UNITED STATES SUPERSTORE DATA ANALYTICS

QUERY	:	SQL (PostgreSQL)
VISUALIZATION	:	TABLEAU

by EMANUEL FEBRIANO DWI SAPUTRA

Hello!

My name is Emanuel Febriano Dwi Saputra. I am an Electrical Engineering graduate from Sanata Dharma University. My GPA is 3.14 to 4.00. I am interested in the data field, especially in the data analytics field.

I have experience as an assistant production supervisor. I am used to processing production data. Because of my interest in the field of data analytics, I broaden my knowledge of process data using SQL and for visualization, I deepen my knowledge of Tableau.

Through this portfolio, I present insights from existing cases using PostgreSQL for query processing as well as final visualization using Tableau. Hopefully from this portfolio there will be insight that has value and we can discuss opportunities that may exist in the future.

febrianoemanuel@gmail.com

Emanuel-Febriano-29021996

SUPERSTORE DATASET

In this dataset, there are 18 columns and 9800 rows of data as shown in the image below

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales
1	CA-2017-152156	8/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset C	261.96

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales
9800	CA-2016-128608	12/1/2016	17/01/2016	Standard Class	CS-12490	Cindy Schnelling	Corporate	United States	Toledo	Ohio	43615	East	TEC-AC-10000487	Technology	Accessories	SanDisk Cruzer 4	10.384

Data in the table is data for 3 years, 2015 to 2018

BUSINESS QUESTION

1. How is the sales data every month for 3 years (2015 – 2018) and in what month was the highest and lowest sales?
2. How much sales did each state for that month?
3. How many sales and quantity by product category for that month?
4. How many quantity of ship mode used for that month?

01

HOW IS THE SALES DATA EVERY MONTH FOR 3 YEARS (2015 – 2018) AND IN WHAT MONTH WAS THE HIGHEST AND LOWEST SALES?

PostgreSQL

```
With
salesjan(january)
    as (select sum(sales)
        from public."SUPERSTORE"
            where order_date like '%/1/%' or order_date like '%/01/%'),
salesfeb(february)
    as (select sum(sales)
        from public."SUPERSTORE"
            where order_date like '%/2/%' or order_date like '%/02/%'),
```

```
salesnov(november)
    as (select sum(sales)
        from public."SUPERSTORE"
            where order_date like '%/11/%' or order_date like '%/011/%'),
salesdes(desember)
    as (select sum(sales)
        from public."SUPERSTORE"
            where order_date like '%/12/%' or order_date like '%/012/%'),
```

I use WITH clause to create a temporary table to calculate sales every month (January – Desember). SUM statement that I use for sum the values each month and LIKE statement for determine what month the sales will be calculated.

PostgreSQL

```
bulan("12")
    as (select order_date,
        case
            when order_date like '%/1/%' or order_date like '%/01/%' then 'JANUARY'
            when order_date like '%/2/%' or order_date like '%/02/%' then 'FEBRUARY'
            when order_date like '%/3/%' or order_date like '%/03/%' then 'MARCH'
            when order_date like '%/4/%' or order_date like '%/04/%' then 'APRIL'
            when order_date like '%/5/%' or order_date like '%/05/%' then 'MAY'
            when order_date like '%/6/%' or order_date like '%/06/%' then 'JUNE'
            when order_date like '%/7/%' or order_date like '%/07/%' then 'JULY'
            when order_date like '%/8/%' or order_date like '%/08/%' then 'AUGUST'
            when order_date like '%/9/%' or order_date like '%/09/%' then 'SEPTEMBER'
            when order_date like '%/10/%' or order_date like '%/010/%' then 'OCTOBER'
            when order_date like '%/11/%' or order_date like '%/011/%' then 'NOVEMBER'
        else 'DESEMBER'
        END as "12month"
    from public."SUPERSTORE"),
```

Then by using the CASE statement, I create the name of the month based on the order date column. I use the LIKE statement to determine the name of the month and giving name to column as 12month

```
bulandtnt("months")
    as (select distinct "12month"
        from bulan),
```

Next, I use the DISTINCT statement to select a unique data name from 12month column, namely January to December. This is done so that the name of the month displayed is not double.

