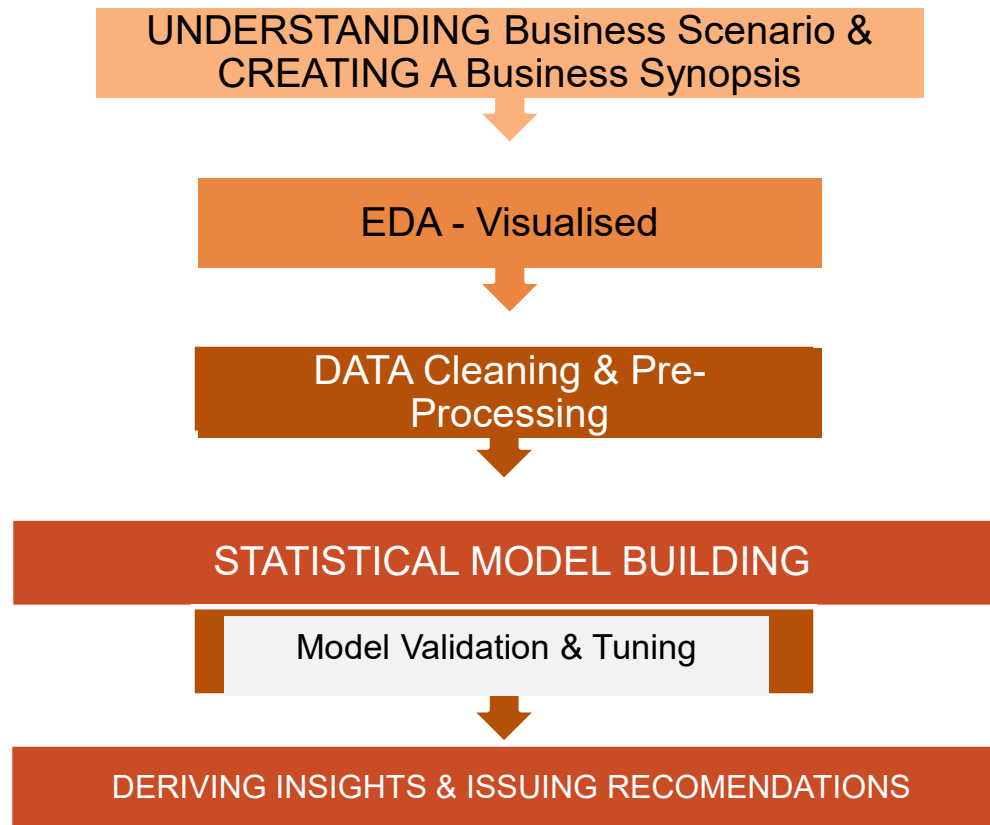# LATE DELIVERY in Supply Chain Management

RISK Analysis & Annihilation

- By Pradeep Panicker

- [A CAPSTONE PROJECT FOR GREATLAKES]

ANALYSIS METHODOLOGY

UNDERSTANDING Business Scenario &
CREATING A Business Synopsis

EDA - Visualised

DATA Cleaning & Pre-Processing

STATISTICAL MODEL BUILDING

Model Validation & Tuning

DERIVING INSIGHTS & ISSUING RECOMENDATIONS

Business Scenario & Business Synopsis

Business Scenario

❑ The company is engaged in TRADING ITEMS for THREE Category of Consumers.

❑ It receives orders from its customers and creates an internal order for issuance of goods from its stock and delivering it to the End-Customers.

❑ While Customer Orders are generated from USA and PUERITO RICO, Delivery is done at 164 Countries across the GLOBE.

❑ Warehouses are suitably placed in all the FIVE regions and FIVE MODE of Shipment is used for Delivery.

❑ LATE DELIVERY is OBSERVED in 54.8% of the Customer Orders.

