

# Business Intelligence

## Short Course

by Muhammad Nashir



# Class Topics

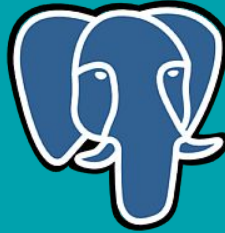
Understanding Business Problem & How to Solve It Using Data	SQL	Data Driven Using Tableau
<ul style="list-style-type: none"><li>• BI Methodology and Fundamental of Bussiness Understanding</li><li>• Data Story Telling</li><li>• Statistic Fundamentals</li></ul>	<ul style="list-style-type: none"><li>• SQL Fundamentals</li><li>• Simple Queries</li><li>• Complex Queries</li></ul>	<ul style="list-style-type: none"><li>• Introduction to Tableau and Data Visualization</li><li>• Analyzing Data with SQL and Tableau</li><li>• Data Aggregation 1</li><li>• Data Aggregation 2</li><li>• Interactive Dashboard for Insight and Business Recommendation</li><li>• Tableau Analytic and Deployment</li></ul>



# Tools



Google Sheets



PostgreSQL



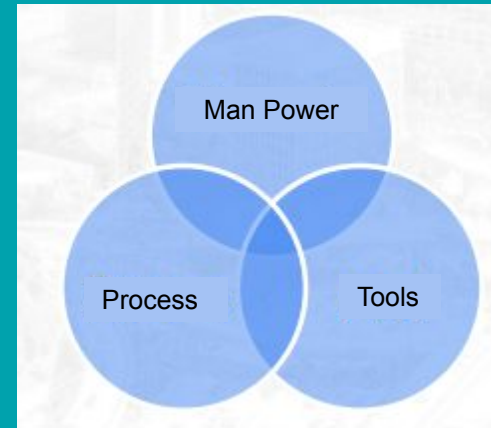
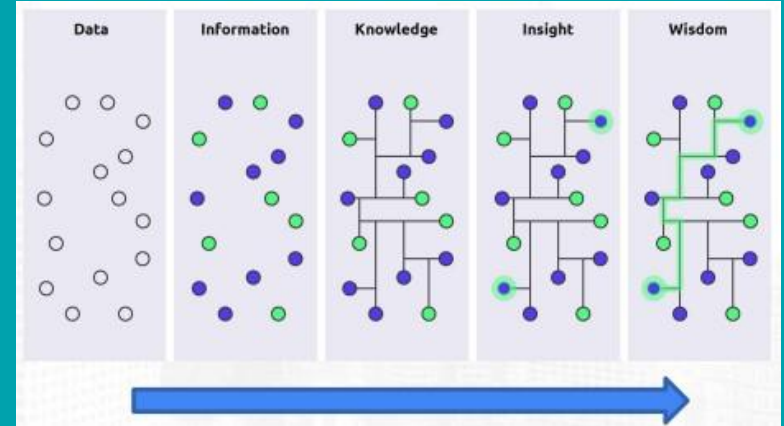
# What is Business Intelligence (BI)?



**Business intelligence** is a mix of man power, processes and tools that help company change:

- Data becomes information
- Information becomes insight
- Until the company can make more accurate decision.

Business intelligence can help companies make better decisions by showing present and historical data within their business context.



# Understanding Business Problem & How to Solve It Using Data

**—Case Study—**



## Problem:

Imagine you have a business in education field, say as ABC Academy.

During the annual evaluation, investors want to know accurately how the teachers are performing, by knowing what the range of all ABC Academy student's score. However, it will take a long time to wait for all students to complete the final project.

To answer investor's questions more quickly but still accurate, a random sampling of 10 students was taken with the following scores:

45, 55, 67, 45, 68, 79, 98, 87, 84, 82.

Calculate the confidence interval of the student's scores at ABC Academy, if the expected confidence level is 98%?

# Formula of CL:

$$CI = \bar{x} \pm t * \frac{s}{\sqrt{n}}$$

- $\bar{x}$  = the sample mean, which is the point estimate.
- $t$  = the critical t-value
- $s$  = the sample standard deviation
- $\sqrt{n}$  = the square root of sample size





## Step 1. Find The Degrees of Freedom

Number of samples - 1  
→  $10 - 1 = 9$

## Step 2. Find The Amount of Area In One Tail

$(1 - \text{confidence level}) / 2$ , where confidence level = 98% = 0.98  
→  $(1 - 0.98) / 2 = 0.01$

## Step 3. Find the Critical Value of t-distribution table

Degrees of Freedom = 9  
Amount of Area in One Tail = 0.01



2.821438

Degrees of freedom ( $\nu$ )	Amount of area in one tail ( $\alpha$ )							
	0.0005	0.001	0.005	0.010	0.025	0.050	0.100	0.200
1	636.6192	318.3088	63.65674	31.82052	12.70620	6.313752	3.077684	1.376382
2	31.59905	22.32712	9.924843	6.964557	4.302653	2.919986	1.885618	1.060660
3	12.92398	10.21453	5.840909	4.540703	3.182446	2.353363	1.637744	0.978472
4	8.610302	7.173182	4.604095	3.746947	2.776445	2.131847	1.533206	0.940965
5	6.868827	5.893430	4.032143	3.364930	2.570582	2.015048	1.475884	0.919544
6	5.958816	5.207626	3.707428	3.142668	2.446912	1.943180	1.439756	0.905703
7	5.407883	4.785290	3.499483	2.997952	2.364624	1.894579	1.414924	0.896030
8	5.041305	4.500791	3.355387	2.896459	2.306004	1.859548	1.396815	0.888890
9	4.780913	4.296806	3.249836	2.821438	2.262157	1.833113	1.383029	0.883404
10	4.586894	4.143700	3.169273	2.763769	2.228139	1.812461	1.372184	0.879058
11	4.436979	4.024701	3.105807	2.718079	2.200985	1.795885	1.363430	0.875530
12	4.317791	3.929633	3.054540	2.680998	2.178813	1.782288	1.356217	0.872609
13	4.220832	3.851982	3.012276	2.650309	2.160369	1.770933	1.350171	0.870152
14	4.140454	3.787390	2.976843	2.624494	2.144787	1.761310	1.345030	0.868055
15	4.072765	3.732834	2.946713	2.602480	2.131450	1.753050	1.340606	0.866245
16	4.014996	3.686155	2.920782	2.583487	2.119905	1.745884	1.336757	0.864667
17	3.965126	3.645767	2.898231	2.566934	2.109816	1.739607	1.333379	0.863279

## Step 4. Find The Standar Deviation Divided by The Square Root of Sample Size

Using Excel Function:

→  $\text{STDEV.S()}/\text{SQRT()}$  => 5.746496517

## Step 5. Step 3 \* Step 4

→  $2.821438 * 5.746496517 = 16.21338364$

## Lower Confidence Interval:

→ Sample Mean - Step 5 71 - 16.21338364 = 54.78661636

## Upper Confidence Interval:

→ Sample Mean + Step 5 71 + 16.21338364 = 87.21338364

# Conclusion

So, after we found the values of the lower and upper confidence interval in this case, we can conclude that the most of scores of all students at ABC Academy are around 54.79 and 87.21.



