



# DATA CO SUPPLYCHAIN REPORT

Data Analytics – SQL, Power BI

# SUMMARY

- Data Cleaning
  - ❖ Normalization the data using:
    - I. 1NF – Removing the Duplicates, Delimited Separated Columns
    - II. 2NF – Separating the Partial Dependence Data
    - III. 3NF – Separating the Transitive Dependence Data
  - ❖ Removing the duplicated Columns
  - ❖ Removing the unwanted Columns
  - ❖ Changing the data types in column in SQL DB
- Data Modelling
  - ❖ Using Star Schema in Power BI.
  - ❖ Changing the data types for columns in power BI
- Data Visualization
- Root Cause Analysis

# DATA CLEANING - NORMALIZATION

- Using 1NF, 2NF, 3NF: Removing the duplicates and creating the fact table.
  - Created the Separated Tables for all below terms:
    1. Category Table: Category\_ID, Category\_Name
      - ✓ Deleted the Category\_Name column in fact table.
    2. Customer Info Table: Customer\_ID, Customer\_City, Customer\_Country, ....
      - ✓ Removed all the customer related columns except customer\_ID in fact table.
    3. Department Info Table: Department\_ID, Department\_Name
      - ✓ Removed the Department name column in fact table.
    4. Customer Location Info Table: Customer\_ID, Latitude, Longitude.
      - ✓ Removed the Latitude, Longitude columns in fact table.
    5. Product Info Table: Product\_ID, Product\_Category\_id, Product\_Name, .....,
      - ✓ Removed all the product related columns except Product\_id and Product\_Category\_ID in fact table.
    6. Shipment Info Table: Order\_ID, Shipping\_Date, Shipping\_Mode, .....,
      - ✓ Removed all the shipping relation columns except the order\_id.
    7. Sales Info Table: Order\_ID, Order\_Date, Order\_Country, Order\_item\_discount, .....,
      - ✓ Removed all the shipping relation columns except the order\_id.

# DATA CLEANING – AFTER NORMALIZATION LOOKS LIKE

Category Table

	Category_Id	Category_Name
1	73	Sporting Goods
2	17	Cleats
3	29	Shop By Sport

Customer Table

	Customer_Id	Customer_City	Customer_Country	Customer_Email	Customer_Fname	Customer_Lname	Customer_Password	Customer_Segment	Customer_State	Customer_Street
1	7193	Caguas	Puerto Rico	XXXXXXXX	Jennifer	Smith	XXXXXXXX	Corporate	PR	1440 Fallen Fox
2	4255	Caguas	Puerto Rico	XXXXXXXX	Brittany	Smith	XXXXXXXX	Consumer	PR	7420 Crystal View

**Fact Table** – After removing all the quantitative values from main table and keeping only Foreign keys of all the dimension tables (primary key) is called fact table. It will be useful in data modelling

Department Table

	Department_id	Department_name
1	7	Fan Shop
2	11	Pet Shop
3	3	Footwear

	Category_Id	Customer_Id	Department_Id	Order_Customer_Id	Order_Id	Order_Item_Id	Product_Card_Id	Product_Category_Id
1	73	20755	2	20755	77202	180517	1360	73
2	73	19492	2	19492	75939	179254	1360	73
3	73	19491	2	19491	75938	179253	1360	73
4	73	19490	2	19490	75937	179252	1360	73

Sales Table

	Order_ID	Days_for_shipping	Days_for_shipment	Delivery_Status	Late_delivery_risk	shipping_date	Shipping_Mode
1	21850	6	4	Late delivery	1	11/21/2015 22:40	Standard Class
2	43511	2	4	Advance shipping	0	9/29/2016 3:29	Standard Class
3	30036	2	4	Advance shipping	0	3/16/2016 10:35	Standard Class
4	57522	3	4	Shipping canceled	0	4/22/2017 16:10	Standard Class
5	40126	2	4	Advance shipping	0	08-10-2016 17:34	Standard Class

Customer Location Table

	Customer_ID	Latitude
1	11255	38
2	4199	37

Product Table

	Product_Card_id	Product_Category_ID	Product_Description	Product_Image	Product_Name	Product_Price
1	1073	48	NULL	http://images.acmesports.sports/Pelican+Sunstream+1...	Pelican Sunstream 100 Kayak	199
2	127	7	NULL	http://images.acmesports.sports/Stiga+Master+Series+...	Stiga Master Series ST3100 Competition Indoor	329
3	134	7	NULL	http://images.acmesports.sports/Nike+Women%27s+L...	Nike Women's Legend V-Neck T-Shirt	25
4	1346	50	NULL	http://images.acmesports.sports/Speed+Boats	Speed Boats	21

# DATA CLEANING – DUPLICATES REMOVING

Removing the columns and data for following:

1. Removing Duplicate columns
2. Removing NULL Columns.
3. Removing the unwanted Columns.
4. Removing the invalid Columns.



# DATA CLEANING – DUPLICATES REMOVING LOOK LIKE

## Duplicated Columns

Order_ID	Benefit_per_order	Order_Profit_Per_Order
16213	29	29
12020	76	76
8872	74	74
54031	97	97
54389	34	34
52958	21	21
65930	96	96
53242	39	39
21460	20	20
33746	13	13
63795	78	78
10028	14	14
63904	34	34
66798	82	82
654	-30	-30
39289	-193	-193
20111	46	46
17138	73	73
56908	23	23

	Sales_Per_Customer	Order_Item_Total
1	124	124
2	245	245
3	176	176
4	287	287
5	165	165
6	149	149
7	284	284
8	134	134
9	227	227
10	129	129
11	278	278
12	44	44
13	272	272
14	230	230
15	127	127
16	110	110
17	132	132

## Invalid Columns

	Customer_ID	Customer_Email	Customer_Password
1	7193	XXXXXXXXXX	XXXXXXXXXX
2	4255	XXXXXXXXXX	XXXXXXXXXX
3	3387	XXXXXXXXXX	XXXXXXXXXX
4	8085	XXXXXXXXXX	XXXXXXXXXX
5	5409	XXXXXXXXXX	XXXXXXXXXX
6	7935	XXXXXXXXXX	XXXXXXXXXX
7	8023	XXXXXXXXXX	XXXXXXXXXX
8	3879	XXXXXXXXXX	XXXXXXXXXX
9	7273	XXXXXXXXXX	XXXXXXXXXX
10	16322	XXXXXXXXXX	XXXXXXXXXX
11	4651	XXXXXXXXXX	XXXXXXXXXX
12	19281	XXXXXXXXXX	XXXXXXXXXX
13	4103	XXXXXXXXXX	XXXXXXXXXX
14	3554	XXXXXXXXXX	XXXXXXXXXX
15	64	XXXXXXXXXX	XXXXXXXXXX
16	8204	XXXXXXXXXX	XXXXXXXXXX

## Unwanted/Null Columns

Order_ID	Order_Zipcode
16213	NULL
12020	NULL
8872	NULL
54031	NULL
54389	NULL
52958	NULL
65930	NULL
53242	NULL
21460	NULL
33746	95123
63795	NULL
10028	NULL
63904	NULL
66798	NULL
654	NULL
39289	33433
20111	NULL
17138	NULL

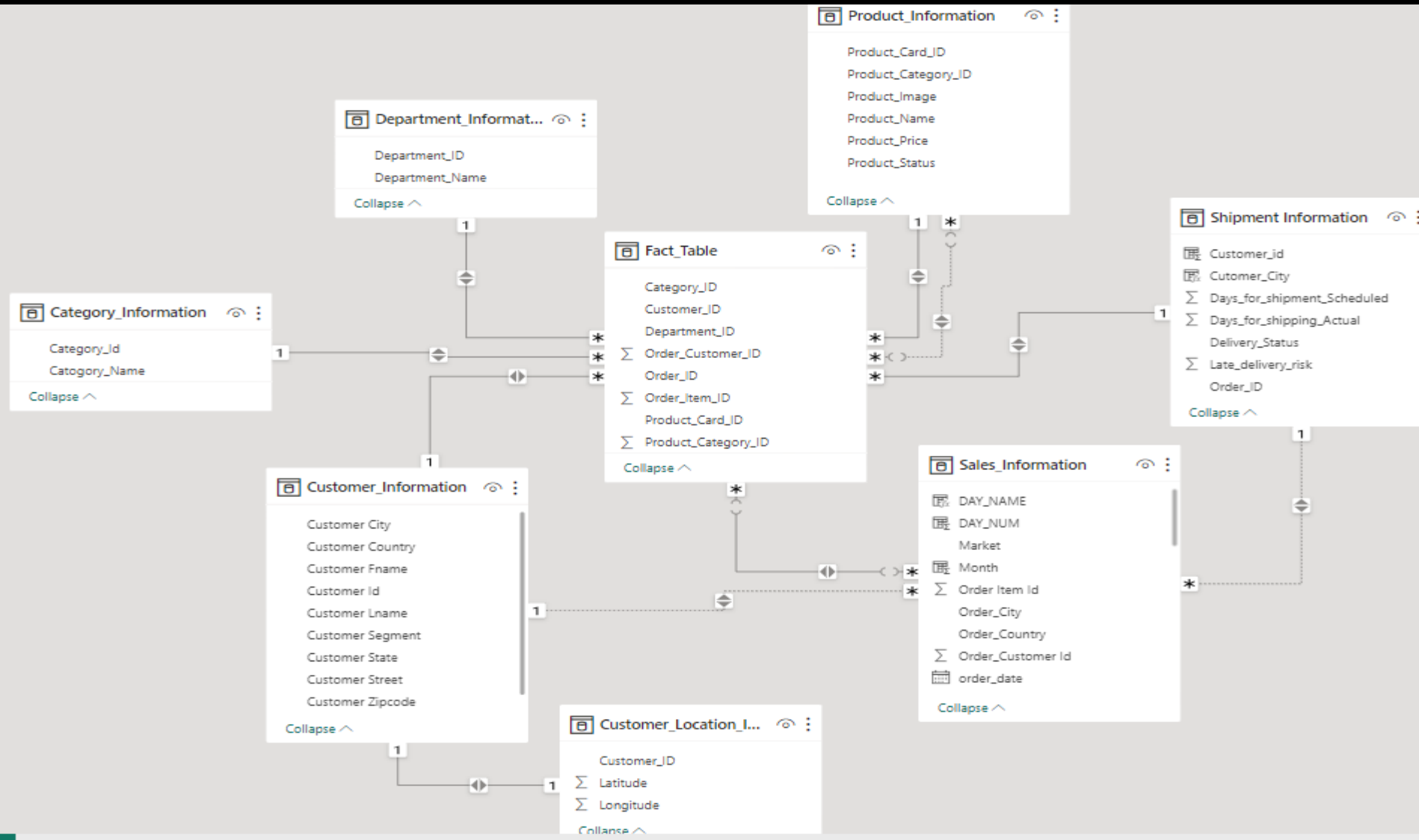
We have to cross verify the duplicated columns using three aggregation to find the columns provide the same value, if every aggregation provides same value then we can delete the duplicated columns

Total_Count_Benefit	Total_Count_Profit	Total_Benefit	Total_Profit	AVG_Benefit	AVG_Profit
180519	180519	3912583	3912583	21	21

	Total_Count_Sales	Total_Count_Order	Total_Sales	Total_Orders
1	180519	180519	32928052	32928052

# DATA MODELLING

**Star Schema** – Connecting all the dimension tables into Fact tables when it looks like star shaped formation. This Schema will provide the interactive report from all the tables. It will link all the tables and provide appropriate result in visualizations.



