MongoDB Project – Google Store Visitor Data

BUAN 6320.00#

Group Members

Contents

Data Review

Assumptions/Notes About Data Collections, Attributes and Relationships between Collections

We made some changes to the data:

- We deleted all values that do not have any meaning such as: ‘not available in demo dataset’
- fullVisitorID’s format was changed so that the whole ID is shown

Then, we uploaded the entire dataset as a single collection.

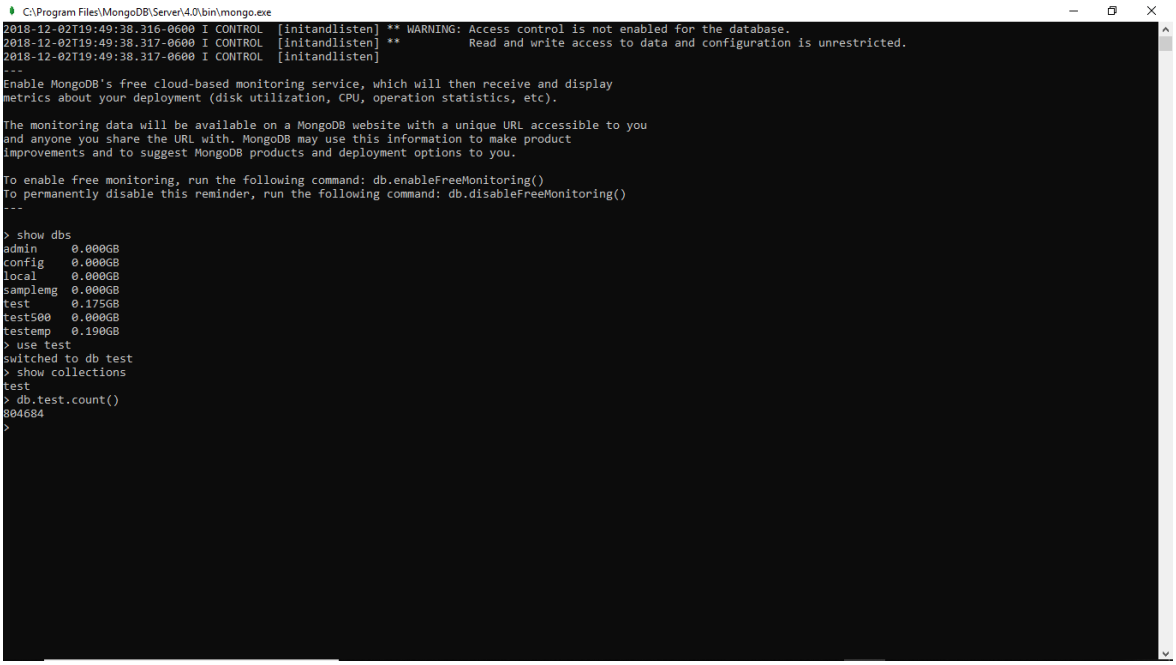
Physical Database

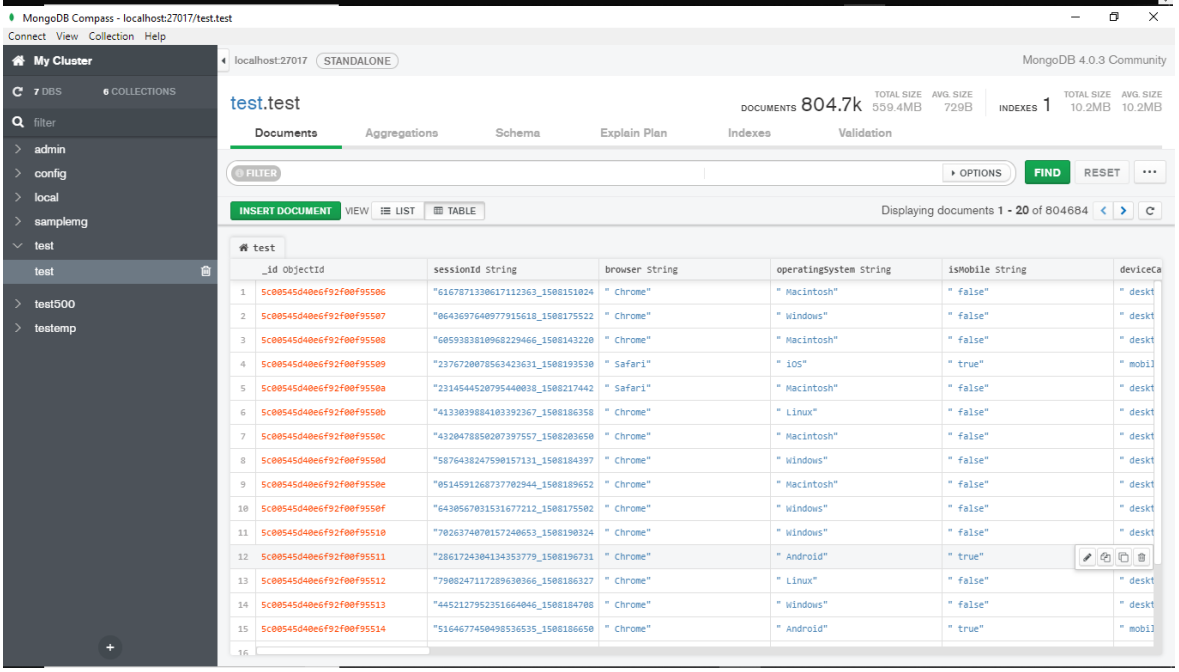
Assumptions/Notes About Data Set

The data was collected in the order of visit date

visitNumber increments with each visit made by the same visitor

Screen shot of Physical Database objects (Database, Collections and Attributes)





Data in the Database