# SQL

## —Case Study—



## Problem:

As a Business Intelligence Analyst at the Superstore, you are asked to do some analysis supports from the internal BI team and from the other teams such as Marketing, Business Development, Sales, etc.

Here are some of the most urgent analysis requests from other teams. As a member of the BI Analyst who works most efficiently, you are asked to do assists to the following problems!

You are provided with the following dataset::

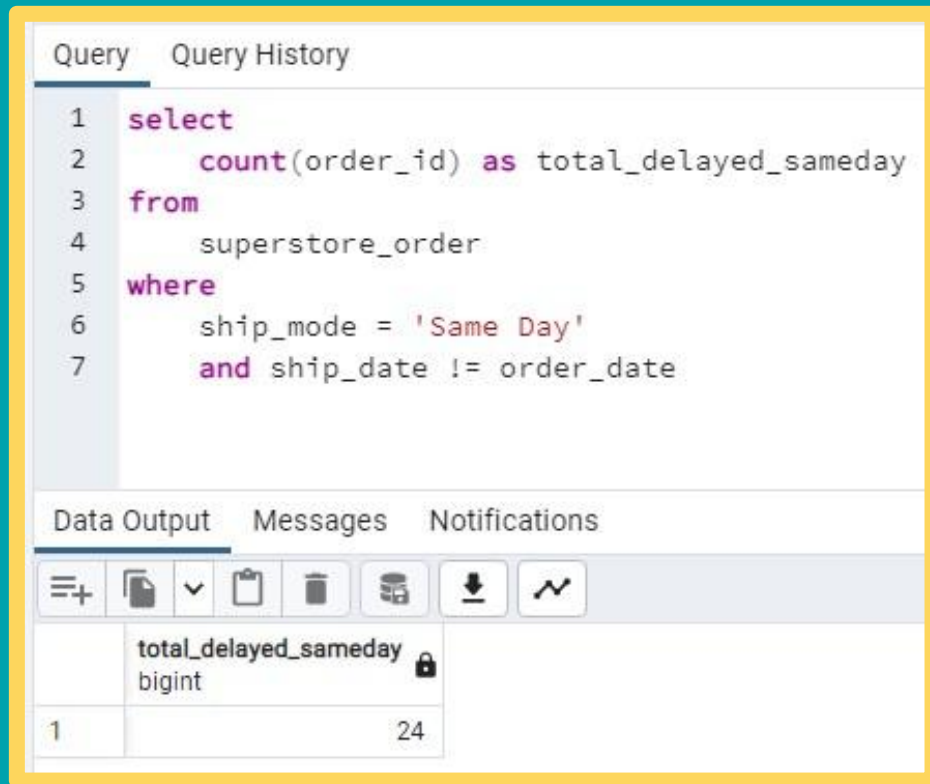
1. Superstore\_customer.csv
2. Superstore\_order.csv
3. Superstore\_product.csv

# Case 1

---

SAME DAY Ship Mode service is a service where the product ordered by the customer can be sent directly on the same day as the day of order. But in reality, not all customers who order receive the benefits of this service well. In other words, there are also some SAME DAY orders that are not delivered on the same day as the order day. The Operations Team would like to analyze this problem to be followed up. You are asked to display the number of SAME DAY orders which is experiencing delays in delivery.

# SQL Query Syntax



The screenshot shows a SQL query editor with two tabs: "Query" and "Query History". The "Query" tab is active, displaying a SQL query with line numbers 1 through 7. The query is as follows:

```
1 select
2     count(order_id) as total_delayed_sameday
3 from
4     superstore_order
5 where
6     ship_mode = 'Same Day'
7     and ship_date != order_date
```

Below the query editor, there are three tabs: "Data Output", "Messages", and "Notifications". The "Data Output" tab is active, showing a table with the results of the query. The table has two columns: "total\_delayed\_sameday" (type: bigint) and a lock icon. The first row shows the value 24.

	total_delayed_sameday	bigint
1	24	

There were 24 same day orders that were delayed



# Case 2

---

The Business team would like to conduct further analysis of the company's profitability. This time, they want to see the relationship between the amount of discount given and profitability received by the company. You are asked to display this relationship with shows the average profit for each discount level, where the discount level criteria are as follows:

- LOW if the discount is below 0.2 (excluding 0.2),
- MODERATE if the discount starts from 0.2 to below 0.4 (excluding 0.4)
- HIGH if the discount starts from 0.4 and above.

# SQL Query Syntax

Query Query History

```
1 select
2     case when discount < 0.2 then 'Low'
3         when discount < 0.4 then 'Moderate'
4         else 'High'
5     end as level_discount,
6     avg(profit) as average_profit
7 from superstore_order
8 group by level_discount
9 order by average_profit desc
```

Data Output Messages Notifications



	level_discount text	average_profit numeric
1	Low	67.03797971278316737
2	Moderate	19.83556402454614475
3	High	-107.6520113257243339

The higher the discount level,  
the lower the profit.

# Case 3

---

The Sales Team asked the Business Intelligence Analyst to analyze the performance of Category and Subcategory of products owned by the company. You are asked to return the following metrics for each of the existing Category-Subcategory pairs:

- Average discount
- Average profit

Don't forget to display the full Category and Subcategory names instead of just displaying only the Product ID to make it easier for the Sales Team to understand the results of your analysis

# SQL Query Syntax

Query Query History

```
1  select
2      category,
3      sub_category,
4      avg(discount) as avg_discount,
5      avg(profit) as avg_profit
6  from superstore_order o
7  join superstore_product p
8  on o.product_id = p.product_id
9  group by 1, 2
10 order by 1, 2
```

