**ASSIGNMENT OF**

**DATA SCIENCE**

Submitted by: - Nishant Jaiswal

Class: - 12

Section: - ‘P2’

Submitted to:- Bibek Ghimire

(Department of data science)

**TABLE OF CONTENT**

TASK 1: Database Development

|  |  |
| --- | --- |
| S.N | PAGE |
| 1 | 4 |
| 2 | 5 |
| 3 | 6 |
| 4 | 7 |
| 5 | 8 |
| 6 | 9 |

TASK 2 : Big Data Processing

|  |  |
| --- | --- |
| S.N | PAGE |
| SCRUB | 10-18 |
| EXPLORE | 19 |
| INTERPRET | 20-25 |

**TABLE OF FIGURE**

**Task 1:**

1. Designing Database …..................................................................................4
2. Displaying all records.....................................................................................5
3. Order details with qty > 5...............................................................................6
4. All equipment's associated with Client “KMC COLLEGE” ............................7
5. Delete “Book” from order id “ST-2356” and put it in order id “ST-1234” ....8
6. Deleting the orders between 01/11/2014 and 20/01/2015................................9

**TASK 2:**

**Scrub:**

1. Process of uploading the file to the SQL server................................(10-13)
2. Formatting the date …........................................................................14
3. Populating the state data using area code.............................................15
4. Populating the product type using product id.......................................16
5. Removing duplicates.............................................................................17
6. Deleting the market size and budget columns.......................................18

**Explore:**

1. Deleting the market size and budget columns....................................19

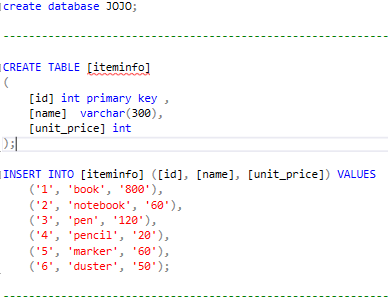
**Interpret:**

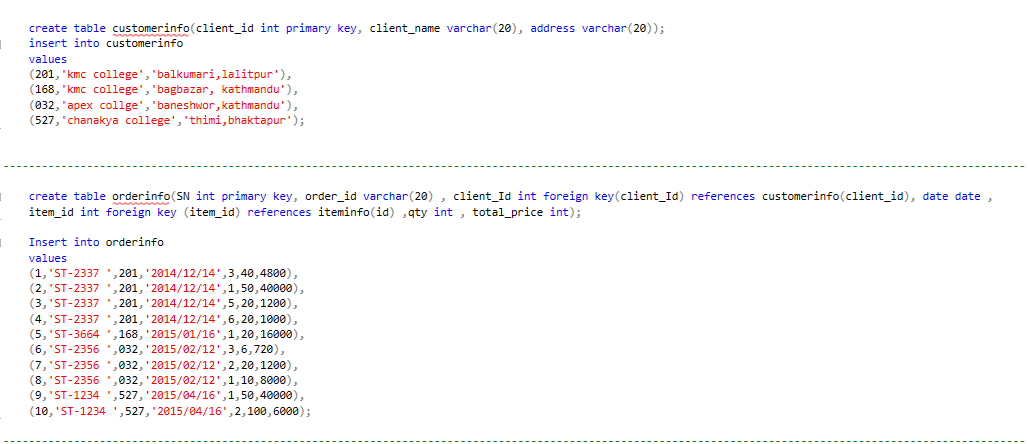
1. Sheet\_1...............................................................................................20
2. Sheet\_2................................................................................................21
3. Sheet\_3................................................................................................22
4. Sheet\_4................................................................................................23
5. Sheet\_5................................................................................................24
6. Dashboard............................................................................................25

