PMG SQL Challenge

Benjamin Ajibade

baajibade@crimson.ua.edu

Submitted GitHub link: https://github.com/BenAji/PMG-

A database of two tables, store_revenue and marketing_data was provided as shown below:

	A	B	C	E
1	date	brand_id	store_location	revenue
2	1/1/2016	1	United States-CA	100
3	1/1/2016	1	United States-TX	420
4	1/1/2016	1	United States-NY	142
5	1/1/2016	2	United States-CA	234
6	1/1/2016	2	United States-TX	234
7	1/1/2016	2	United States-NY	142
8	1/2/2016	1	United States-CA	231
9	1/2/2016	1	United States-TX	2342

1	date	geo	impressions	clicks
2	1/1/2016	TX	2532	45
3	1/1/2016	CA	3425	63
4	1/1/2016	NY	3532	25
5	1/1/2016	MN	1342	784
6	1/2/2016	TX	3643	23
7	1/2/2016	CA	1354	53
8	1/2/2016	NY	4643	85

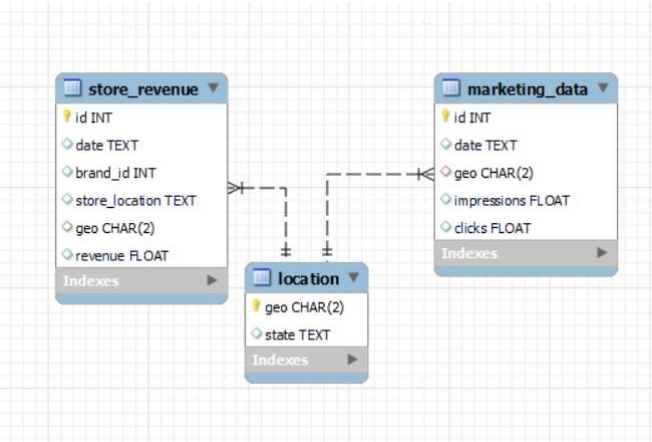
Using Excel, another column ***geo*** was created from store_location column from the store_location table, got the unique values and created a fact table for location (having a one-to-many relationship).

CA	California		=UNIQUE(A1:A20,,FALSE)	=XLOOKUP(D2,A1:A20,B1:B20,,0)
CA	California		CA	California
CA	California		MN	Minesotta
CA	California		NY	New York
CA	California		TX	Texas
MN	Minesotta			
MN	Minesotta			
MN	Minesotta			
MN	Minesotta			
MN	Minesotta			
NY	New York			

D2 =RIGHT(C2,2)

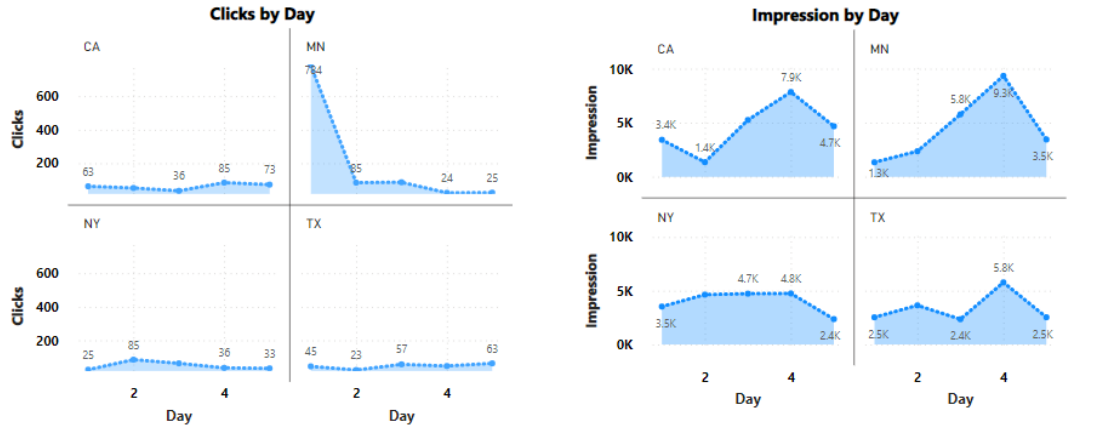
	A	B	C	D	E
1	date	brand_id	store_location	geo	revenue
2	1/1/2016	1	United States-CA	CA	100
3	1/1/2016	1	United States-TX	TX	420
4	1/1/2016	1	United States-NY	NY	142
5	1/2/2016	1	United States-CA	CA	231
6	1/2/2016	1	United States-TX	TX	2342
7	1/2/2016	1	United States-NY	NY	232
8	1/3/2016	1	United States-CA	CA	100