# SQL Query Syntax

Output Query:

	category character varying	sub_category character varying	avg_discount numeric	avg_profit numeric
1	Furniture	Bookcases	0.21114035087719298246	-15.2305087719298380
2	Furniture	Chairs	0.17017828200972447326	43.0958935170178139
3	Furniture	Furnishings	0.13834900731452455590	13.64591807732497215
4	Furniture	Tables	0.26128526645768025078	-55.5657714733542497
5	Office Supplies	Appliances	0.16652360515021459227	38.92275836909870306
6	Office Supplies	Art	0.07487437185929648241	8.20073743718592866
7	Office Supplies	Binders	0.37229152987524622456	19.8435740643466793
8	Office Supplies	Envelopes	0.08031496062992125984	27.4180185039370051
9	Office Supplies	Fasteners	0.08202764976958525346	4.37565990783410111
10	Office Supplies	Labels	0.06868131868131868132	15.2369615384615381
11	Office Supplies	Paper	0.07489051094890510949	24.8566199270072976
12	Office Supplies	Storage	0.07470449172576832151	25.15227706855791304
13	Office Supplies	Supplies	0.07684210526315789474	-6.25841842105263644
14	Technology	Accessories	0.07845161290322580645	54.11178799999999635
15	Technology	Copiers	0.16176470588235294118	817.9091897058822724
16	Technology	Machines	0.30608695652173913043	29.4326686956520713
17	Technology	Phones	0.15455568053993250844	50.07393768278964770

# SQL Query Syntax

Order by average profit and take limitation to 5 to see the top 5 category and subcategory

QueryQuery History

```
1 select
2     category,
3     sub_category,
4     avg(profit) as avg_profit,
5     avg(discount) as avg_discount
6 from superstore_order o
7 join superstore_product p
8 on o.product_id = p.product_id
9 group by 1, 2
10 order by 3 desc
11 limit 5
```

Data OutputMessagesNotifications

	category character varying	sub_category character varying	avg_profit numeric	avg_discount numeric
1	Technology	Copiers	817.9091897058822724	0.16176470588235294118
2	Technology	Accessories	54.11178799999999635	0.07845161290322580645
3	Technology	Phones	50.07393768278964770	0.15455568053993250844
4	Furniture	Chairs	43.0958935170178139	0.17017828200972447326
5	Office Supplies	Appliances	38.92275836909870306	0.16652360515021459227

The **Technology** category dominate the highest average profit with the Copiers, Accessories, and Phones subcategories.

# SQL Query Syntax

Order by average discount and take limitation to 5 to see the top 5 category and subcategory

Query


Query History

```
1 select
2     category,
3     sub_category,
4     avg(profit) as avg_profit,
5     avg(discount) as avg_discount
6 from superstore_order o
7 join superstore_product p
8 on o.product_id = p.product_id
9 group by 1, 2
10 order by 4 desc
11 limit 5
```

Data Output

Messages

Notifications



	category character varying	sub_category character varying	avg_profit numeric	avg_discount numeric
1	Office Supplies	Binders	19.8435740643466793	0.37229152987524622456
2	Technology	Machines	29.4326686956520713	0.30608695652173913043
3	Furniture	Tables	-55.5657714733542497	0.26128526645768025078
4	Furniture	Bookcases	-15.2305087719298380	0.21114035087719298246
5	Furniture	Chairs	43.0958935170178139	0.17017828200972447326

The highest average discount is around 0.37 is owned by the BINDERS subcategory of the OFFICE SUPPLIES category.

# Case 4

---

The Business Development team is considering further expansion in California, Texas and also Georgia. As material for their consideration, you are asked to display the performance of each of the Customer Segments in that states on 2016 only. The requested performance metrics are as follows:

- Total sales
- Average profit














# SQL Query Syntax

Query    Query History

```
1  select
2      c.segment,
3      sum(o.sales) as total_sales,
4      avg(o.profit) as avg_profit
5  from superstore_order o
6  join superstore_customer c
7  on o.customer_id = c.customer_id
8  where c.state in ('California', 'Texas', 'Georgia')
9  and date_part('year', o.order_date) = '2016'
10 group by 1
11 order by 3 desc
```

# SQL Query Syntax

Output Query:

Data Output Messages Notifications			
       			
	segment character varying 	total_sales numeric 	avg_profit numeric 
1	Home Office	34897.9529999999997967	34.66199395973154455
2	Corporate	50951.91100000000016284	33.57349056603773146
3	Consumer	90982.31960000000012181	30.32965562913906877

The HOME OFFICE segment has the highest average profit but the lowest total sales.

# Case 5

---

The Business team is interested to see which region has the most number of customers who loves discounts. Therefore, the Business Team asks you as a Business Intelligence Analyst to display the number of people/customers who have an average discount above 0.4 for each existing region.

# SQL Query Syntax

## Part 1:

Query	Query History
1	with
2	temp as
3	(select
4	customer_id,
5	avg(discount)
6	from superstore_order
7	group by 1
8	having avg(discount) >= 0.4)
9	

Created temporary table (temp) to find customer\_id that have average discount above 0.4

# SQL Query Syntax

Output Query:

	customer_id character varying 🔒	avg numeric 🔒
1	SG-20890	0.400000000
2	GH-14485	0.466666666
3	LH-17020	0.466666666
4	TS-21085	0.700000000
5	MG-18205	0.450000000
6	AG-10765	0.460000000
7	CD-11980	0.433333333
8	RH-19555	0.533333333
9	VS-21820	0.475000000
10	TP-21415	0.462000000

There are 10 customers who have average discount above 0.4.

# SQL Query Syntax

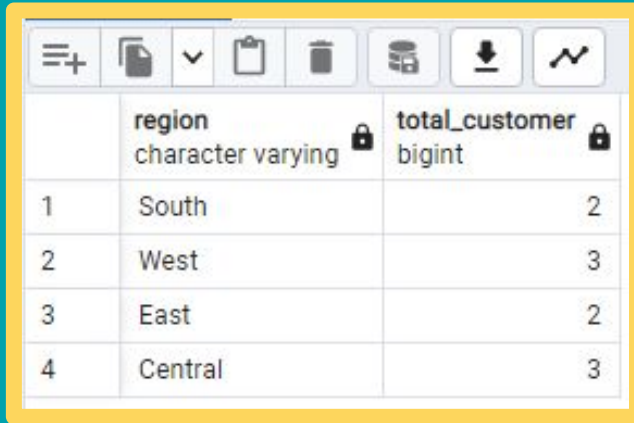
## Part 2:

```
10 select
11     region,
12     count(1) as total_customer
13 from temp
14 join superstore_customer c
15 on temp.customer_id = c.customer_id
16 group by 1
```

After finding total customer\_id that have avg discount above 0.4, then combine it with the customer table to count the total customer\_id for each region.

# SQL Query Syntax

Output Query:



The screenshot shows a database query result interface. At the top is a toolbar with icons for menu, save, dropdown, clipboard, delete, database, download, and chart. Below the toolbar is a table with two columns: 'region' (character varying) and 'total\_customer' (bigint). The table contains four rows of data, numbered 1 to 4 in the first column.

	region character varying	total_customer bigint
1	South	2
2	West	3
3	East	2
4	Central	3

The West & Central Region has the most customers who have average discount above 0.4

# Tableau Part 1

## —Case Study—





## Problem:

You are a Product Manager exploring data from the Sample Superstore. Do a simple exploration into the data to get the information you need. For increase profits, we must reduce the number of order returned. You have to present regarding any items returned by the buyer.