Collection Name		Relationshps With Other Collections (if any)	# of Rows in Table
Test			804684

MongoDB Queries/Code

Query 1

Question: Which users had the minimum number of visits and when?

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

I assumed that the minimum number of visit would be 1.

The below screen shot displays the Users with only 1 visit and the time they visited

# of rows: 403

Translation

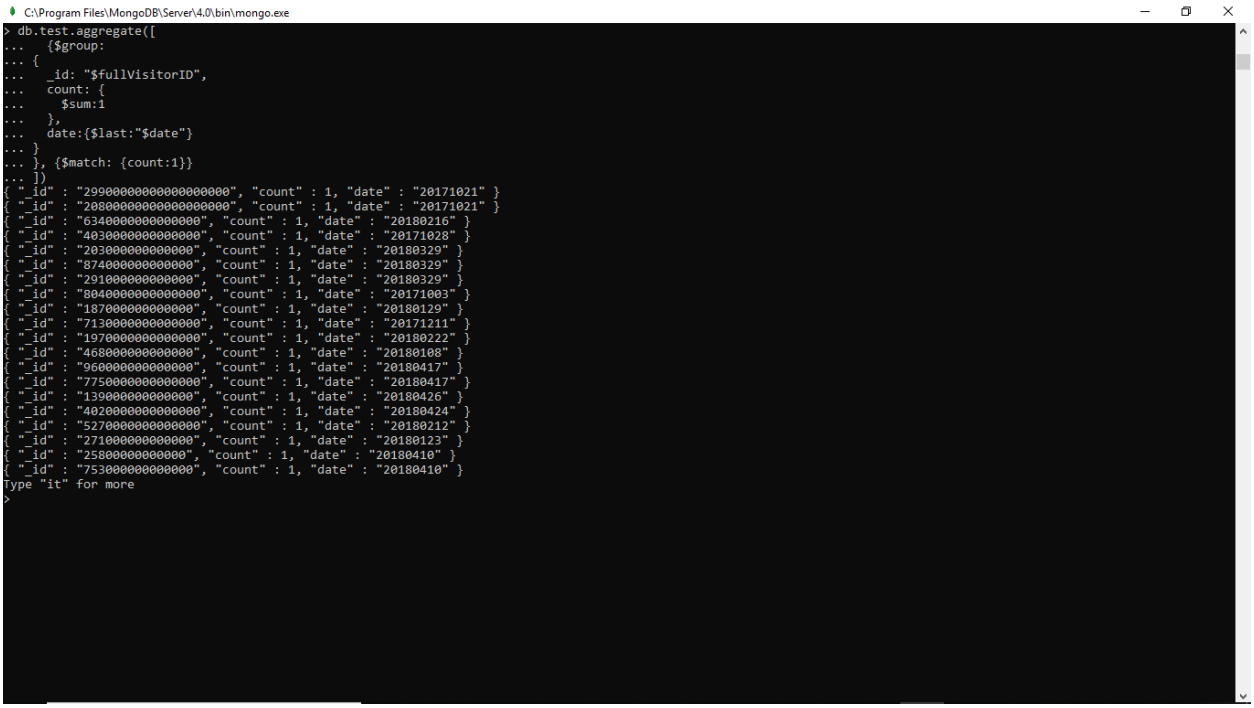
Group the data based on FullVisitorID

Count the number of result in each group (the number of visit by each visitor), then take the date they visited

Filter results that have count = 1

➔ result will display fullvisitorID, count and date

Screen Shot of MongoDB Query/Code and Results



Query 2

Question: Which operating system (devices) was the most popular amongst store visitors with non-mobile devices?

