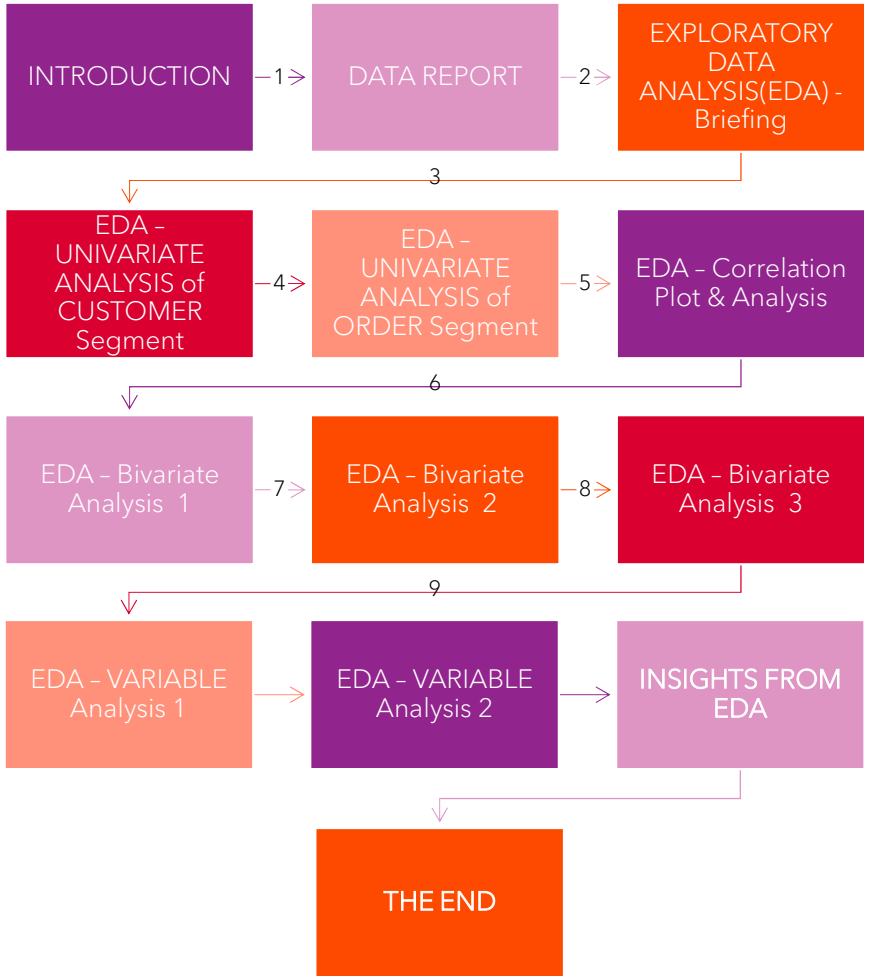
DELIVERY RISK ANALYSIS IN SUPPLY CHAIN

- CAPSTONE PROJECT FOR PG-BABI

INDEX-


INTRODUCTION
1
DATA REPORT
2
EXPLORATORY DATA ANALYSIS(EDA) - Briefing
3
EDA – UNIVARIATE ANALYSIS of CUSTOMER Segment
4
EDA – UNIVARIATE ANALYSIS of ORDER Segment
5
EDA – Correlation Plot & Analysis
6
EDA – Bivariate Analysis 1
7
EDA – Bivariate Analysis 2
8
EDA – Bivariate Analysis 3
9
EDA – VARIABLE Analysis 1
EDA – VARIABLE Analysis 2
INSIGHTS FROM EDA
THE END

INTRODUCTION

-

DATA with respect to orders placed by customers for Three category of goods and its delivery status to the End-Users is given. Also captured is the SALES and BENEFIT details – order wise.

The 'DELIVERY STATUS' shows that Supply Chain Management failed in 59% of the orders placed – failed orders resulting into 'LATE DELIVERY' or ' ORDER CANCELLED'. This is ALARMING situation, wherein now-a-days we have tools in SCM by which Inventory management and logistics can be optimized with machine learning applications.