**TASK 1:**

**1: Design the database for the information above.**

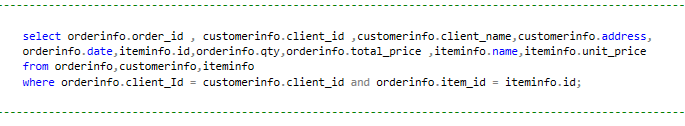
**SQL:**



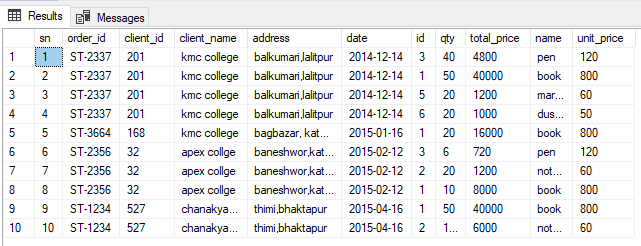


**2. Display all the order details with its client name and item associated with it.**

**SQL:**

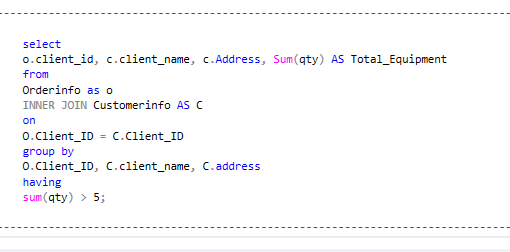


**OUTPUT:**

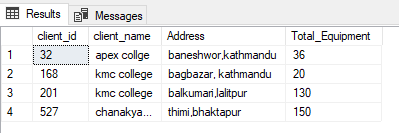


**3. Display all the client details whose total equipment quantity is more than 5.**

**SQL:**

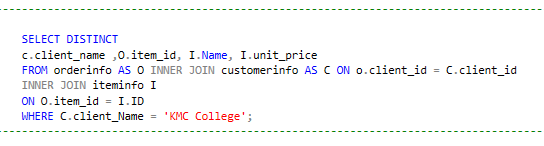


**OUTPUT:**

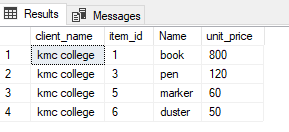


**4. Display all the equipment associated with client ‘KMC College’.**

**SQL:**

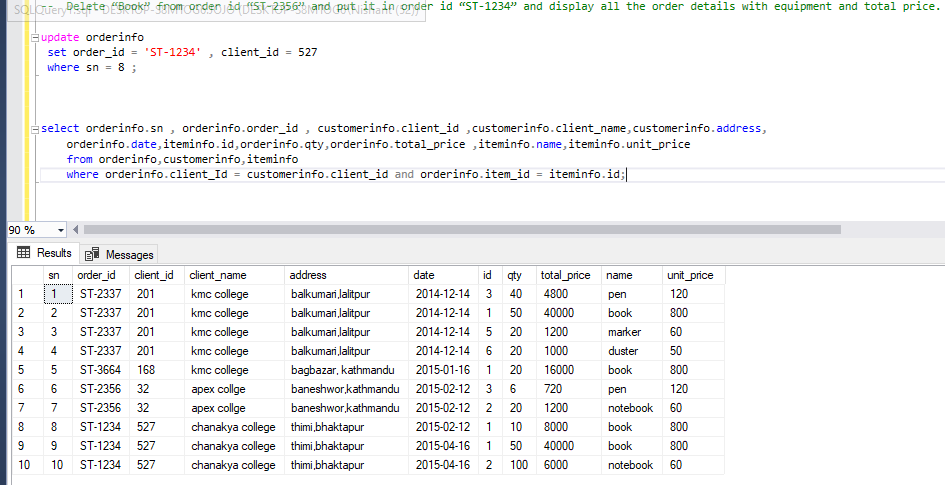


OUTPUT:



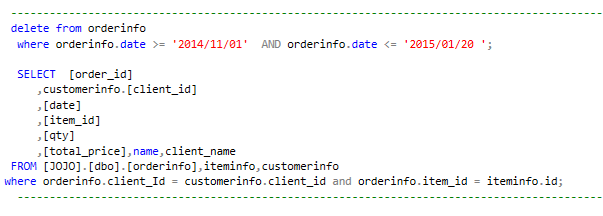
**5. Delete “Book” from order id “ST-2356” and put it in order id “ST-1234” and display all the order details with equipment and total price.**