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

Windows is the most popular amongst store visitors with non-mobile devices.

We can understand this question in 2 ways:

- Count & compare the number of people (unique fullvisitorID) that used each type of operating system among visitors with non-mobile devices.
- Count & compare the number of times each operating system was used to visit the store by visitors with non-mobile devices.

I chose the second approach because a single visitor might use different operating system to visit the store. Therefore, we should take into consideration each visit, not just the number of visitors that used that operating system.

# of rows: 13

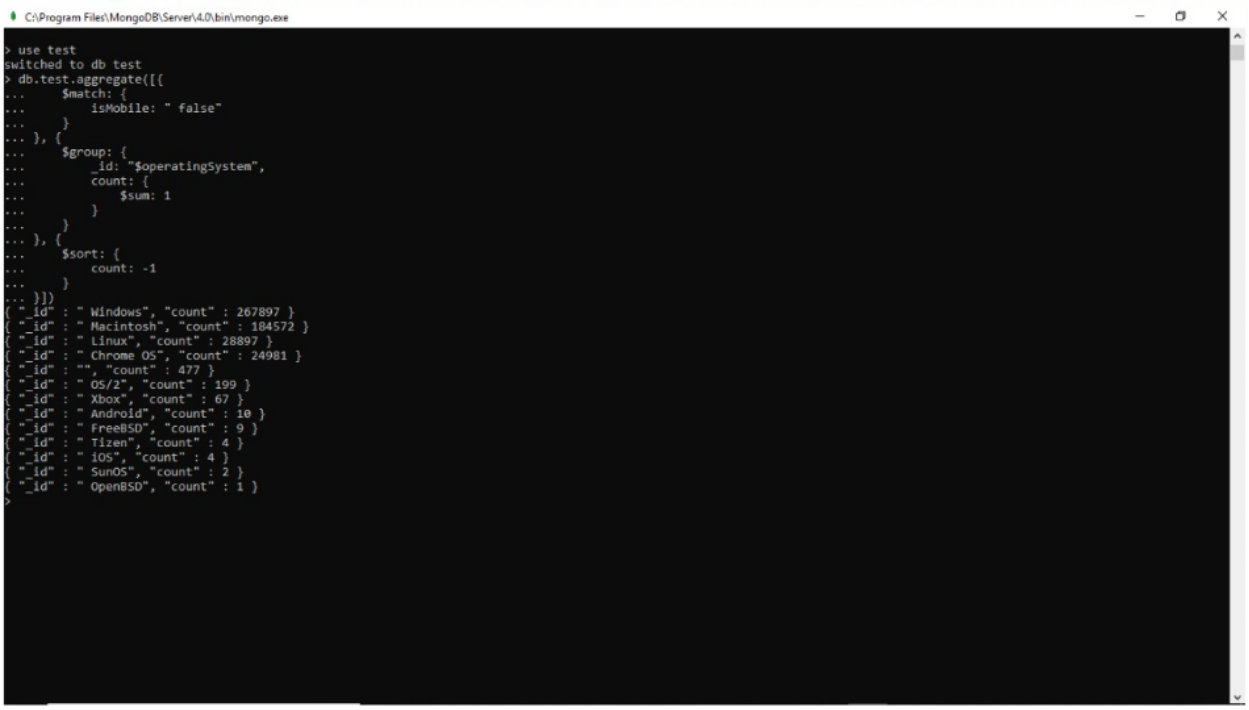
Translation

Filter the data for visits that used Non-Mobile Devices.

Group the data based on operating system and count the number of documents in the group.

Sort the result in the descending order.

Screen Shot of MongoDB Query/Code and Results



Query 3

Question: Which date had the least and most number of visitors with mobile devices?