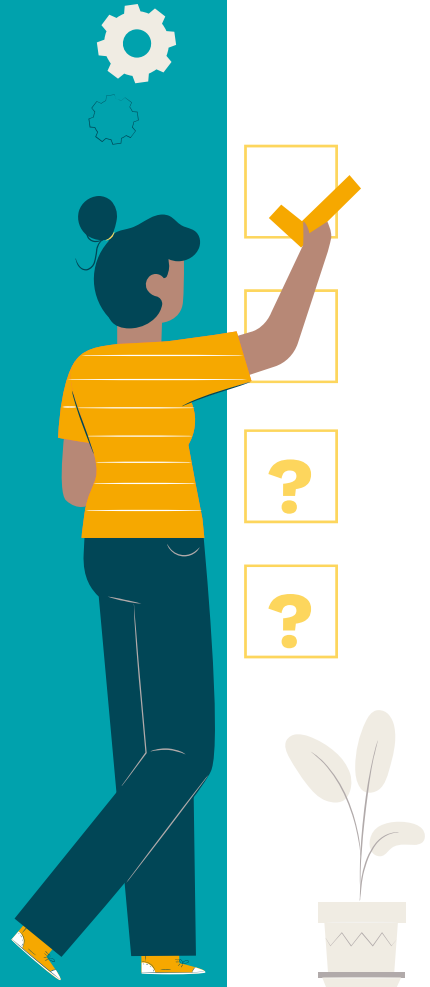
Make appropriate charts to answer the following questions:

1. Which Category and Ship Mode has the highest number of Order Returns?
2. In what year did the number of Order Returns in July increase from June?
3. Find the percentage of Order Return from United States compared to total Order return in 2015, and what Quarter is the highest Order Return in the US?

# Data Preparation

Dataset:

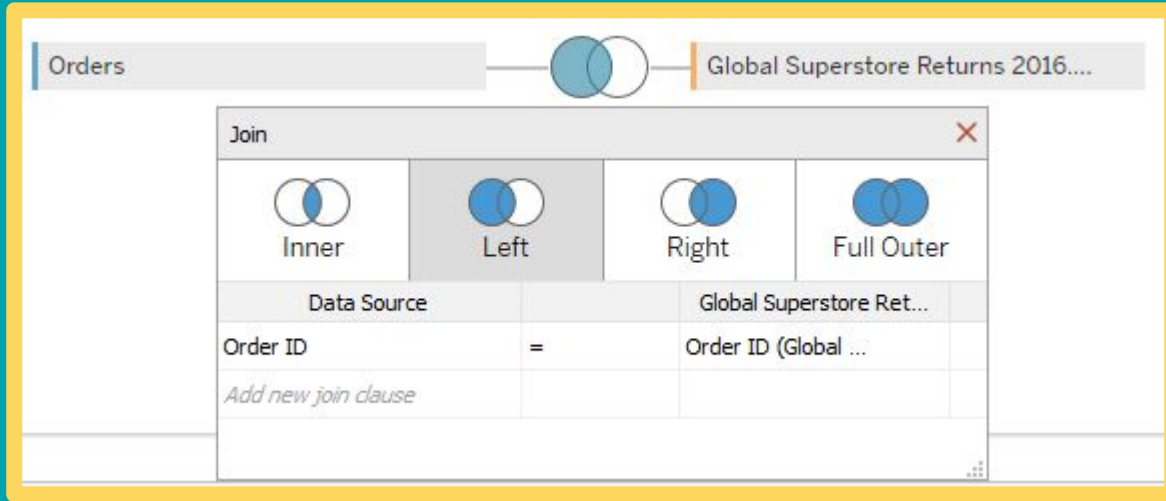
1. Global Superstore Orders records in 2012 - 2015
2. Global Superstore Returns record in 2012 - 2015



# Data Preparation

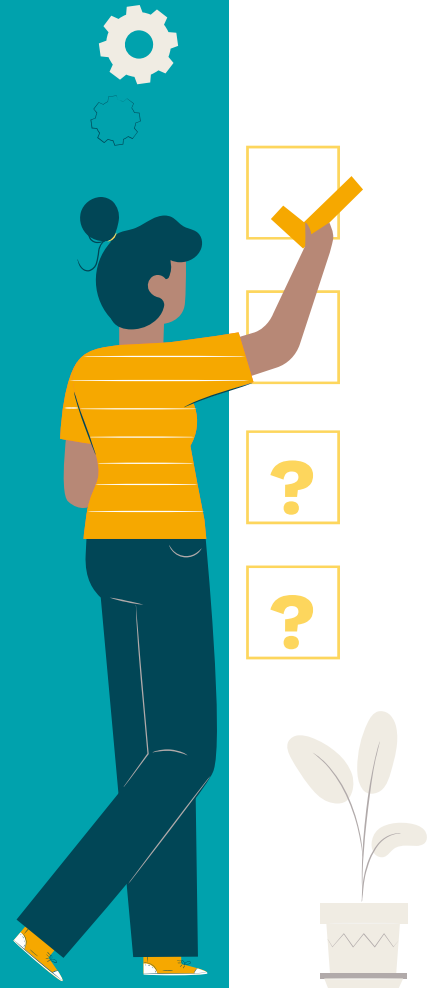
Connection Dataset:

Choose Left Join, and connect the key column using Order ID on both tables.



The screenshot shows a data tool interface with a join configuration window. The window is titled "Join" and has a close button (X). It displays two data sources: "Orders" and "Global Superstore Returns 2016....". The "Left" join type is selected, indicated by a blue circle and a white circle. The "Data Source" column is mapped to "Global Superstore Ret..." and the "Order ID" column is mapped to "Order ID (Global ...)".

Join Type	Data Source	Global Superstore Ret...
Inner		
Left		
Right		
Full Outer		
	Order ID	Order ID (Global ...)
	<i>Add new join clause</i>	



## Case 1

**Which Category and Ship Mode has the highest number of Order Returns?**

## Columns & Rows

- Add count distinct Order ID from Returns table to columns -
- Add Category and Ship Mode to rows



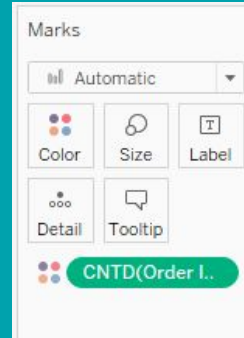
## Charts

Use horizontal bars chart for categorical series



## Marks

Add count distinct Order ID from Returns table to Color



# Result of Case 1



Office Supplies Category & Standard Class Ship Mode has a highest number of order returns.

