

DAB203 Graded Lab#2 [Marks 60]

Use the SQLBook database for this lab.

The Customers, Orders, and ZipCensus tables will be used for the queries.

Your query along with the results of the query should be pasted in Excel.

There are 6 questions, and each question should be answered on a different worksheet in Excel. Thus each worksheet will have query, results, and corresponding chart, if applicable. The worksheets should be named as Q1, Q2 and so on. The first worksheet should have your Name and Student ID written.

Make sure to properly format the charts with Axis titles and legend where appropriate.

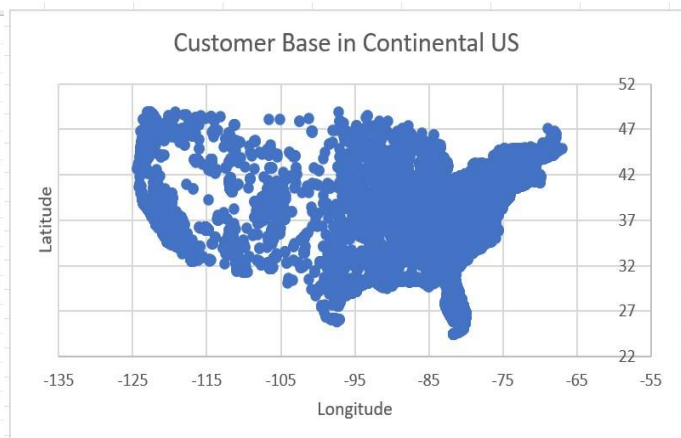
Upload the Excel file and pdf doc. with query, sql result output and excel chart.

Q1. Write a query to display the customer's name, gender, corresponding zip code, longitude and latitude.

Make sure your query removes all the rows that have blank in the first name column. Copy and paste the results in Excel and draw a scatter plot for the customers locations. This map is for continental United States.

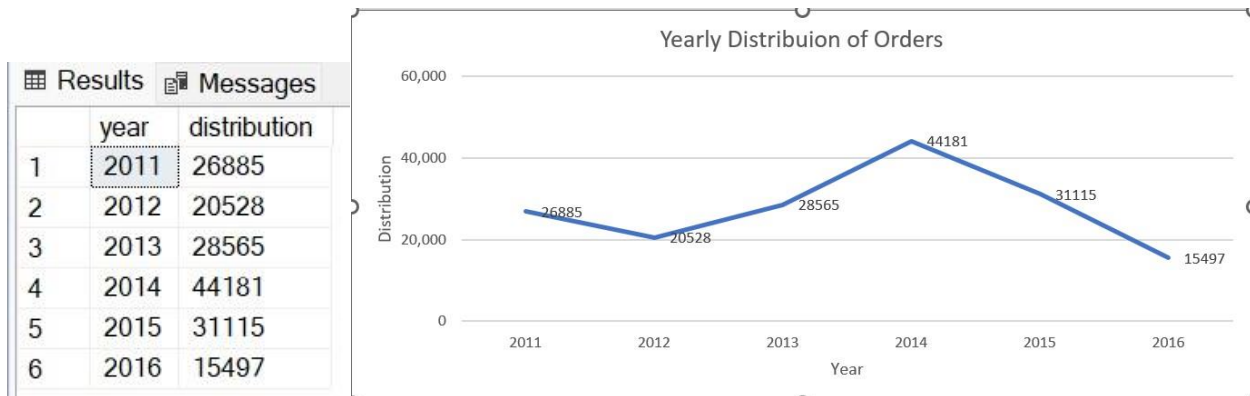
```
SELECT C.[firstname], C.[gender], O.[zipcode], Z.[Latitude], Z.[Longitude]
FROM [dbo].[customers] C
JOIN [dbo].[orders] O ON C.[customerid]= O.[customerid]
JOIN [dbo].[ZipCensus] Z ON O.[zipcode] = Z.[zcta5]
WHERE C.[firstname] IS NOT NULL AND C.[firstname] <> '';
```

