Group #: 14

Vikrant Dev Rathore

Krishna Darakh

Ritish Gupta

Rohit Gurjar

**SQL Project – Google Store Visitor Data**

BUAN 6320.006

Contents

[Data Model 3](#_Toc528400248)

[Assumptions/Notes About Data Entities and Relationships 3](#_Toc528400249)

[Entity-Relationship Diagram 3](#_Toc528400250)

[Physical Database 4](#_Toc528400251)

[Assumptions/Notes About Data Set 4](#_Toc528400252)

[Screen shot of Physical Database objects 4](#_Toc528400253)

[Data in the Database 4](#_Toc528400254)

[SQL Queries 5](#_Toc528400255)

[Query 1 5](#_Toc528400256)

[Question 5](#_Toc528400257)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 5](#_Toc528400258)

[Translation 5](#_Toc528400259)

[Screen Shot of SQL Query and Results 5](#_Toc528400260)

[Query 2 6](#_Toc528400261)

[Question 6](#_Toc528400262)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 6](#_Toc528400263)

[Translation 6](#_Toc528400264)

[Screen Shot of SQL Query and Results 6](#_Toc528400265)

[Query 3 7](#_Toc528400266)

[Question 7](#_Toc528400267)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 7](#_Toc528400268)

[Translation 7](#_Toc528400269)

[Screen Shot of SQL Query and Results 7](#_Toc528400270)

[Query 4 8](#_Toc528400271)

[Question 8](#_Toc528400272)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 8](#_Toc528400273)

[Translation 8](#_Toc528400274)

[Screen Shot of SQL Query and Results 8](#_Toc528400275)

[Query 5 9](#_Toc528400276)

[Question 9](#_Toc528400277)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 9](#_Toc528400278)

[Translation 9](#_Toc528400279)

[Screen Shot of SQL Query and Results 9](#_Toc528400280)

[Query 6 10](#_Toc528400281)

[Question 10](#_Toc528400282)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 10](#_Toc528400283)

[Translation 10](#_Toc528400284)

[Screen Shot of SQL Query and Results 10](#_Toc528400285)

[Query 7 11](#_Toc528400286)

[Question 11](#_Toc528400287)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 11](#_Toc528400288)

[Translation 11](#_Toc528400289)

[Screen Shot of SQL Query and Results 11](#_Toc528400290)