# CATEGORY OF VISUALIZATION

I have analysis for the below topics

1. Sales Overview
2. Customer Analysis
3. Shipment Analysis
4. Product Analysis
5. Sales Analysis
6. Time Series Analysis
7. Root Cause Analysis

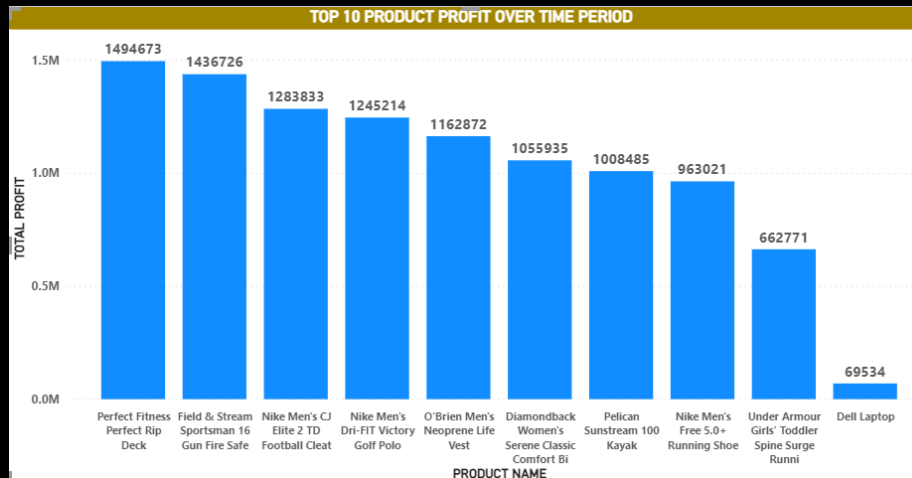
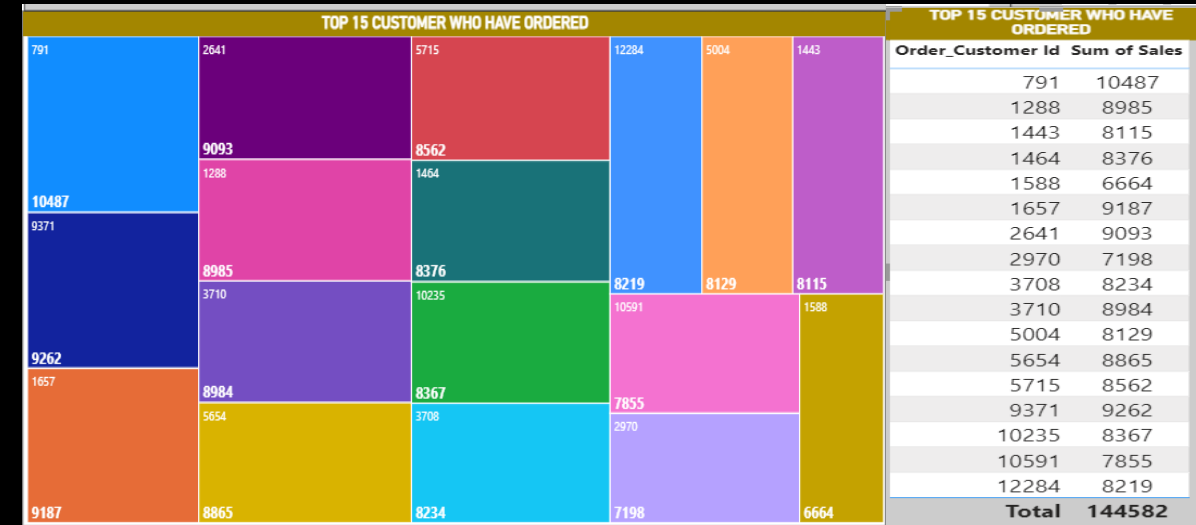
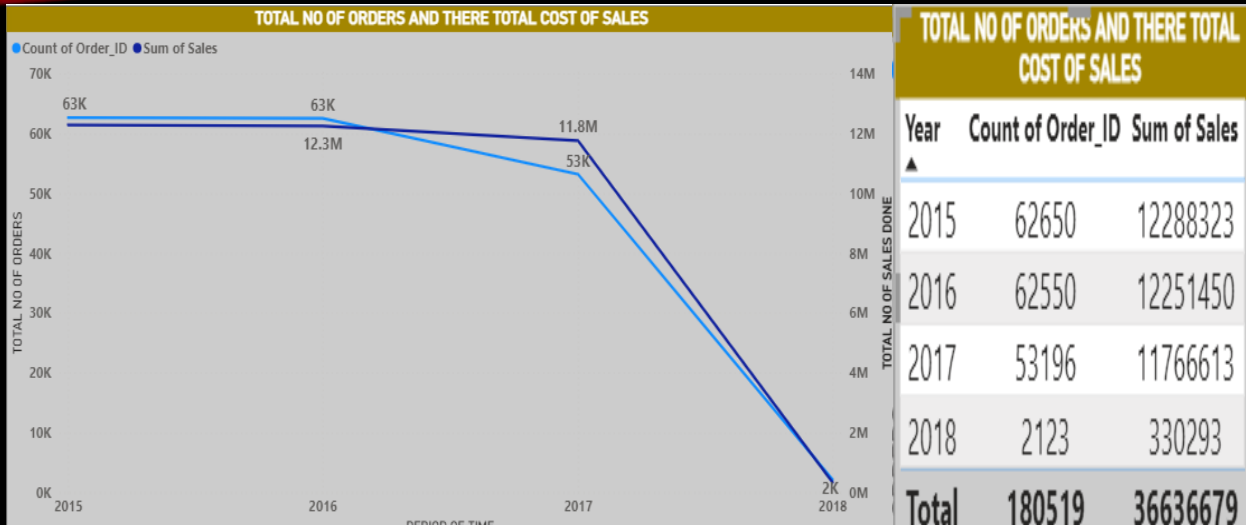


# SALES OVERVIEW

## The Insights for Sales over view:

- Total cost of Sales entirely over 4 years : \$36.63 Million.
- Total cost of Profits entirely over 4 years : \$3.91 Million.
- Total No of Order quantities over 4 years : 180,519.
- Total No of unique Orders over 4 years : 65,752.
- No of Country were received orders : 164.
- No of states were received orders : 1089.
- No of cities were received orders : 3597.

# SALES OVERVIEW VISUALIZATION



**TOP 10 PRODUCT PROFIT OVER TIME PERIOD**

Product_Name	Sum of Order_Profit_Per_Order
Perfect Fitness Perfect Rip Deck	1494673
Field & Stream Sportsman 16 Gun Fire Safe	1436726
Nike Men's CJ Elite 2 TD Football Cleat	1283833
Nike Men's Dri-FIT Victory Golf Polo	1245214
O'Brien Men's Neoprene Life Vest	1162872
Diamondback Women's Serene Classic Comfort Bi	1055935
Pelican Sunstream 100 Kayak	1008485
Nike Men's Free 5.0+ Running Shoe	963021
Under Armour Girls' Toddler Spine Surge Runni	662771
Dell Laptop	69534
<b>Total</b>	<b>3695829</b>

# CUSTOMER ANALYSIS INSIGHTS

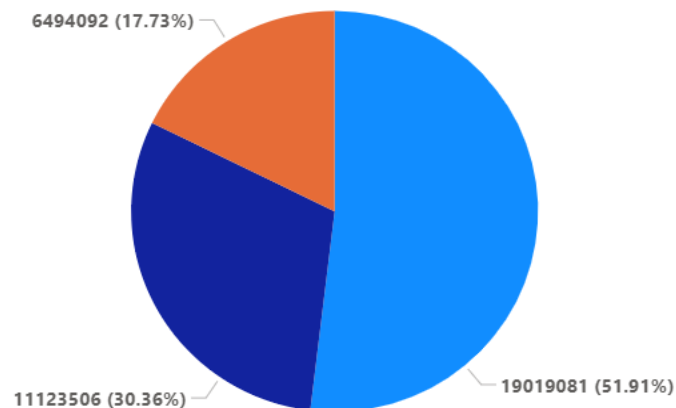
## The Insights for Customer Analysis:

- Total No of Customer: 20,652.
- Total No of County the customer stated: 2.
- Total No of State the customer stated: 46.
- Total No of Cities the customer stated: 563.
- Total No of Customer Segments: 3.

# CUSTOMER ANALYSIS VISUALIZATION

TOTAL SALES BY CUSTOMER SEGMENT

Customer Segment ● Consumer ● Corporate ● Home Office

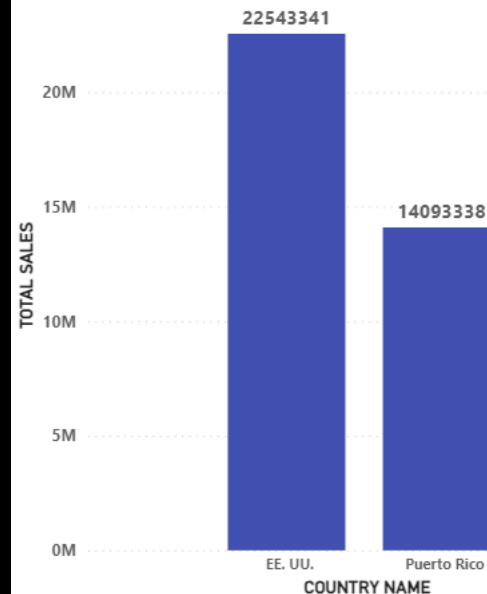


TOTAL SALES BY CUSTOMER SEGMENT

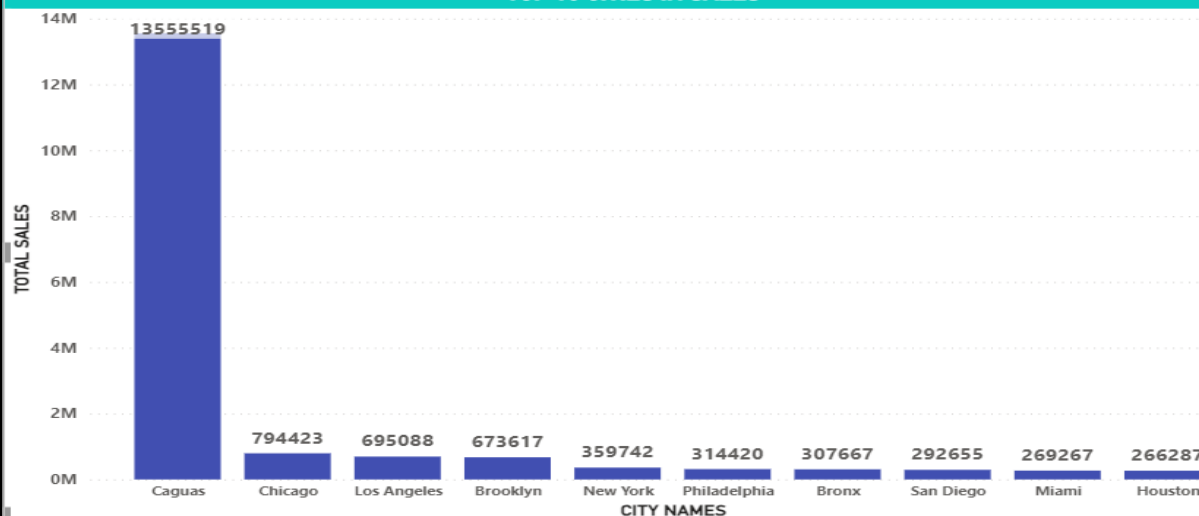
Customer Segment Sum of Sales

Consumer	19019081
Corporate	11123506
Home Office	6494092
Total	36636679

TOP COUNTRY IN SALES



TOP 10 CITIES IN SALES



TOP 10 CITIES IN SALES

Customer City Sum of Sales

Caguas	13555519
Chicago	794423
Los Angeles	695088
Brooklyn	673617
New York	359742
Philadelphia	314420
Bronx	307667
San Diego	292655
Miami	269267
Houston	266287
Total	17528685