With high competition in online trading, SCM has become the essence and influential factor for SALES Growth. Predicting a satisfactory and competitive delivery plan for his purchased product, the customer's satisfaction increases in multiplicity when prediction becomes factual.

DATA REPORT -

- The current data details out to be –
  a. Three-year transactions – JAN 2015 to JAN 2018.
  b. 1,80,519 transaction happened during these period.
  c. 65,752 Orders were generated for 118 Items .
  d. 20652 Customers from TWO Countries – USA & PURITO RICO - placed these orders to be delivered across cities in 164 Countries around the globe.

- The data also reveals –
  I. The Customer wise Sales and Customer Order wise Benefit generated across 65752 Orders.
  II. The total goods are segregated into THREE Categories
  III. The complete SCM status is assessed with FOUR distinctive Delivery Status.
  IV. The payment modes used by the customers are FOUR in nature.
  V. Customer details w.r.t. address, Zip Code is available. Contact details are missing. Customer location can be zoomed-in on a map using Lat-Long coordinates.
  VI. For Customer Orders received; Internal ORDER IDs are cross-generated by the system for inventory management. Delivery locations across the globe is mapped and registered as 'Order Country/City'.
  VII. As per 'Product Status [Stock Status]', shipping dates with method of shipment is finalised.
  VIII. Invoicing is done with DISCOUNTS included – NO TREND OBSERVED ON HOW DISCOUNT IS CALCULATED !!!

EXPLORATORY DATA ANALYSIS(EDA)

- UNIVARIATE Analysis –

We will be attempting to analyse the CUSTOMER Segment and the ORDERING & SHIPMENT to END USERS segment, separately.

A dashboard of interactive graphs will depict the analysis in a meaningful way.

Renaming –

a. Country EE.UU(as in data) is known as USA and renamed accordingly.

- BIVARIATE Analysis –

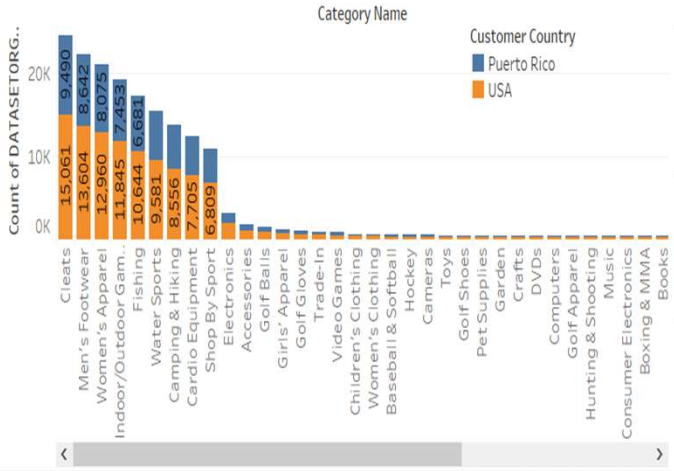
Here cross-analysis will be done within the variables related to the Customer Segment as well as the ORDERING Segment [system generated] and get insights as to how to improve the SCM mechanism.

A dashboard will be presented with graphical representations of analysis made followed by insights.