PostgreSQL

```
sales(smonths)
    as (select months,
        case
            when months = 'JANUARY' then "january"
            when months = 'FEBRUARY' then "february"
            when months = 'MARCH' then "march"
            when months = 'APRIL' then "april"
            when months = 'MAY' then "may"
            when months = 'JUNE' then "june"
            when months = 'JULY' then "july"
            when months = 'AUGUST' then "august"
            when months = 'SEPTEMBER' then "september"
            when months = 'OCTOBER' then "october"
            when months = 'NOVEMBER' then "november"
        else "desember"
        END AS "sales ($)"
    from bulandtnt, salesjan, salesfeb, salesmar, salesap, salesmay,
        salesjun, salesjul, salesaug, salessep, salesoct, salesnov, salesdes
    group by months, january, february, march, april, may, june, july,
        august, september, october, november, desember)
```

Then in the last WITH statement, I use the CASE statement to match the sales value according to the month name and give that column name as sales ($)

PostgreSQL

```
select smonths, "sales ($)"
from sales
order by "sales ($)" desc;
```

Finally, to display sales data every month, I choose the smonth column to display the name of the month and the sales ($) column to display sales data each month from a temporary table called sales.

The data displayed will be like the picture on the right

	smonths text	sales ($) numeric
1	NOVEMBER	350169.6
2	DESEMBER	321488.6
3	SEPTEMBER	300111.1
4	OCTOBER	199502.2
5	MARCH	197577.1
6	AUGUST	157320.0
7	MAY	154091.4
8	JUNE	145841.2
9	JULY	145538.8
10	APRIL	136286.9
11	JANUARY	94293.0
12	FEBRUARY	59372.8

ANSWER

How is the sales data every month for 3 years (2015 – 2018) and in what month was the highest and lowest sales?

Based on the data in the table on the right, it can be concluded that the highest sales were in November and the lowest sales were in February

	smonths text	sales ($) numeric
1	NOVEMBER	350169.6
2	DESEMBER	321488.6
3	SEPTEMBER	300111.1
4	OCTOBER	199502.2
5	MARCH	197577.1
6	AUGUST	157320.0
7	MAY	154091.4
8	JUNE	145841.2
9	JULY	145538.8
10	APRIL	136286.9
11	JANUARY	94293.0
12	FEBRUARY	59372.8

2

HOW MUCH SALES DID EACH STATE FOR THAT MONTH?

PostgreSQL

```
WITH
salesnov(nov)
as (select distinct "state", sum(sales) as "november_sales ($)"
    from public."SUPERSTORE"
    where order_date like '%/11/%' or order_date like '%/011/%'
    group by state
    order by "november_sales ($)" asc),
salesfeb(feb)
as (select distinct "state", sum(sales) as "february_sales ($)"
    from public."SUPERSTORE"
    where order_date like '%/2/%' or order_date like '%/02/%'
    group by "state"
    order by "february_sales ($)" asc)
```

To find the sales value in November and February, I used the WITH statement to create a temporary table which I named salesnov and salesfeb. I use the LIKE statement to determine the months that sales will be calculated are November and February

PostgreSQL

```
select feb as "state", "november_sales ($)", "february_sales ($)"
from
(salesfeb left join salesnov
on salesfeb.feb = salesnov.nov)
order by "november_sales ($)", "february_sales ($)";
```

After getting the sales value of each state in the two temporary tables. I use the LEFT JOIN statement to join the two tables whose results can be seen in the image on the right

	state	november_sales ($)	february_sales ($)
2	Arizona	4097.7	897.4
3	Arkansas	372.5	30
4	California	53614.4	11077.7
5	Colorado	4003.6	9.4
6	Connecticut	968.9	612.1
7	Delaware	16859.2	1558.4
8	Florida	12219.6	789.7
9	Georgia	4335.9	4762
10	Illinois	6271.2	2544.5
11	Indiana	9439.1	82.6
12	Kansas	540.3	21.1
13	Kentucky	9282.1	1190.3
14	Louisiana	683	29.2
15	Maryland	4974.2	3265.1
16	Massachusetts	1747.8	25.9
17	Michigan	9434.6	79.4
18	Minnesota	3368.9	90.6
19	Mississippi	173.8	264.2
20	Missouri	2176.2	535.2
21	Nebraska	90	178.4
22	Nevada	346	14.6
23	New Hampshire	NULL	35.9
24	New Jersey	3701.1	808.5
25	New Mexico	1363.1	49.1
26	New York	65007.3	3481.9
27	North Carolina	19421.2	1556.7
28	Ohio	19892.4	2790.4
29	Oklahoma	4681.4	89.6
30	Oregon	3940.5	55.7
31	Pennsylvania	18247.3	1603.9
32	Rhode Island	53	599.9
33	Tennessee	3215.1	700
34	Texas	24026.9	2599.4
35	Utah	464.1	12.1
36	Virginia	3673.1	11660
37	Washington	18676.2	3280.9
38	Wisconsin	3961.7	1991

ANSWER

HOW MUCH SALES DID EACH STATE FOR THAT MONTH?