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

Least: 2017/12/31 count: 543

Most: 2017/12/12 count: 6499

To get unique visitors that use mobile devices, I filtered for documents that have visitNumber: 1 and isMobile: “true”

Then, I grouped the data by date, count the number of rows in each group and sort by count

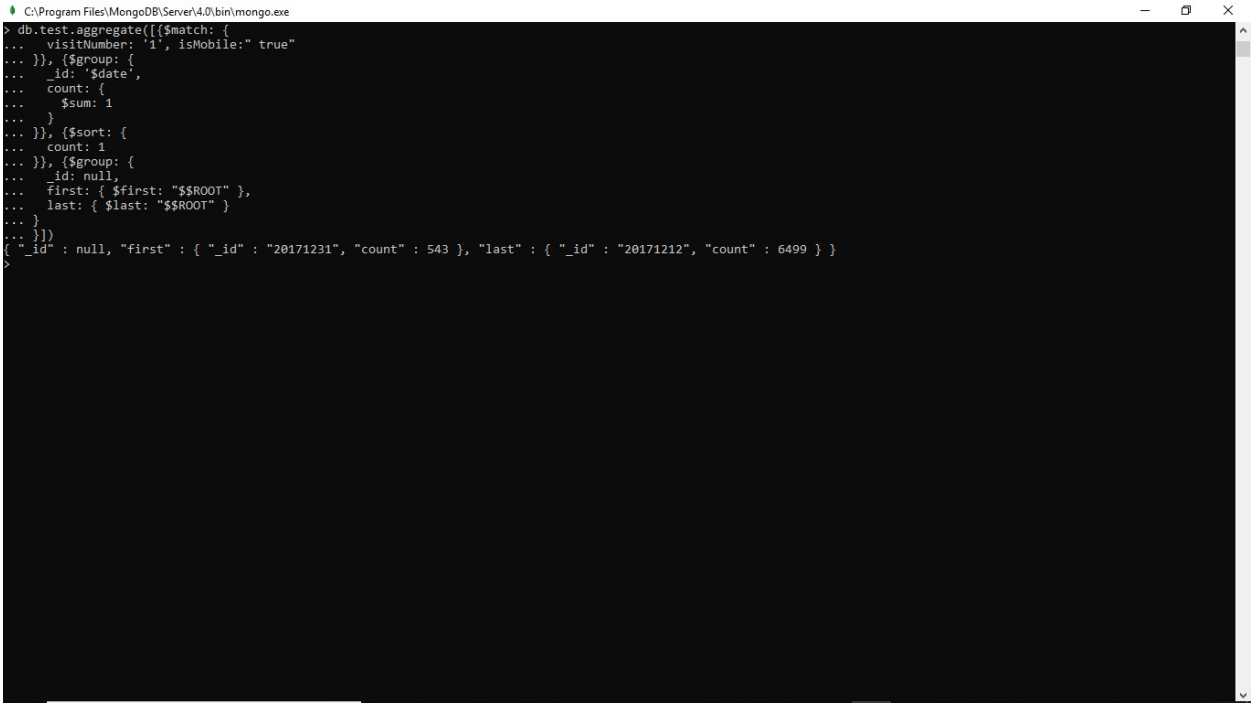
After that, I get the first and last rows of the result set. These are the days with least and most visitors

# of rows: 1

Translation

Count the number of visitors using mobile devices to visit each day, then sort by count and show the first and last results

Screen Shot of MongoDB Query/Code and Results



Query 4:

Question: were macintosh users more socially engaged than windows users?

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

Both the Macintosh users and windows users are 100% not socially engaged.

If there are visits that are socially engaged with Mac and Windows, there'll be 4 lines of results for 4 groups and we can compare the proportion to see which type of user is more socially engaged