## Case 2

**In what year did the number of Order Returns in July increase from June?**

## Columns & Rows

- Add Order date extracted by year & month to columns
- Add count distinct Order ID from Returns table to rows

Columns	YEAR(Order Date)	MONTH(Order Da..
Rows	CNTD(Order ID (Glob..	

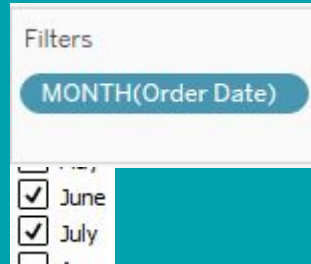
## Charts

Use line chart for time series



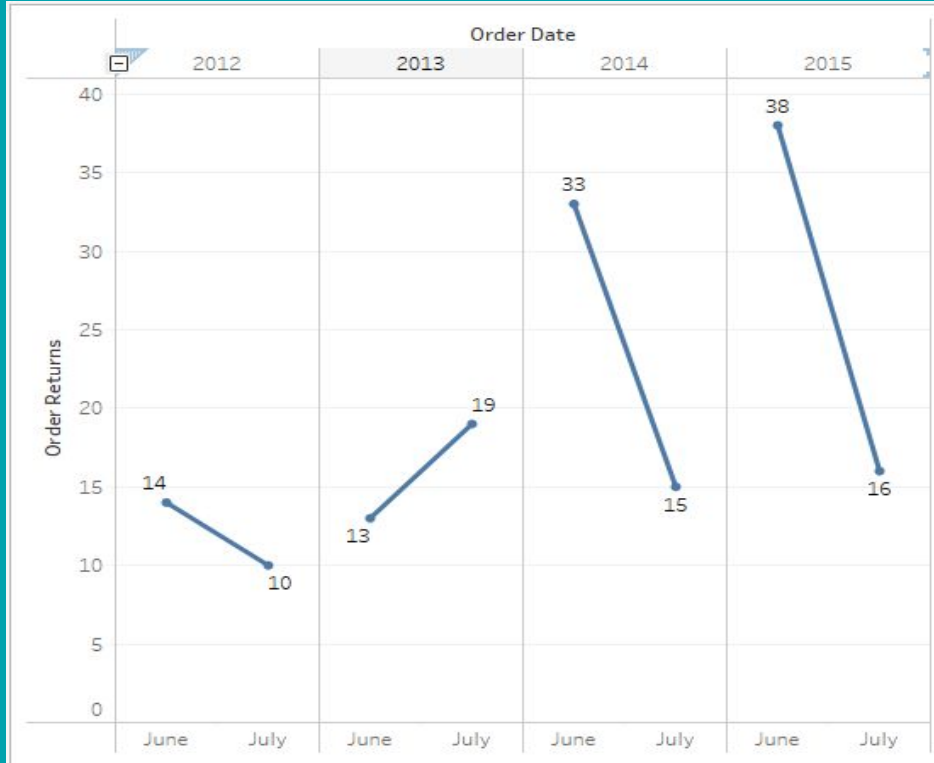
## Filters

Add filter by Month of Order Date,  
choose June and July





## Result of Case 2



The order return that increased from June to July is in 2013 only.

## Case 3

Find the percentage of Order Return from United States compared to total Order return in 2015, and what Quarter is the highest Order Return in the US?

## Columns & Rows

- Add Order date extracted by quarter to columns
- Add count distinct Order ID from Returns table to rows

Columns	QUARTER(Order Date)
Rows	CNTD(Order ID (Global))

## Charts

Use line chart for time series



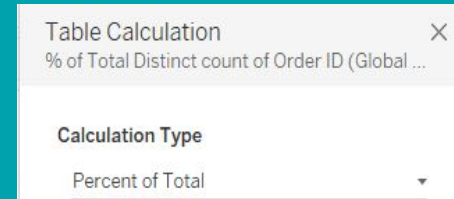
## Filters

- Add Year of Order Date, choose 2015
- Add Country, choose United States

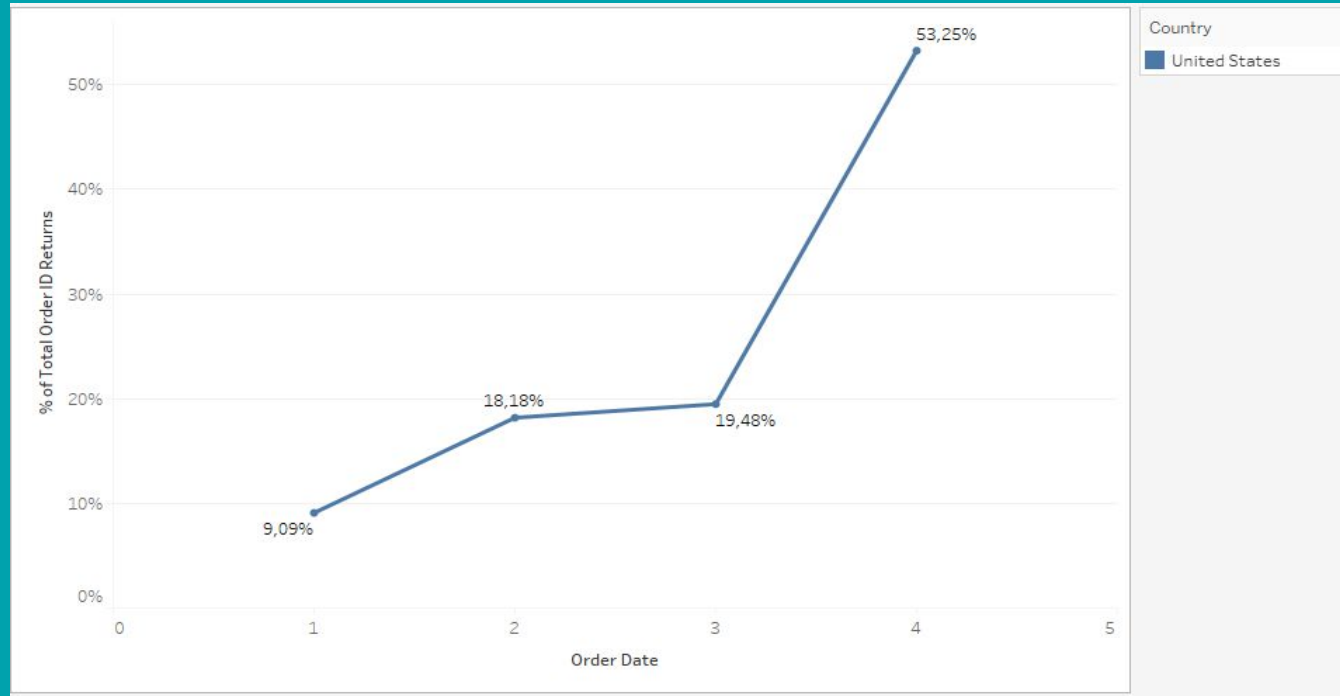


## Table Calculation

Click count distinct Order ID Returns in table rows, click quick table calculation choose Percent of Total.