Based on the data obtained, there are 38 states that make sales in November and February. The sales value for each state can be seen in the table on the right.

	state	november_sales ($)	february_sales ($)
2	Arizona	4097.7	897.4
3	Arkansas	372.5	30
4	California	53614.4	11077.7
5	Colorado	4003.6	9.4
6	Connecticut	968.9	612.1
7	Delaware	16859.2	1558.4
8	Florida	12219.6	789.7
9	Georgia	4335.9	4762
10	Illinois	6271.2	2544.5
11	Indiana	9439.1	82.6
12	Kansas	540.3	21.1
13	Kentucky	9282.1	1190.3
14	Louisiana	683	29.2
15	Maryland	4974.2	3265.1
16	Massachusetts	1747.8	25.9
17	Michigan	9434.6	79.4
18	Minnesota	3368.9	90.6
19	Mississippi	173.8	264.2
20	Missouri	2176.2	535.2
21	Nebraska	90	178.4
22	Nevada	346	14.6
23	New Hampshire	NULL	35.9
24	New Jersey	3701.1	808.5
25	New Mexico	1363.1	49.1
26	New York	65007.3	3481.9
27	North Carolina	19421.2	1556.7
28	Ohio	19892.4	2790.4
29	Oklahoma	4681.4	89.6
30	Oregon	3940.5	55.7
31	Pennsylvania	18247.3	1603.9
32	Rhode Island	53	599.9
33	Tennessee	3215.1	700
34	Texas	24026.9	2599.4
35	Utah	464.1	12.1
36	Virginia	3673.1	11660
37	Washington	18676.2	3280.9
38	Wisconsin	3961.7	1991

3

HOW MUCH SALES AND QUANTITY BY PRODUCT CATEGORY FOR THAT MONTH?

PostgreSQL

```
WITH
catnov(cnov)
as (select distinct category, count(category) as qty_category_nov, sum(sales) as "sales_nov ($)"
    from public."SUPERSTORE"
    where
        order_date like '%/11/%' or order_date like '%/011/%'
    group by category),
catfeb(cfeb)
as (select distinct category, count(category) as qty_category_feb, sum(sales) as "sales_feb ($)"
    from public."SUPERSTORE"
    where
        order_date like '%/2/%' or order_date like '%/02/%'
    group by category)
```

To get the quantity and sales values, I use the WITH statement which in the sub-queries there is a DISTINCT category column to find out what product categories exist, the COUNT statement to find out the quantity value of each category, and the SUM statement to find the sales value.

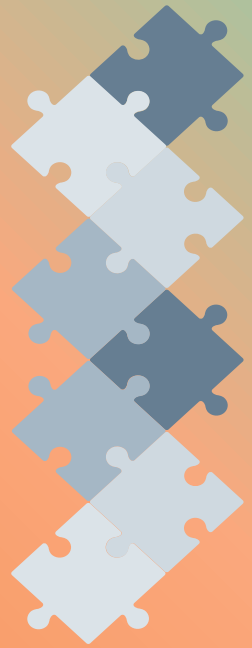
PostgreSQL

```
select cfeb as product_category, qty_category_nov, qty_category_feb, "sales_nov ($)", "sales_feb ($)"
from
(catfeb left join catnov
on catnov.cnov = catfeb.cfeb);
```

When the two temporary tables containing the desired data have been created, then my next step is to combine the two tables using the LEFT JOIN statement.