**SQL:**

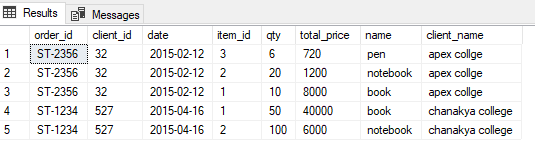


**6. Delete the orders between 01/11/2014 and 20/01/2015 and display all the order details with client name and equipment.**

**SQL:**



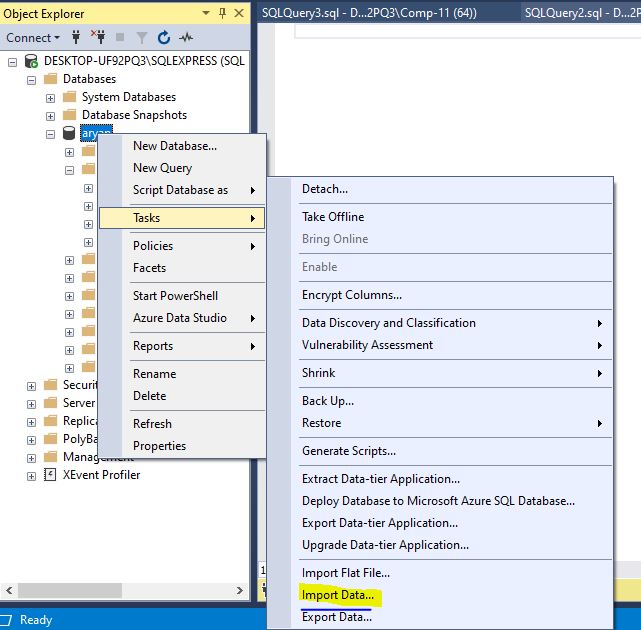
