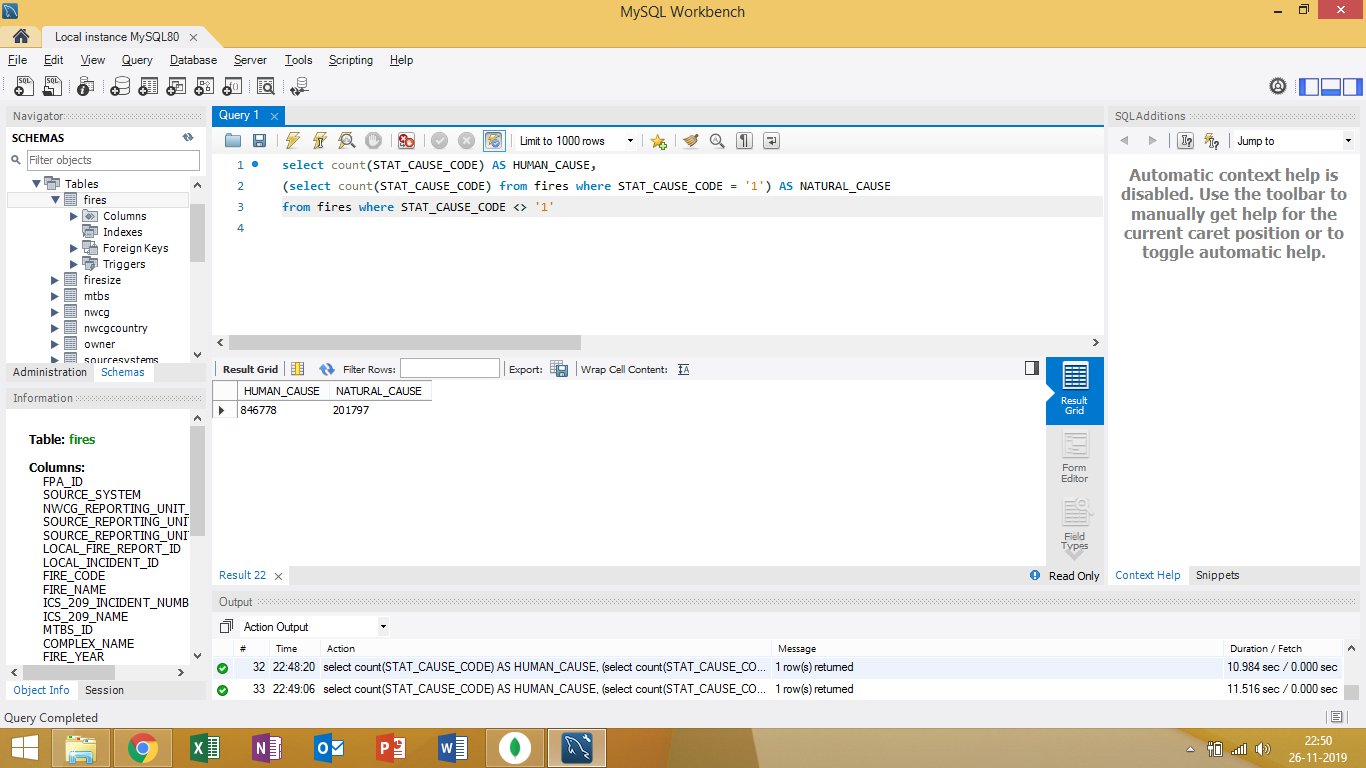
### Screen Shot of SQL Query and Results

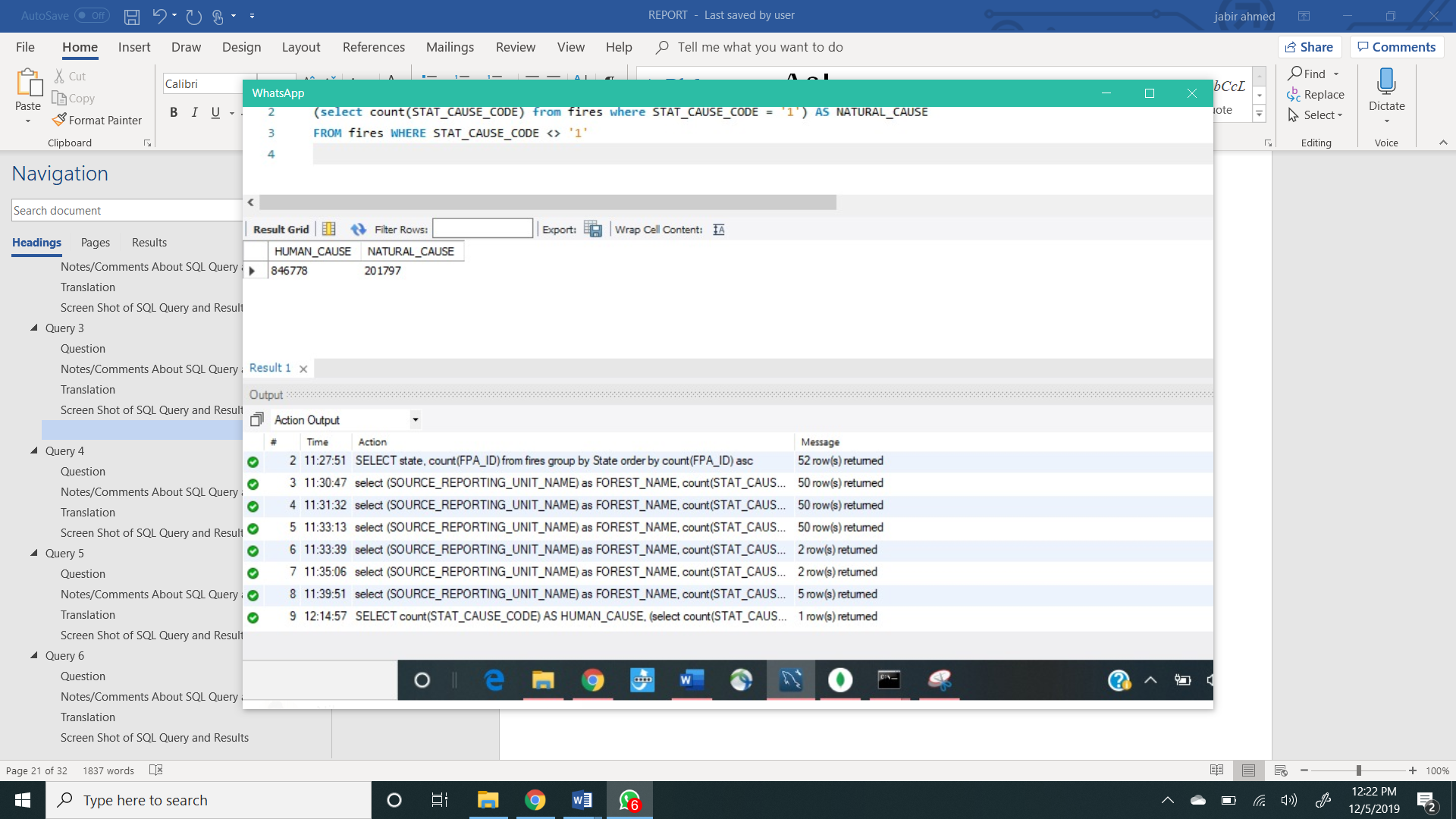
SELECT count(STAT\_CAUSE\_CODE) AS HUMAN\_CAUSE,