Business Synopsis

TRANSACTION Period Captured for Analysis	JAN 2015 ~ JAN 2018
NUMBER OF TRANSACTION	180,519.00
NUMBER OF CUSTOMER ORDERS	65,752.00
CATEGORY OF ITEMS	50.00
NUMBER OF ITEMS SOLD	118.00
ORDERS DELIVERED TO	05 MARKET REGIONS- Africa; Europe; Latin America; Pacific Asia & United States of Central America.
ORDERS DELIVERED TO	164 COUNTRIES
TOTAL SALES VALUE	$ 33,054,402.00
TOTAL PROFIT VALUE	$ 3,966,903.00
TOTAL DISCOUNT PROVIDED	$ 3,730,378.00

EDA – through visualization

Time Series Analysis- ORDER GENERATION

Time Series Analysis- SALES Generation

### Time Series Analysis - PROFIT Trend

Time Series Analysis- DISCOUNT Shelled Out

Insights -

TOTAL Volume Of -	JAN 2015 ~ OCT 2017	NOV 2017 ~ DEC 2017	Jan-18
SALES	NO TREND - STATIONARY	100% RAISE	150% REDUCTION
ORDER QTY.	SLIGHT DIP IN EVERY JANUARY. ELSE STATIONARY.	50% RAISE	STATIONARY AT NEW PEAK
PROFIT	SLIGHT DIP IN EVERY JANUARY. ELSE STATIONARY.	50% RAISE	150% REDUCTION
DISCOUNT	SLIGHT DIP IN EVERY JANUARY. ELSE STATIONARY.	50% RAISE	150% REDUCTION

EDA – through visualization
[contd .....]

Time Series Analysis - SHIPMENT DELIVERY Trend
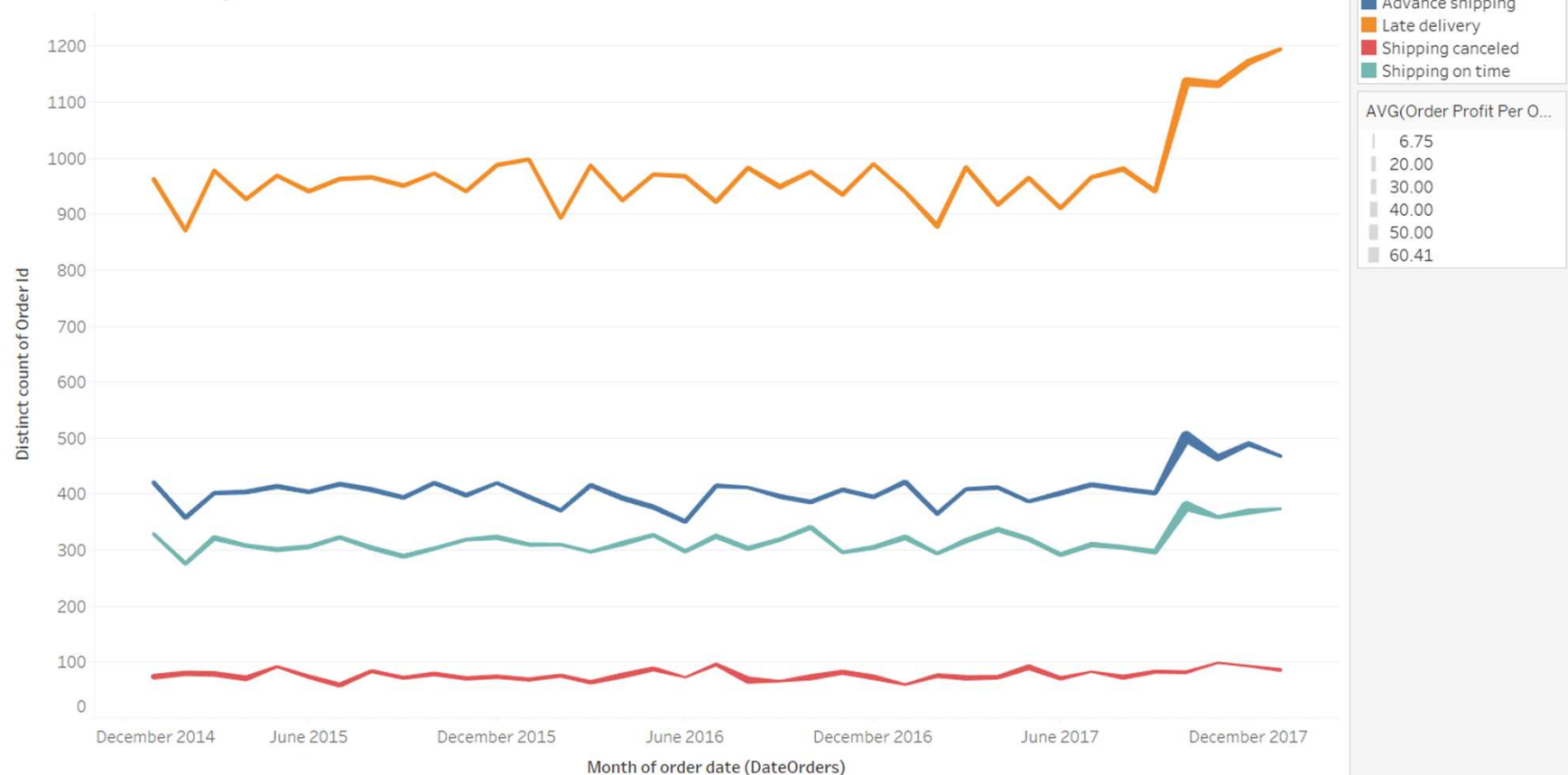
Insights - It is evident from the above graph –
a) The Late Delivery of orders are present thought out the transactional period – indicating it as a feature independent of Volume of - ORDERs / Sales/ Profitability / Discount. Hence these four factors do not influence LATE DELIVERY of Orders.
b) Thereby, it has become necessary to factor out the reason for 'LATE DELIVERY' – through a statistical analysis of Data.