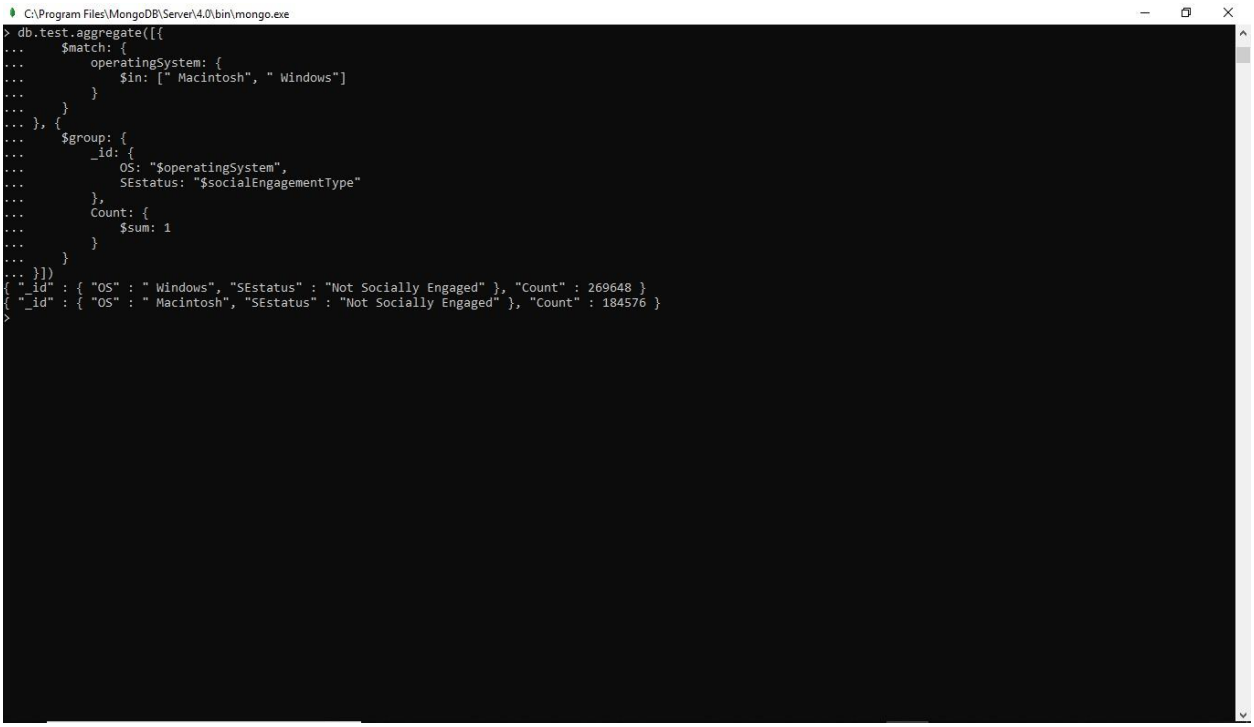
# of rows: 2

Translation

Filter the data for visits with Macintosh and Windows operating systems.

Group by operating system and social engagement status, then count the number of users.

Screen Shot of MongoDB Query/Code and Results



Query 5

Question: Provide a breakdown of unique visitors by country?

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

The below screen shot provides the screen shot of unique visitors by country.