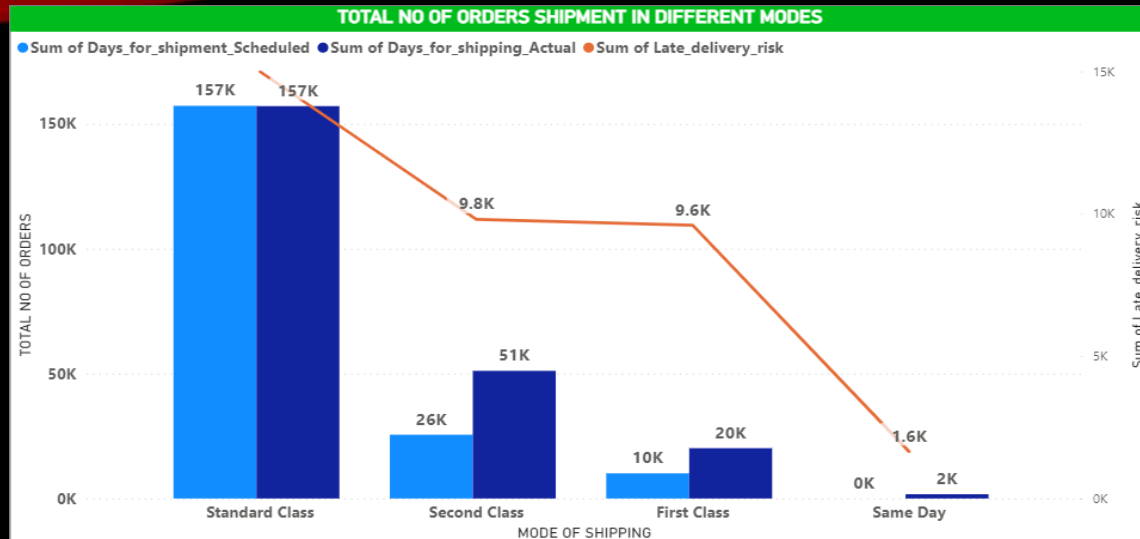
# SHIPMENT ANALYSIS INSIGHTS

## The Insights for Shipment Analysis:

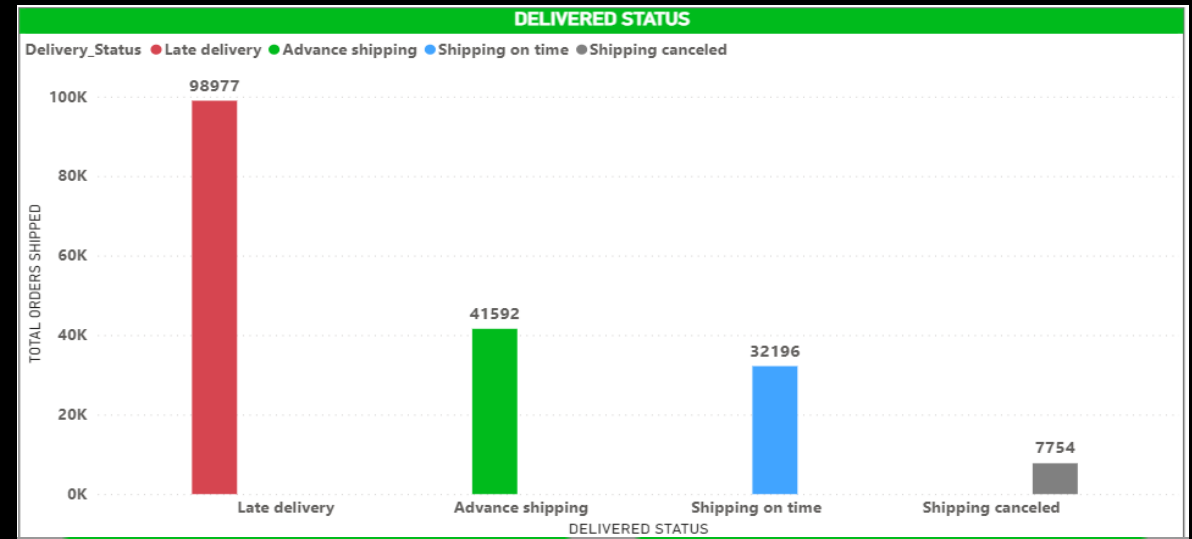
- Total no of Shipment Delivered: 65,752.
- Total no of Shipment Delivered Late: 36,048.
- Total no of Shipment Mode: 4
- Total no of Delivery Status: 4



# SHIPMENT ANALYSIS VISUALIZATION



TOTAL NO OF ORDERS SHIPMENT IN DIFFERENT MODES			
Shipping_Mode	Sum of Days_for_shipment_Scheduled	Sum of Days_for_shipping_Actual	Sum of Late_delivery_risk
Standard Class	157296	157174	14995
Second Class	25556	51132	9803
First Class	10079	20158	9602
Same Day	0	1727	1648
<b>Total</b>	<b>192931</b>	<b>230191</b>	<b>36048</b>



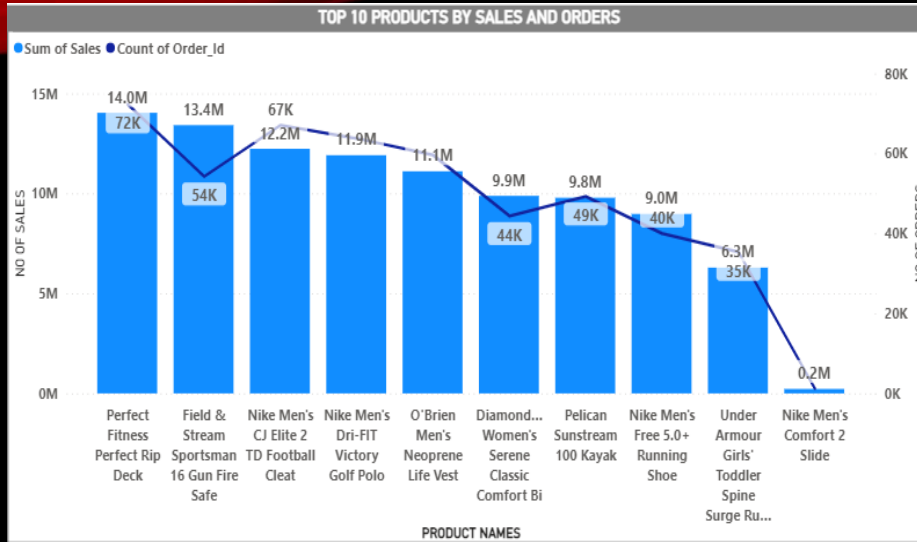
DELIVERED STATUS	
Delivery_Status	Count of Order_ID
Advance shipping	41592
Late delivery	98977
Shipping canceled	7754
Shipping on time	32196
<b>Total</b>	<b>180719</b>

# PRODUCT ANALYSIS INSIGHTS

## The Insights for Product Analysis:

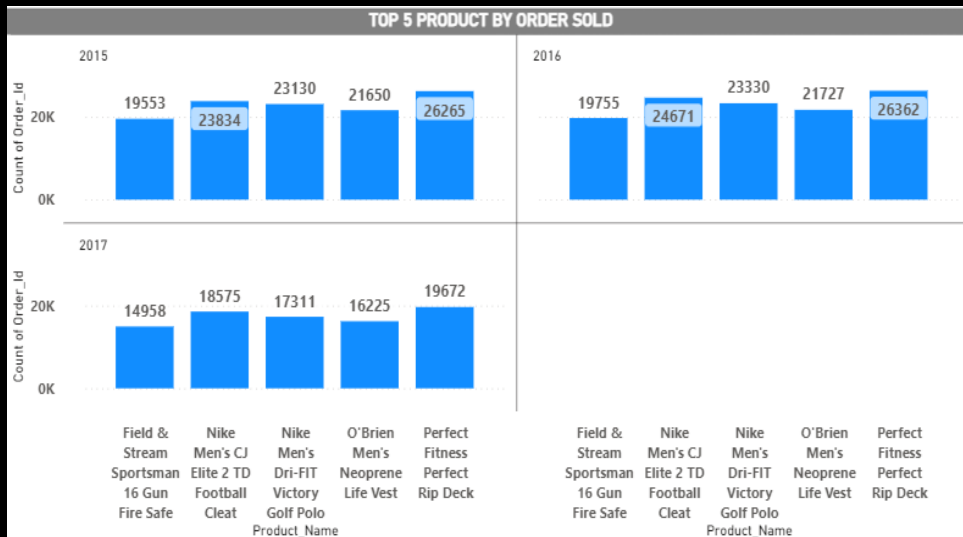
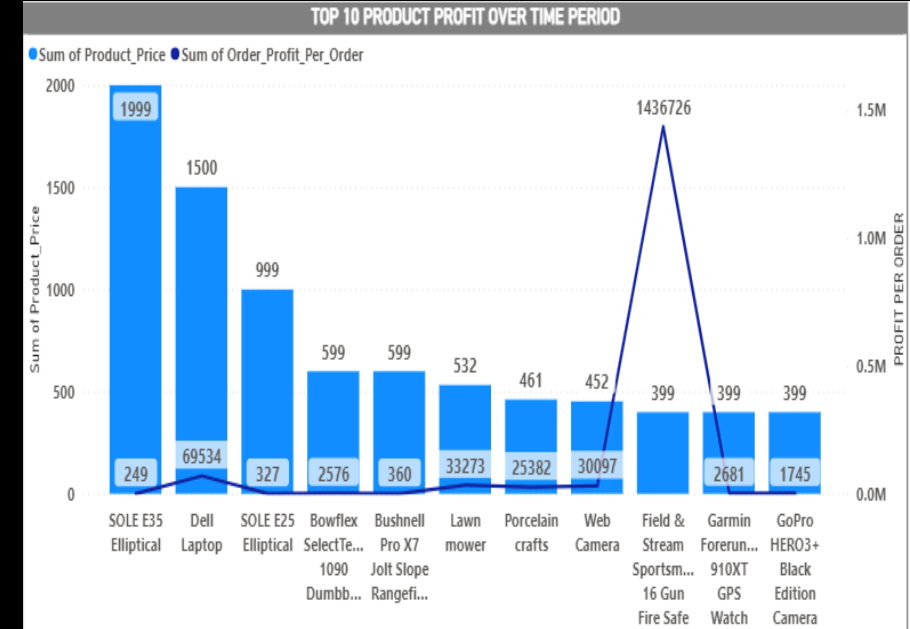
- Total no of Product Category: 51.
- Total no of Product Items Sold: 118.
- Total Product Prices per unit cost: \$19,545.
- Total no of product status: 1.