EDA – through visualization
[contd .....]

Insights -

From the correlation plot depicted here, A HIGH CORELATION EXISTS BETWEEN THE VARIABLES –

1. SALES & ORDER Item Product Price.
2. ORDER Item Discount & Sales.
3. PROFIT per Item & Sale Price per Item.

These correlations DO NOT EXPLAIN the reason for 'LATE DELIVERY' of Order.

Hence the need for a statistical evaluation of DATA to find out reasons for LATE DELIVERY of Orders.
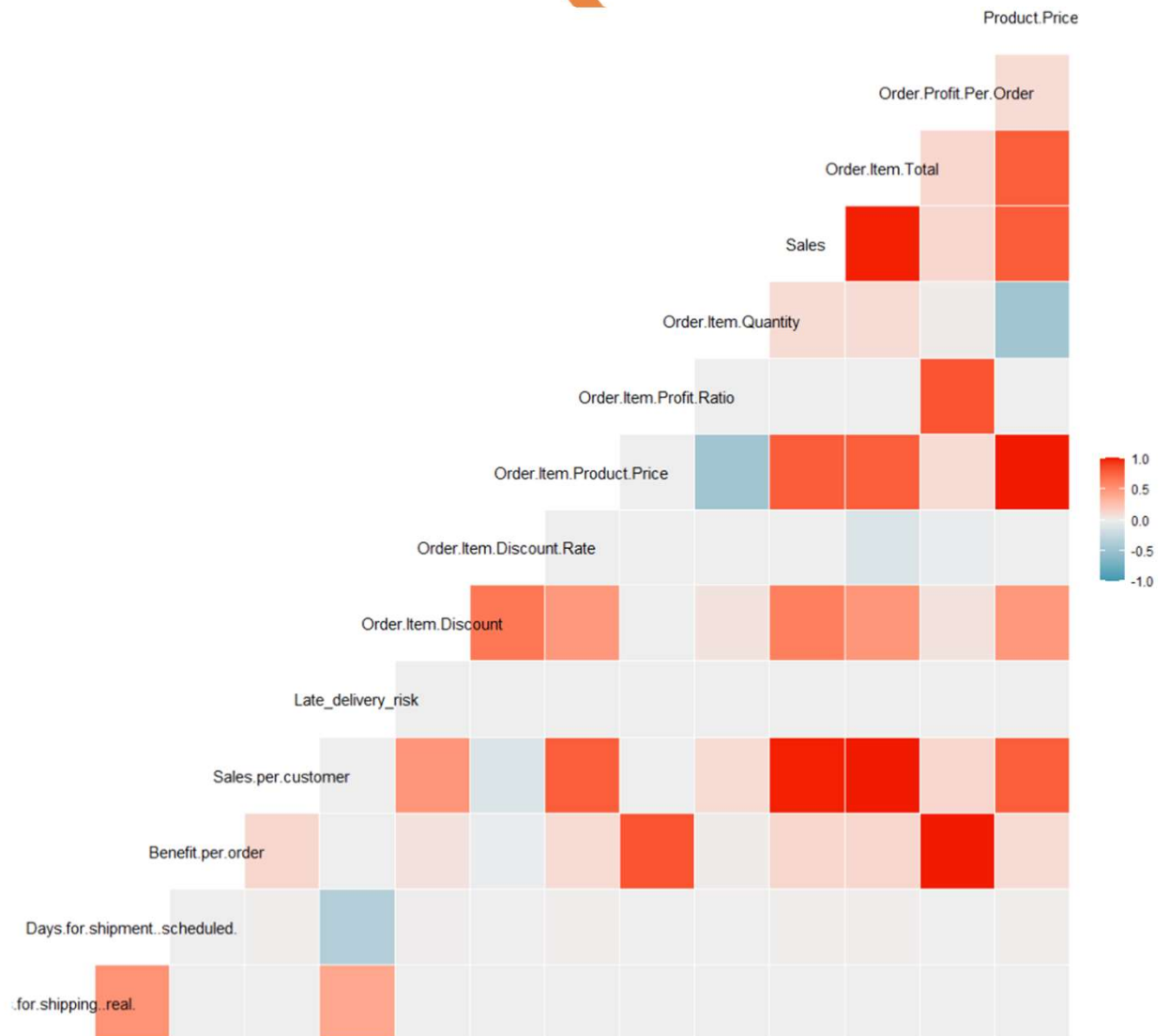
DATA Cleaning & Pre-Processing