# Result of Case 3



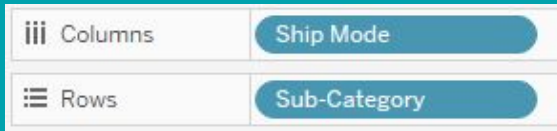
The highest order returns in USA is occurred in Q4 of 2015.

## **Additional Case - 1**

Create a Crosstab that displays the Sub Category and Ship Mode. Choose the following Sub-category and Shipmode field that has the highest number of Order Returns!

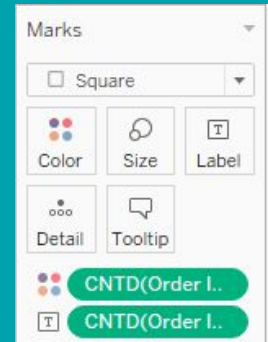
## Columns & Rows

- Add Ship Mode to columns
- Add Sub-Category to rows



## Marks

- Choose Square marks
- Add count distinct Order ID from Returns table to Color
- Choose Label & Tick the "Show mark labels"



# Result of Additional Case - 1

Sub-Category	Ship Mode			
	First Class	Same Day	Second Class	Standard Class
Accessories	15	7	18	89
Appliances	3	2	11	43
Art	25	10	31	121
Binders	33	12	50	144
Bookcases	19	3	15	62
Chairs	11	9	25	87
Copiers	17	6	12	59
Envelopes	15	7	11	58
Fasteners	9	3	19	67
Furnishings	15	5	14	90
Labels	28	4	26	69
Machines	15	3	11	30
Paper	15	10	25	82
Phones	21	2	27	83
Storage	24	14	37	111
Supplies	13	7	23	53
Tables	8	3	6	21

CNTD(Order ID (Global ...  
2144

Binders sub-category with Standard Class ship mode has the highest number of Order Returns.

## **Additional Case - 2**

**Make a Line Chart with Order Date and Order Return Amount, how many the Order Return Amount and Quantity in July 2013?**



## Columns & Rows

- Add Order date extracted by month to columns
- Add count distinct Order ID from Returns table and sum of Quantity to rows

Columns	MONTH(Order Da..
Rows	CNTD(Order ID (Glob.. SUM(Quantity)

## Charts

Use line chart for time series



## Filters

- Add Year of Order Date, choose 2013
- Add Returned, choose "Yes"

Filters
YEAR(Order Date): 2013
Returned: Yes

## Result of Additional Case - 2



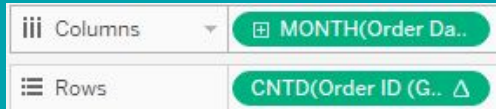
There are 19 Order Returns and 156 Quantity Returned on July 2013.

## **Additional Case - 3**

**What is the Total Accumulated Order Return throughout 2015?**

## Columns & Rows

- Add Order date extracted by quarter to columns
- Add count distinct Order ID from Returns table to rows



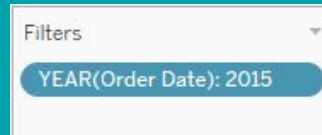
## Charts

Use line chart for time series



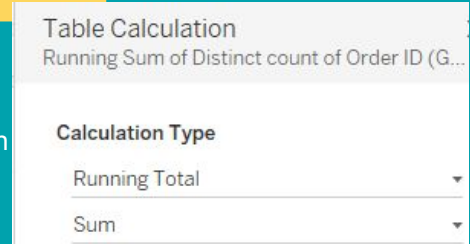
## Filters

- Add Year of Order Date, choose 2015

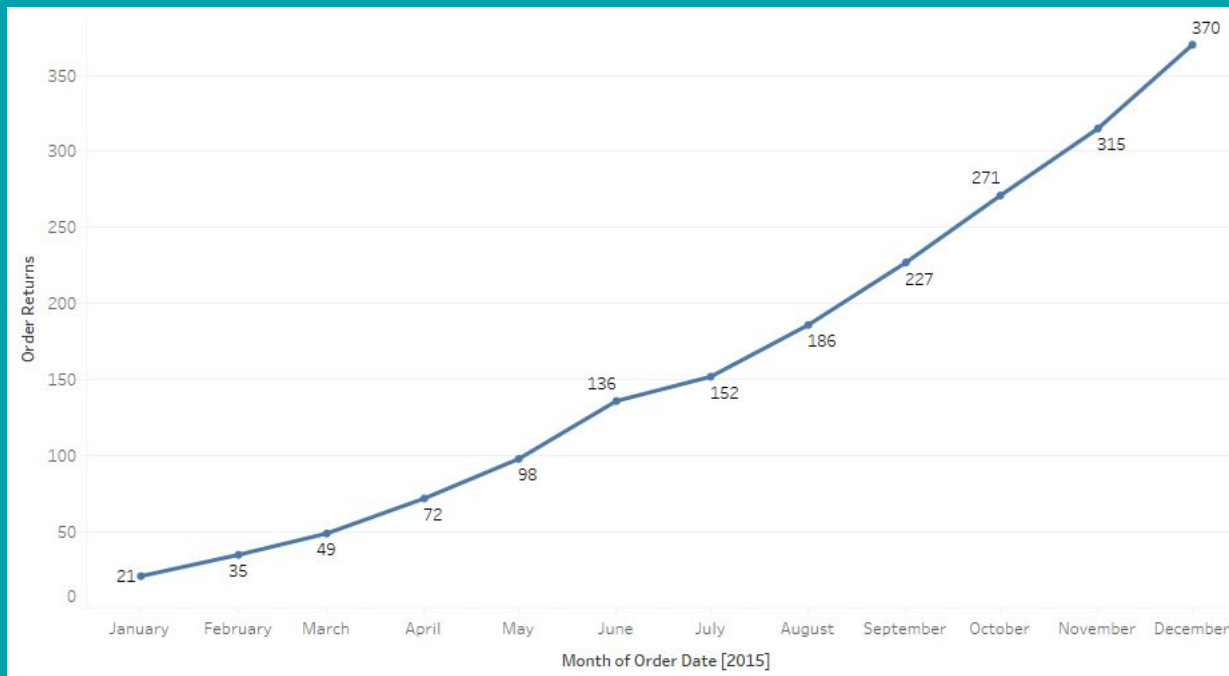


## Table Calculation

Click count distinct Order ID Returns in table rows, click quick table calculation choose Running Total.