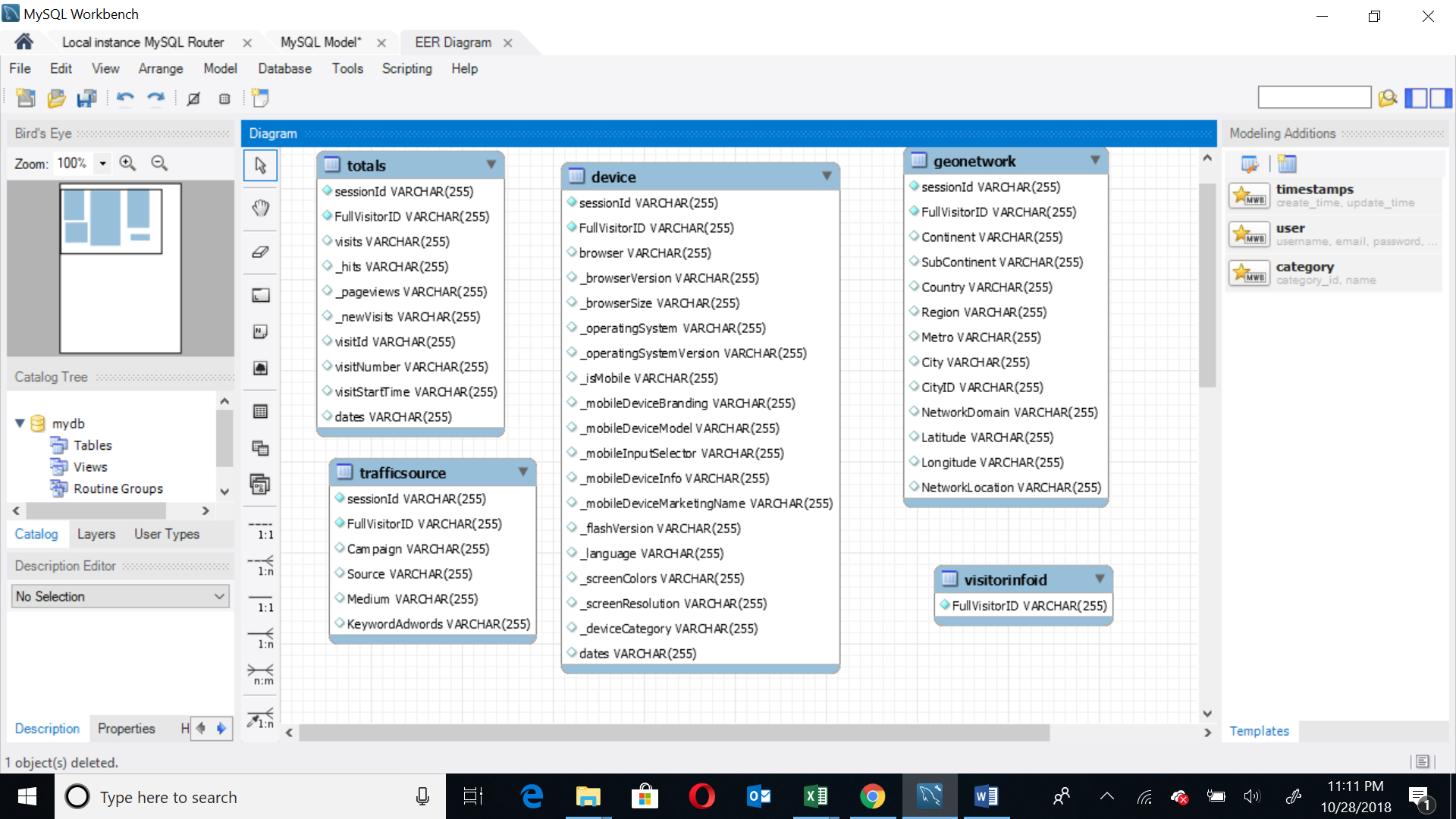
# Data Model

## Assumptions/Notes About Data Entities and Relationships

We have assumed that there are 5 tables that are device, geonetwork ,totals ,trafficsource ,visitorid. All the columns in each table are mentioned. We were not able to insert relationships because when we create primary keys, the value of the primary key changes for some reason that we do not know. Hence we added the sessionID,FullVisitorID to each table.

For making the FullVisitorID, we were able to remove the ‘E’ values by using the LEFT command in Excel which was LEFT(A2,FIND("\_",A2)-1) where A2 is SessionID because sessionID was formed by a combination of FullVisitorID and VisitID.