# PRODUCT ANALYSIS VISUALIZATION



**TOP 10 PRODUCTS BY SALES AND ORDERS**

Product_Name	Sum of Sales	Count of Order_Id
Perfect Fitness Perfect Rip Deck	14034634	72299
Field & Stream Sportsman 16 Gun Fire Safe	13418380	54266
Nike Men's CJ Elite 2 TD Football Cleat	12234076	67080
Nike Men's Dri-FIT Victory Golf Polo	11915012	63771
O'Brien Men's Neoprene Life Vest	11110346	59602
Diamondback Women's Serene Classic Comfort Bi	9882964	44363
Pelican Sunstream 100 Kayak	9787355	49272
Nike Men's Free 5.0+ Running Shoe	8977609	39992
Under Armour Girls' Toddler Spine Surge Runni	6290344	35409
Nike Men's Comfort 2 Slide	221707	1228
Total	34038617	170965



**TOP 5 PRODUCT BY ORDER SOLD**

Product_Name	Count of Order_Id	Year
Perfect Fitness Perfect Rip Deck	26362	2016
Perfect Fitness Perfect Rip Deck	26265	2015
Nike Men's CJ Elite 2 TD Football Cleat	24671	2016
Nike Men's CJ Elite 2 TD Football Cleat	23834	2015
Nike Men's Dri-FIT Victory Golf Polo	23330	2016
Nike Men's Dri-FIT Victory Golf Polo	23130	2015
O'Brien Men's Neoprene Life Vest	21727	2016
O'Brien Men's Neoprene Life Vest	21650	2015
Field & Stream Sportsman 16 Gun Fire Safe	19755	2016
Perfect Fitness Perfect Rip Deck	19672	2017
Field & Stream Sportsman 16 Gun Fire Safe	19553	2015
Nike Men's CJ Elite 2 TD Football Cleat	18575	2017
Nike Men's Dri-FIT Victory Golf Polo	17311	2017
O'Brien Men's Neoprene Life Vest	16225	2017
Field & Stream Sportsman 16 Gun Fire Safe	14958	2017
Total	160307	

**TOP 10 PRODUCT PROFIT OVER TIME PERIOD**

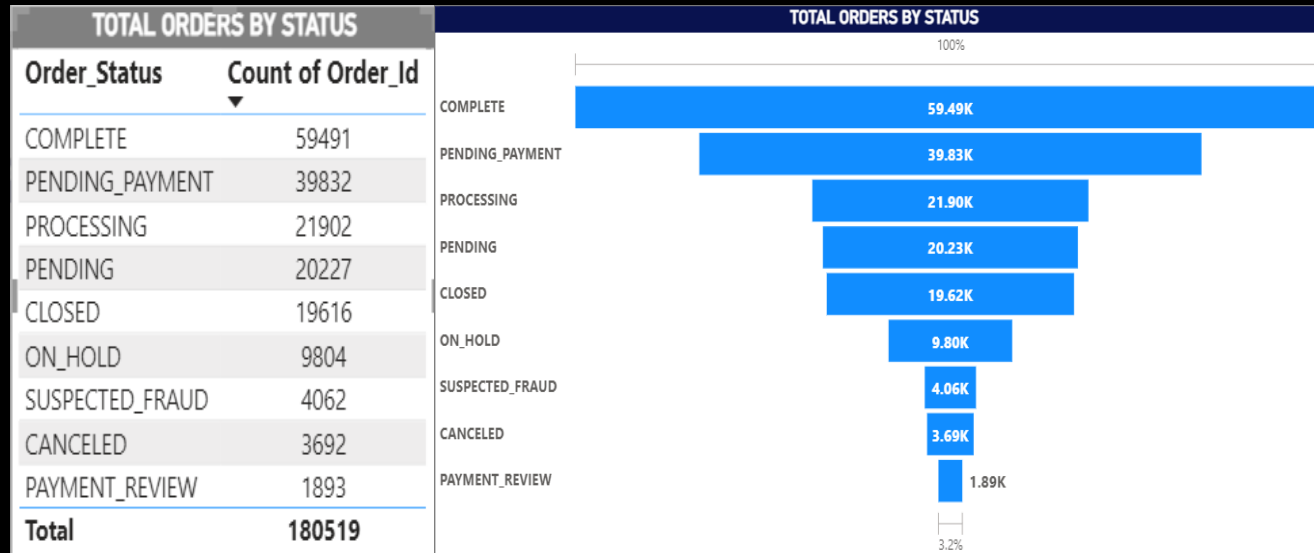
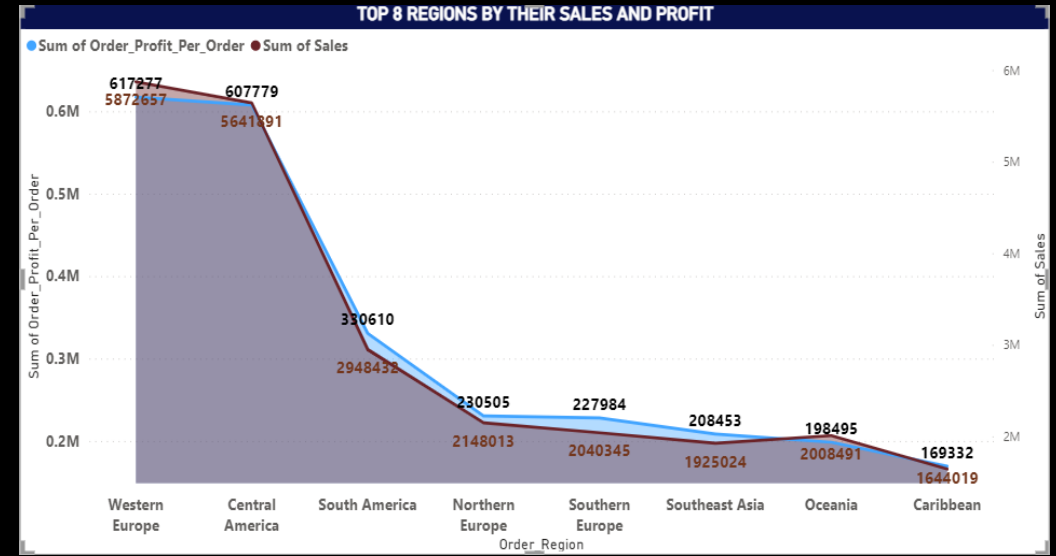
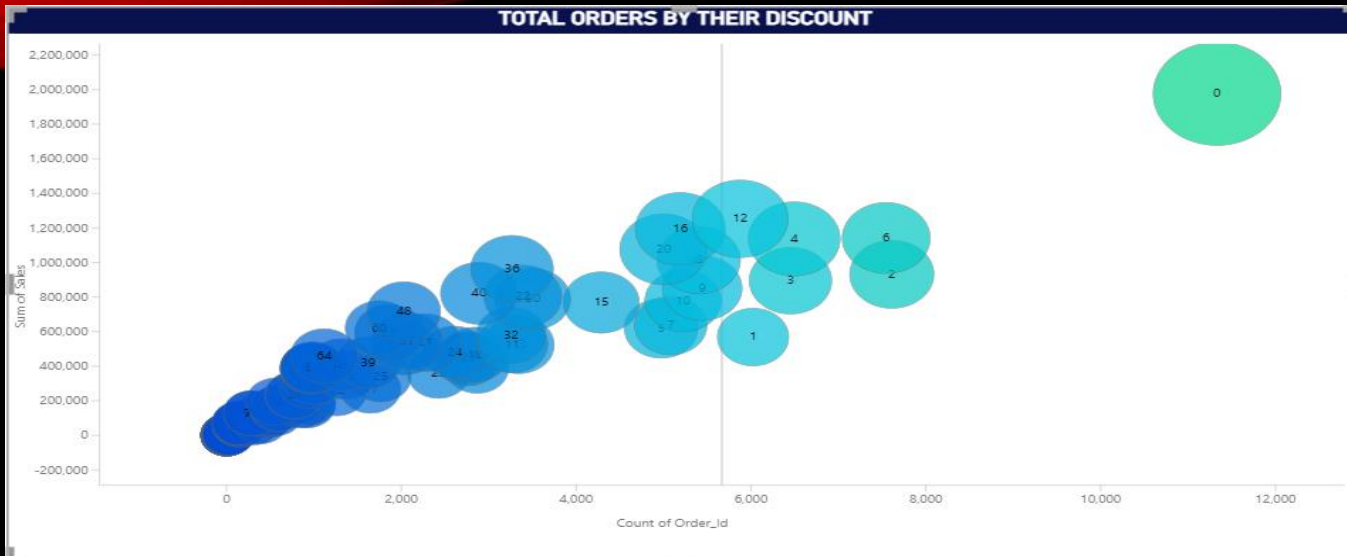
Product_Name	Sum of Product_Price	Sum of Order_Profit_Per_Order
SOLE E35 Elliptical	1999	249
Dell Laptop	1500	69534
SOLE E25 Elliptical	999	327
Bowflex SelectTech 1090 Dumbbells	599	2576
Bushnell Pro X7 Jolt Slope Rangefinder	599	
Lawn mower	532	
Porcelain crafts	461	
Web Camera	452	
Field & Stream Sportsman 16 Gun Fire Safe	399	
Garmin Forerunner 910XT GPS Watch	399	
GoPro HERO3+ Black Edition Camera	399	
Total	833	

# SALES ANALYSIS INSIGHTS

## The Insights for Sales Analysis:

- Total no of Department Integrated: 11.
- Total no of Department Positions: 50.

# SALES ANALYSIS VISUALIZATION



**TOP 8 REGIONS BY THEIR SALES AND PROFIT**

Order_Region	Sum of Sales	Sum of Order_Profit_Per_Order
Western Europe	5872657	617277
Central America	5641891	607779
South America	2948432	330610
Northern Europe	2148013	230505
Southern Europe	2040345	227984
Southeast Asia	1925024	208453
Oceania	2008491	198495
Caribbean	1644019	169332
<b>Total</b>	<b>24228872</b>	