The dataset contains 53 VARIABLES which are classified as PREDICTOR variables . One Variable ' Late Delivery Risk' will be used as the TARGET variable . Statistical model will be built upon the TARGET Variable – to find out the most significant variable that influences the TARGET Variable.

a) To reduce multi-collinearity between predictor variables, a screening of all variables w.r.t. Target variable is done using the CHI. SQ. test for significance.
   Doing this exercise, helped us reduce our predictor variables to 28 from 53.

b) EIGHT transactions were found to have MISSING VALUES which were eliminated from the dataset.

c) FEATURE ENGINEERING was done to create a new variable – ORDER PROCESSING TIME = Shipping Date - Order Date. This was done to capture the possibility of Delay happening due to Manual Labor in Order Shipment.

d) OUTLIER Treatment was kept at ABEYANCE since it was found that HIGH Sales Orders are the ones that get delivered LATE. OUTLIER Treatment will affect the values of these orders and eventually affect regression analysis on ' DELIVERY RISK'

e) For further dimensionality reduction, 'Principal Component Analysis [PCA]' was implied on the 28 variables. For this-
   1) All categorical variables were encoded to numerical classification.
   2) Thereafter, the complete data-set, excluding the Target Variable, was normalised.

f) Results of PCA helped to join similar variables and form a new dataset consisting of TEN Predictor variables.

THIS RE-ENGINEERED DATASET WILL NOW BE USED FOR FURTHER STATISCAL MODEL BUILDING.

BEFORE ' PCA' - Numerical variables	AFTER ' PCA'
1 Days.for.shipping..real	
2 Days.for.shipment..scheduled	Shipment.Scheduling
3 orderprocessingtime	
4 Benefit.per.Order	Profitability
5 Sales.per.Customer	
6 Order.Item.Product.Price	Sales.Order.Value
BEFORE ' PCA' - Categorical variables	**AFTER ' PCA'**
1 Customer.Id	
2 Order.Customer.Id	Customer.ID
3 Type	
4 Order.Status	Order.Fullfilment
5 Customer.State	Customer.Order.Origin
6 Order.City	
7 Order.State	Delivery.Destination
8 Market	
9 Order.Region	Delivery. Region
10 Customer.Fname	
11 Customer.Street	Customer.Name
12 Shipping.Mode	Shipping. Class

STATISTICAL MODEL BUILDING-

MODEL SELECTION –

With a descriptive analysis in mind, LOGISTICS REGRESSION & CART models will be best suitable for the given scenario.

LOGISTICS Regression- enables us to carry out Prediction as well as find out explicitly and quantify – which all factors contribute to LATE DELIVERY.

CART – Classifies the significant variables and STEP-WISE each leaf of the CART TREE defines and quantifies the Loss od Delivery made due to Each predictor variable.