The resulting data is displayed as follows

	product_category character (1000)	qty_category_nov bigint	qty_category_feb bigint	sales_nov ($) numeric	sales_feb ($) numeric
1	Furniture	316	62	120510.8	15674.2
2	Office Supplies	849	176	99158.0	20612.8
3	Technology	284	59	130500.8	23085.8



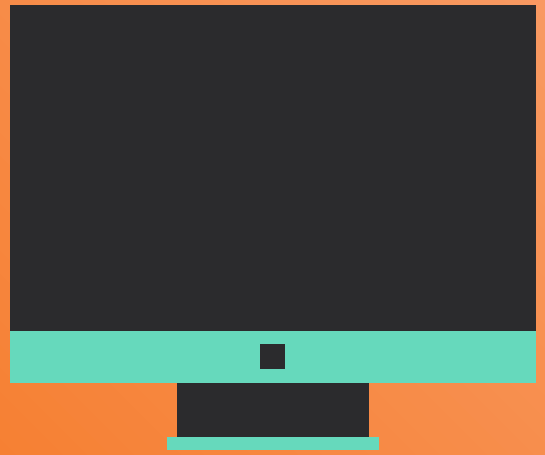
ANSWER

HOW MUCH SALES AND QUANTITY BY PRODUCT CATEGORY FOR THAT MONTH?

	product_category character (1000)	qty_category_nov bigint	qty_category_feb bigint	sales_nov ($) numeric	sales_feb ($) numeric
1	Furniture	316	62	120510.8	15674.2
2	Office Supplies	849	176	99158.0	20612.8
3	Technology	284	59	130500.8	23085.8

Based on the data in the table, in November and February the sales volume in the office supplies category was the highest, at 849 and the lowest was technology, at 284. Meanwhile in February, office supplies were 176 and technology was 59.

Then for sales, in November the largest was in the technology category, which was $130500.8 and the lowest was in the office supplies category at $99158. For February, the largest was technology at $23085.8 and the lowest was in the furniture category at $15674.2



4

HOW MANY QUANTITY OF EACH SHIP MODE USED FOR THAT MONTH?

PostgreSQL

```
WITH
shipnov(nov)
    as (SELECT
        ship_mode, count(ship_mode) AS qty_nov
        FROM
        public."SUPERSTORE"
        WHERE
            (ship_mode = 'Standard Class'  OR
            ship_mode = 'Second Class'  OR
            ship_mode = 'First Class'  OR
            ship_mode = 'Same Day') and
            order_date like '%/11/%' or order_date like '%/011/%'
    Group by country, ship_mode
    Order by qty_nov desc),
shipfeb(feb)
    as (SELECT
        ship_mode, count(ship_mode) AS qty_feb
        FROM
        public."SUPERSTORE"
        WHERE
            (ship_mode = 'Standard Class'  OR
            ship_mode = 'Second Class'  OR
            ship_mode = 'First Class'  OR
            ship_mode = 'Same Day') and
            order_date like '%/2/%' or order_date like '%/02/%'
    Group by country, ship_mode
    Order by qty_feb desc)
```

To find out how many quantity ship modes are used, I created two temporary tables, namely shipnov and shipfeb, each with a column of ship mode and quantity.

PostgreSQL

```
select nov as ship_mode, qty_nov, qty_feb
from
(shipfeb left join shipnov
on shipnov.nov = shipfeb.feb);
```

Then the two temporary tables are combined with a LEFT JOIN statement and the final table will look like in the picture

	ship_mode character (1000)	qty_nov bigint	qty_feb bigint
1	Standard Class	876	175
2	Second Class	297	62
3	First Class	204	46
4	Same Day	72	14

ANSWER

HOW MANY QUANTITY OF EACH SHIP MODE USED FOR THAT MONTH?

	ship_mode character (1000)	qty_nov bigint	qty_feb bigint
1	Standard Class	876	175
2	Second Class	297	62
3	First Class	204	46
4	Same Day	72	14

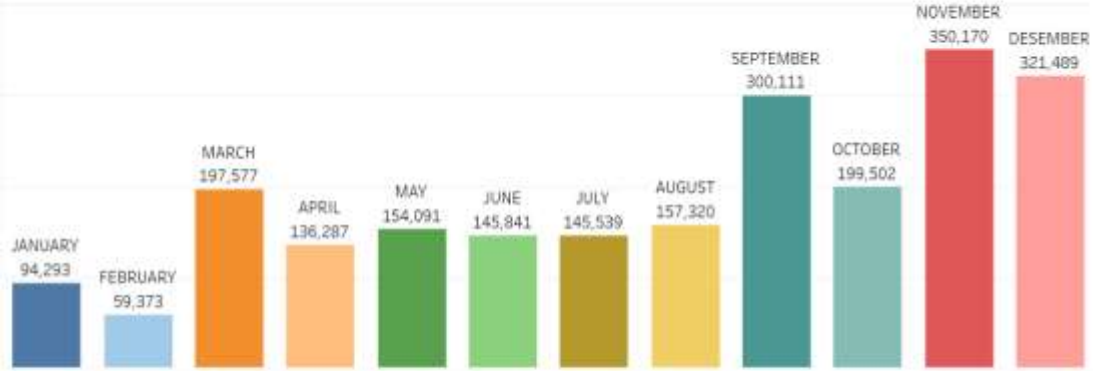
From the existing data, the ship mode that was used the most in November and February was the standard class, which was 876 and in February it was 175. Meanwhile, the ship mode that was used the least in both months was the same day, which was 72 in the month of November and 14 in February