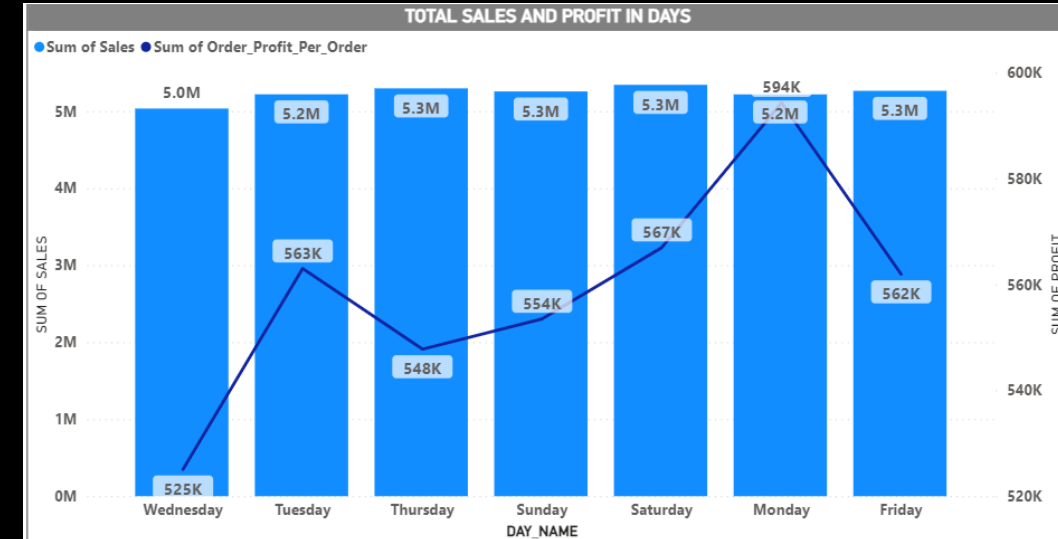
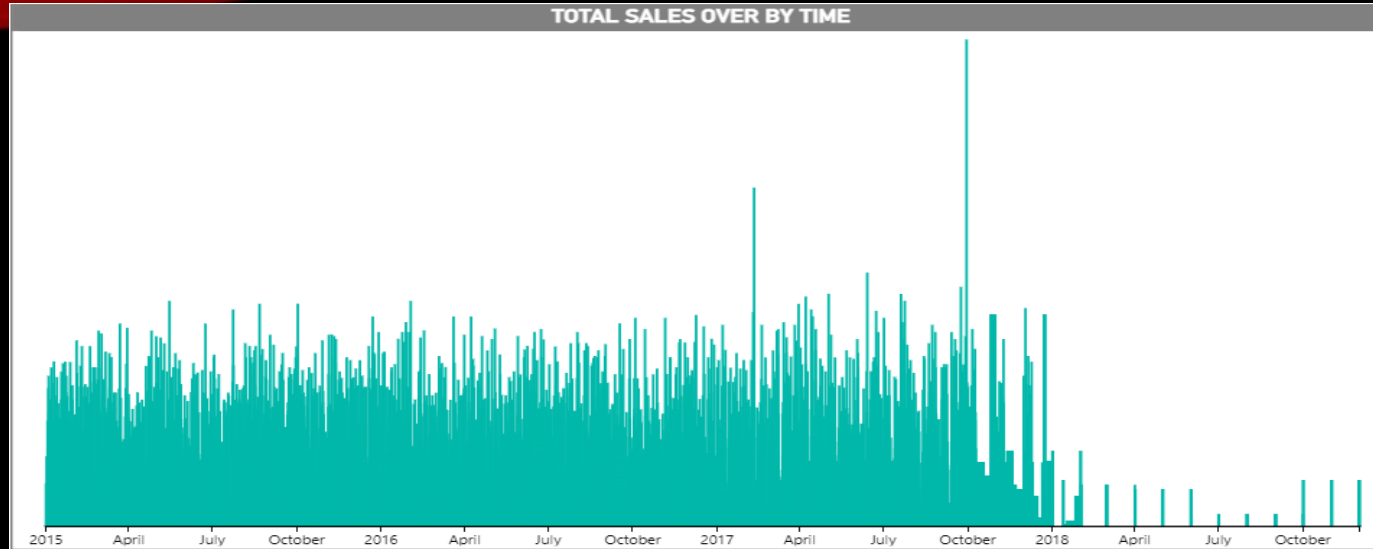


# TIME SERIES ANALYSIS INSIGHTS

## The Insights for Time Series Analysis:

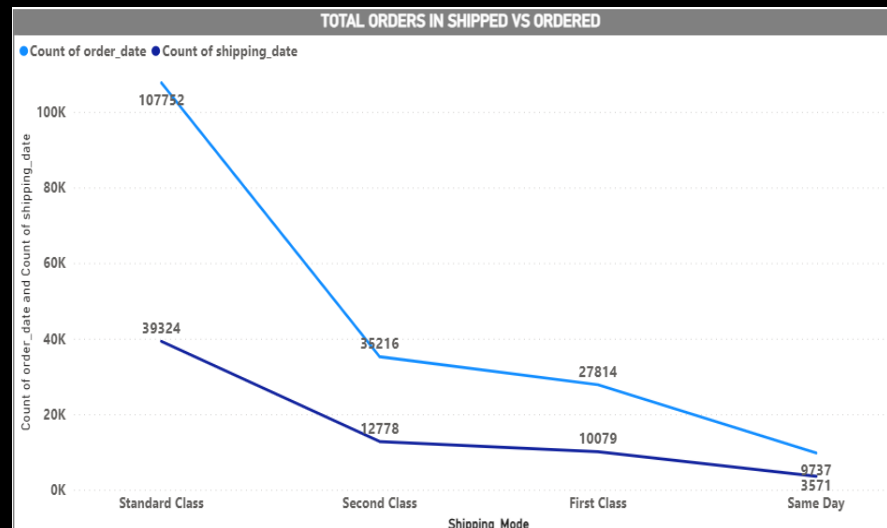
- Total No of Ordered Days: 65752.
- Total No of Business Years: 4.
- Total No of Business Months: 48.
- Total No of Market Occupied: 5.

# TIME SERIES ANALYSIS VISUALIZATION



**TOTAL ORDERS IN SHIPPED VS ORDERED**

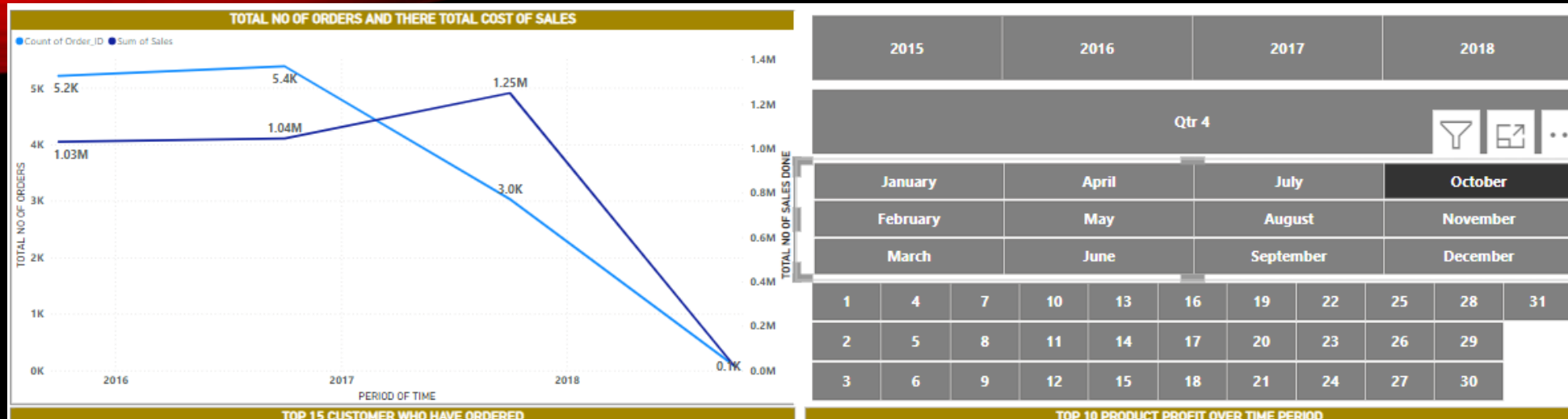
Shipping_Mode	Count of order_date	Count of shipping_date
Standard Class	107752	39324
Second Class	35216	12778
First Class	27814	10079
Same Day	9737	3571
Total	180519	65752



**TOTAL SALES AND PROFIT IN DAYS**

DAY_NAME	Sum of Order_Profit_Per_Order	Sum of Sales
Wednesday	525066	5035147
Tuesday	563014	5220001
Thursday	547781	5297104
Sunday	553506	
Saturday	566990	
Monday	594302	
Friday	561924	
Total	3912583	

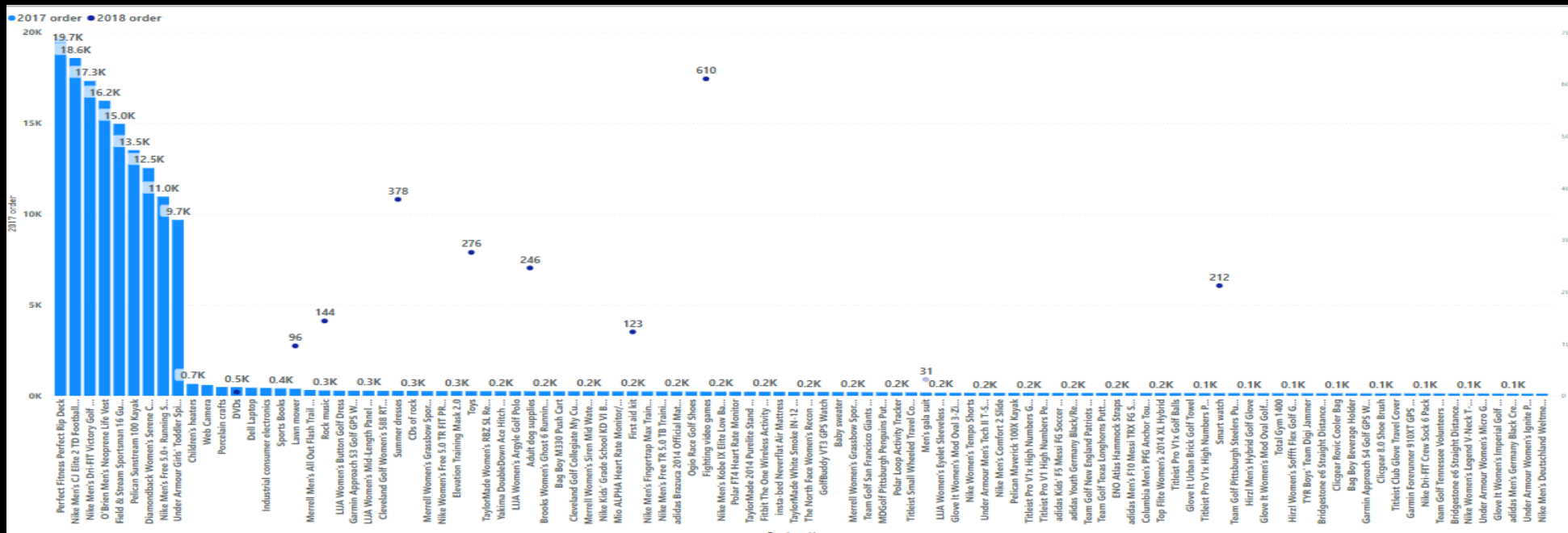
# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD



## Problem 1:

Negative Impact: We need to analysis: what just happen in the October Month of orders happen after 2017. – Sales Overview Report.

1<sup>st</sup> WHY: Why the October Month order reduced on 2018.



## Answer:

Nearly Top and best selling 12 ordered products are missing on 2018. So we have to analysis Vendor feedback on the analysis in vendor dataset.

# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD

## Problem 1: 2<sup>nd</sup> WHY

Negative Impact: We need to analysis: Why Top and best selling 12 ordered products are missing on 2018. – Sales Overview Report.