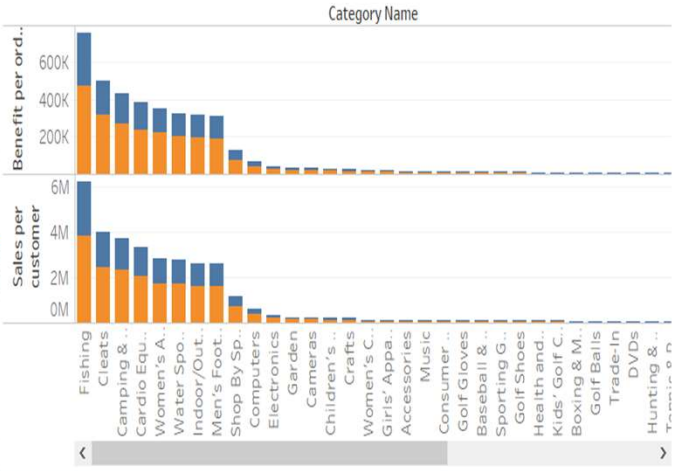
EDA – UNIVARIATE ANALYSIS

1. Graph I – Details out the number of transactions generated country wise – complete transactions are within TWO Countries – USA & PURITO RICO. Major transactions are limited to NINE Category of items – 'CLEATS' claiming to be the top transacted category.
2. Graph II – For both the countries, items related to the 'FISHING' category has maximum SALES and thereby maximum PROFIT per order.
3. Graph III – Among both the countries, Highest Sales & Profit was recorded in 'Caguas' city of Puerto Rico and was way-ahead from its second in line city of 'Chicago' of USA.
4. Graph IV – Interesting to note that whatever be the method of payment, LATE DELIVERY affects majority of the Customer Orders.

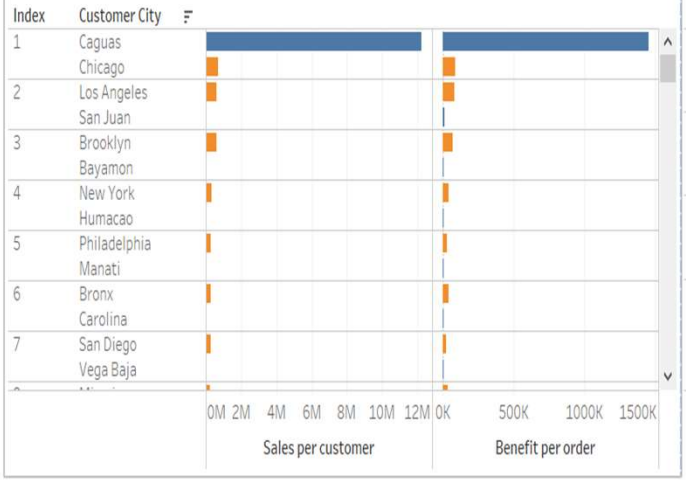
Category of goods- Coutry wise Transactions [Graph I]



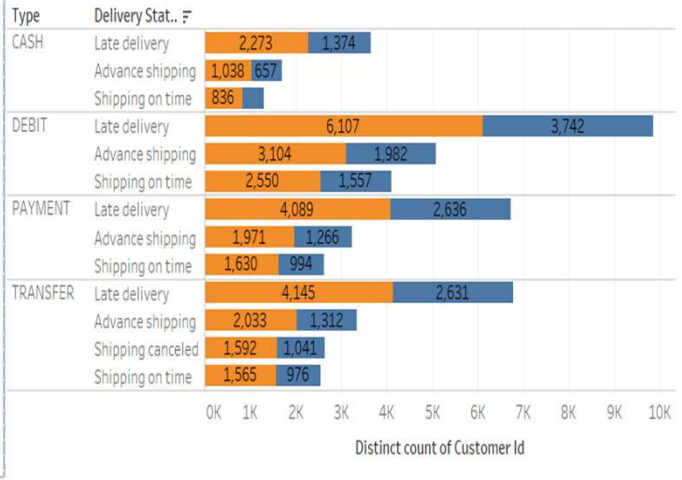
Category of goods- Volume of SALES[per customer] & BENIFIT[per order] [Graph II]



TOP SEVEN CITIES - Sales Volume & Profit wise [Graph III]



ORDERS category - Customer Payment & Delivery Status [Graph IV]