| | firstname | gender | zipcode | Latitude | Longitude |
|----|-----------|--------|---------|-----------|------------|
| 1 | LOUIS | M | 02459 | 42.321472 | -71.192846 |
| 2 | ANTHONY | M | 10804 | 40.946896 | -73.788879 |
| 3 | MERRILL | NULL | 33137 | 25.817259 | -80.18879 |
| 4 | MARIA | F | 07073 | 40.828525 | -74.091647 |
| 5 | T. | NULL | 21218 | 39.327711 | -76.604083 |
| 6 | CRAIG | M | 06853 | 41.0661 | -73.438006 |
| 7 | SUSAN | F | 46240 | 39.903931 | -86.128958 |
| 8 | JOE | M | 10570 | 41.132189 | -73.788959 |
| 9 | ROBERT | M | 18301 | 41.024366 | -75.183117 |
| 10 | BILL | M | 60073 | 42.358724 | -88.101722 |
| 11 | WINIFRED | F | 07003 | 40.802874 | -74.188646 |
| 12 | ANA | F | 62704 | 39.777741 | -89.681535 |
| 13 | JUNE | F | 76710 | 31.533567 | -97.186237 |
| 14 | ROBERT | M | 07004 | 40.881606 | -74.29534 |



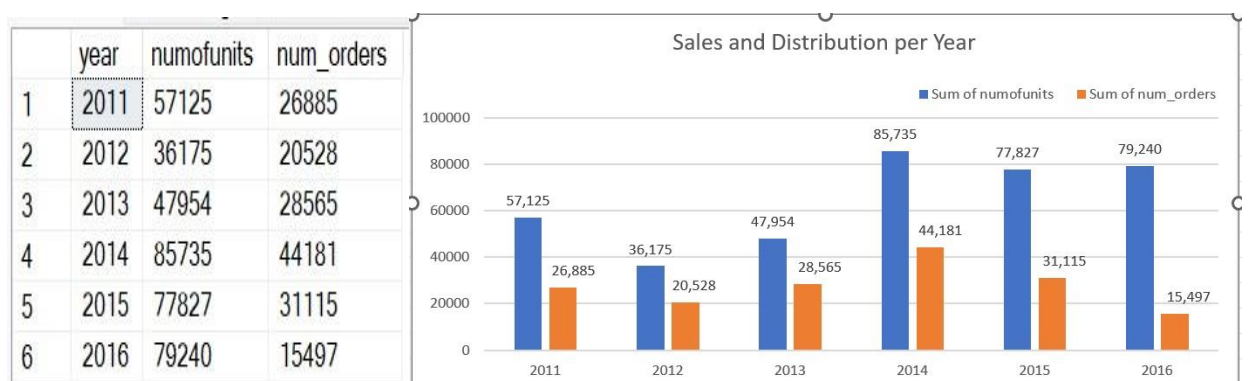
Q2. Write a query to display the yearly distribution of number of orders (year>2010). With the results create a corresponding line chart in Excel.

```
SELECT YEAR([orderdate]) AS year, COUNT([orderid]) AS distribution
FROM [dbo].[orders] WHERE YEAR([orderdate]) > 2010
GROUP BY YEAR([orderdate])
ORDER BY YEAR([orderdate])
```



Q3. Write a query to display the yearly distribution of number of orders (year>2010) and sum of number of units. With the results create a clustered column chart in Excel.

```
SELECT YEAR([orderdate]) AS year, sum([numunits])as numofunits, COUNT(*) AS num_orders
FROM [dbo].[orders] WHERE YEAR([orderdate]) > 2010
GROUP BY YEAR([orderdate])
ORDER BY YEAR([orderdate])
```

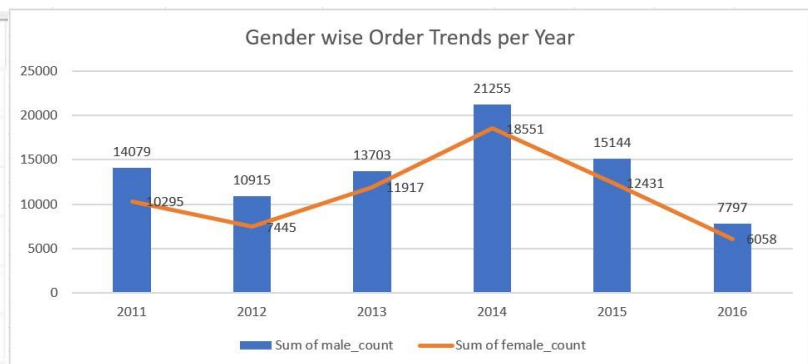


Q4. Write a query to display the distribution of gender (female and male) by years in Order (year>2010).