**OUTPUT:**

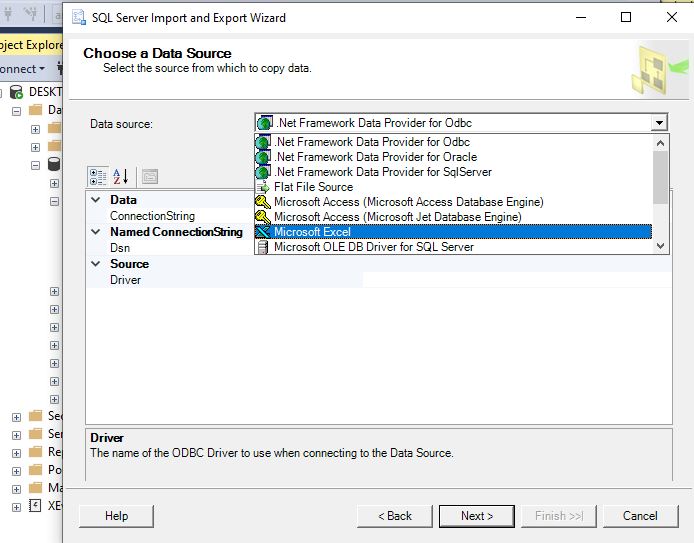


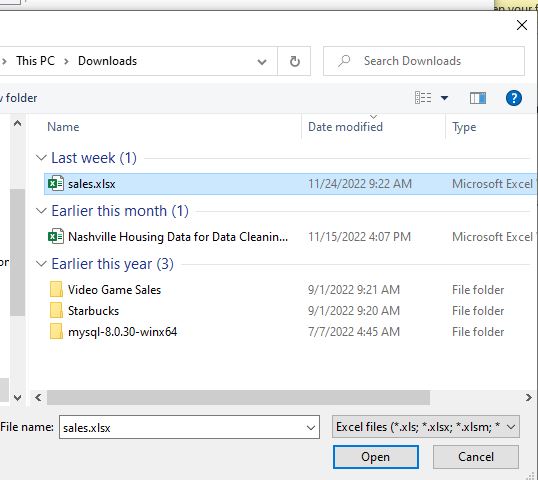
**TASK 2:**

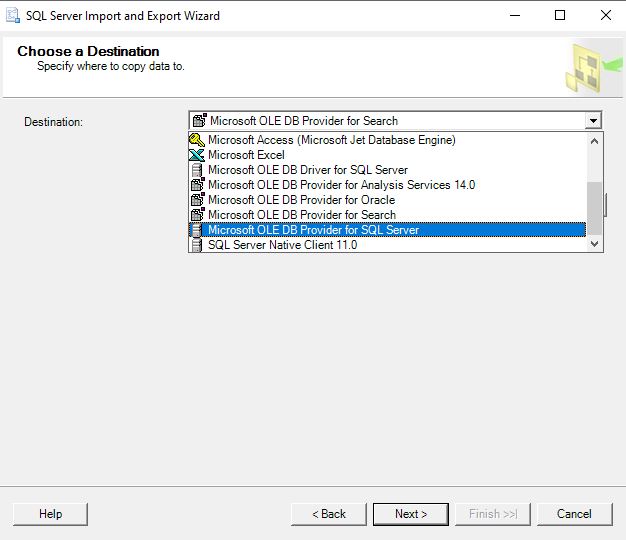
**SCRUB:**

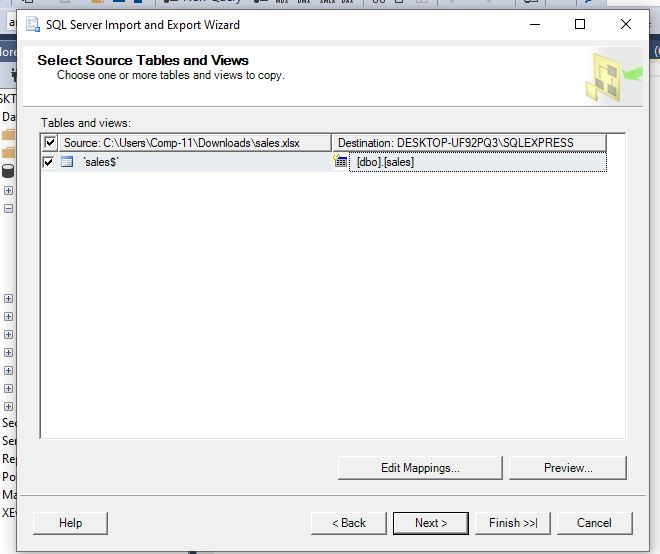
**1. Show the complete process of uploading the file to the sql server.**