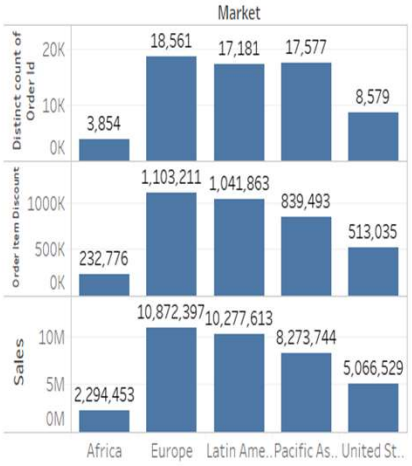


Click here for Live Visualization

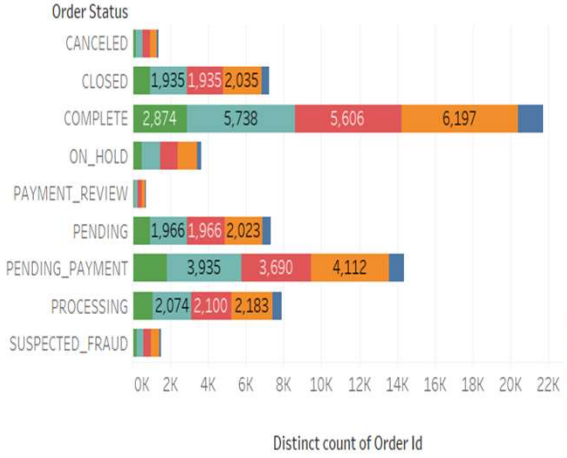
EDA – UNIVARIATE ANALYSIS OF ORDER SEGMENT

1. Graph I- Based on the customers orders, system generated Orders are created for supply-chain activities. Here, complete orders are complied into FIVE Global market regions – Africa; Europe; Latin America; Pacific Asia & United states of Central America. EUOPE Tops the list w.r.t. High Sales Volume; Highest number of Orders placed as well as Highest Discount offered against each Order.
2. Graph II- ORDER Status can be monitored under NINE categories – here maximum orders fall in the COMPLETION process.
3. Graph III – indicates that all the FOUR Mode of dispatches are available by the company – in all FIVE Market regions.
4. Graph IV – Delivery wise, HIGHEST ordered item is 'Perfect Fitness Perfect Rip Deck'- a shoulder exercise equipment.
5. Graph V- The HIGHEST Priced item in the 118 items available is 'SOLE E35 Elliptical'- a flywheel based elliptical Fitness Trainer equipment- valued at US$2000.00

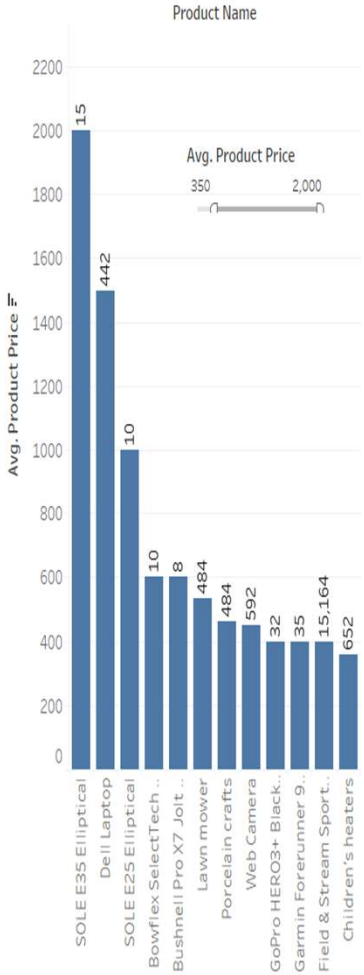
Market wise -ORDERS; SALES & DISCOUNTs [Graph I]



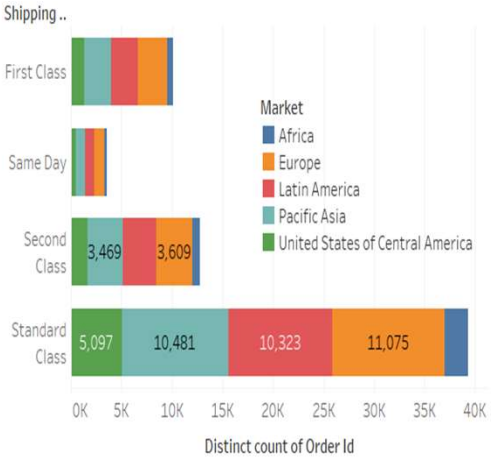
Market wise - ORDER Status [Graph II]



TOP 12 Highest Priced Items [Graph V]



MODE of Shipment - MARKET wise distribution pattern [Graph III]



TOP 10 Highest Ordered Items [Graph IV]