(select count(STAT\_CAUSE\_CODE) from fires where STAT\_CAUSE\_CODE = '1') AS NATURAL\_CAUSE

FROM fires WHERE STAT\_CAUSE\_CODE <> '1'

## 





## Query 4

### Question

How many wildfires were reported by at least two units/agencies?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

### 

### Translation

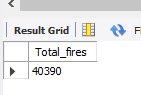
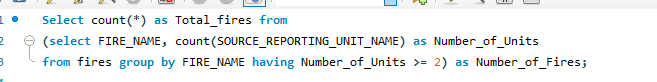
Select the Unit type, state and nwcg reporting unit ID from the nwcg table, and the country from the nwcg country table by joining the nwcg and nwcg country table. Group by the Unit type, country and the state. Order the results in ascending order for the state and descending order for the nwcg reporting unit ID.

### Screen Shot of SQL Query and Results

Select count(\*) as Total\_fires from

(select FIRE\_NAME, count(SOURCE\_REPORTING\_UNIT\_NAME) as Number\_of\_Units

from fires group by FIRE\_NAME having Number\_of\_Units >= 2) as Number\_of\_Fires;





## 

## Query 5

### Question

Which state had fires only in the second half of the calendar years?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Assumptions: As there are 365 days in an year, the second half of the calendar years means that the day of year is greater than or equal to 182. The CONT\_DOY is the day of the year (range: 1-366) that fire happened.   
 We had difficulty filtering distinct states in the WHERE clause to limit the ones ONLY in the second half of the year. We would us WHERE CONT\_DOY > 181 and it would still include states that had fires in the first half with their second half numbers.   
 To solve this we used the MIN operator. The rational was that the minimum CONT\_DOY of a certain state should be equal to or higher than 182.

As you can see below, CT (Connecticut) was the only state with fires in the second half of the year since its minimum of 206 is above 182.

### Translation

Select the state and minimum containment day of the year from the fires table. Group by the state. Order the result in descending order according to the minimum containment day of year.

### Screen Shot of SQL Query and Results

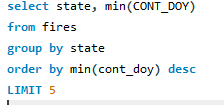
select state, min(CONT\_DOY)

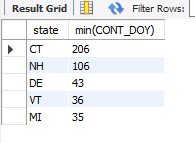
from fires

group by state

order by min(cont\_doy) desc

LIMIT 5







## Query 6

### Question

Which forest had the number of fires equal to the average number of wild fires in the US?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

The average number of fires in the U.S is calculated by averaging the count of fires in each U.S state then comparing that to the count of fires in each forest.

SOURCE\_REPORTING\_UNIT\_NAME Is the closest column to represent forests.

Results: We determined based on the Zero result output that there are forests that are EQUAL to the average number for the U.S we calculated.

\*\*This query has 2 subqueries inside, below you will see the query and it broken into parts to better explain answer.\*\*

### Translation

Select the number of fires for individual forests and the average number of wildfires in the US from the fires table.

### Screen Shot of SQL Query and Results

Initially cam back with zero results with this query:

SELECT F.SOURCE\_REPORTING\_UNIT\_NAME, count(F.SOURCE\_REPORTING\_UNIT\_NAME) AS "firecount" FROM fires F

where "firecount" =

(

SELECT avg(Count) FROM

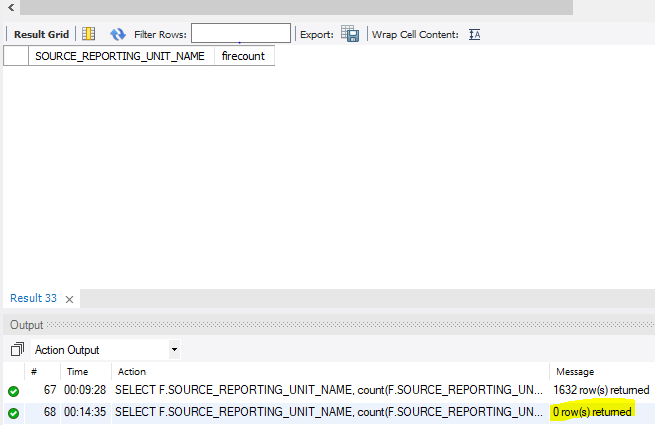
(SELECT F. STATE, NC.Country, COUNT(F.FPA\_ID) AS Count FROM fires F

JOIN nwcgcountry NC ON NC.state = F.state

where country = "US"

GROUP BY state) as Counts)

group by F.SOURCE\_REPORTING\_UNIT\_NAME



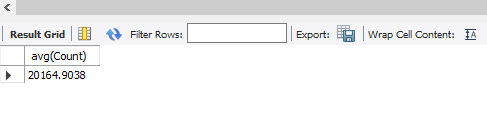
The average was calculated to be: 20,164  
 Using this query:  
SELECT avg(Count) FROM

(SELECT F. STATE, NC.Country, COUNT(F.FPA\_ID) AS Count FROM fires F

JOIN nwcgcountry NC ON NC.state = F.state

where country = "US"

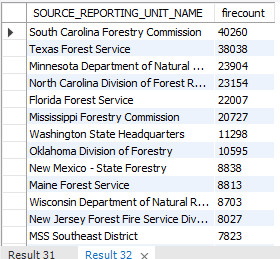
GROUP BY state) as Counts



The query below shows that none of the forests had **EQUAL** number of fires to the U.S National Average  
SELECT F.SOURCE\_REPORTING\_UNIT\_NAME, count(F.SOURCE\_REPORTING\_UNIT\_NAME) AS "firecount" FROM fires F

Group by F.SOURCE\_REPORTING\_UNIT\_NAME

ORder BY "firecount" DESC



# MongoDB Queries/Code

## Query 1

### Question

A leading beverage company has announced a billion-dollar fund for removing debris from forests, rivers and mountains in the US. All states are interested. Which 2 states have the least chance to win a share of the fund?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

Assumptions: It can be assumed the states with the most debris will receive the fund. Many fires are caused by debris and intuition leads us to believe the states with the most debris have the most fires caused by debris and the that the state with the least amount of fires caused by debris have the least amount of debris and are the least likely to receive the fund. Therefore, the state with the least amount of fires caused by debris will have the least chance to receive the fund. It wouldn’t make sense for a state with no debris to receive the money.