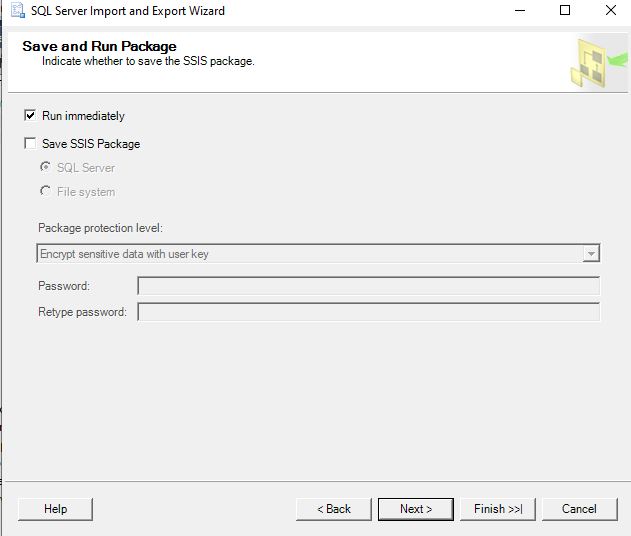


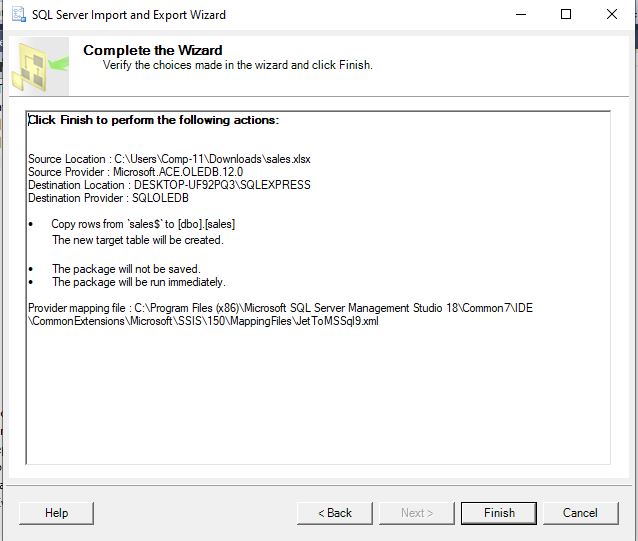






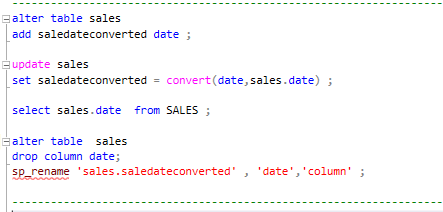




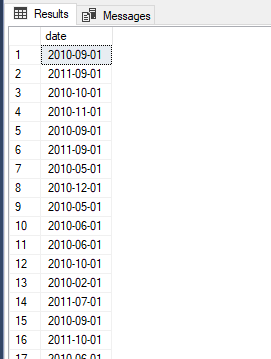


**2. Format the date into a useable format.**

**SQL:**

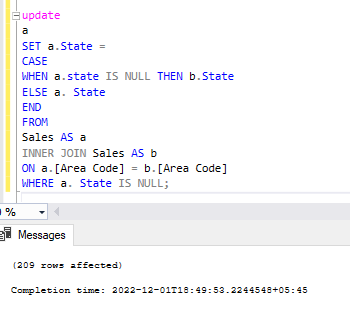


**OUTPUT:**

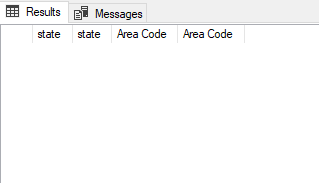


**3. Populate the state data using area code.**

**SQL:**

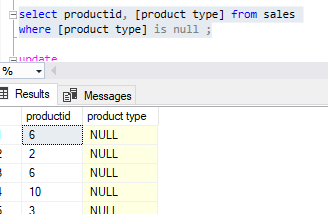


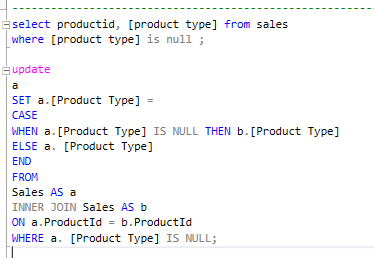
**OUTPUT:**

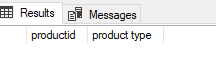


**4. Populate the product type using product id.**

**SQL:**

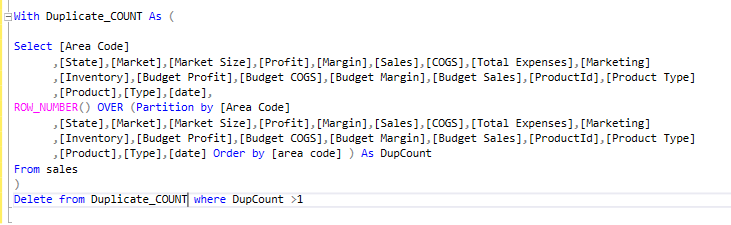




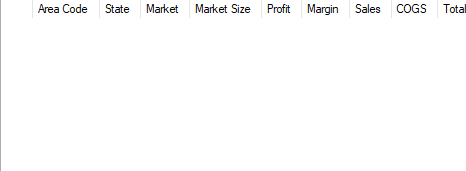


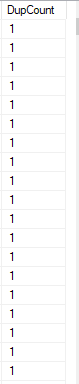
**5. Remove Duplicates.**

**SQL:**



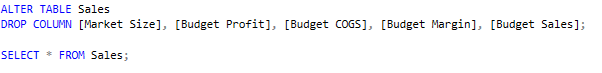
**OUTPUT:**



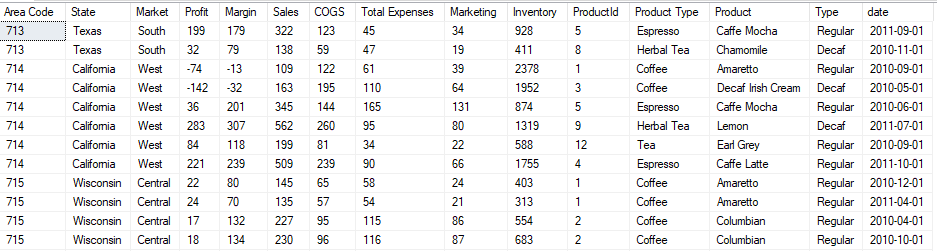