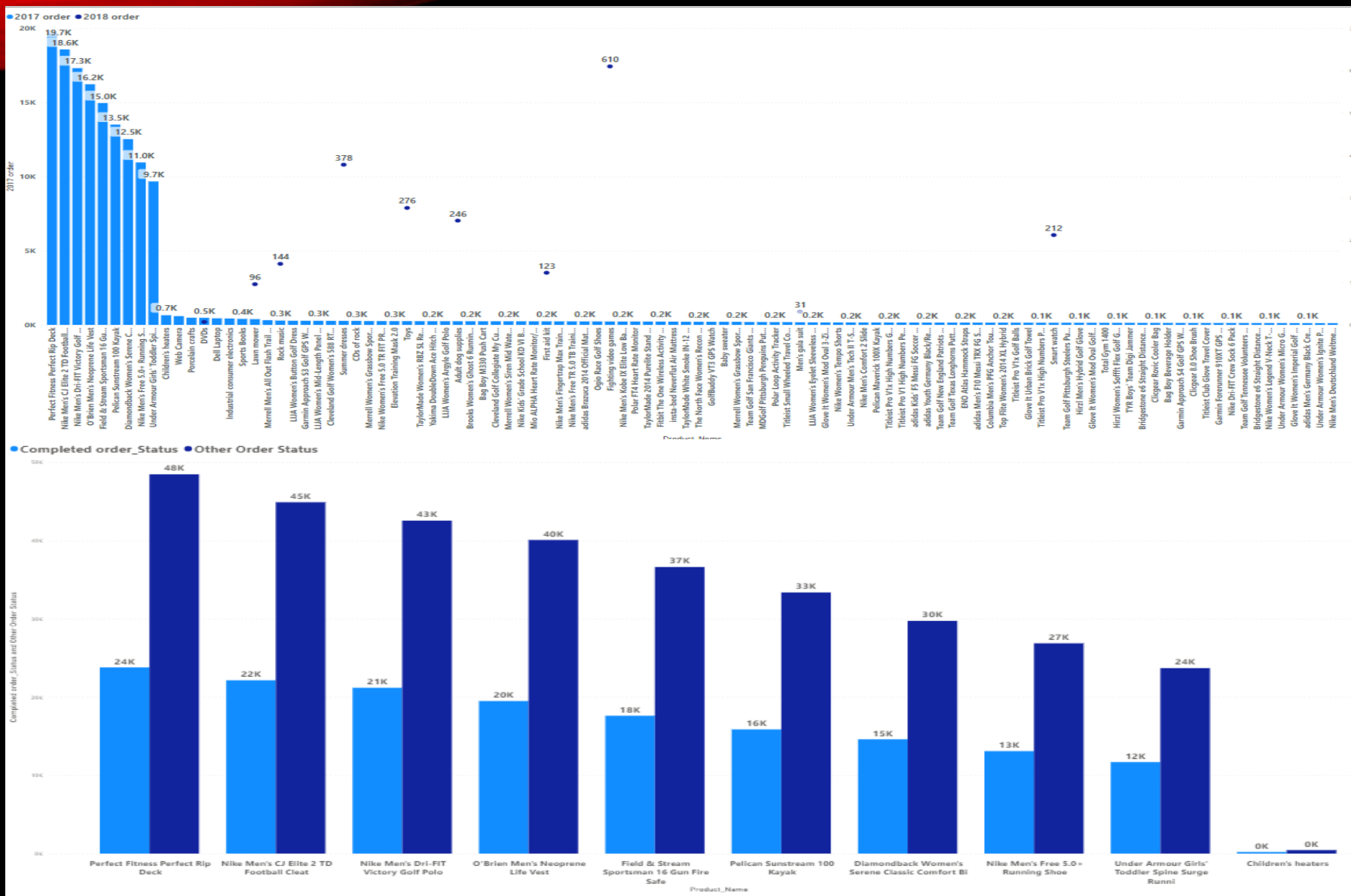
2<sup>nd</sup> WHY: Why all the top products are missing from 2018 Sales.

## Answer:

As we can able to see the Other Status(Pending, On-Hold, Processing, Pending, Closed Payments are twice of completed payment. So the payment system problem leads the major impact on 2018 sales.

3<sup>rd</sup> WHY : why the payment processing have issues.

Que: We need to ask to payment gateway team to analysis. What is the real cause.





# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD

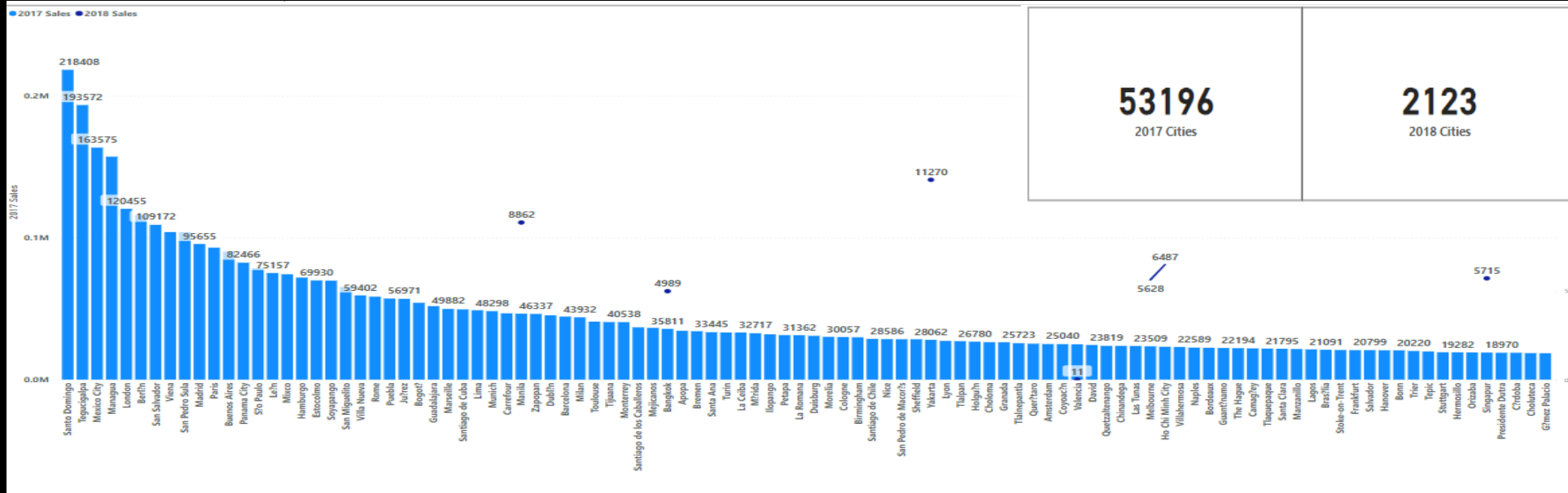


## Problem 2:

Negative Impact:

We need to analysis, what happen to other cities where sales happen before 2017 December. – Customer Analysis.

1<sup>st</sup> WHY: Why did the sales are dropped more cities than past year.

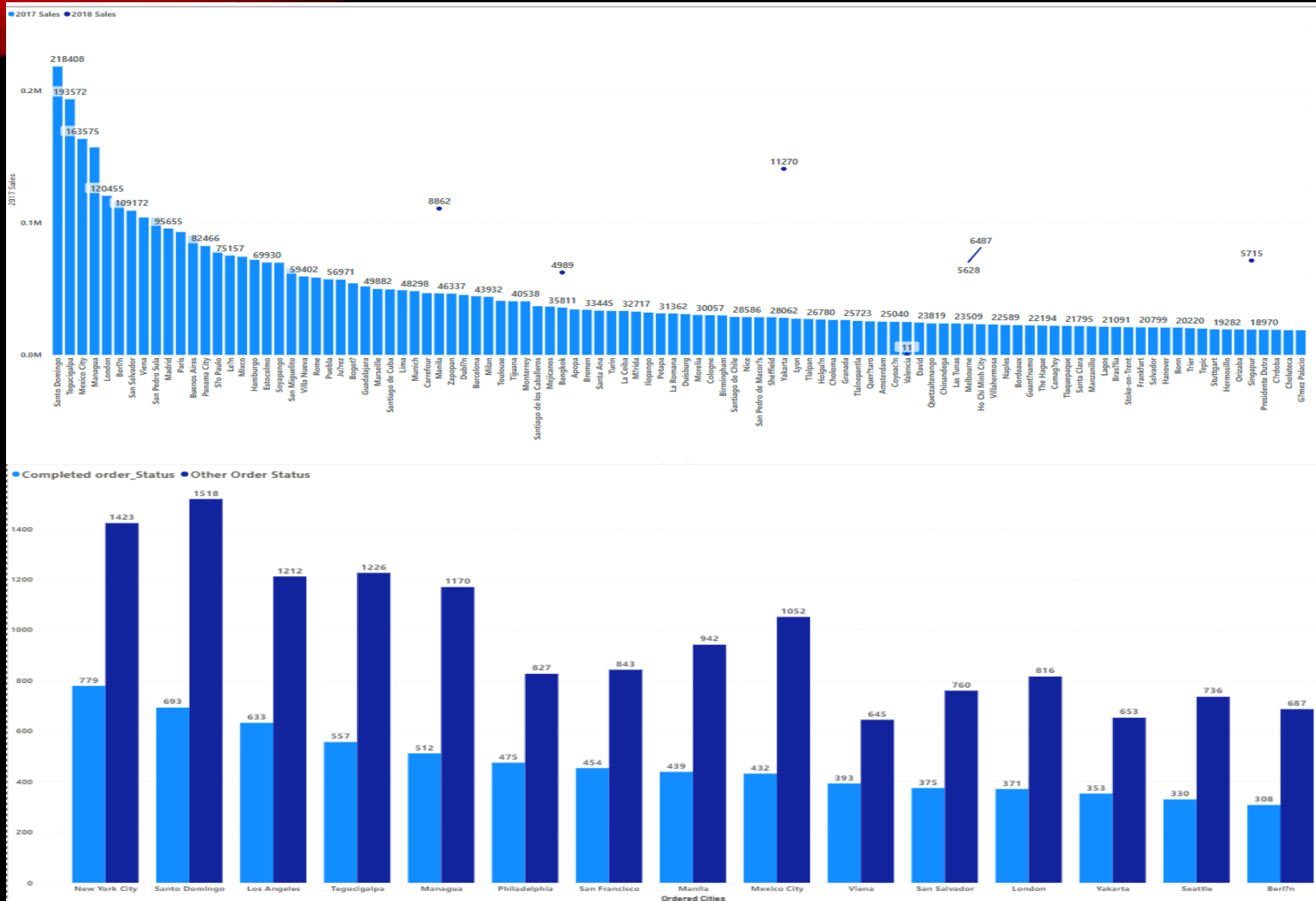


## Answer:

As I seen there is a TOP 31 Cities are dropped there Sales, So obviously the sales and profit dropped for the major share hold cities. And as I can able to see there is the 53196 sold cities are decreased to 2123, nearly 51,073 cities are stopped their sales.



# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD



## Problem 2: 2<sup>nd</sup> WHY

Negative Impact : Why the TOP 31 Cities are dropped there Sales, So obviously the sales and profit dropped for the major share hold cities. And as I can able to see there is the 53196 sold cities are decreased to 2123, nearly 51,073 cities are stopped their sales.

2<sup>nd</sup> WHY: Why all the top Selling Cities are dropped there sales in year 2018.

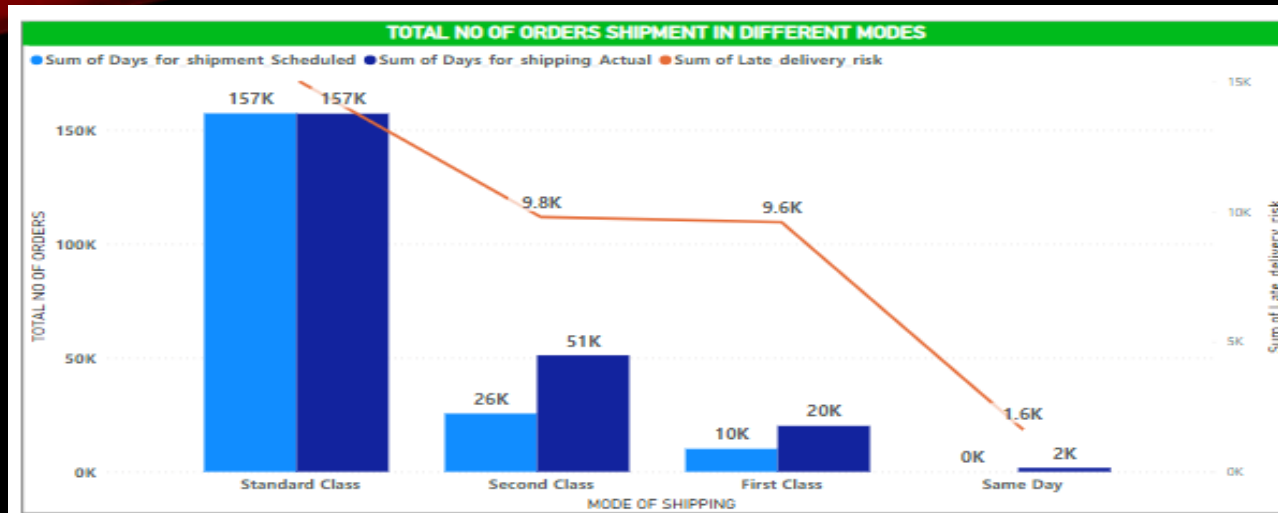
## Answer:

As we can able to see the Other Status(Pending, On-Hold, Processing, Pending, Closed Payments) are twice of completed payment in city wise. So the payment system problem leads the major impact on 2018 sales.