## Entity-Relationship Diagram



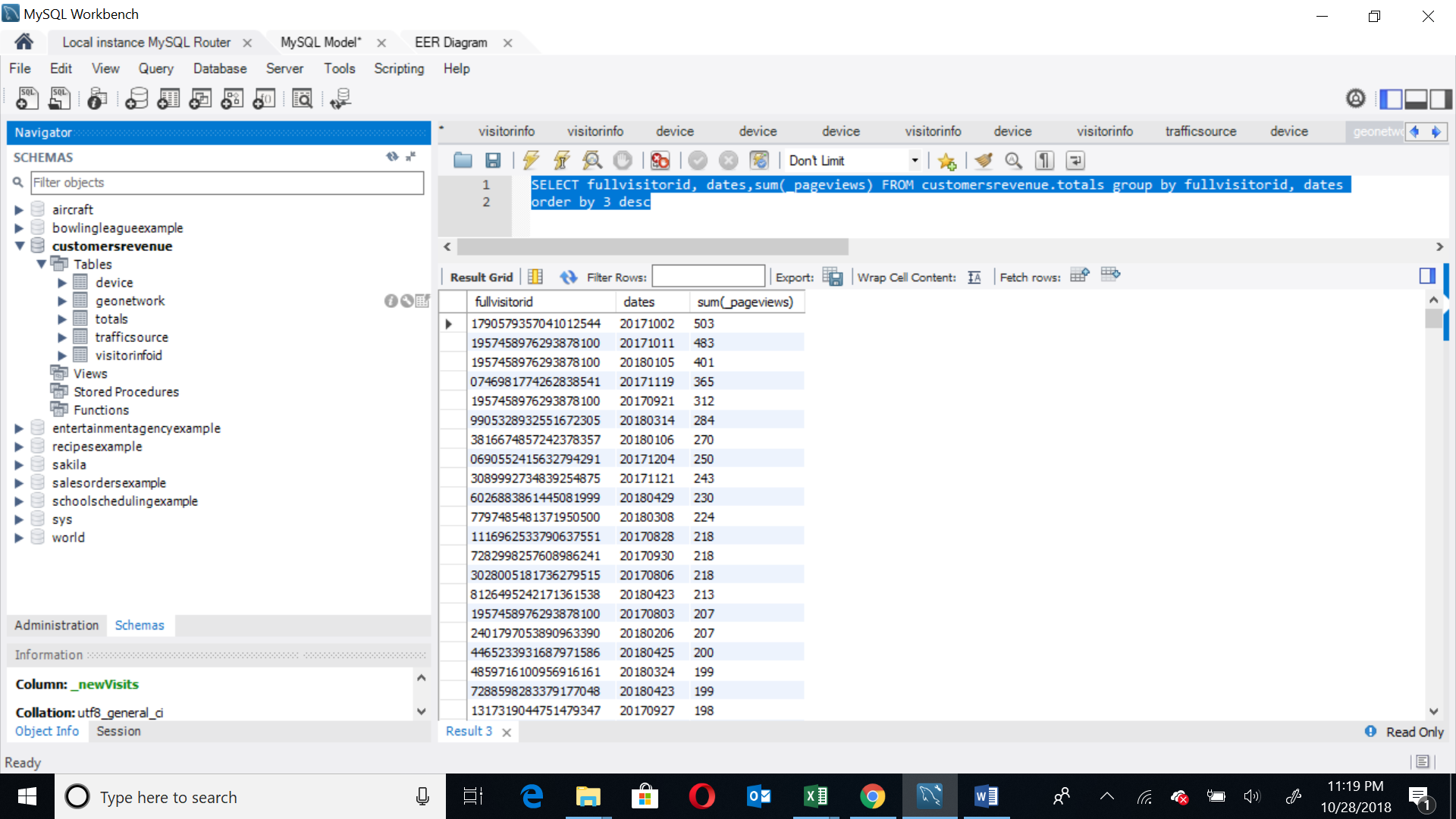
# Physical Database

## Assumptions/Notes About Data Set

For making the FullVisitorID, we were able to remove the ‘E’ values by using the LEFT command in Excel which was LEFT(A2,FIND("\_",A2)-1) where A2 is SessionID because sessionID was formed by a combination of FullVisitorID and VisitID.

We used the Load Infile query in mysql and changed the default security settings. We imported data from multiple excel sheets into multiple tables.

## Screen shot of Physical Database objects



## Data in the Database

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table** |
| Device | - | - | 804684 |
| Geonetwork | - | - | 804684 |
| Totals | - | - | 804684 |
| trafficsource | - | - | 804684 |
| Visitorinfoid | FullVisitorID( not primary key but unique) | - | 617242 |

# SQL Queries

## Query 1

### Question

Which user had the minimum number of visits and when?

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

This question is ambiguous and not clear in its meaning. Threre are multiple users with 1 visit and on multiple dates. Result has 681619 rows which are minimum number of visits.

### Translation

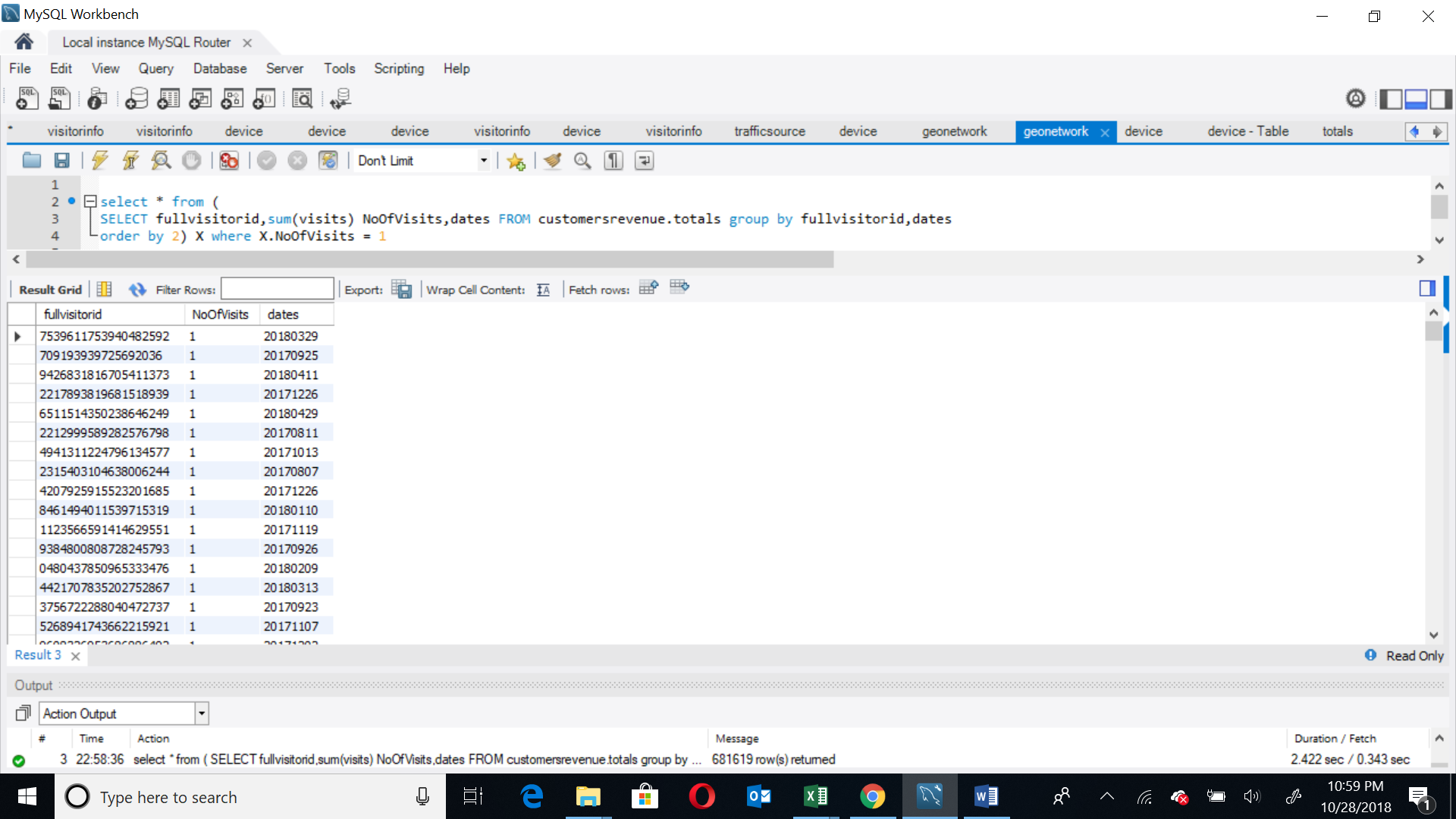
Select visitorId from totals table with minimum mumber of visits grouping by dates and visitorId