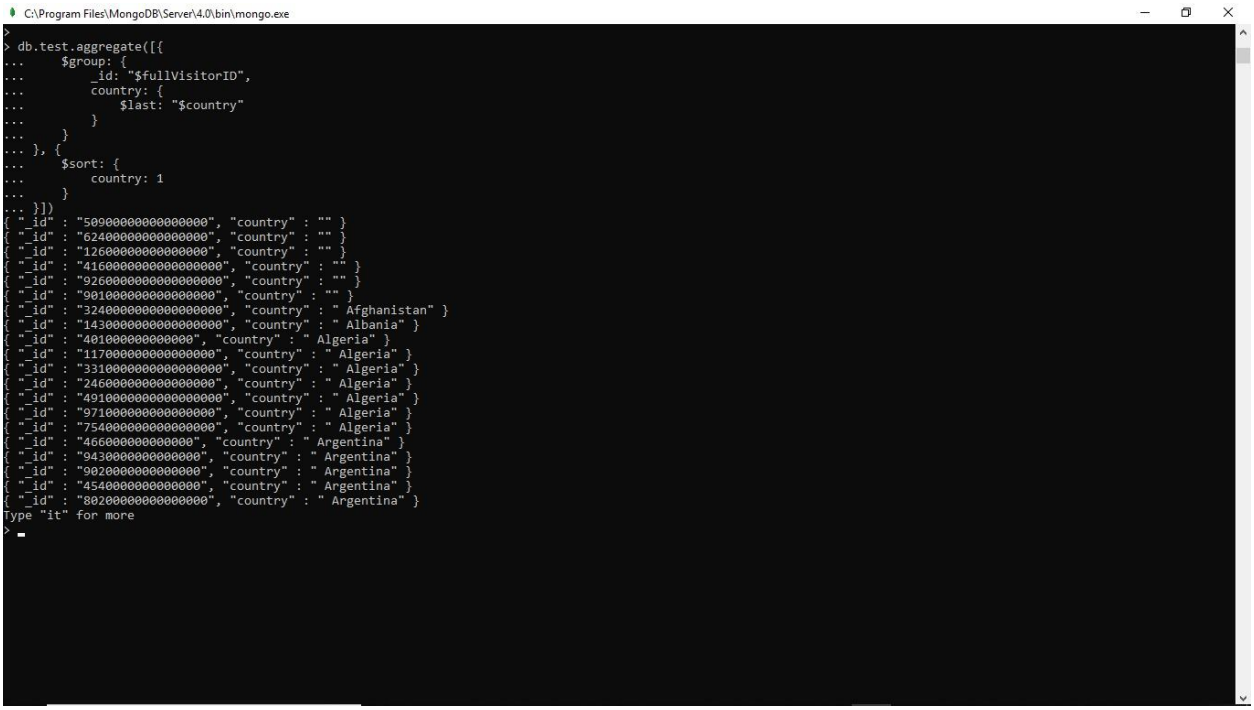
# of rows: 3827

Translation

Group the dataset by FullVisitorID and get the last country that visitor is in.

Sort the result in the ascending order.

Screen Shot of MongoDB Query/Code and Results



Query 6

Question: Which country had the most number of macintosh users?

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

United States had the most number of macintosh users.

# of rows: 82

Translation

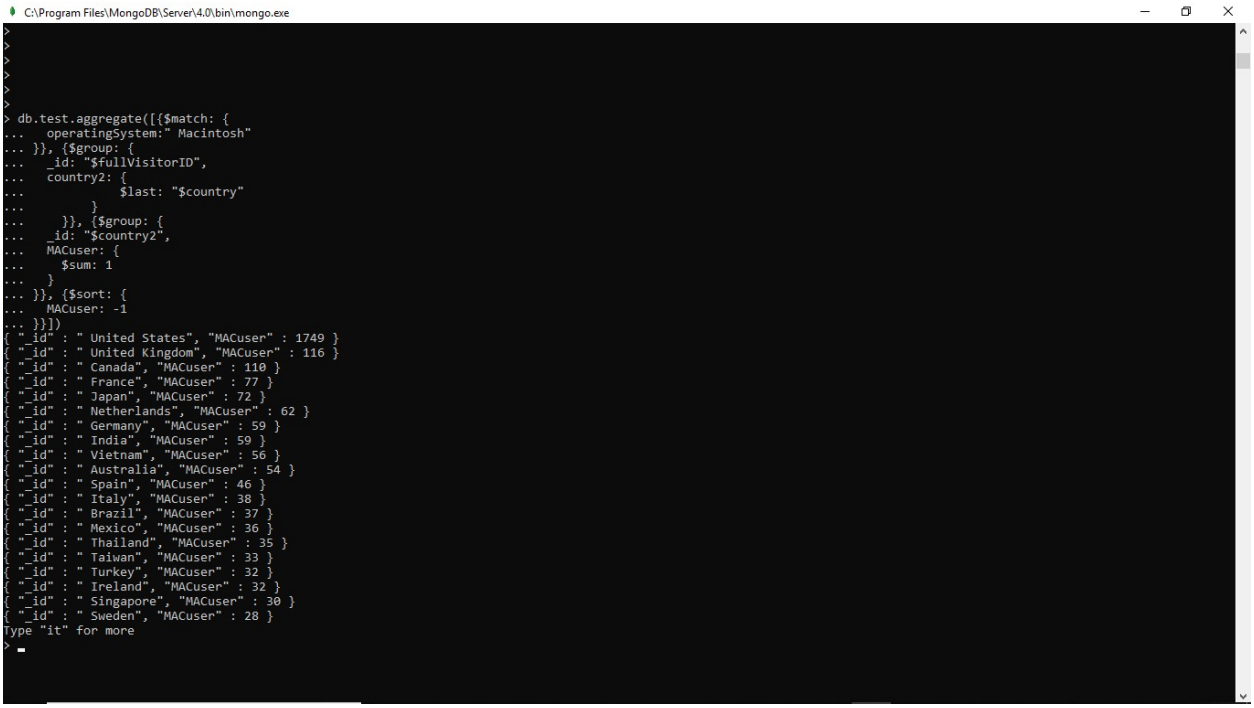
Filter the data for visits with Macintosh operating system.

Group the data based on the FullVisitorID, get the last country that visitor is in.

Group the result set by country, then count the rows in each group. This is the number of Mac users

Sort the result set to get the highest count

Screen Shot of MongoDB Query/Code and Results



Query 7

Question: Which visitor generated the highest amount of pageviews and when?

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

I compared the pageviews in each visits for the visit with highest pageviews and projected the visitorID, number of pageviews and the date that visit was made

# of rows: 804684

Translation

Project fullVisitorID, pageviews and date

Sort by the number of pageviews in descending order

Screen Shot of MongoDB Query/Code and Results