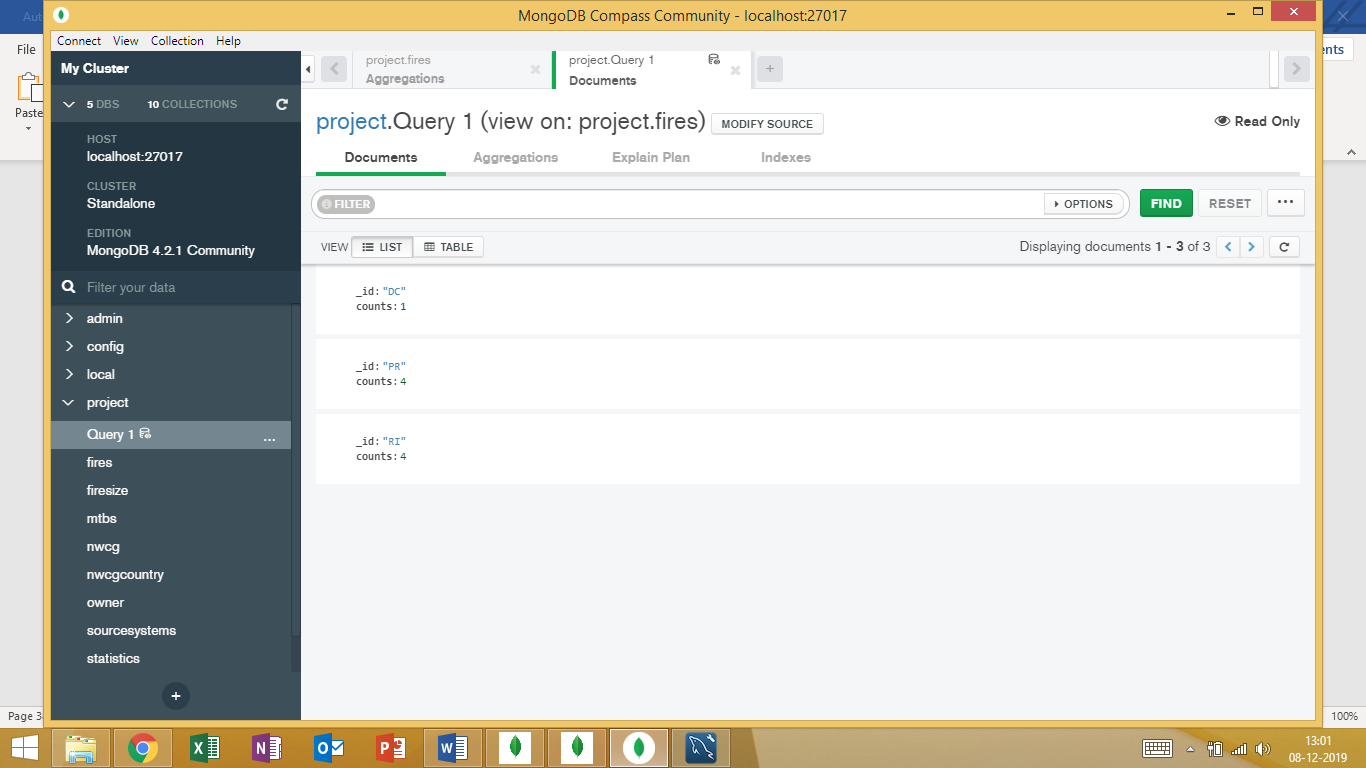
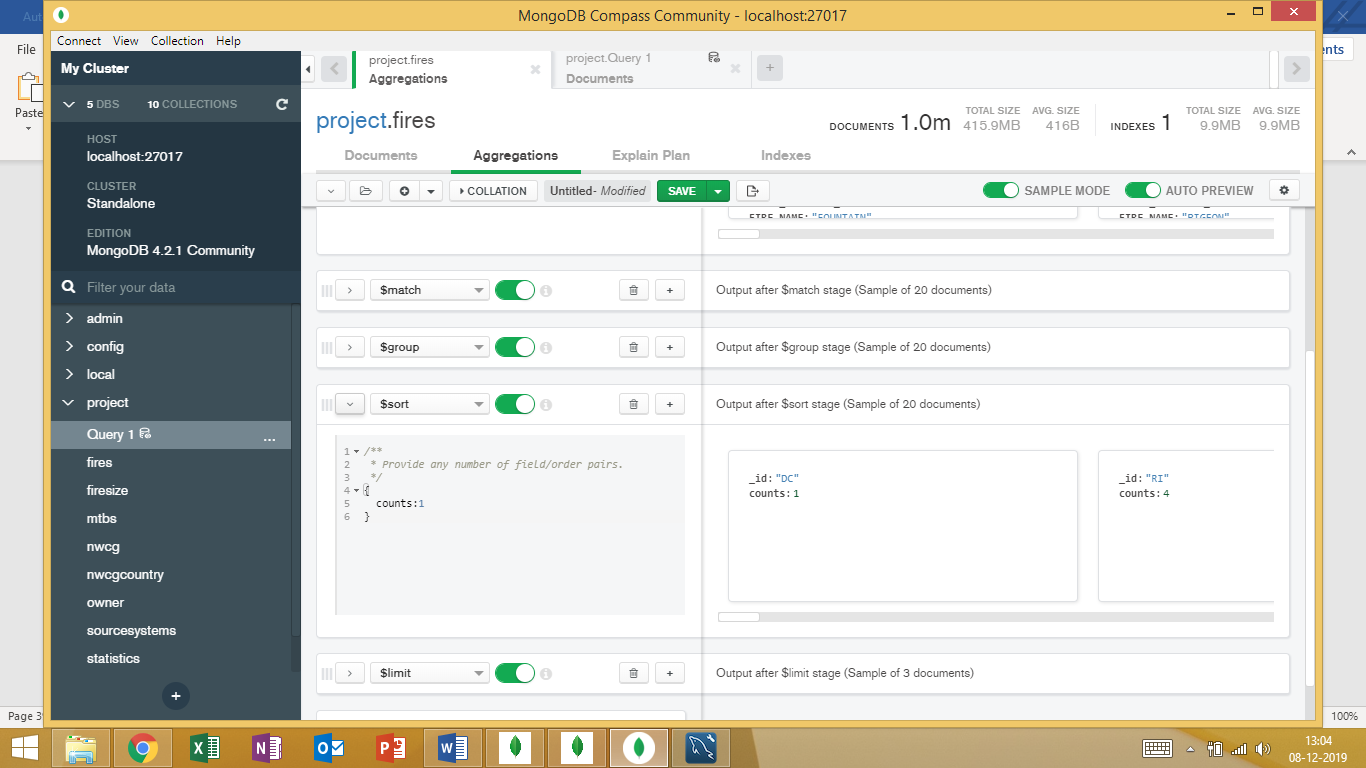
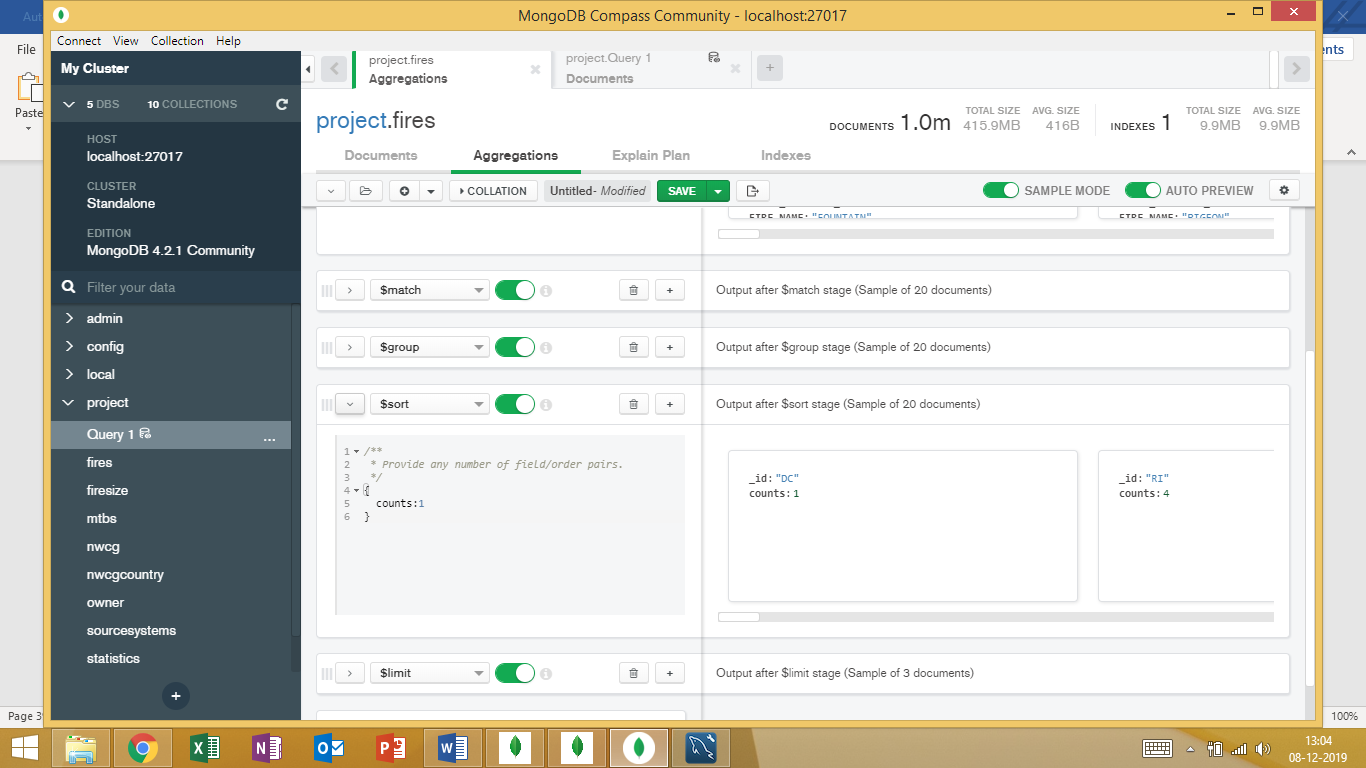