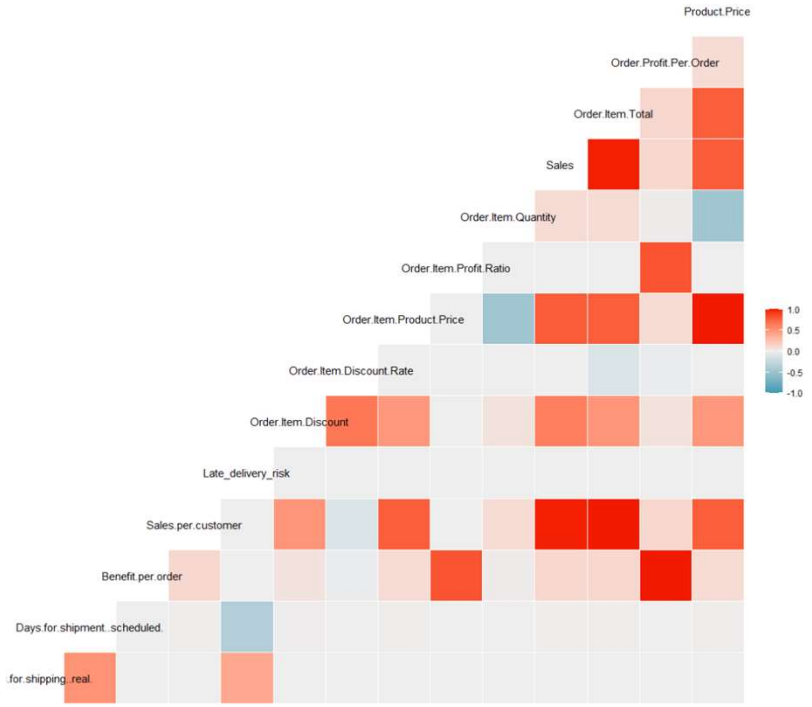
[Click here for Live Visualization]

EDA – CORRELATION PLOT AND ANALYSIS

<u>The plot indicates –</u>

- Sales volume is positively correlated to the 'Product Pricing' as well as 'Order Item Discount' – Larger the Sales , larger Discounts are offered . Higher the 'Product Pricing' naturally increases 'Sales'.
- Higher priced products generate higher Benefit per Order value.
- Sales per customer tends to increase based on the REAL-TIME Shipment time.
- The current Scheduled Shipment Delivery time is negatively impacting the Sales.
- A negative correlation is shown between 'Product Price' and 'Profit Ratio' and between 'Sales' and 'Profit Ratio' – indicating that Increase in 'Sales' is resulting into decrease in Profitability.

The correlation plot indicates to investigate the reason behind NEGETIVE IMPACT of 'Profit Ratio' upon Product Price & Discount Rate.