### Screen Shot of SQL Query and Results

select \* from (

SELECT fullvisitorid,sum(visits) NoOfVisits,dates FROM customersrevenue.totals group by fullvisitorid,dates

order by 2) X where X.NoOfVisits = 1



## Query 2

### Question:

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

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

We have taken \_isMobile=false so that we can get customers which are not using mobile devices.

### Translation

Select the operating system which is most popular amongst the users who are not accessing from mobile devices from the device table.

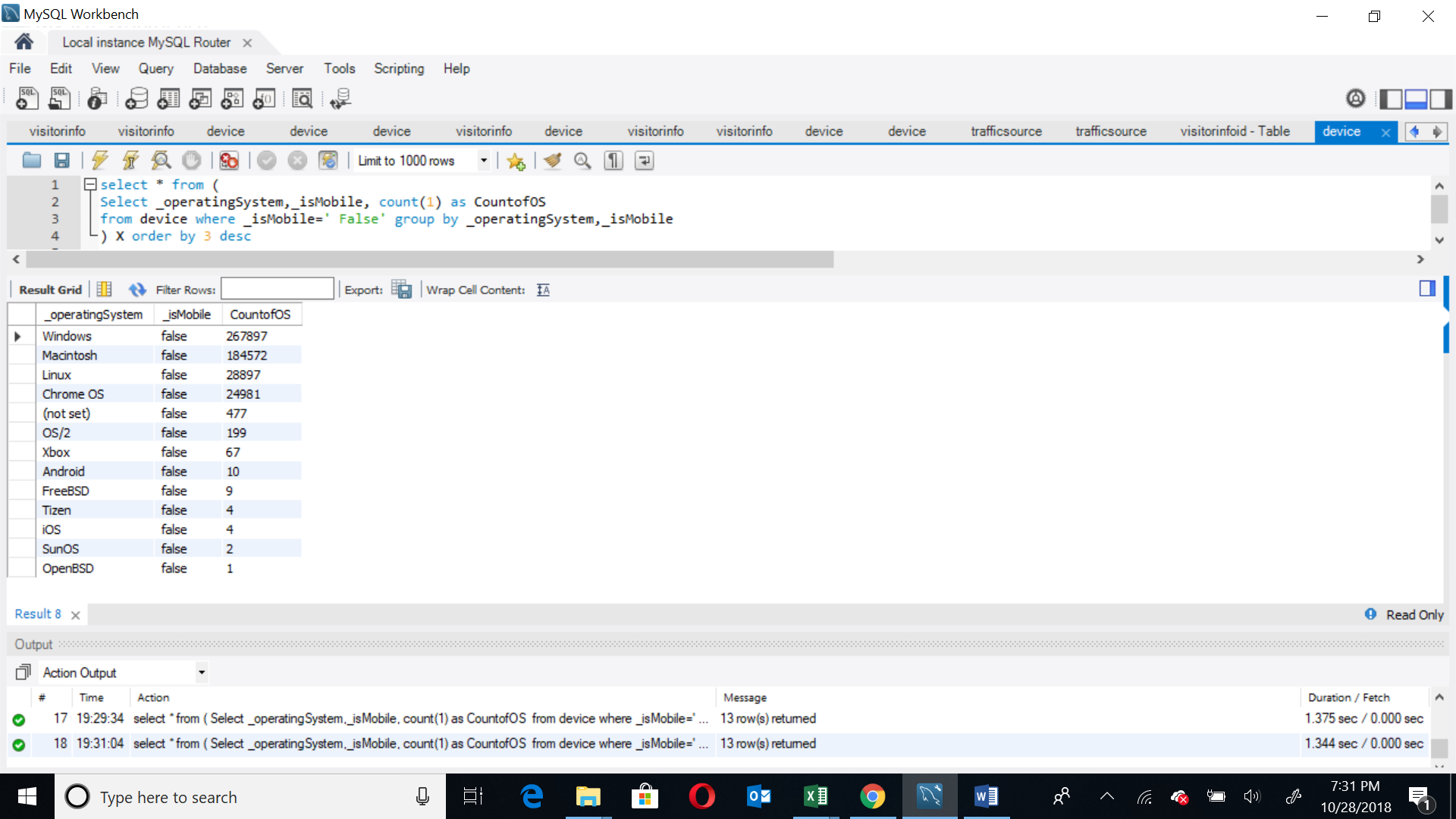
### Screen Shot of SQL Query and Results

