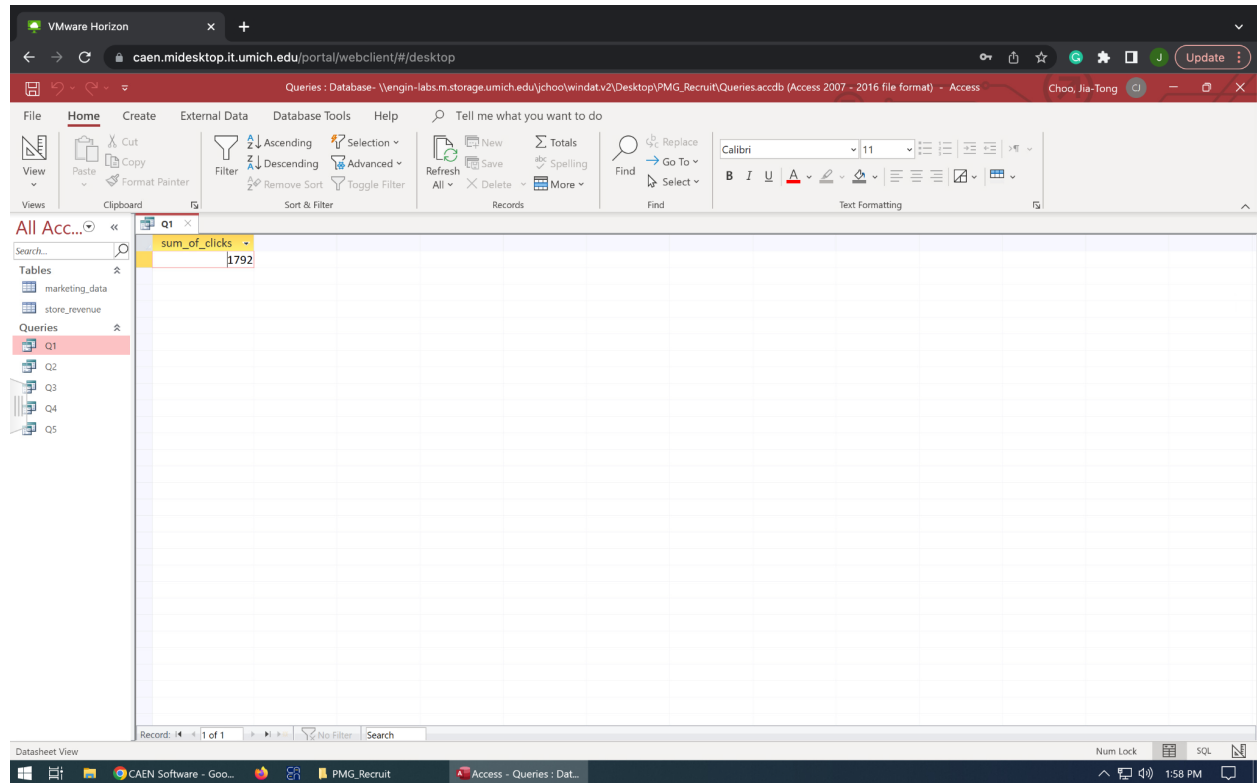


1. Generate a query to get the sum of the clicks of the marketing data

SQL query:

```
SELECT SUM(clicks) AS sum_of_clicks  
FROM marketing_data;
```

Output:



sum_of_clicks
1792

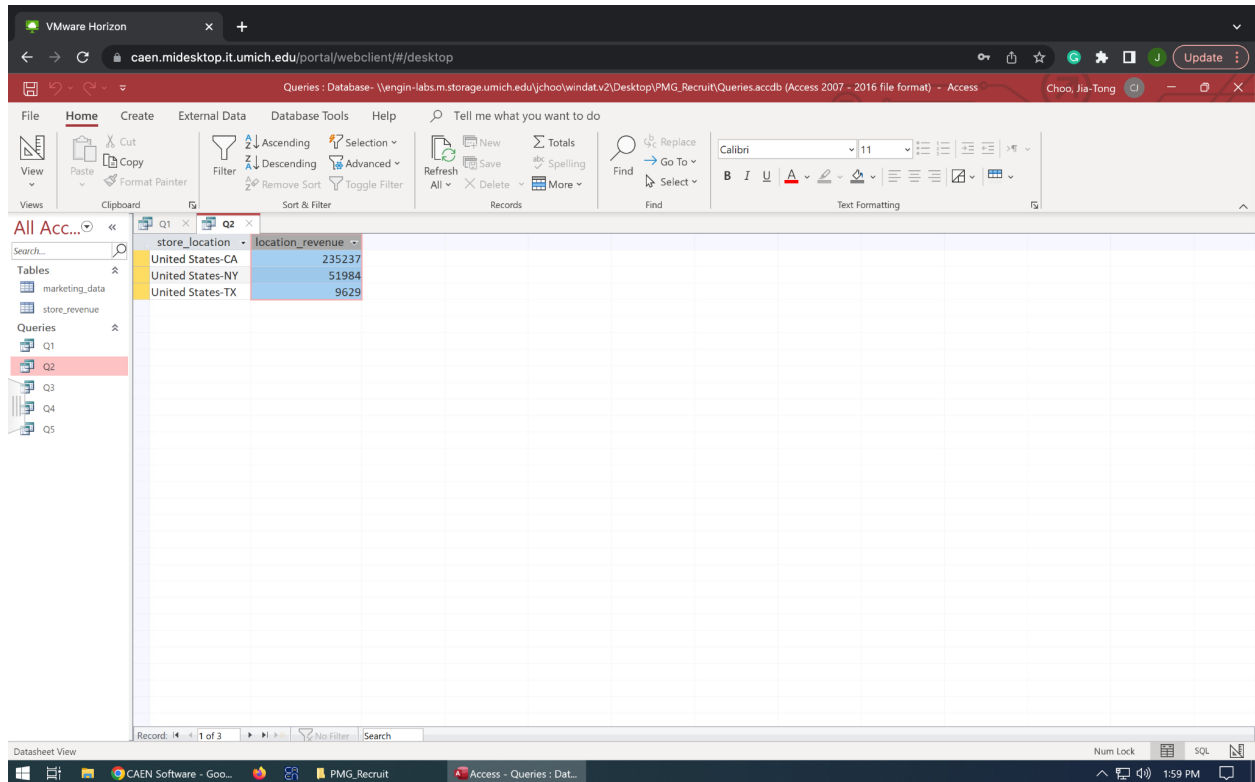
The sum of clicks of the marketing data is 1792, which means that there are 1792 clicks in total from all the documented locations.