Important variable Identification –

a. Probability of occurrence of a significant predictor variable can be evaluated – under the given circumstances. – LOGISTICS REGRESSION Model & CART Model was built and Tuned.

b. Each significant variable can be classified according to its importance in creating variance on the target variable – ODDS Coefficient Analysis for Logit Model & IMPORANT VARIABLE Plot for CART was analyzed

'CART' MODEL WAS CHOSEN OVER 'LOGIT' MODEL DUE TO BETTER MODEL VALIDATION AS WELL AS GIVING OPTIMUM SIGNIFICANT VARIABLES SELCTION.
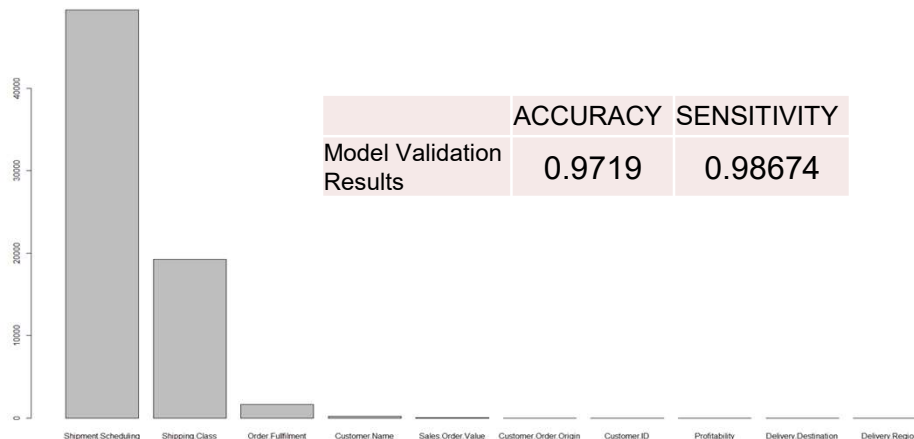
	ACCURACY	SENSITIVITY
Model Validation Results	0.9719	0.98674

CART MODEL – Development and results.

Initial model is built to allow maximum nodes to the TREE. Control parameters are kept to its minimum.

Thereafter the CP Value - & CROSS-Validation error is derived .

It is found that THERE IN NO OVER-FITTING Observed – by which the model developed is suggested to be robust.

Overfitting observed when the 'xerror' value reduces and then tends to increase.

	CP VALUE	n split	Rel. Error	x error	xstd
1	0.423057	0	1	1	0.00310
2	0.321274	1	0.576943	0.577	0.00274
4	0.016835	3	0.208863	0.209	0.00182
5	0.015774	8	0.106624	0.107	0.00133
6	0.014889	10	0.075075	0.075	0.00113
7	0.01	11	0.060186	0.06	0.00101

Using the above parametric values, the tree is pruned further. The important variables as well as the Model validation values are as shown in the LHS -

DERIVING INSIGHTS from the Statistical Model –

1. Importance variable plot reveals that 'Shipment Scheduling' ,'Shipping Class' & 'Order Fulfillment' are the most significant factors for causing the 'Late Delivery Risk'

2. The CART TREE plotting also indicates the level at which the customers go into Late Delivery Risk due to these three factors.

Note -

a) 'Shipment Scheduling' comprises of 'Date of Shipment(real) + Date of Shipment(scheduled) + Order Processing Time.
b) 'Shipping Class' is 'Shipping Mode'.
c) 'Order Fulfilment' consists of 'TYPE of Payment' & 'Order Status'.
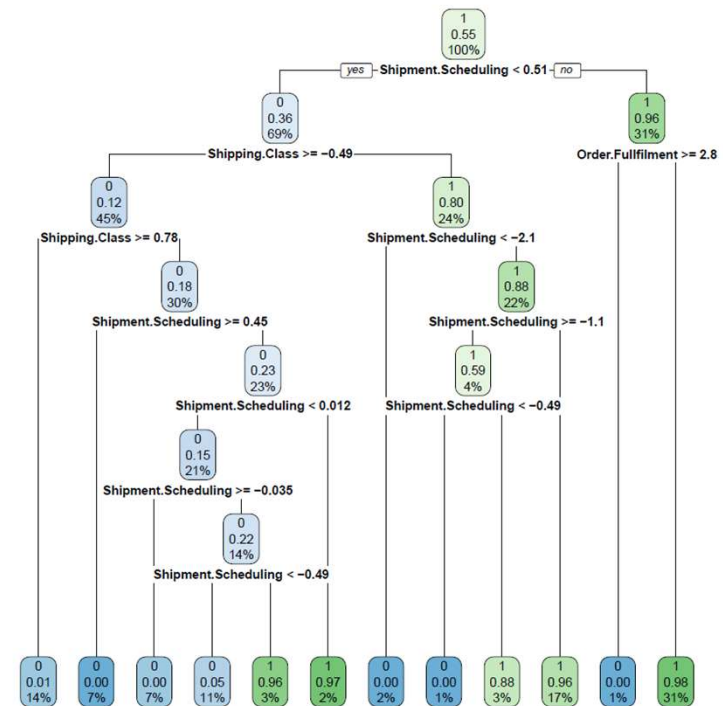
SUMMARISING INSIGHTS -

