Name: Xinyue Li
Email: lxinyue@umich.edu

Questions answered using MySQL

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

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

clickSum
1792

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

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

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

Question #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.

Here I use substring to extract geo information for store_revenue table. MySQL don't have full outer join, use left join and union instead

```
(SELECT md.date AS marketDate, sr.date AS revenueDate,
        md.geo, sr.store_location, md.impressions,
        md.clicks, sr.revenue, sr.brand_id
 FROM store_revenue sr
 LEFT OUTER JOIN marketing_data md
 ON md.geo = SUBSTRING(sr.store_location, 15, 2)
 AND md.date = sr.date)
UNION
(SELECT md.date AS marketDate, sr.date AS revenueDate,
        md.geo, sr.store_location, md.impressions,
        md.clicks, sr.revenue, sr.brand_id
 FROM marketing_data md
 LEFT OUTER JOIN store_revenue sr
 ON md.geo = SUBSTRING(sr.store_location, 15, 2)
```

AND md.date = sr.date);

marketDate	revenueDate	geo	store_location	impressions	clicks	revenue	brand_id
2016-01-01 00:00:00	2016-01-01 00:00:00	CA	United States-CA	3425	63	100	1
2016-01-01 00:00:00	2016-01-01 00:00:00	TX	United States-TX	2532	45	420	1
2016-01-01 00:00:00	2016-01-01 00:00:00	NY	United States-NY	3532	25	142	1
2016-01-02 00:00:00	2016-01-02 00:00:00	CA	United States-CA	1354	53	231	1
2016-01-02 00:00:00	2016-01-02 00:00:00	TX	United States-TX	3643	23	2342	1
2016-01-02 00:00:00	2016-01-02 00:00:00	NY	United States-NY	4643	85	232	1
2016-01-03 00:00:00	2016-01-03 00:00:00	CA	United States-CA	5258	36	100	1
2016-01-03 00:00:00	2016-01-03 00:00:00	TX	United States-TX	2353	57	420	1
2016-01-03 00:00:00	2016-01-03 00:00:00	NY	United States-NY	4735	63	3245	1
2016-01-04 00:00:00	2016-01-04 00:00:00	CA	United States-CA	7854	85	34	1
2016-01-04 00:00:00	2016-01-04 00:00:00	TX	United States-TX	5783	47	3	1
2016-01-04 00:00:00	2016-01-04 00:00:00	NY	United States-NY	4754	36	54	1
2016-01-05 00:00:00	2016-01-05 00:00:00	CA	United States-CA	4678	73	45	1
2016-01-05 00:00:00	2016-01-05 00:00:00	TX	United States-TX	2535	63	423	1
2016-01-05 00:00:00	2016-01-05 00:00:00	NY	United States-NY	2364	33	234	1
2016-01-01 00:00:00	2016-01-01 00:00:00	CA	United States-CA	3425	63	234	2
2016-01-01 00:00:00	2016-01-01 00:00:00	TX	United States-TX	2532	45	234	2
2016-01-01 00:00:00	2016-01-01 00:00:00	NY	United States-NY	3532	25	142	2
2016-01-02 00:00:00	2016-01-02 00:00:00	CA	United States-CA	1354	53	234	2
2016-01-02 00:00:00	2016-01-02 00:00:00	TX	United States-TX	3643	23	3423	2
2016-01-02 00:00:00	2016-01-02 00:00:00	NY	United States-NY	4643	85	2342	2
2016-01-03 00:00:00	2016-01-03 00:00:00	CA	United States-CA	5258	36	234234	2
NULL	2016-01-06 00:00:00	NULL	United States-TX	NULL	NULL	3	3
2016-01-03 00:00:00	2016-01-03 00:00:00	TX	United States-TX	2353	57	3	2
2016-01-03 00:00:00	2016-01-03 00:00:00	NY	United States-NY	4735	63	234	2
2016-01-04 00:00:00	2016-01-04 00:00:00	CA	United States-CA	7854	85	2	2
2016-01-04 00:00:00	2016-01-04 00:00:00	TX	United States-TX	5783	47	2354	2
2016-01-04 00:00:00	2016-01-04 00:00:00	NY	United States-NY	4754	36	45235	2
2016-01-05 00:00:00	2016-01-05 00:00:00	CA	United States-CA	4678	73	23	2
2016-01-05 00:00:00	2016-01-05 00:00:00	TX	United States-TX	2535	63	4	2
2016-01-05 00:00:00	2016-01-05 00:00:00	NY	United States-NY	2364	33	124	2
2016-01-01 00:00:00	NULL	MN	NULL	1342	784	NULL	NULL
2016-01-02 00:00:00	NULL	MN	NULL	2366	85	NULL	NULL
2016-01-03 00:00:00	NULL	MN	NULL	5783	87	NULL	NULL
2016-01-04 00:00:00	NULL	MN	NULL	9345	24	NULL	NULL
2016-01-05 00:00:00	NULL	MN	NULL	3452	25	NULL	NULL

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

Here, I utilize the revenue per click to represent the efficiency for each store. This metric is calculated using the sum of the revenue for each store divided by the sum of the clicks for each store. From the output we could see that a store in CA for brand 2 has the biggest revenue per click, as a result, in my opinion, **the store with brand 2 in United States-CA** is the most efficient store.

```
SELECT sr.brand_id, sr.store_location, SUM(sr.revenue)/SUM(md.clicks) AS
revenuePerClick
FROM store_revenue sr
INNER JOIN marketing_data md
ON md.geo = SUBSTRING(sr.store_location, 15, 2)
AND md.date = sr.date
```

```
GROUP BY sr.brand_id, sr.store_location
Order By revenuePerClick DESC;
```

brand_id	store_location	revenuePerClick
2	United States-CA	757.1838709677419
2	United States-NY	198.66528925619835
2	United States-TX	25.608510638297872
1	United States-NY	16.144628099173552
1	United States-TX	15.353191489361702
1	United States-CA	1.6451612903225807

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

```
SELECT store_location, SUM(revenue) AS revenueSum
FROM store_revenue
GROUP BY store_location
ORDER BY revenueSum DESC
LIMIT 10;
```

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

As there are only 3 states in this dataset, the sql code could be applied to a larger dataset.