2. Generate a query to gather the sum of revenue by store_location from the store_revenue table

SQL query:

```
SELECT store_location, SUM(revenue) AS location_revenue  
FROM store_revenue  
GROUP BY store_location;
```

Output:



store_location	location_revenue
United States-CA	235237
United States-NY	51984
United States-TX	9629

After grouping by the store location, we can see the sum of each location - CA with 125237, NY with 51984, and TX with 9629. By having this result, we are able to state that California has the most revenue, which can be caused by multiple factors including population, average income, etc.

3. Merge these two datasets so we can see impressions, clicks, and revenue together by date and geo. Please ensure all records from each table are accounted for.

SQL query:

```
SELECT marketing_data.date, marketing_data.geo, marketing_data.impressions,
marketing_data.clicks, store_revenue.store_location, store_revenue.revenue
FROM store_revenue INNER JOIN marketing_data ON (marketing_data.geo =
RIGHT(store_revenue.store_location,2)) AND (marketing_data.date =
store_revenue.date);
```

Output:

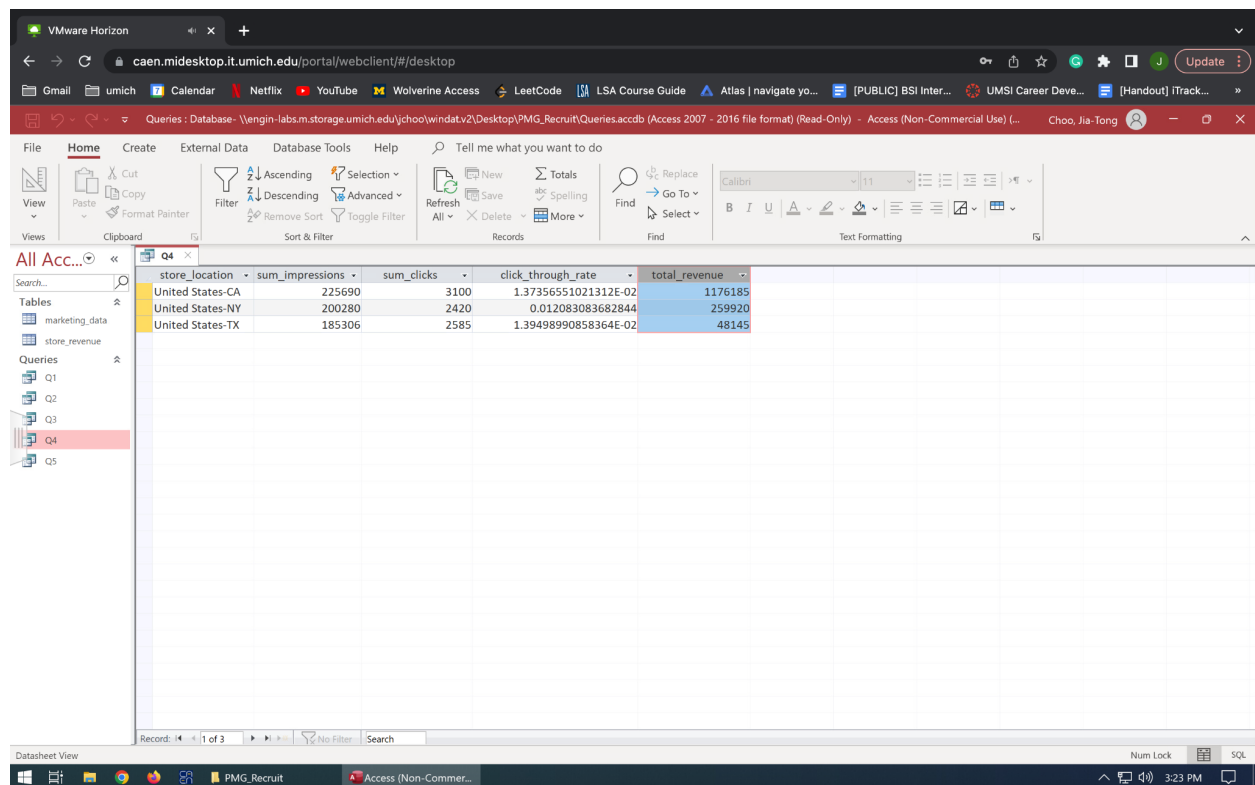
date	geo	impressions	clicks	store_location	revenue
1/1/2016	TX	2532	45	United States-TX	420
1/1/2016	TX	2532	45	United States-TX	234
1/1/2016	CA	3425	63	United States-CA	234
1/1/2016	CA	3425	63	United States-CA	100
1/1/2016	NY	3532	25	United States-NY	142
1/1/2016	NY	3532	25	United States-NY	142
1/2/2016	TX	3643	23	United States-TX	3423
1/2/2016	TX	3643	23	United States-TX	2342
1/2/2016	CA	1354	53	United States-CA	231
1/2/2016	CA	1354	53	United States-CA	234
1/2/2016	NY	4643	85	United States-NY	2342
1/2/2016	NY	4643	85	United States-NY	232
1/3/2016	TX	2353	57	United States-TX	3
1/3/2016	TX	2353	57	United States-TX	420
1/3/2016	CA	5258	36	United States-CA	234234
1/3/2016	CA	5258	36	United States-CA	100
1/3/2016	NY	4735	63	United States-NY	3245
1/3/2016	NY	4735	63	United States-NY	234
1/4/2016	TX	5783	47	United States-TX	3
1/4/2016	TX	5783	47	United States-TX	2354
1/4/2016	CA	7854	85	United States-CA	2
1/4/2016	CA	7854	85	United States-CA	34
1/4/2016	NY	4754	36	United States-NY	45235
1/4/2016	NY	4754	36	United States-NY	54
1/5/2016	TX	2535	63	United States-TX	4
1/5/2016	TX	2535	63	United States-TX	423
1/5/2016	CA	4678	73	United States-CA	23
1/5/2016	CA	4678	73	United States-CA	45
1/5/2016	NY	2364	33	United States-NY	234