select \* from (

Select \_operatingSystem,\_isMobile, count(1) as CountofOS

from device where \_isMobile=' False' group by \_operatingSystem,\_isMobile

) X order by 3 desc



## Query 3

### Question

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

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

We took union to give max and min of dates with visitors for mobile devices. Number of rows is 2 which display the maximum and minimum visitors by dates.

### Translation

Select the dates which have the maximum visitors and the minimum visitors from the device table.

### Screen Shot of SQL Query and Results

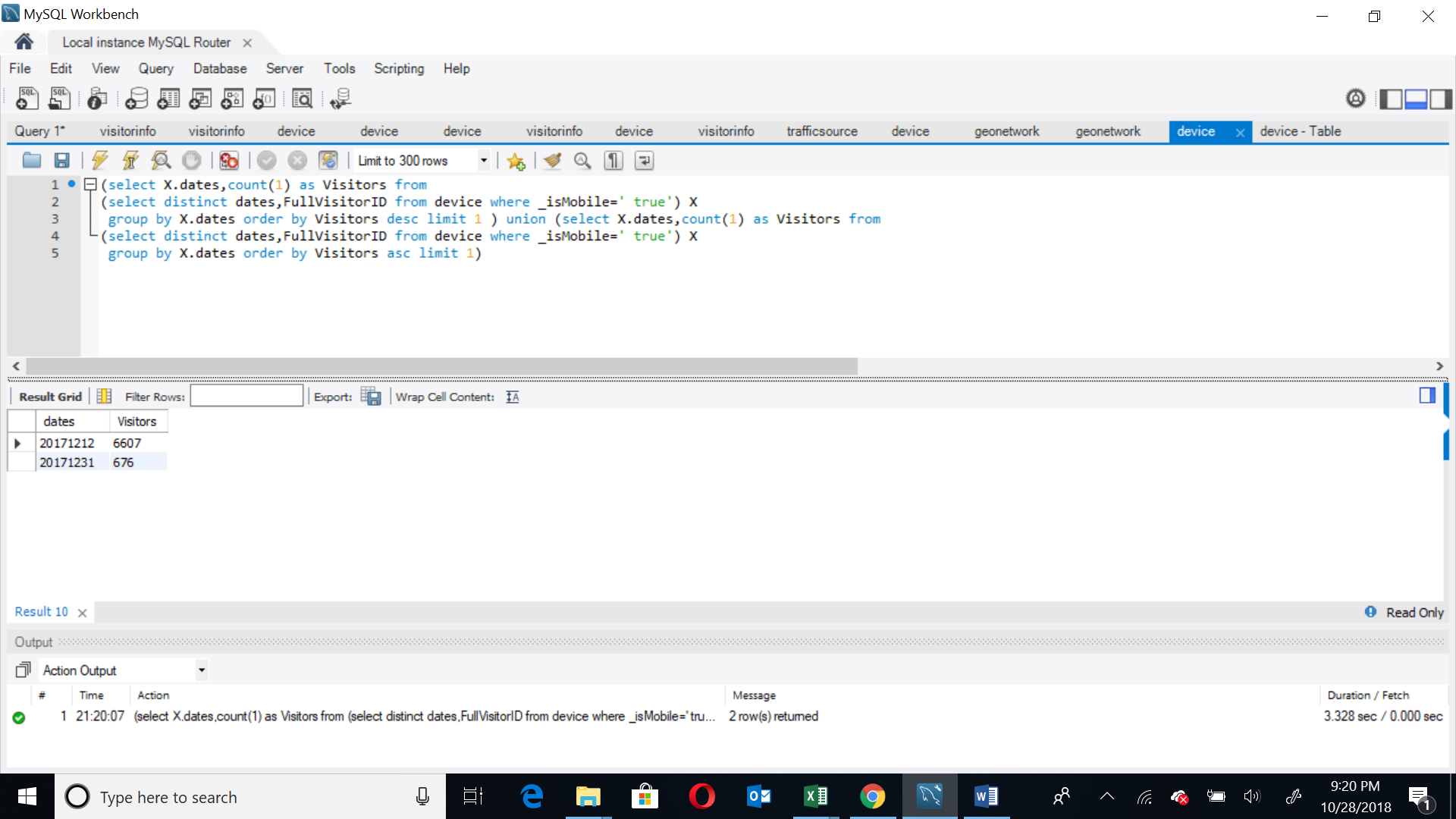
(select X.dates,count(1) as Visitors from