3<sup>rd</sup> WHY : why the payment processing have issues.

Que: We need to ask to payment gateway team to analysis. What is the real cause.

# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD



## Problem 3:

Negative Impact:

We need to analysis: we are not good in on time delivery for second class delivery orders. – Shipment Analysis.

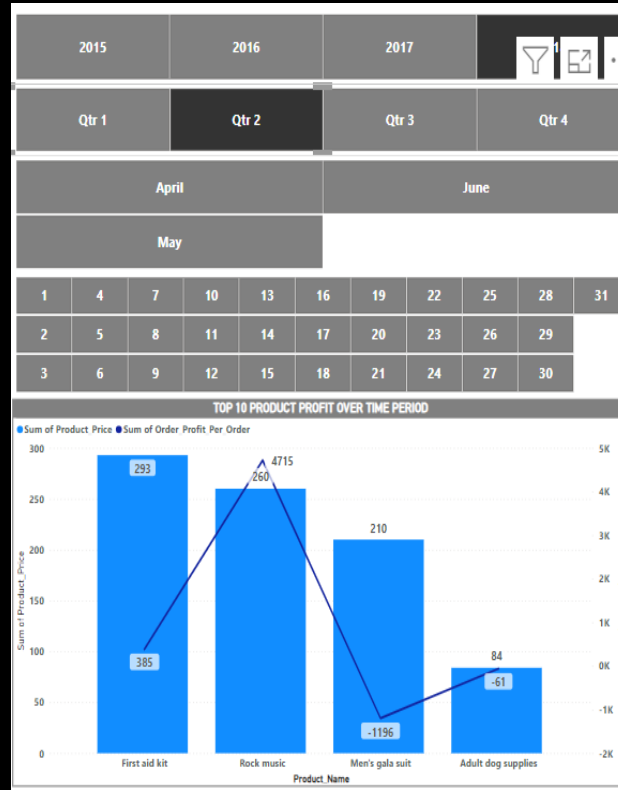
1<sup>st</sup> WHY: Why only 2<sup>nd</sup> and 1<sup>st</sup> class has a late delivery.

## Answer:

As we can able to see there is the high order received cities are the most Late delivered cities. So, the conclusions are we will not have any high impact on orders because the late delivery is usual on all high order received cities. But we have to focus on delivery system or need to change the delivery system to increase the Customer satisfaction.



# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD



## Problem 4: Negative Impact:

We can able to see there is TOP selling category products are gradually decreases on 2018 every quarter, it shows there is a vendor downgrade/Decreases or any other vendor related issue cause this major loss.

We need to analysis: Why the Top selling category products are gone down.

1<sup>st</sup> WHY: Why only TOP selling category products are gone down.

Answer : Again the Payment gateway method problem may caused the major impact on this product reduce.

# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD

## Problem 5:

Negative Impact:

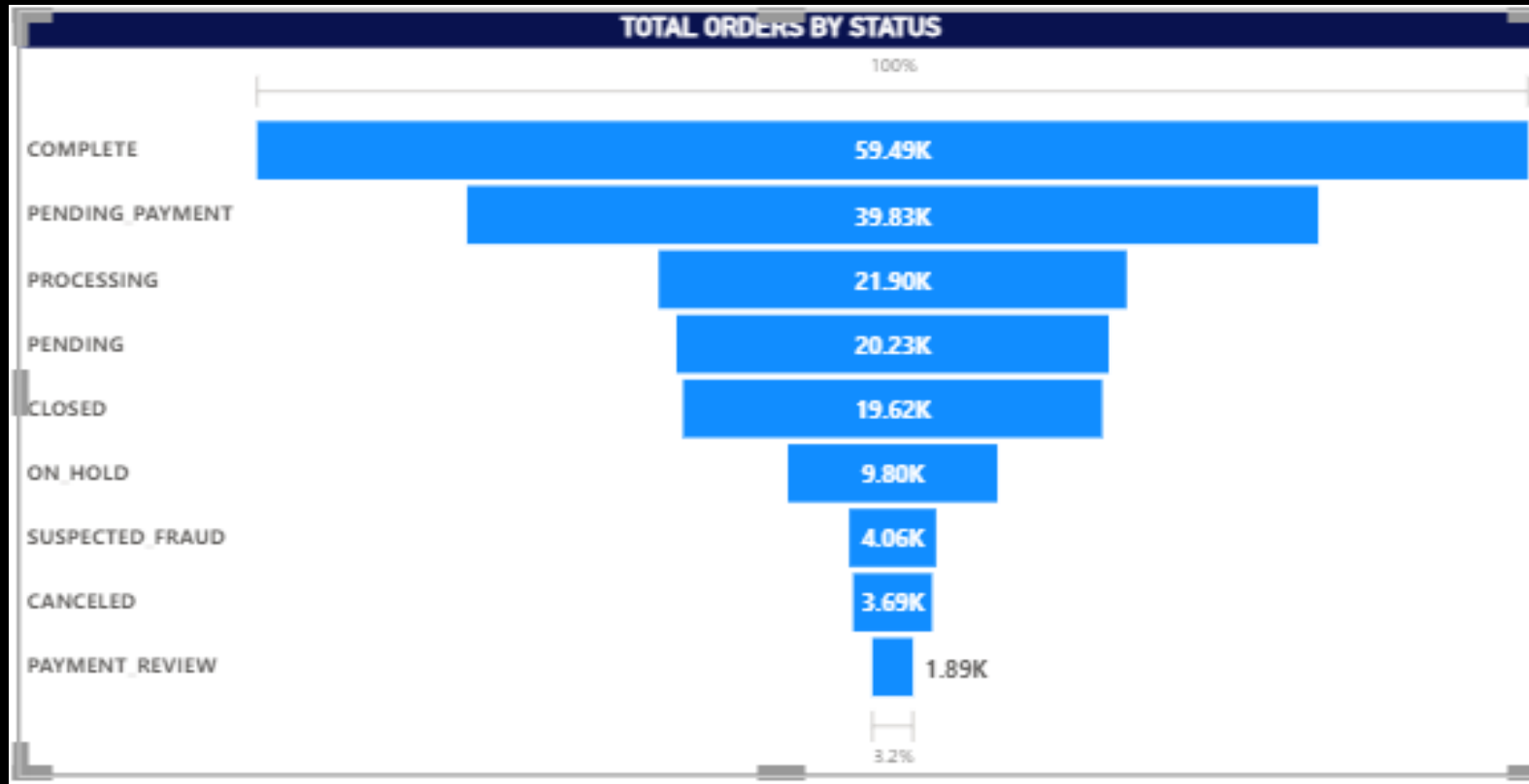
We need to analysis: There are more than 70% of order status is in the pending, On-hold and others. – Shipment Analysis.

1<sup>st</sup> WHY: Why this many orders are not done there payments.

## Answer:

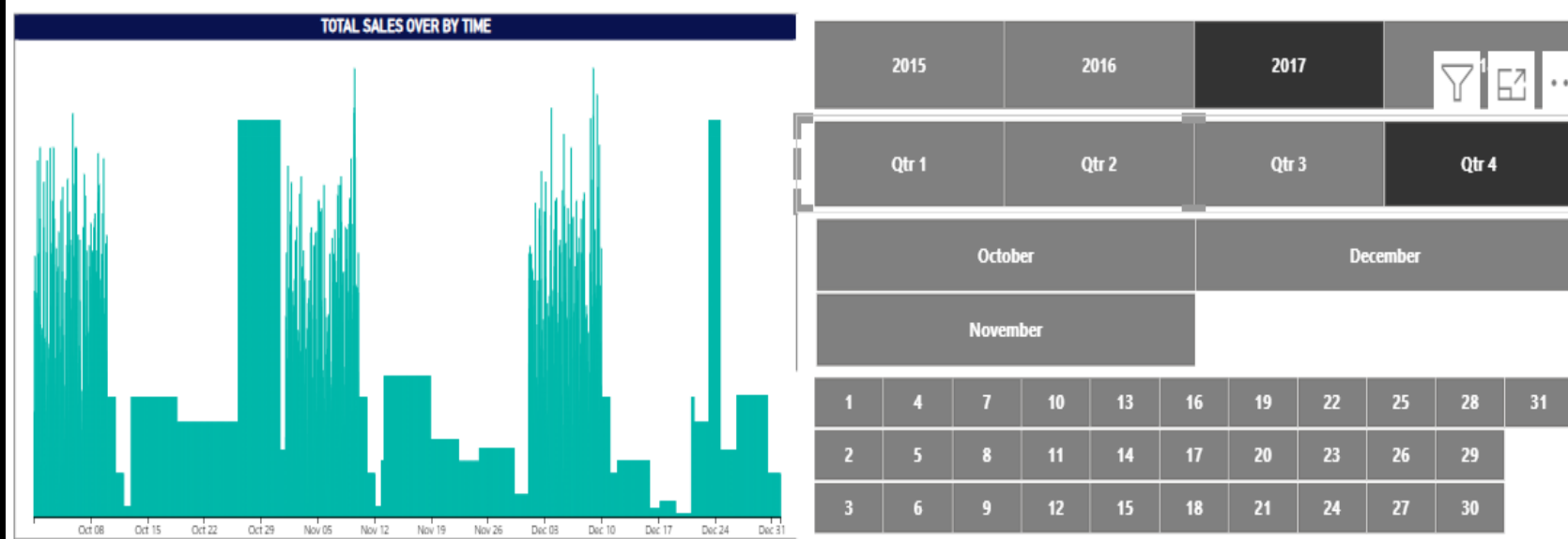
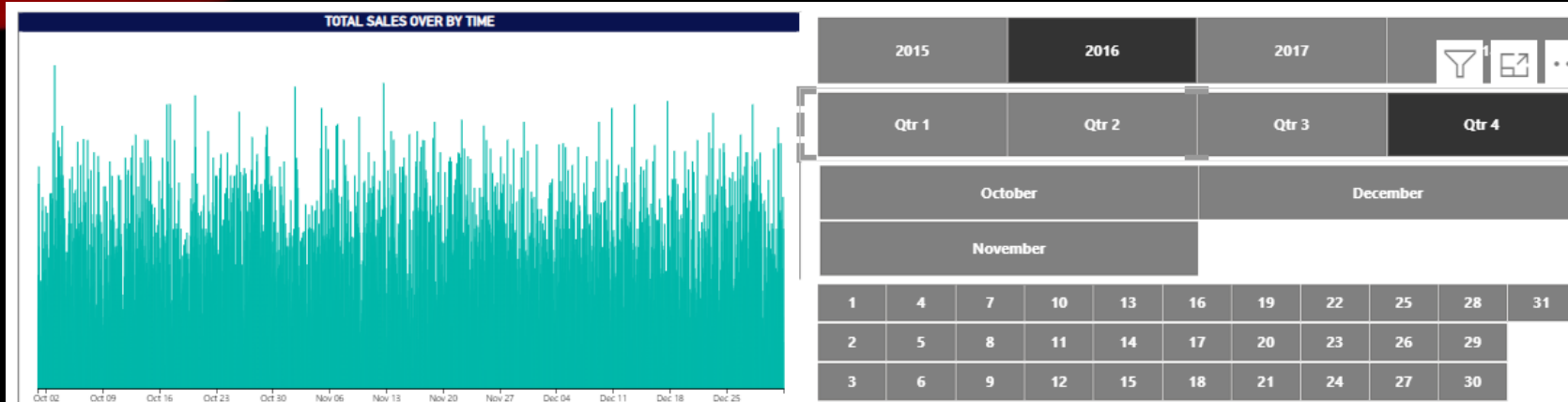
This may the major cause for the 2018 loss and product missed to sale and cities are missing.