With the results create a line chart and a stacked column in Excel.

```
SELECT YEAR(orderdate) AS yearly,
       SUM(CASE WHEN c.gender = 'M' THEN 1 ELSE 0 END) AS male_count,
       SUM(CASE WHEN c.gender = 'F' THEN 1 ELSE 0 END) AS female_count
FROM orders o
JOIN customers c
ON o.customerid = c.customerid
GROUP BY year(orderdate)
HAVING year(orderdate) > 2010
ORDER BY yearly
```

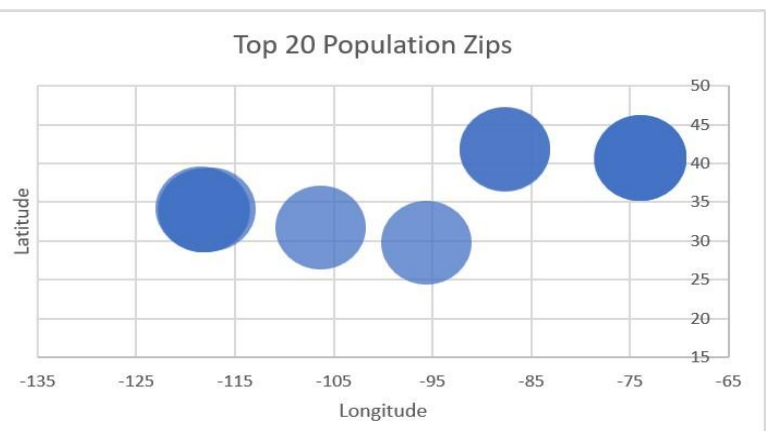
| | yearly | male_count | female_count |
|---|--------|------------|--------------|
| 1 | 2011 | 14079 | 10295 |
| 2 | 2012 | 10915 | 7445 |
| 3 | 2013 | 13703 | 11917 |
| 4 | 2014 | 21255 | 18551 |
| 5 | 2015 | 15144 | 12431 |
| 6 | 2016 | 7797 | 6058 |



Q5. Write a query for displaying the longitude, latitude, total populations of 20 zipcode areas that has the highest total population (totpop). With the results, create a bubble chart in Excel. This map is for continental United States.

```
SELECT TOP 20 zcta5, latitude, longitude, TotPop
FROM ZipCensus
ORDER BY TotPop DESC
```

| | zcta5 | latitude | longitude | TotPop |
|----|-------|-----------|-------------|--------|
| 1 | 79936 | 31.766296 | -106.295673 | 115538 |
| 2 | 60629 | 41.777722 | -87.709167 | 111893 |
| 3 | 90650 | 33.9076 | -118.083453 | 105603 |
| 4 | 11368 | 40.749742 | -73.853078 | 104486 |
| 5 | 90201 | 33.970343 | -118.171368 | 101647 |
| 6 | 90011 | 34.00709 | -118.258681 | 100671 |
| 7 | 11220 | 40.641027 | -74.016399 | 100476 |
| 8 | 11385 | 40.703722 | -73.893161 | 99508 |
| 9 | 11226 | 40.645708 | -73.956456 | 99285 |
| 10 | 77084 | 29.837373 | -95.663136 | 98697 |
| 11 | 92335 | 34.089124 | -117.451774 | 98228 |
| 12 | 11373 | 40.738749 | -73.878207 | 97430 |
| 13 | 60618 | 41.947568 | -87.702647 | 97389 |
| 14 | 11236 | 40.640802 | -73.903552 | 97084 |



Q6. Write a query for showing the zipcodes that in the Orders table but not in the Zipcensus table.

```
SELECT zipcode
FROM orders
EXCEPT
SELECT zcta5 FROM ZipCensus
ORDER BY zipcode
```

| | zipcode |
|----|---------|
| 1 | NULL |
| 2 | - |
| 3 | -- |
| 4 | * |
| 5 | . |
| 6 | |
| 7 | _____ |
| 8 | 0 |
| 9 | 0000 |
| 10 | 00000 |
| 11 | 00120 |
| 12 | 00157 |
| 13 | 00162 |
| 14 | 00177 |