(select distinct dates,FullVisitorID from device where \_isMobile=' true') X

group by X.dates order by Visitors desc limit 1 ) union (select X.dates,count(1) as Visitors from

(select distinct dates,FullVisitorID from device where \_isMobile=' true') X

group by X.dates order by Visitors asc limit 1)



## Query 4

### Question

Which visitor generated the highest amount of pageviews and when?

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

The rows in the result are 736391 rows. Date 2017-10-02 has the maximum pageviews which are 503.

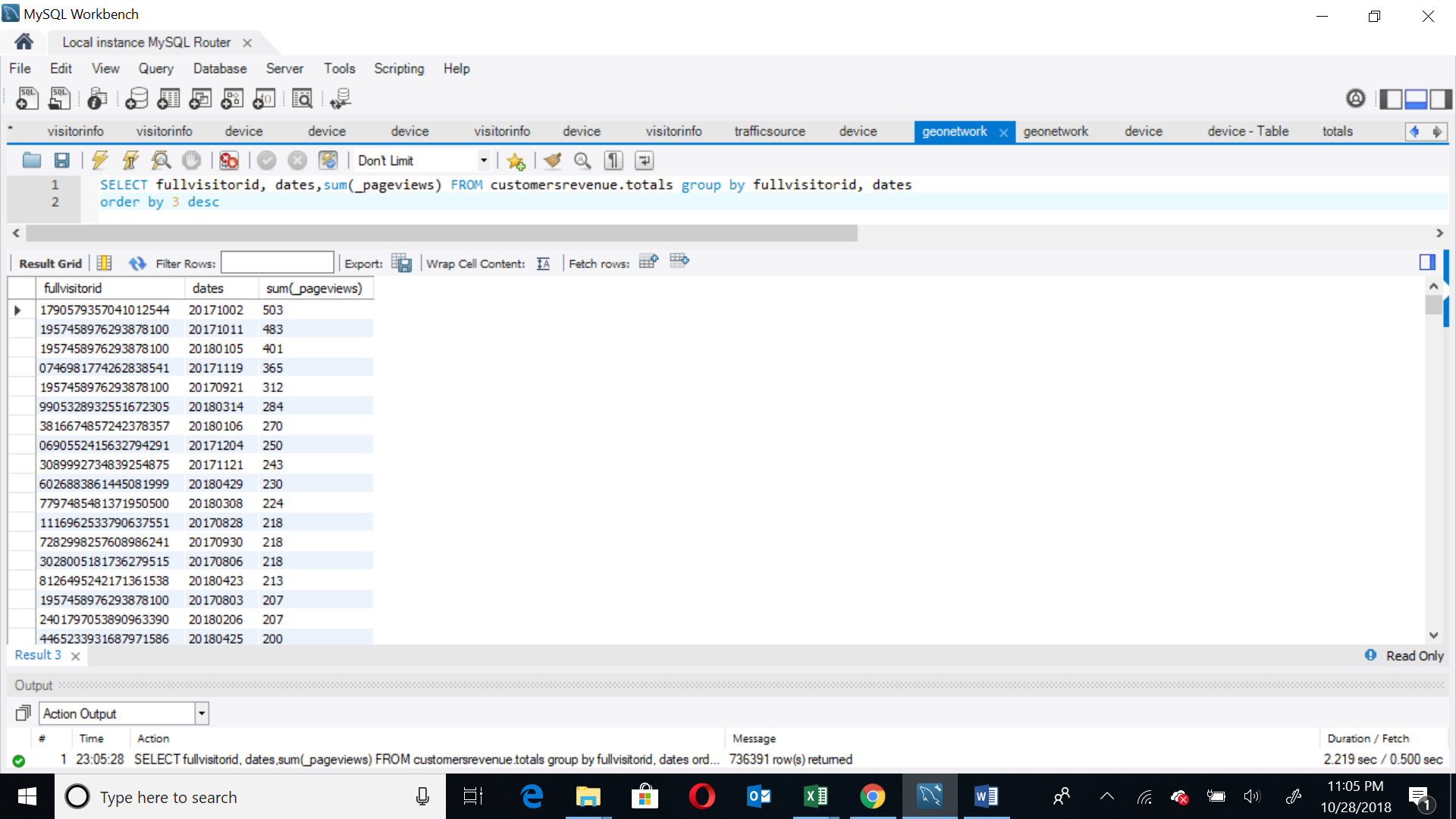
### Translation

Select visitorid,date,sum of pagesviews from totals table and group by visitorid and dates and arrange pageviews in descending order.