We can able to see there is the PENDING PAYMENT, PROCESSING. PENDING, OH-HOLD are the major share of our orders status. So, we need to focus on why the pending and on-hold closely because we have nearly 50% of orders or in this category. – Sales Analysis





# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD



## Problem 6: Miscellaneous Visualization

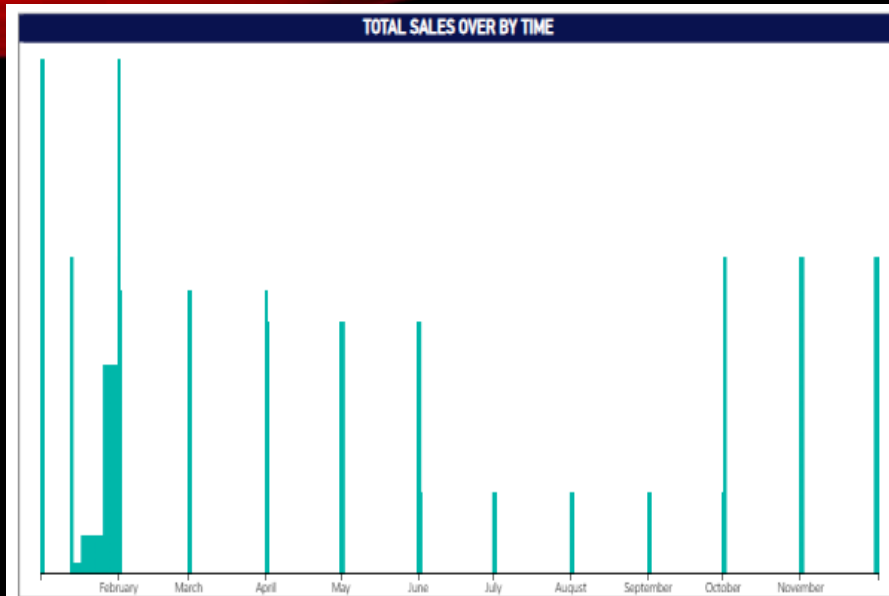
### Miscellaneous Area of Analysis:

There was a some very strange pattern we can able to see in the 2017's Last quarter. We can able to see same sales count over long period of time and it repeats with different level periodically.

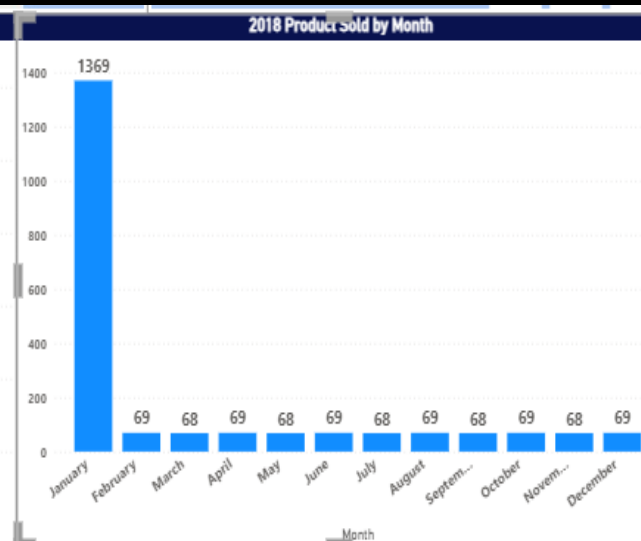
We have analysis this miscellaneous spike in risk analysis. It may a Glitch or data error. We need to find and make a proper change accordingly to ensure the data a



# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD



2015			2016			2017			<div><div><div></div></div><div><div></div></div><div><div></div></div></div>		
Qtr 1			Qtr 2			Qtr 3			Qtr 4		
January			April			July			October		
February			May			August			November		
March			June			September			December		
1	4	7	10	13	16	19	22	25	28	31	
2	5	8	11	14	17	20	23	26	29		
3	6	9	12	15	18	21	24	27	30		



Year	Month	Product_Name	Sum of Product_Category_ID
2018	January	DVDs	67
2018	January	Fighting video games	75
2018	January	Lawn mower	68
2018	January	Smart watch	73
2018	January	Summer dresses	76
2018	January	Toys	74
2018	February	First aid kit	69
2018	February	Lawn mower	68
2018	March	First aid kit	69
2018	April	First aid kit	69
2018	April	Men's gala suit	70
2018	April	Rock music	71
2018	May	Rock music	71
2018	June	Adult dog supplies	72
2018	June	Rock music	71
2018	July	Adult dog supplies	72
2018	August	Adult dog supplies	72
2018	September	Adult dog supplies	72
2018	October	Adult dog supplies	72
2018	October	Smart watch	73
2018	November	Smart watch	73
Total			715

## Problem 6 – Page 2: Miscellaneous Visualization

Miscellaneous Area of Analysis:

2018 is then worst then previous year. As per my analysis the product category is decreases rapidly so the sales amount would become same in lesser period of time frame. But we required the detailed analysis for these miscellaneous curves and we need the clear document analysis or else will leads various chain link issue over the period.

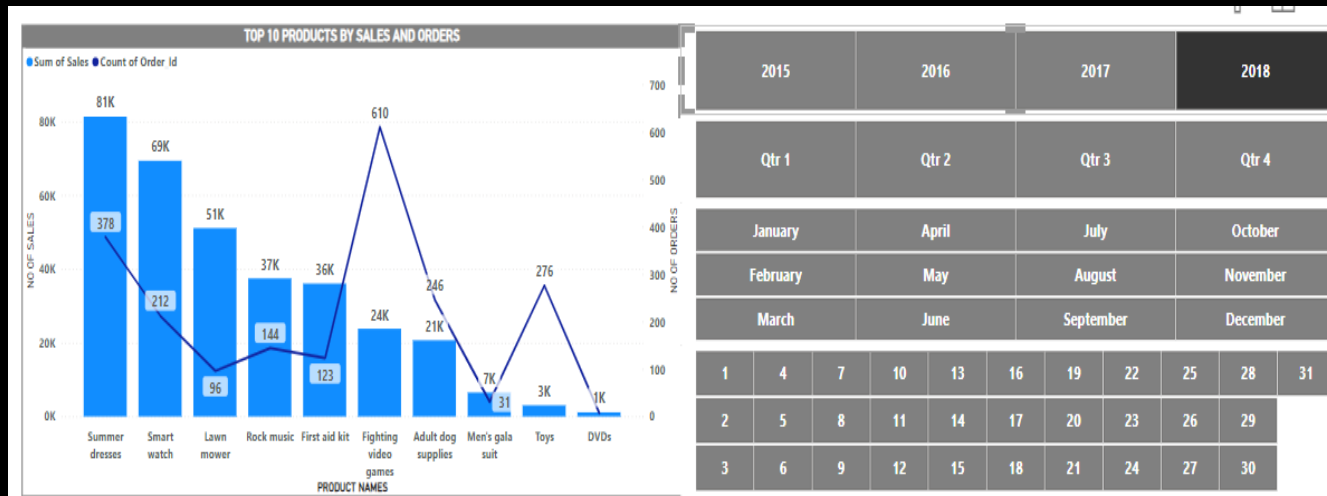
## Answer:

As I can able to see the trend is not any miscellaneous data and the actual trend the managed with the product maintained is for 11 months over 68 and 69 products though out the year.

# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD

## Positive Impact:

We can able to see the order count is spiked on 'Fighting Video Games' even the sales is normal in the product category. So, it will lead the next profit zone. – Product Analysis

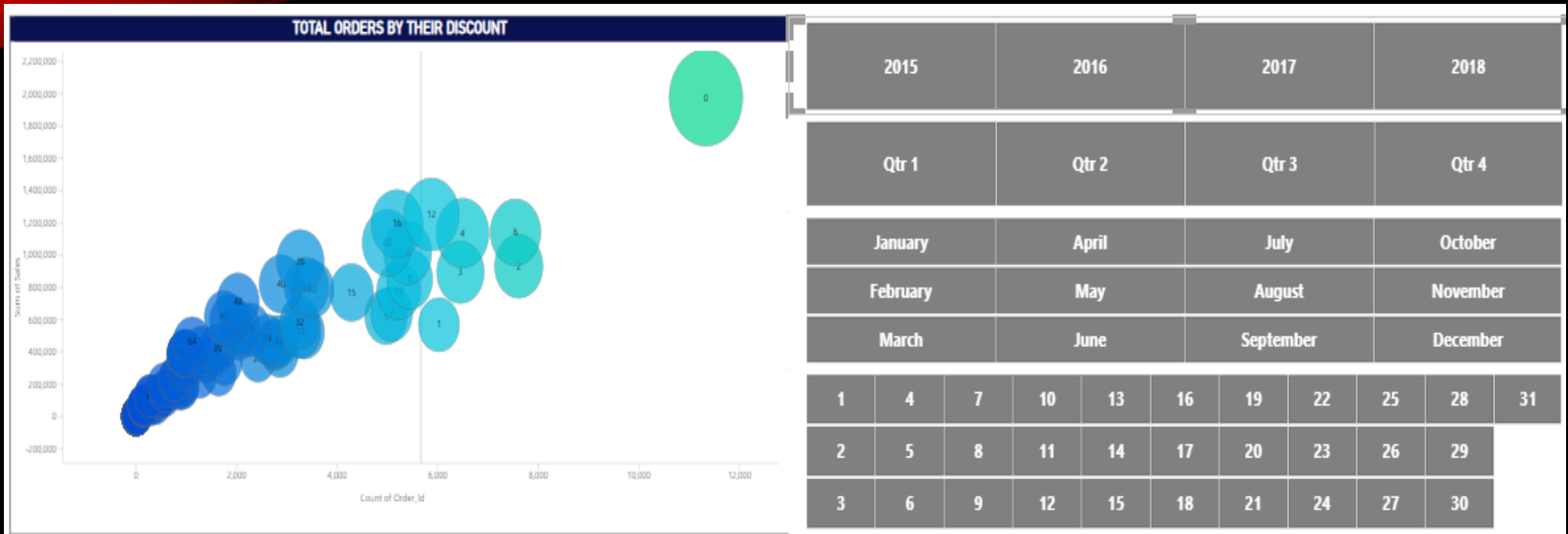


## Positive Impact:

We can able to see the Profit is spiked for "Summer Dresses" then even the sales are normal in the product category. So, it is the profitable zone when coming to product pricing. – Product Analysis



# ROOT CLAUSE ANALYSIS – USING 5 WHY METHOD



## Positive Impact:

The Discount plays a major role on sales. Because the sales curve is mostly depending on the discount part a we are not able to see any minus curves in profit margin as well. Good work we have done we with discount Sales Analytics

# END

Sriram P – Senior Process Associate