INSIGHTS from EDA –

1. It was evident from the analysis of data that 'SALES' , 'NUMBER OF ORDERS', 'PROFIT & 'DISCOUNT' – ALL THESE FOUR FACTORS DID NOT BECOME A CAUSE FOR 'LATE DELIVERY' OF ORDERS.

2. The Correlation plot also indicated that the 'Late Delivery Risk' variable was not corelated to any evident variable in the dataset.

INSIGHTS from Statistical MODEL ' CART' –

1. The statistical model – CART , predicted the importance of variables 'Shipping Date(real), Shipping Date (actual) & Order Processing Time' as the most important variable that influenced the LATE DELIVERY RISK factor of orders.
2. 'SHIPPING CLASS' was defined as SECOND MOST important variable , followed by
3. 'ORDER FULLFILMENT' consisting of 'TYPE of Payment & ORDER STATUS'.

Note -

a) Variables indicated as significant through the statistical modelling approach COULD NOT be identified with bare-eyes while EDA, thereby proclaiming the significance of Statistical Analysis of data for actionable insights.

ACTIONABLE RECOMMENDATIONS for Minimizing 'LATE DELIVERY RISK' factor in SCM. -
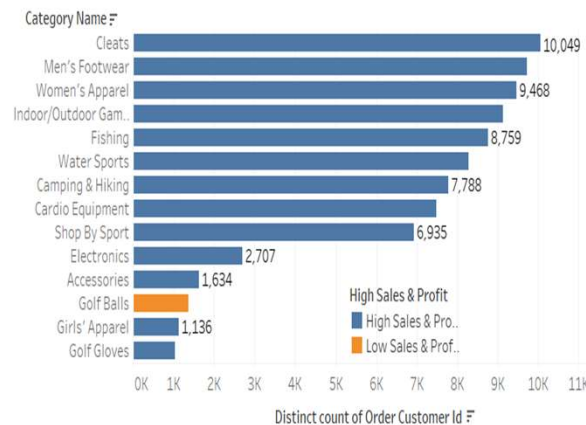
Analyzing the 'High Sales & Profitable' category of items,

Findings [F 1/2]–

1. All these items have more than 50% of their orders delivered LATE due to HIGH 'Order Processing Time' & HIGH 'Shipment Time'.
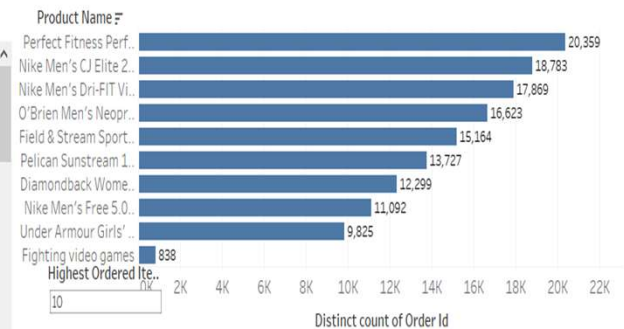
Recommendations for [F 1/2]-

A. **High Order Processing Time & High Shipment Time can be due to lack of facilities at such warehouses from where the goods are dispatched. City-wise mapping may be done for Transport infrastructure & Workforce infrastructure improvement, at the Warehouse level.**

B. **After optimizing the warehouse performance, re-visit the 'SHIPMENT Schedule days' data declared to customer and revise it to factual shipment schedule days, if necessary.**