DATA ANALYTICS
PORTFOLIO

VISUALIZATION

UNITED STATES SUPERSTORE DATA ANALYTICS

SALES (2015 - 2018)

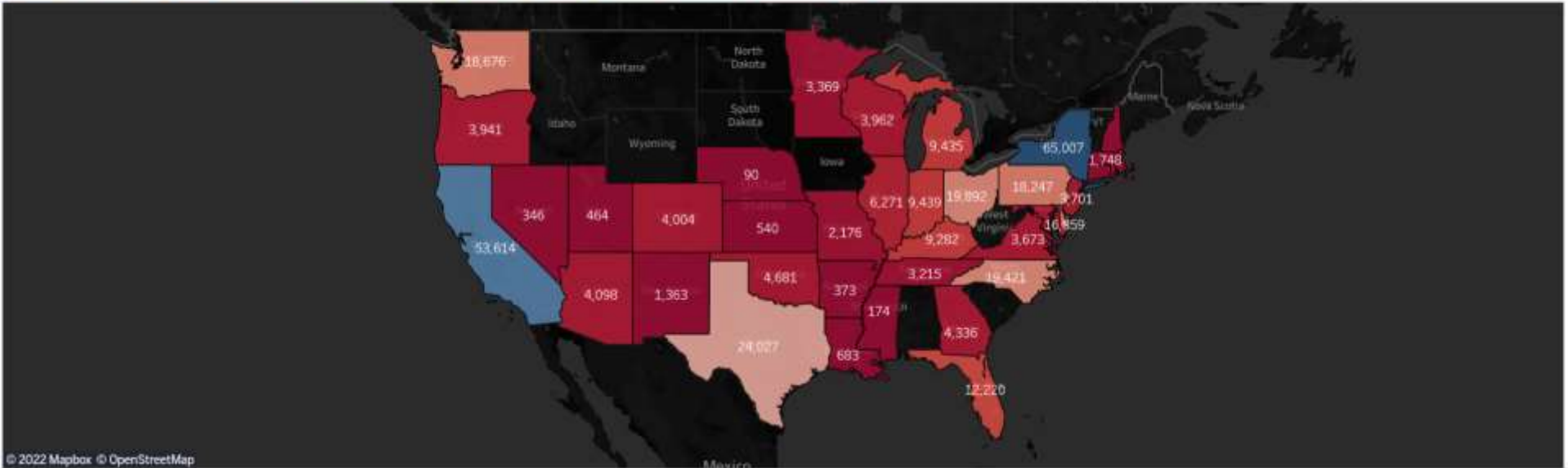


QTY & SALES NOVEMBER

Produ..	QTY NOV	Sales NOV ($)
Technology	284	130,501
Furniture	316	120,513
Office Supplies	849	99,158

SHIP MODE NOVEMBER

NOVEMBER SALES (MAP)



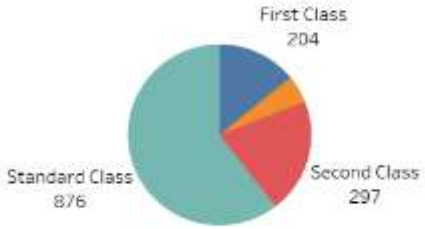
QTY & SALES NOVEMBER

Produc..	QTY NOV	Sales NOV ($)
Technology	284	130,501
Furniture	316	120,511
Office Supplies	849	99,158

Those buttons can be used to change sales data from November to February and vice versa