### Screen Shot of SQL Query and Results

SELECT fullvisitorid, dates,sum(\_pageviews) FROM customersrevenue.totals group by fullvisitorid, dates

order by 3 desc



## Query 5

### Question

Provide a breakdown of unique visitors by country.

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

The result contains 219 rows with the breakdown of unique visitors for each country.

### Translation

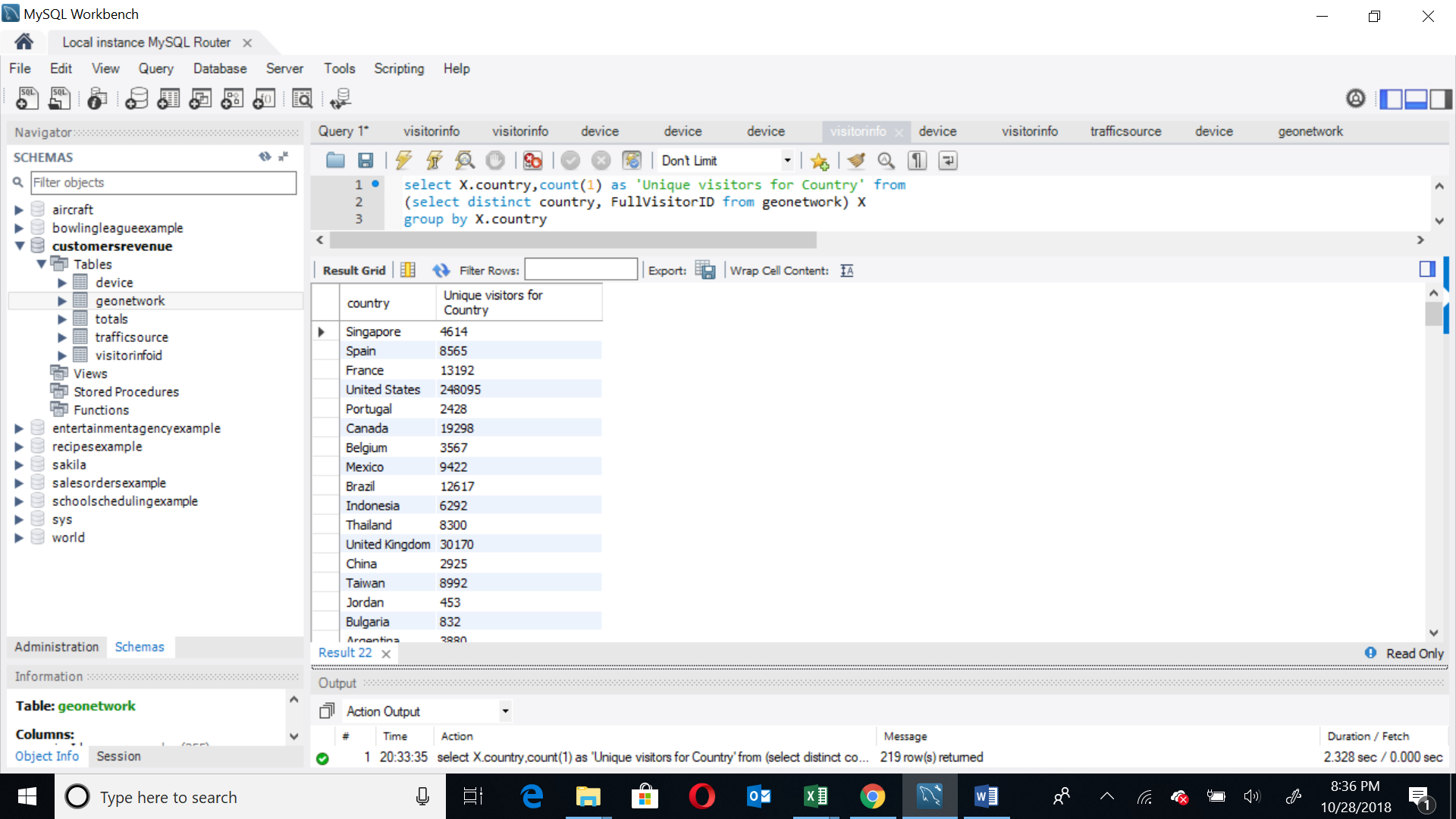
Select unique visitors for each country

### Screen Shot of SQL Query and Results

select X.country,count(1) as 'Unique visitors for Country' from

(select distinct country, FullVisitorID from geonetwork) X

group by X.country



## Query 6

### Question

How many users from Brazil used only iOS devices to visit the store?

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

The result contains 512 rows.