## Result of Additional Case - 3



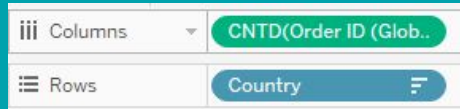
**Total Accumulated Order Return throughout 2015 is 370.**

## **Additional Case - 4**

**Make a Bar Chart that displays the Country and Order Return Amount. Show the top 5 Countries with the highest number of Returned orders in 2015!**

## Columns & Rows

- Add Country to rows
- Add count distinct Order ID from Returns table to columns



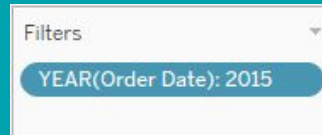
## Charts

Use horizontal bars chart for categorical series



## Filters

- Add Year of Order Date, choose 2015

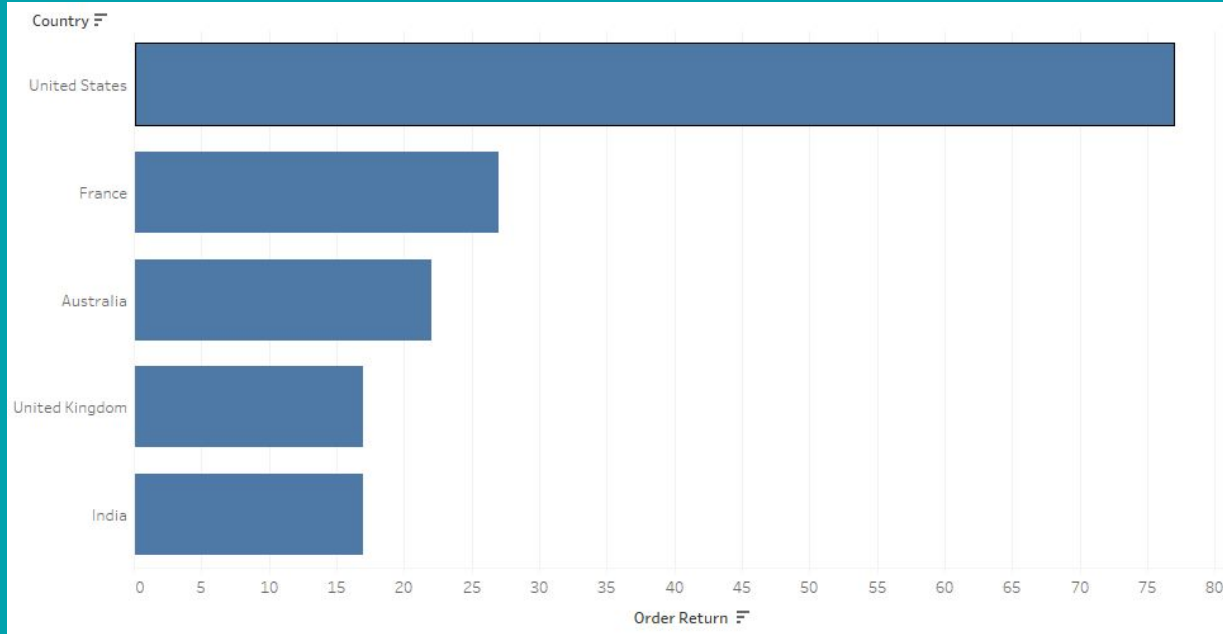


## Calculated Field

Create a calculated field named "Index" to help for filtering with Top function.



# Result of Additional Case - 4



Top 5 Countries with the highest number of return order in 2015.



## **Additional Case - 5**

**Create a Calculated Field that calculates Time to Ship (Order Date - Ship Date). Show the top 5 countries with the longest average Time to Ship in 2015!**

## Columns & Rows

- Add Country to rows
- Add Add average of Time to Ship to columns

Columns	AVG(time_to_ship)
Rows	Country

## Charts

Use horizontal bars chart for categorical series



## Filters

- Add Year of Order Date, choose 2015
- Add Returned, choose "Yes"
- Add Index, set the range of value from 1 to 5

Filters

YEAR(Order Date): 2015

Returned: Yes

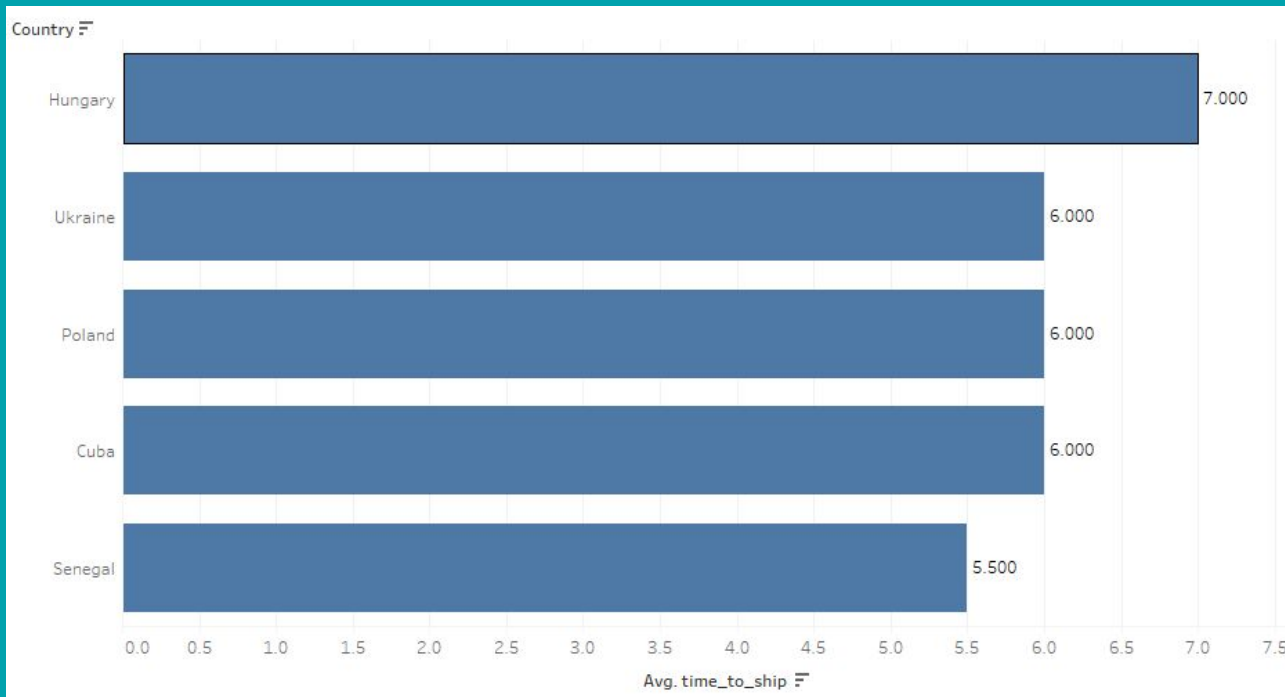
Index

## Calculated Field

Create a calculated field named "Time to Ship", use DATEDIFF() function.