**6. Delete the market size and budget columns.**

**SQL:**



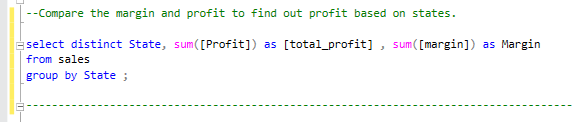
**OUTPUT:**



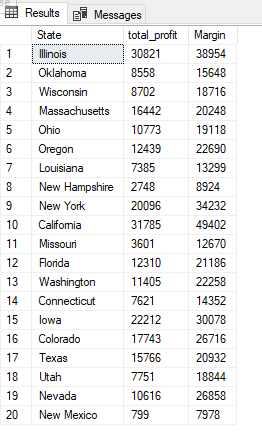
**Explore:**

**1. Compare the margin and profit to find out profit based on states.**

**SQL:**



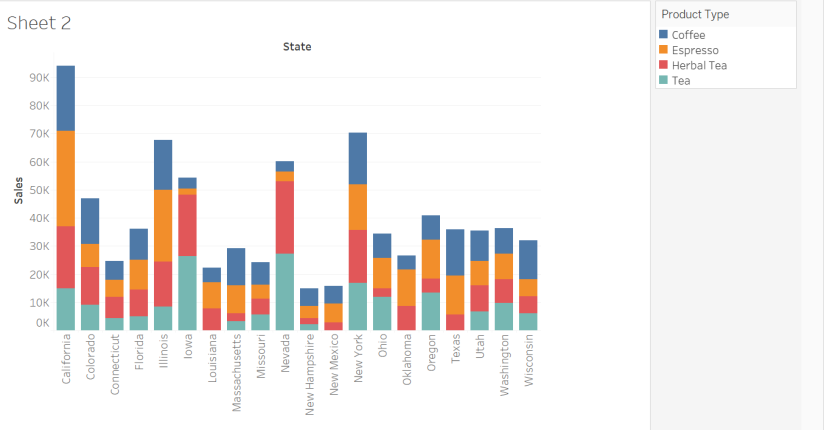
**OUTPUT:**



**Interpret:**

**1. Use tools like tableau to visualize the given data into five different graphs and explain the data accordingly.**

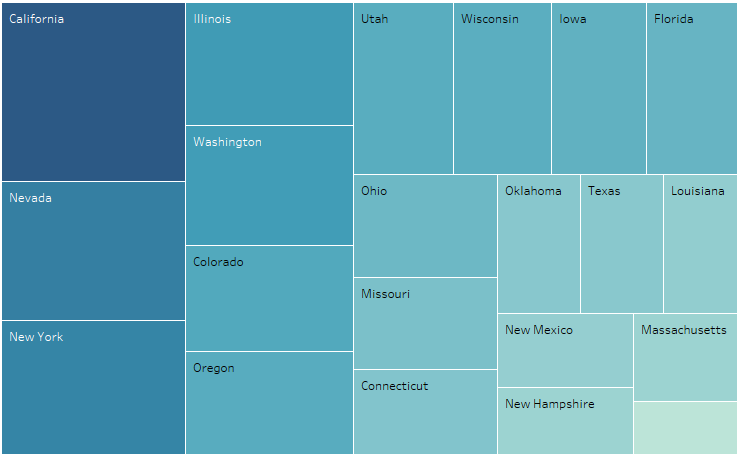
**SHEET 1:**



EXPLANATION:

From this bar graph we can conclude that we are getting most sales from “California” which is almost 300k sales and least sales from “New Hampshire” and “New Mexico “which is below 50k sales, where most selling products are coffee and Espresso. Overall, the most selling product is coffee as it is getting most sales in all states with highest sales of 72k.