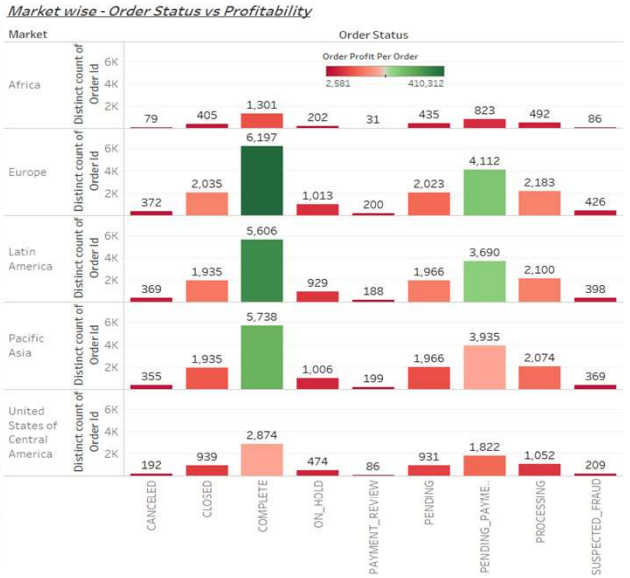
EDA - BIVARIATE ANALYSIS

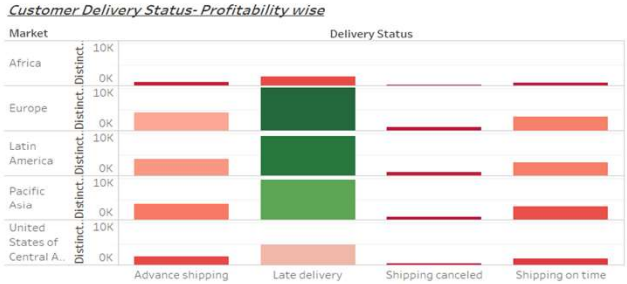
TOP Graph –

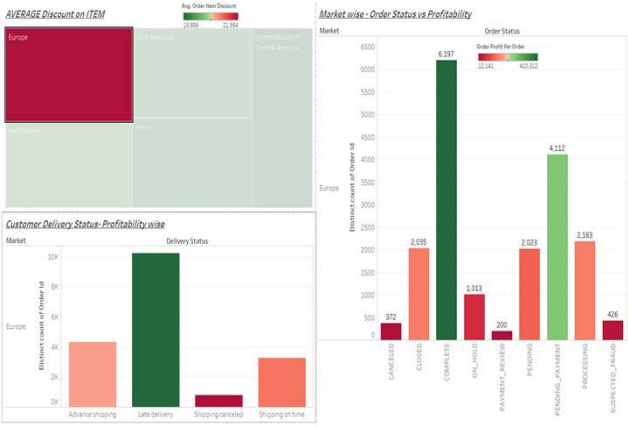
a. The average discount per item ranges from US$22 to US$19.9 and is this discount is present in all five market sectors.
b. The High Profitable orders fall in the Order Status – 'COMPLETE' , with the Orders Second-In-Line w.r.t. Profitability falling in the 'Payment Pending' status. All Low Profitable Orders are given third priority w.r.t. Order processing.
c. Astonishing to see that High Profitable orders are ALWAYS DELIVERED LATE, while the LEAST PROFITABLE Orders are most likely Cancelled from Shipping out.

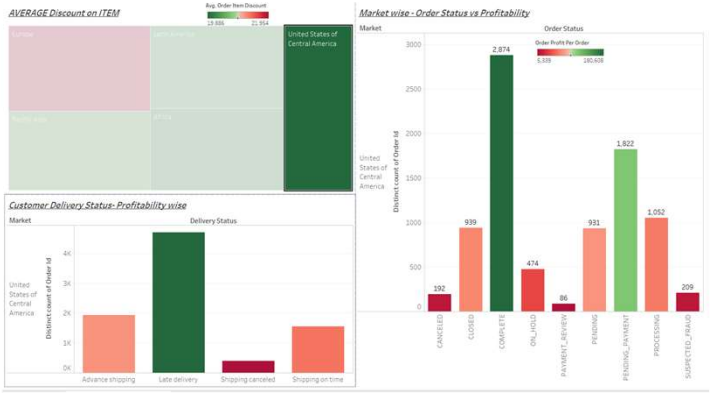
BOTTOM Two Graphs -

a. Whether it be the market sector with Highest Discount(EUROPE) or with Least Discount(USCA), the high-profitable orders are always getting Late Delivered, whereas Least Profitable orders get Cancelled.









Click here for Live Visualization

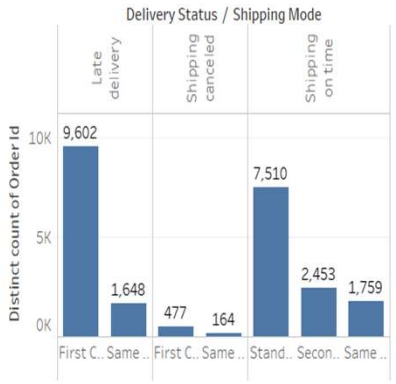
EDA – BIVARIATE ANALYSIS

ORDER Processing Time – is the time taken for the company to process the Order generated. This is the Number of Days between 'Shipping Date' and 'Order Date'. This variable was engineered for analysis.