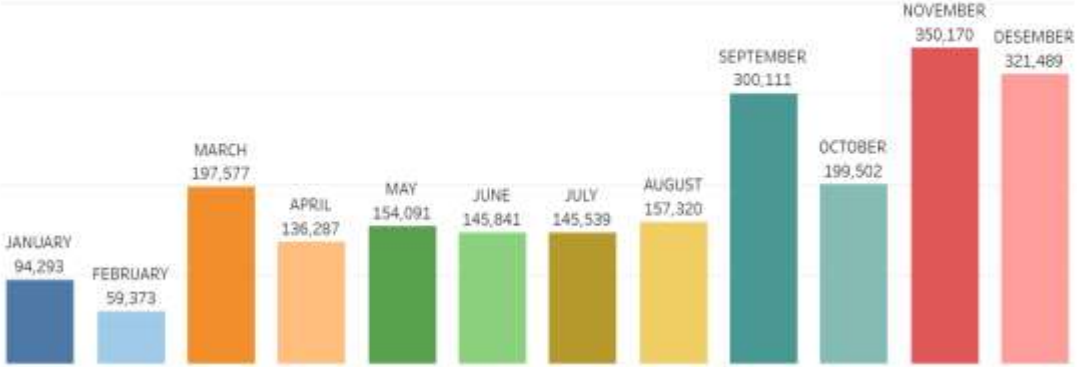
NOVEMBER SALES (MAP)

SHIP MODE NOVEMBER



UNITED STATES SUPERSTORE DATA ANALYTICS

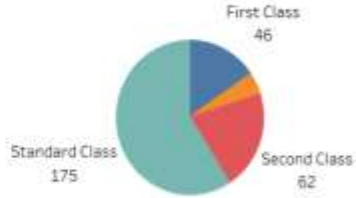
SALES (2015 - 2018)



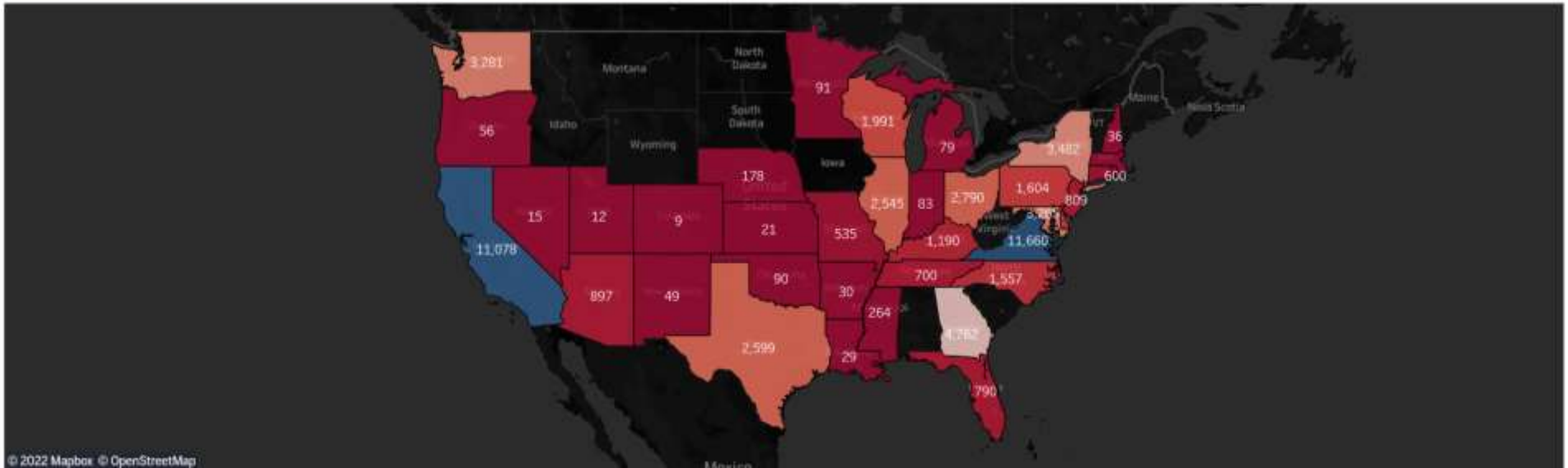
QTY & SALES FEBRUARY

Produ...	QTY FEB	Sales FEB ($)
Technology	59	23,086
Office Supplies	176	20,613
Furniture	62	15,674

SHIP MODE FEBRUARY



FEBRUARY SALES (MAP)



THANK YOU

My Tableau Public Page