4. In your opinion, what is the most efficient store and why?

SQL query:

```
SELECT store_revenue.store_location, SUM(marketing_data.impressions) AS  
sum_impressions, SUM(marketing_data.clicks) AS sum_clicks,  
(sum_clicks/sum_impressions) AS click_through_rate, SUM(store_revenue.revenue) AS  
total_revenue  
FROM marketing_data INNER JOIN store_revenue ON marketing_data.geo =  
RIGHT(store_revenue.store_location,2)  
GROUP BY store_revenue.store_location;
```

Output:



store_location	sum_impressions	sum_clicks	click_through_rate	total_revenue
United States-CA	225690	3100	1.37356551021312E-02	1176185
United States-NY	200280	2420	0.012083083682844	259920
United States-TX	185306	2585	1.39498990858364E-02	48145

I calculated the clickthrough rate (CTR) for each store location with $CTR = \text{sum of clicks} / \text{sum of impressions}$. Clickthrough rates tells us how well the ads or the marketing plan are performing, which can also means the marketing efficiency of the store. As we can see, TX has the highest clickthrough rate (1.39) compared to the other two sites with CA having just a slightly lower rate (1.37). Therefore, we can have a conclusion that TX is the most efficient store. The total revenue of each store is also shown above in addition to the CTR to show that not the store that has the highest revenue is the most efficient store.

5. (Challenge) Generate a query to rank in order the top 10 revenue producing states

SQL query:

```
SELECT RIGHT(store_location,2) AS location_state, SUM(revenue) AS total_revenue
FROM store_revenue
GROUP BY store_location
ORDER BY SUM(revenue) DESC;
```

Output:

The screenshot shows the Microsoft Access application window. The title bar indicates it's a VMware Horizon session. The browser address bar shows the URL 'caen.midesktop.it.umich.edu/portal/webclient/#/desktop'. The Access ribbon is visible with tabs for File, Home, Create, External Data, Database Tools, and Help. The 'Database Tools' tab is active, showing options like Filter, Sort & Filter, and Tell me what you want to do. The left sidebar shows the 'All Access Objects' pane with a search bar and a list of tables (marketing_data, store_revenue) and queries (Q1, Q2, Q3, Q4, Q5). The main area displays the results of query 'Q5' in Datasheet View. The table has two columns: 'location_state' and 'total_revenue'. The data is sorted in descending order of total revenue. The states shown are CA (235237), NY (51984), and TX (9629). The status bar at the bottom shows 'Record: 1 of 3' and 'No Filter'.

location_state	total_revenue
CA	235237
NY	51984
TX	9629

From my output, CA has the highest revenue (235237) following by NY with a revenue of 51984. TX has the lowest revenue out of all the states.