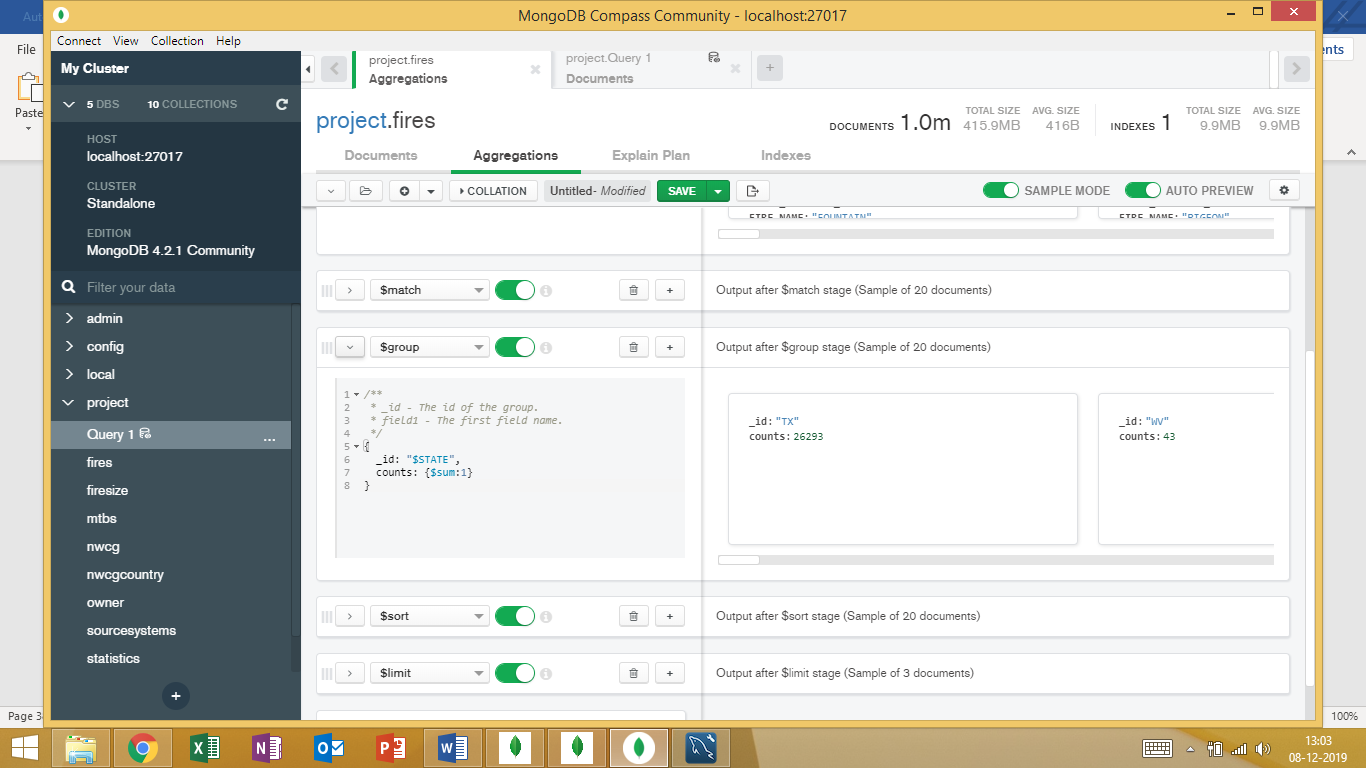
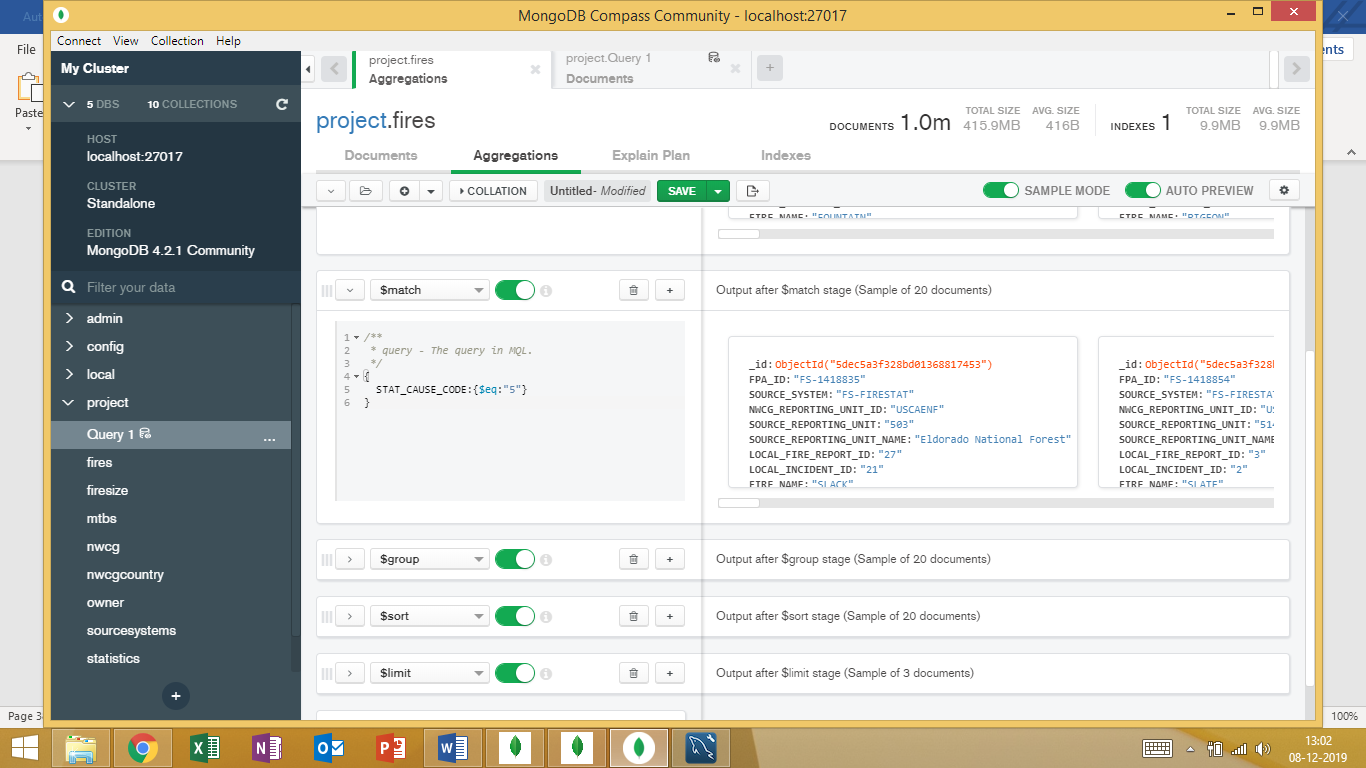
The statistical cause code for fires caused by debris is 5. So, the count of statistical cause code where the Stat\_Cause\_Code is equal to 5, is displayed as fires caused by debris.