time_to_ship
<code>DATEDIFF('day', [Order Date], [Ship Date])</code>

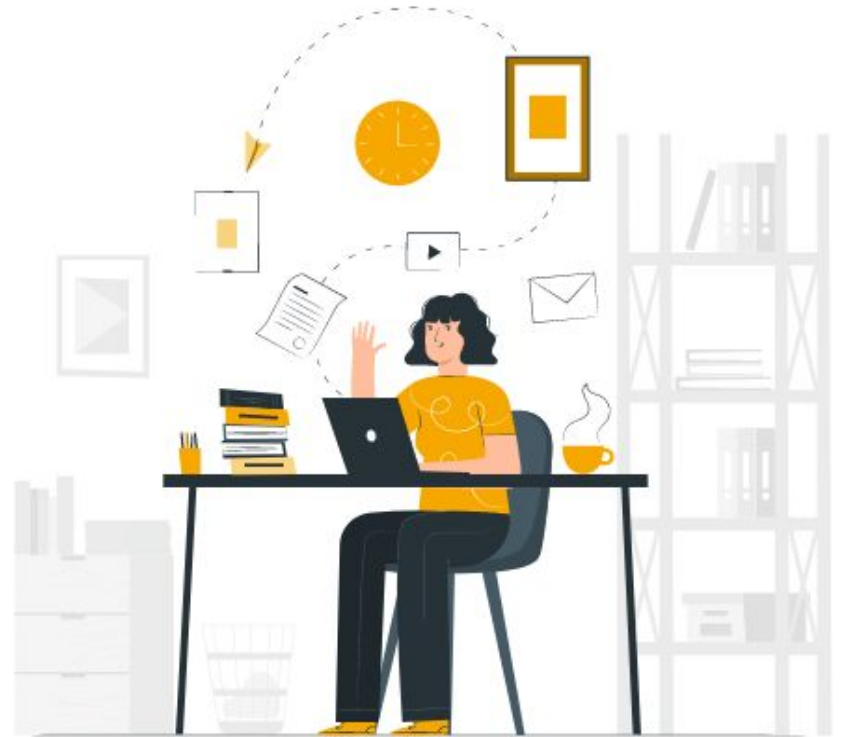
# Result of Additional Case - 5



Top 5 Countries with the longest average time to ship in 2015.

# Tableau Part 2

## —Case Study—



## Problem:

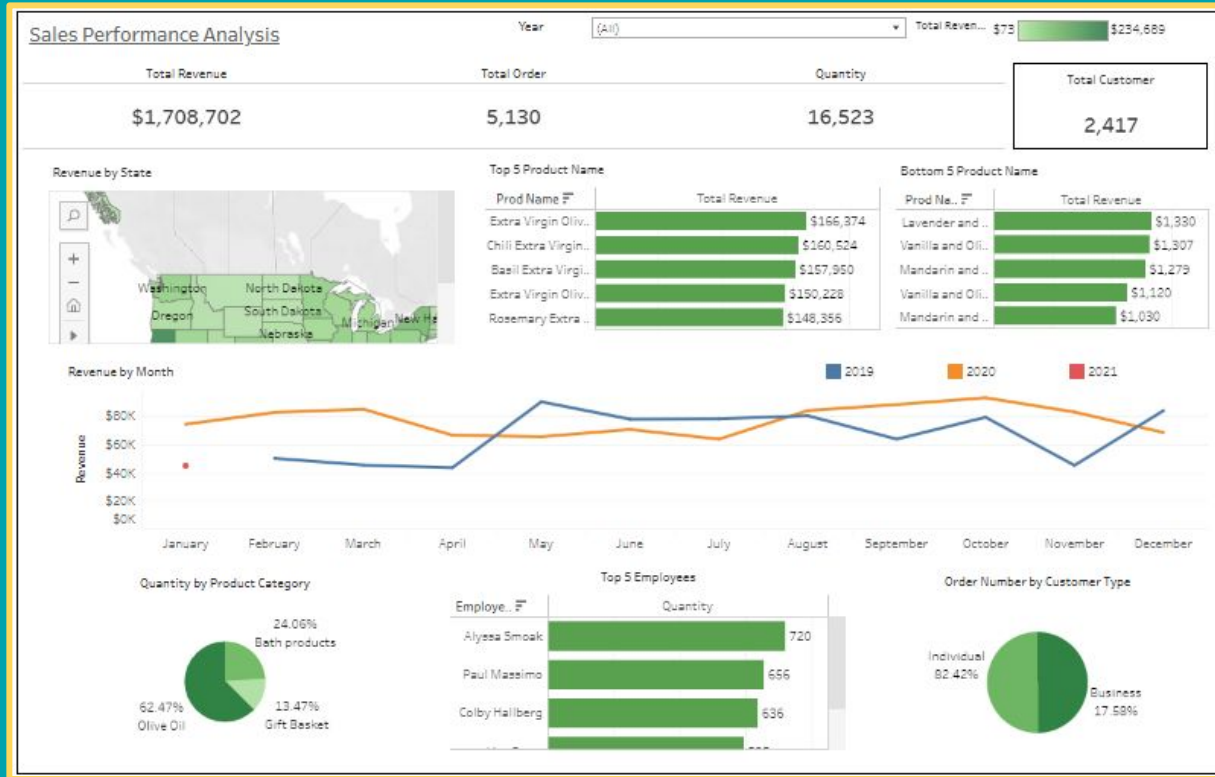
You will do an analysis on data with 3 main views: Employees, Customers and Products. Generate your unique analytics dashboard.

Determine the following things before create dashboards:

1. Determine your goals and audience
2. Create a simple mock-up
3. Create and do testing
4. Make sure your dashboard is interactive
5. Apply best practices

**Dataset contains sales records in January 2019 - January 2021. It has 17 fields and 5130 rows.**

# Result Dashboard



[https://public.tableau.com/app/profile/mnashir/viz/BI1\\_16747242942100/Dashboard1?publish=yes](https://public.tableau.com/app/profile/mnashir/viz/BI1_16747242942100/Dashboard1?publish=yes)

# Thanks!