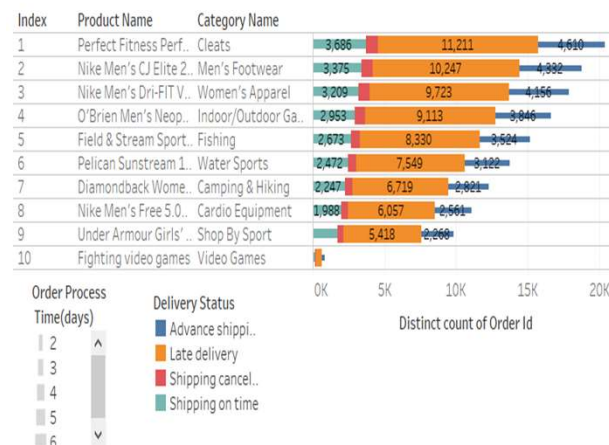**C.** **Allocation of 'Shipment Mode' to be revised to meet customer expectations as well as match**

Category Wise - Sales & Profit =ABOVE $10000 each

Category Name	Distinct count of Order Customer Id
Cleats	10,049
Men's Footwear	
Women's Apparel	9,468
Indoor/Outdoor Gam..	
Fishing	8,759
Water Sports	
Camping & Hiking	7,788
Cardio Equipment	
Shop By Sport	6,935
Electronics	2,707
Accessories	1,634
Golf Balls	
Girls' Apparel	1,136
Golf Gloves	

High Sales & Profit
- High Sales & Pro..
- Low Sales & Prof..

TOP 10 Highest Ordered Items [Graph IV]

Product Name	Distinct count of Order Id
Perfect Fitness Perf..	20,359
Nike Men's CJ Elite 2..	18,783
Nike Men's Dri-FIT Vi..	17,869
O'Brien Men's Neopr..	16,623
Field & Stream Sport..	15,164
Pelican Sunstream 1..	13,727
Diamondback Wome..	12,299
Nike Men's Free 5.0..	11,092
Under Armour Girls' ..	9,825
Fighting video games	838

Highest Ordered Ite..
10

TOP 10 Highest Ordered Items with Delivery Status Vs Order Process Time

Index	Product Name	Category Name	Distinct count of Order Id
1	Perfect Fitness Perf..	Cleats	3,686 / 11,211 / 4,610
2	Nike Men's CJ Elite 2..	Men's Footwear	3,375 / 10,247 / 4,332
3	Nike Men's Dri-FIT V..	Women's Apparel	3,209 / 9,723 / 4,156
4	O'Brien Men's Neop..	Indoor/Outdoor Ga..	2,953 / 9,113 / 3,846
5	Field & Stream Sport..	Fishing	2,673 / 8,330 / 3,524
6	Pelican Sunstream 1..	Water Sports	2,472 / 7,549 / 3,122
7	Diamondback Wome..	Camping & Hiking	2,247 / 6,719 / 2,821
8	Nike Men's Free 5.0..	Cardio Equipment	1,988 / 6,057 / 2,561
9	Under Armour Girls' ..	Shop By Sport	5,418 / 2,268
10	Fighting video games	Video Games	

Order Process Time(days)

Delivery Status
- Advance shippi..
- Late delivery
- Shipping cancel..
- Shipping on time

TOP 10 Highest Ordered Items with Delivery Status vs Actual Shipment Time

Index	Product Name	Category Name	Distinct count of Order Id
1	Perfect Fitness Perfect Rip Deck	Cleats	11,211 / 4,610
2	Nike Men's CJ Elite 2 TD Football Cleat	Men's Footwear	10,247 / 4,332
3	Nike Men's Dri-FIT Victory Golf Polo	Women's Apparel	9,723 / 4,156
4	O'Brien Men's Neoprene Life Vest	Indoor/Outdoor Ga..	9,113
5	Field & Stream Sportsman 16 Gun Fire Safe	Fishing	8,330
6	Pelican Sunstream 100 Kayak	Water Sports	7,549
7	Diamondback Women's Serene Classic Comfort Bi	Camping & Hiking	6,719
8	Nike Men's Free 5.0+ Running Shoe	Cardio Equipment	6,057
9	Under Armour Girls' Toddler Spine Surge Runni	Shop By Sport	5,418
10	Fighting video games	Video Games	

Days for shipping (real)
2, 3, 4, 5, 6

# ACTIONABLE RECOMMENDATIONS for Minimizing 'LATE DELIVERY RISK' factor in SCM. -

Analyzing the 'TYPE of Payment made' & 'ORDER Status',

Findings [F 2/2]–

1. A Time Series analysis of the 'Order Fulfillment' factor is done with respect to 'Late Delivered Orders'.

2. It is found that Orders for which payment has been received through DEBIT Mode and have been processed- ARE LATE DELIVERED.