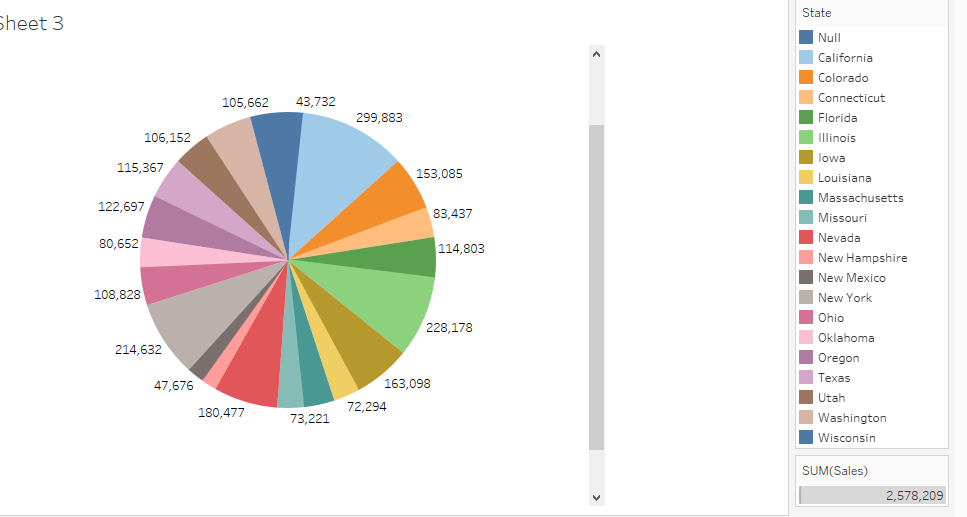
**SHEET 2:**



**EXPLANATION:**

This graph shows the data of state and the total expenses of each state. From this graph we can observe that the most expenses are from California and Nevada. We have least expenses in Massachusetts. The total expenses of all state are more than 721k.

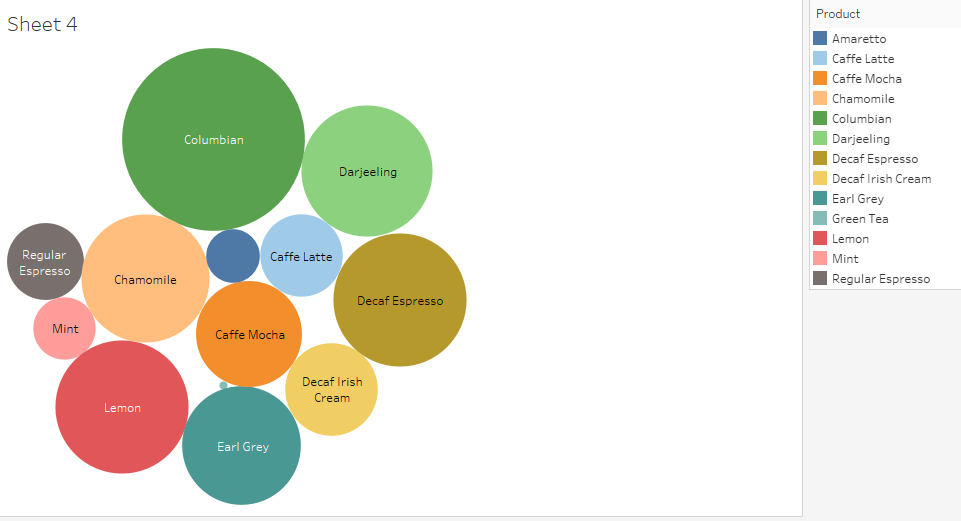
**SHEET 3:**



**EXPLANATION:**

This pie chart shows the data of state and their total sales. As we can see from the above pie chart, we are getting most sales from California and least sales from New Mexico and New York. The total sales of all state are 2,578,209. The average sale is more than 122k.