### Translation

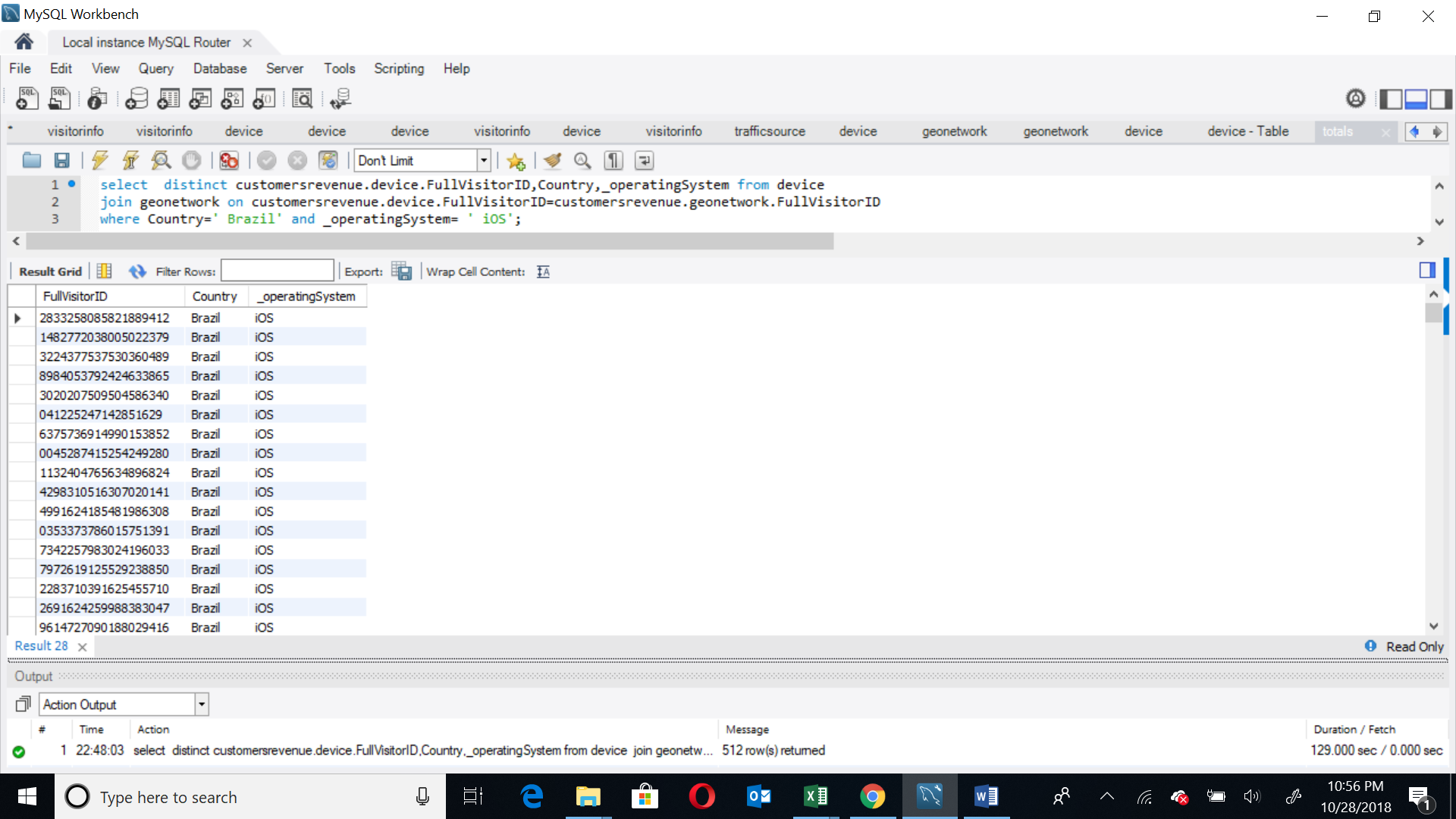
Select distinct visitorid from the device table joined with geonetwork table on geonetwork visitorId matching the device visitorID where the country is Brazil and Operating System is IOS.

### Screen Shot of SQL Query and Results

select distinct customersrevenue.device.FullVisitorID,Country,\_operatingSystem from device

join geonetwork on customersrevenue.device.FullVisitorID=customersrevenue.geonetwork.FullVisitorID

where Country=' Brazil' and \_operatingSystem= ' iOS';



## Query 7

### Question

Which country had the most number of macintosh users?

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

There are 175 rows in the result. United States has the maximum macintosh users with 65495 users.

### Translation

Select countries, from the geonetwork table and the count of operating system from the device table for those users who have operating system as macintosh and order by count of operating system in descending order.

### Screen Shot of SQL Query and Results

select b.country, count(1) as cntofcountry from (select distinct country, fullvisitorid from

geonetwork a where a.fullvisitorid in (select distinct(FullVisitorID) from device where \_operatingSystem = ' Macintosh') ) b

group by country order by cntofcountry desc;