3. Also ALL Orders that has taken the 'STANDARD CLASS' of shipment method, HAS BEEN LATE DELIVERED.

Recommendations for [F 2/2] -

A. **Internal administrative review of procedure followed for ORDER PROCESSING for orders received against DEBIT payment – NEED TO BE DONE.**

B. **'STANDARD CLASS' shipment method should be made more efficient.**
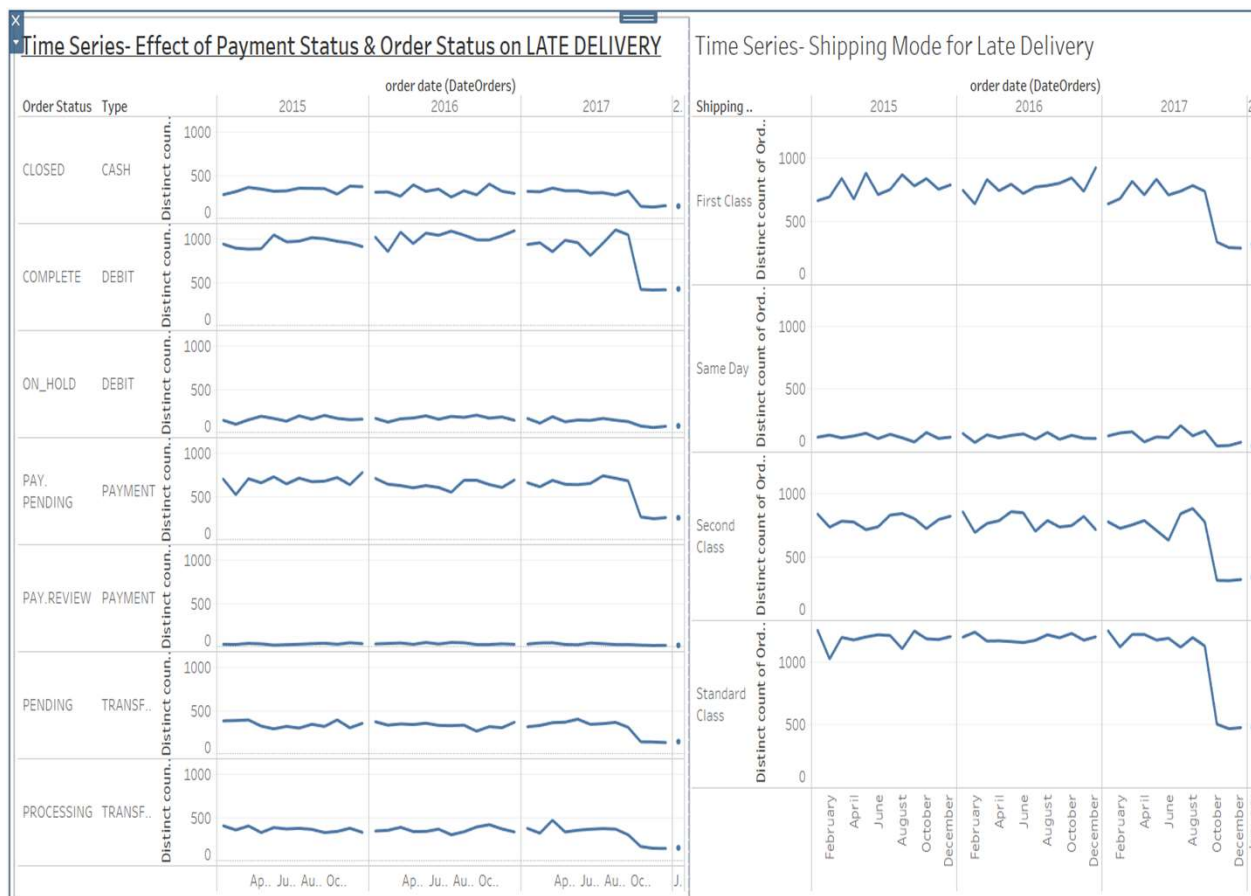
NOTE – Statistical Validation to be done every Three Months – MARKET Wise, for SCM optimization check.