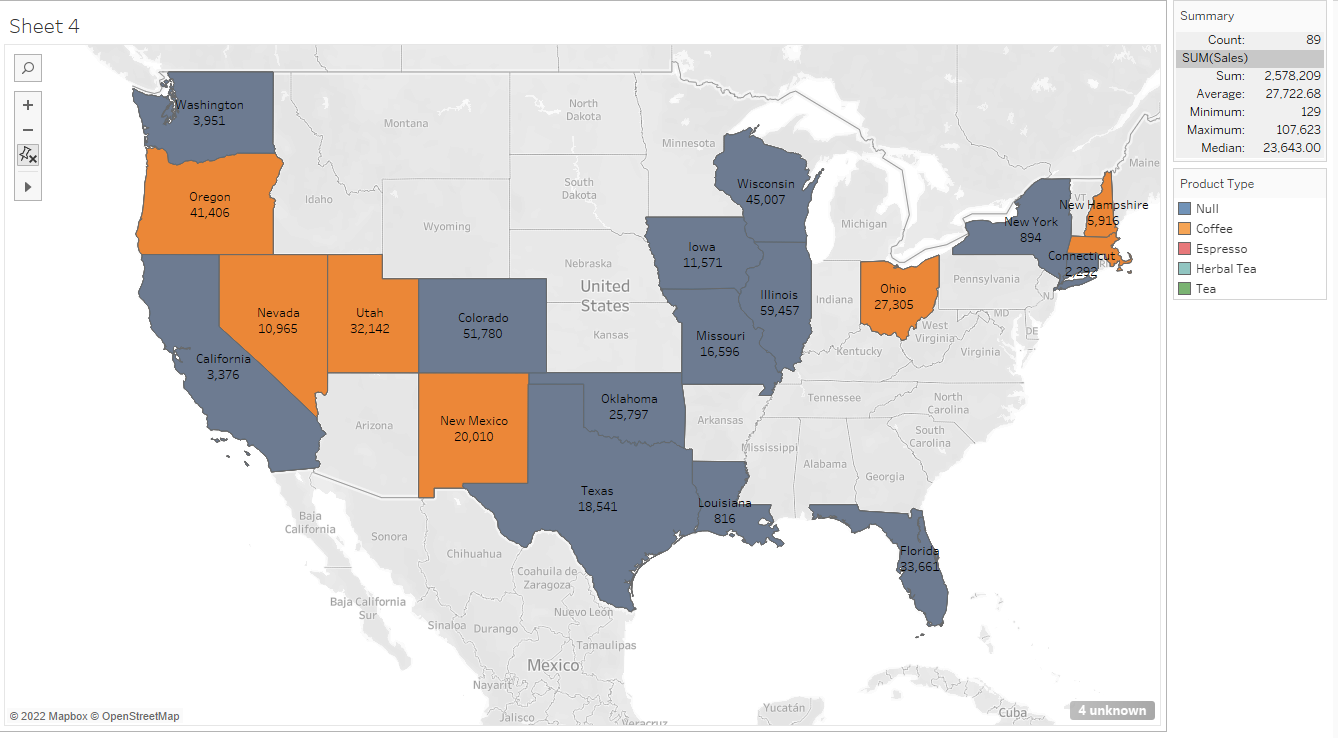
**SHEET 4 :**



**EXPLANATION:**

The above packed bubbles shows the data of the product and their sales. So from this graph can see that the most selling product is Columbian and decaf espresso with the average sales of 327k and 198k being the average sale of all product.

**SHEET 5:**



**EXPLANATION:**

The above map shows the data of the product type, sales and state. From this map we can see that the most selling product type is coffee and espresso where most sales have been done in Colorado with total sales of 51k.

**2. Create a dashboard with all the sheets created.**