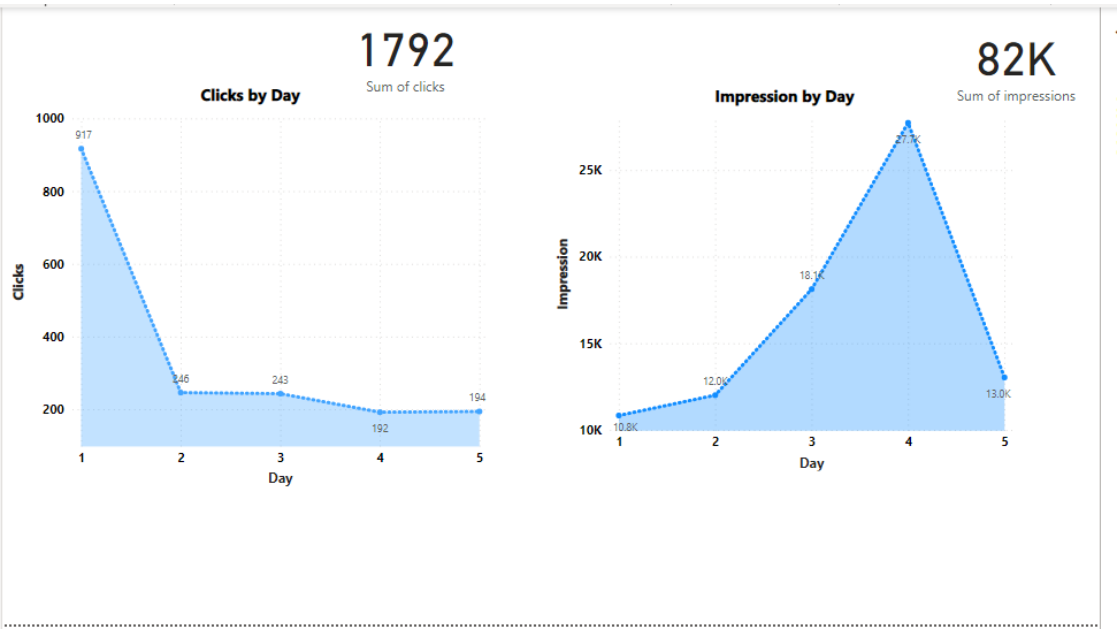
MySql workbench was used to create the Database and Tables. There are three tables store_revenue ,marketing_data and location . All constraints and datatype were updated using ALTER Table command , table was loaded the schema is as shown below:



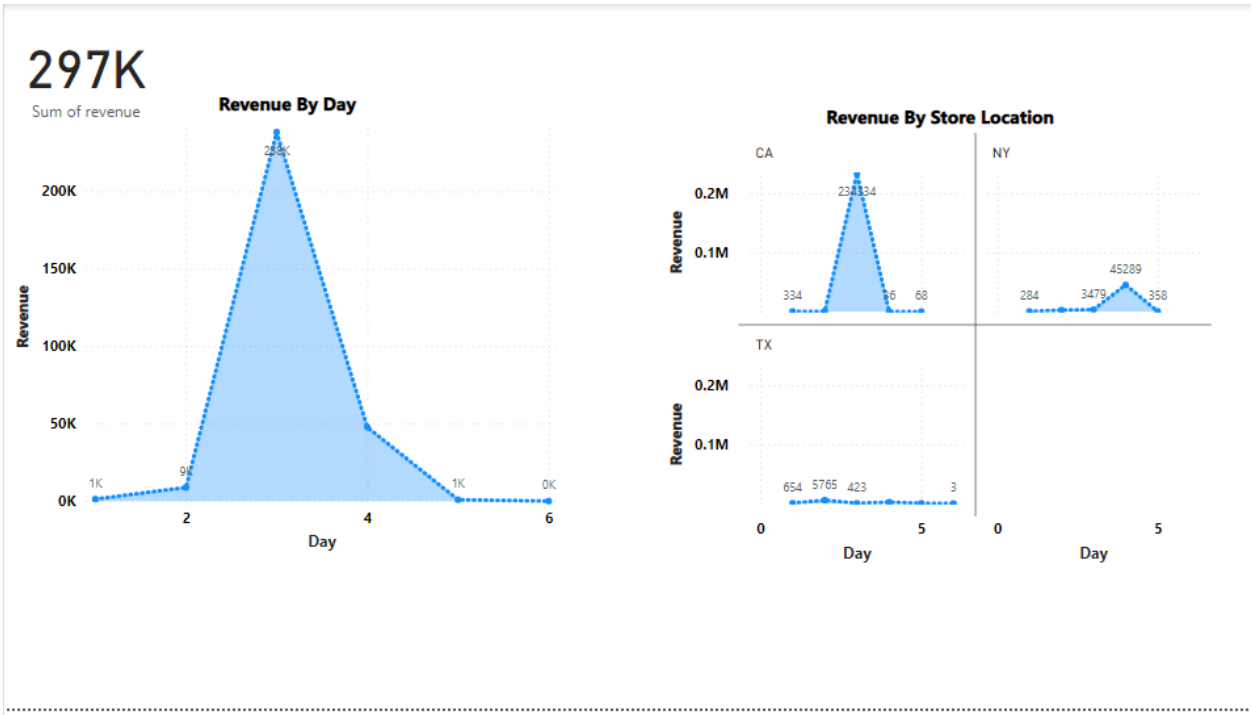
Here are some questions of interest, PowerBI was used for visualization:

- Generate a query to get the sum of the clicks of the marketing data

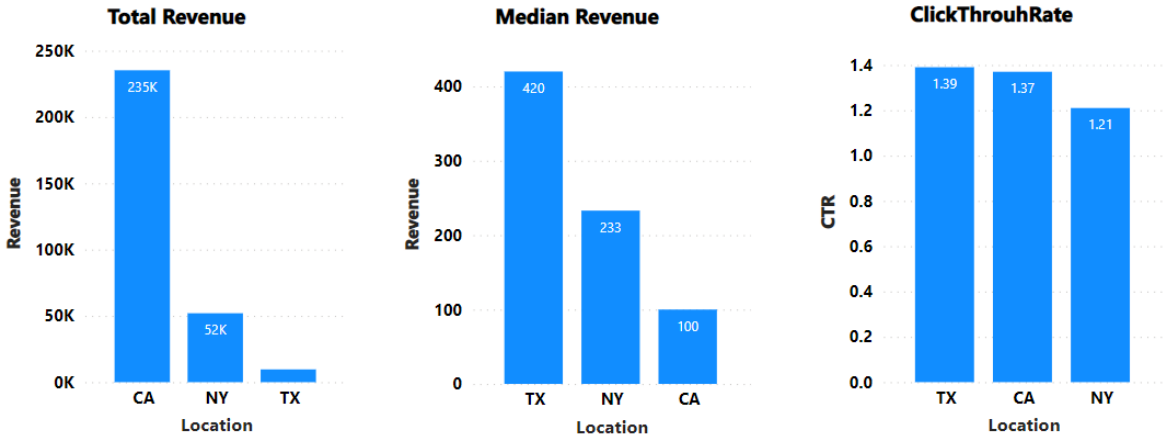




- Most efficient store and why
  -- The total revenue (daily) was checked.
  --The Clickthrough(daily) rate was calculated.



-- The Clickthrough rate for the overall stores in each state was also calculated.
From this analysis, although Texas store seems to be leading using the Clickthrough ratio but just within 0.02% (this is not too statistically significant due to the other factor that could be further considered) however the total revenue generated by the store in California is more (although this needs to be investigated as there is a potential outlier which skewed the distribution) and the difference is statistically significant , therefore California store is more efficient.



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

  --Although we do not have up to 10 states, intuitively I calculated the ranking based on the revenue each brand in for each date.

  -- We have California first on the list with more than 200K , more than4 times than the next on the list which is New York.

  -- It is important to notice that New York appears most in the list.

- **Conclusion**

I am strongly of the opinion that we should further investigate the outliers in California store before arriving at a decisive conclusion. Below is the comparison of the median value and the total value of Revenue and clickthrough for each state.