The states with least number of fires caused by debris are Washington DC, Puerto Rico and Rhode Island. The result displays three states because, Puerto Rico and Rhode Island have equal number of fires caused due to debris.

### Translation

Select the count of fires caused due to debris and the state of the forest from the fires document. Group by the state. Order the number of fires caused due to debris in ascending order to display the forests with least number of fires. Limit the result to 3.

### Screen Shot of MongoDB Query/Code and Results



## Query 2

### Question

One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 4 children in a group visiting a forest. Name top 5 forests where this would be the least appropriate.

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

Assumptions: It is least appropriate to ban children from forests that have the highest number of fires that did not involve children.

The statistical cause code for fires caused due to children is 8. So, the count of statistical cause code where the Stat\_Cause\_Code is not equal 8, is the number of fires not involving children.

Since the following five forests have the highest number of fires that did not involve any children, the

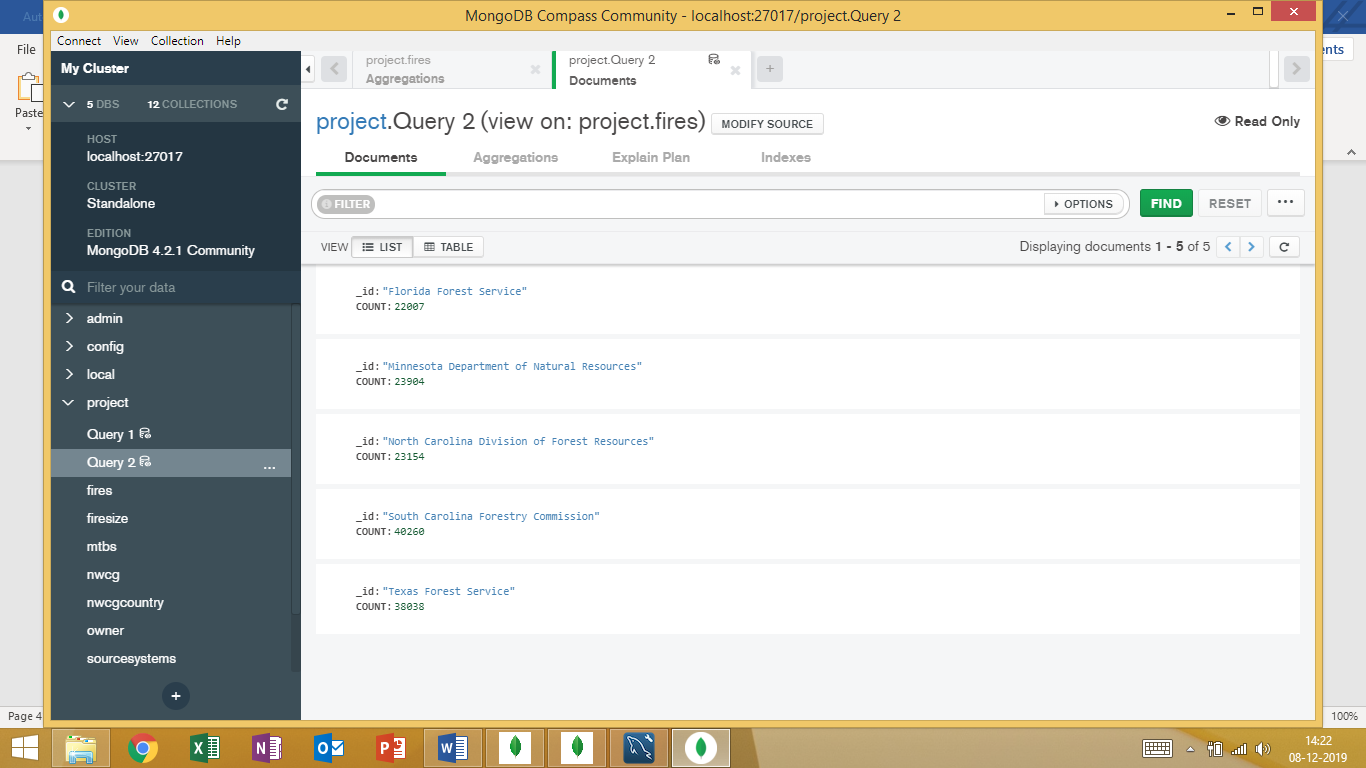
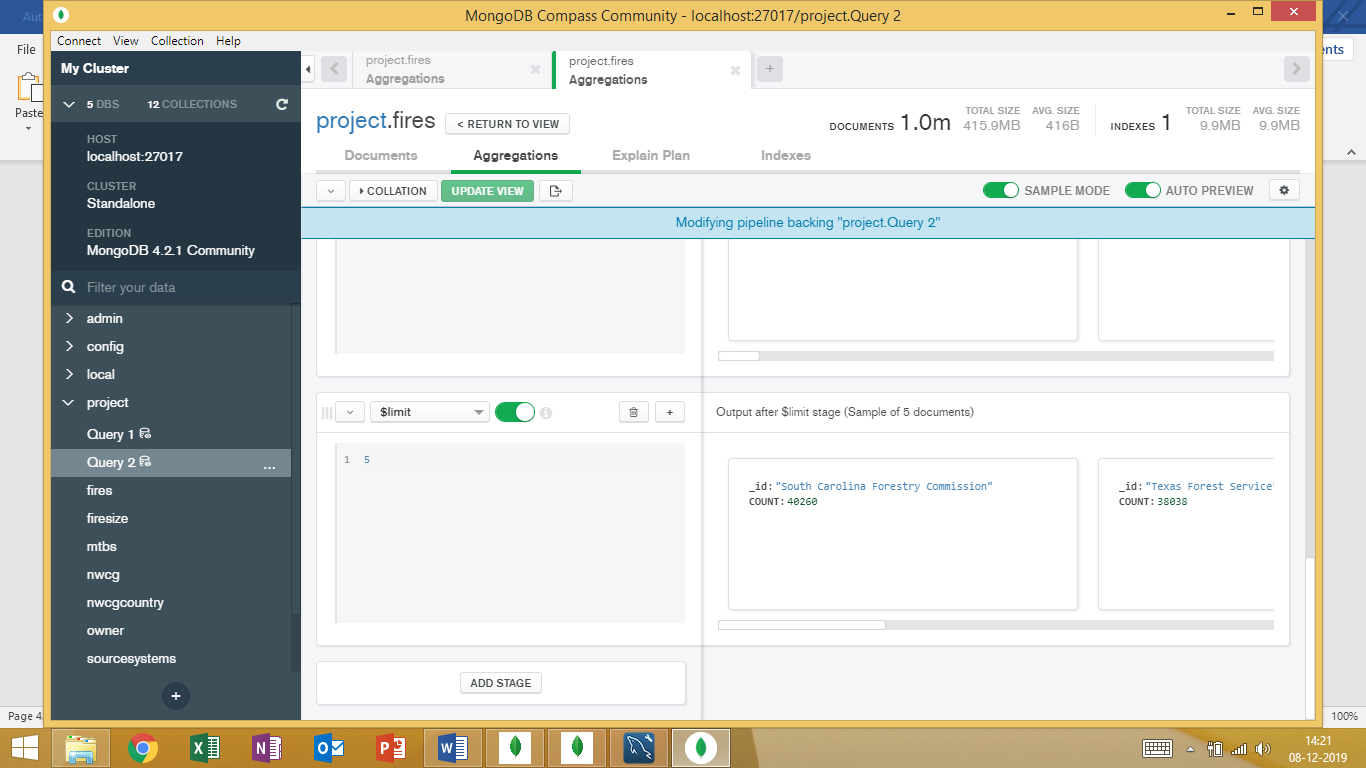
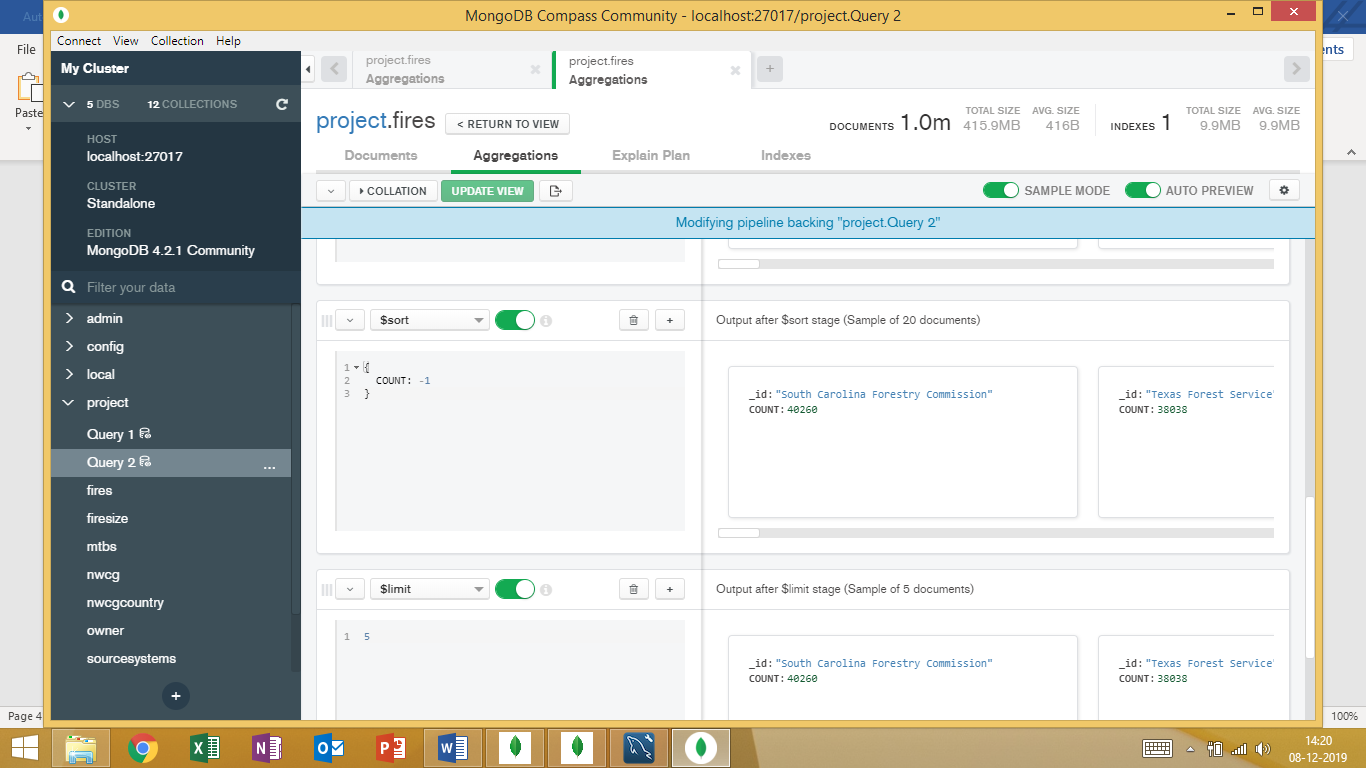
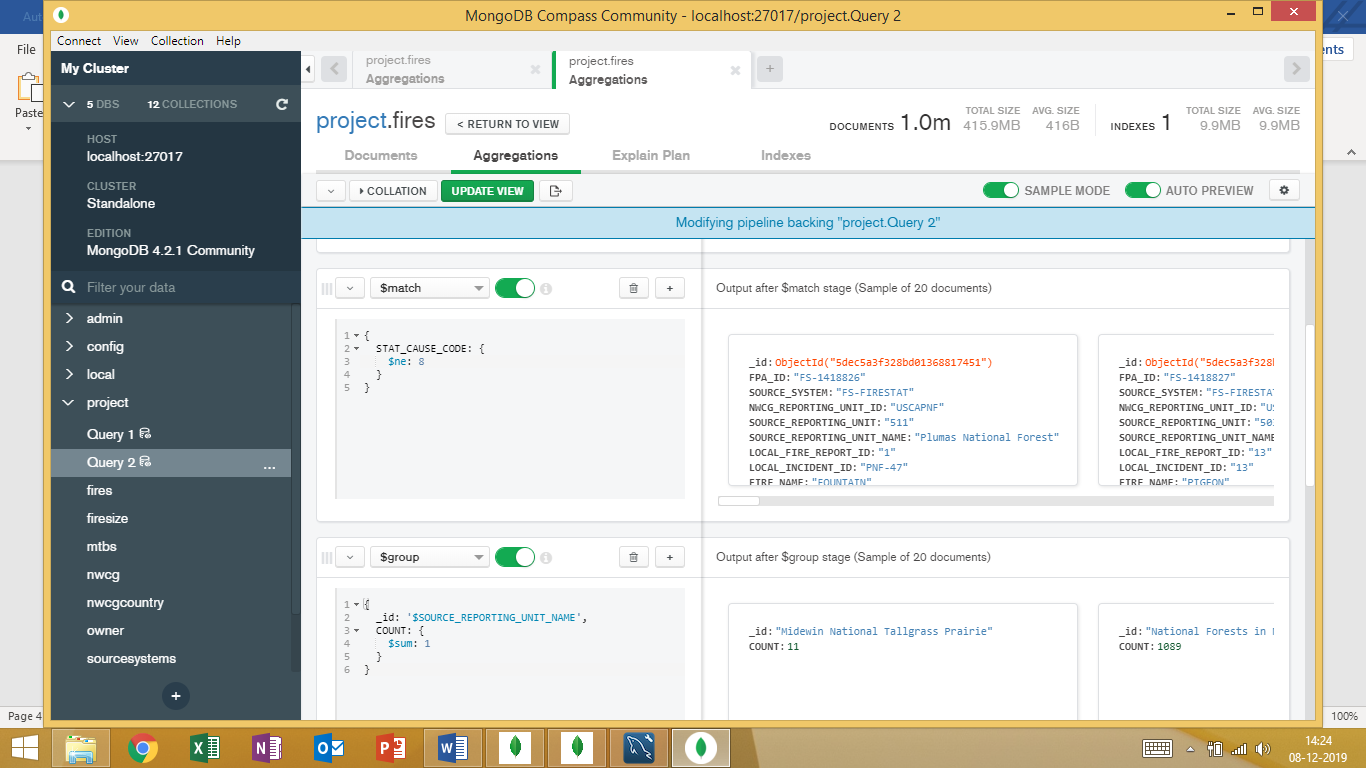
ban on children will be the least appropriate for these forests.

Forests with most fires not involving children: South Carolina Forestry Commission, Texas forest Service, Minnesota Department of Natural Resources, North Carolina Division of Forest Resources, Florida Forest Service.

### Translation

Select the Source reporting unit name as forest name and the number of fires that were not caused by children as fires not involving children from the fires table. Group by the source reporting unit name. Order the number of fires that were not caused by children in descending order. Limit the result to 5.

### Screen Shot of MongoDB Query/Code and Results



## Query 3

### Question

One advocacy group says human actions and nature are equally to blame for most wildfires. Write a query that can help determine the truth of this statement.

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

Assumptions: In the database the only natural cause of fire is lightning. All the other fires were caused due to human actions.

The statistical cause code for fires caused due to lightning is 1. So, the count of statistical cause code where the Stat\_Cause\_Code is not equal to 1, is displayed as *HUMAN*, and the count of statistical cause code where the Stat\_Cause\_Code is equal to 1, is displayed as *NATURAL*.

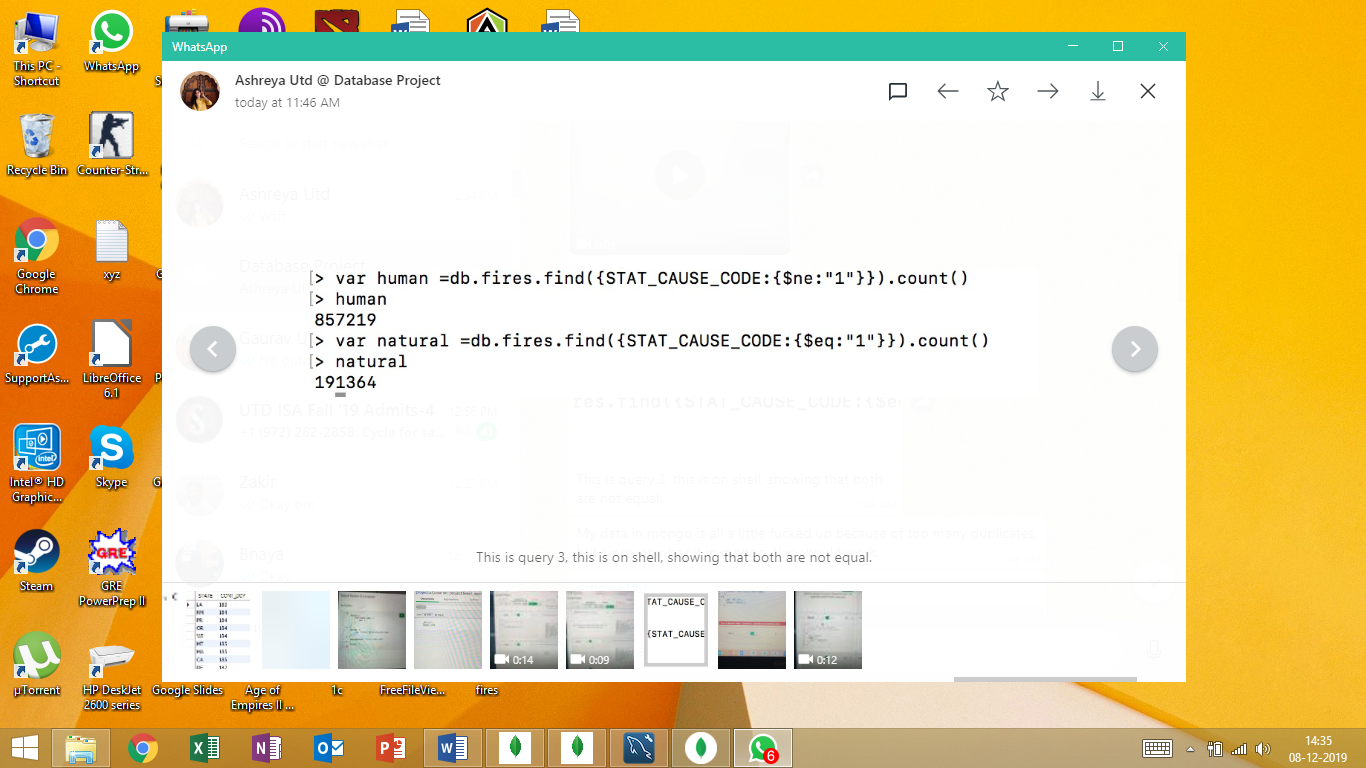
Since the number of fires caused by human actions is significantly greater than natural causes, human actions and nature are not equally to blame for most wildfires.

\*\* As you will notice, the numbers are slightly off compared to what was shown in the SQL output. We strongly believe our query is correct. As mentioned in the data assumptions, we believe the data may be corrupted/lacks integrity. However, the story still holds true in both narratives drawn by the queries and we believe this output would still help determine the truth of the above statement.

### Translation

Select the number of fires caused due to human actions and the number of fires caused due to natural causes from the fires table.

### Screen Shot of MongoDB Query/Code and Results



## Query 4

### Question

How many wildfires were reported by at least two units/agencies?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

The above question mentions “units/agencies” and in our data we have columns for both units and agencies. We chose to use units for the simplicity of staying in the fires table.

Our output resulted in a zero which is not what we had anticipates since we got 40,390 in SQL. After trying many different solutions, we believe our logic and query are correct and attribute this corrupted data in the fires table.

### Translation

Select the number of wildfires reported from fires table where the count of source reporting unit with respect to the fire name is greater than 1.

### Screen Shot of MongoDB Query/Code and Results



## Query 5

Which state had fires only in the second half of the calendar years?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

Assumptions: As there are 365 days in an year, the second half of the calendar years means that the day of year is greater than or equal to 182. The CONT\_DOY is the day of the year (range: 1-366) that fire happened.   
 We had difficulty filtering distinct states in the WHERE clause to limit the ones ONLY in the second half of the year. We would use WHERE CONT\_DOY > 181 and it would still include states that had fires in the first half with their second half numbers.   
 To solve this, we used the $MIN operator. The rational was that the minimum CONT\_DOY of a certain state should be equal to or higher than 182.

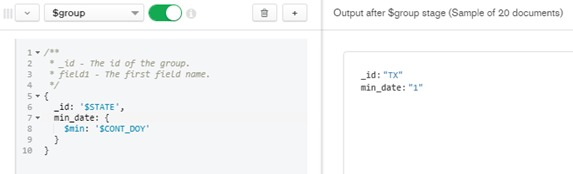
As you can see below, NY was determined to be the state to be in the correct range, which is different than what we found in SQL. However, as we mention the lack of integrity in the fires table leads us to believe that this is why the answers are different and that the SQL is correct.

\*\*Used Skip 6 because the first 6 values were null\*\*

### Translation

Select the state and minimum containment day of the year from the fires table. Group by the state. Order the result in descending order according to the minimum containment day of year.

### Screen Shot of MongoDB Query/Code and Results







## Query 6

### Question

Which forest had the number of fires equal to the average number of wild fires in the US?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

The average number of fires in the U.S is calculated by averaging the count of fires in each U.S state then comparing that to the count of fires in each forest.

SOURCE\_REPORTING\_UNIT\_NAME Is the closest column to represent forests.

We included a collection that was created based off one of the subqueries outputs. Since the initial SQL written had two subqueries, we thought that creating a collection that already included the count of the fires in each state would make the Mongo query much easier to handle. The average number of fires in the U.S could not be calculated without breaking it into states. This collection is included in list of tables.

Results: The output is different than SQL and we result this to a corruption of data in the fires table.

### Translation

Select the number of fires for individual forests and the average number of wildfires in the US from the fires table.

### Screen Shot of MongoDB Query/Code and Results