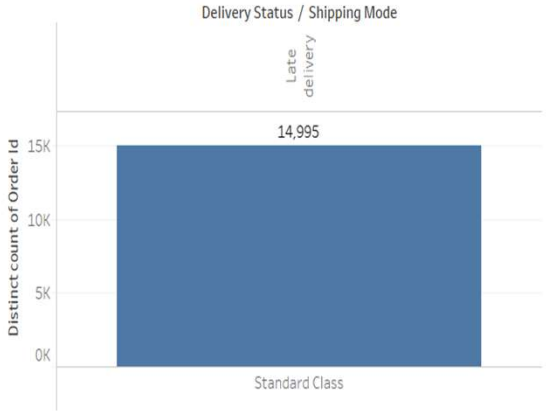
'Product Status' reveals that 100% of the items ordered ARE AVAILABLE IN STOCK. Thereby, a relationship plot between the 'Order Processing Time' & Delivery – MODE & STATUS was created. ***Except ONE Order which was processed 13.5 Days prior to placement of Order***, all Order Processing duration varied from 01 Day to 05 Days.

a. Orders which were processed within ONE Day – was dispatched using Three different modes. Even then 47.5% of orders were delivered LATE.
b. Orders processed with TWO Days time – were dispatched using 'Standard Class' but ended up being Late Delivered.
c. Orders processed within THREE Days were dispatched using 'Standard Class' and was delivered under the 'Advance Shipping' terms.
d. Orders processed with FOUR Days of processing time were dispatched using the 'Second Class' mode of dispatch but ended up as Late Delivery.
e. Orders which took the maximum processing time – FIVE Days were ultimately cancelled from Shipment.

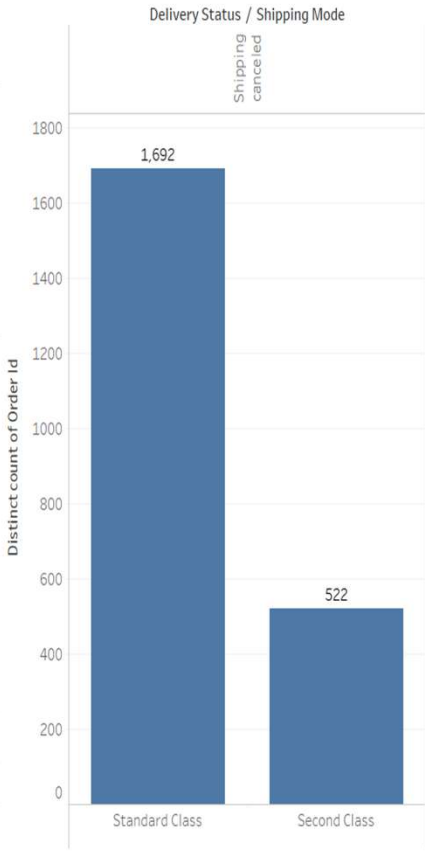
Delivery- Mode & Status for Single Day Order Processing



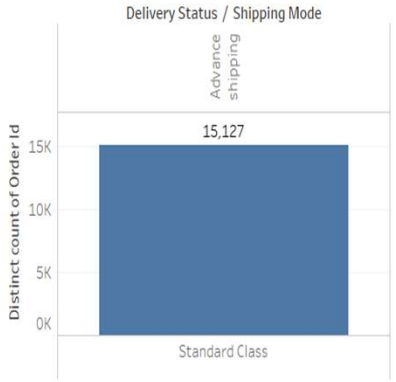
Delivery- Mode & Status for TWO Day Order Processing



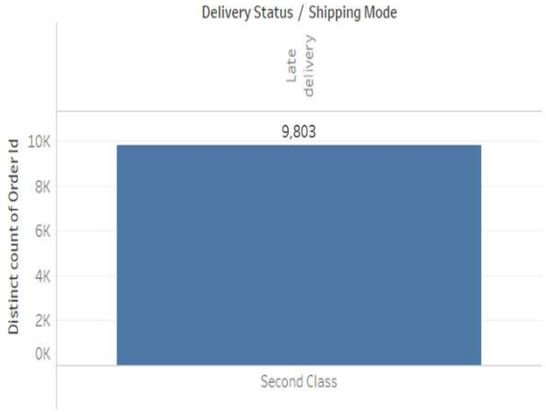
Delivery- Mode & Status for FIVE Day Order Processing



Delivery- Mode & Status for Three Day Order Processing



Delivery- Mode & Status for FOUR Day Order Processing



Click for Live Visualization

EDA - BIVARIATE ANALYSIS

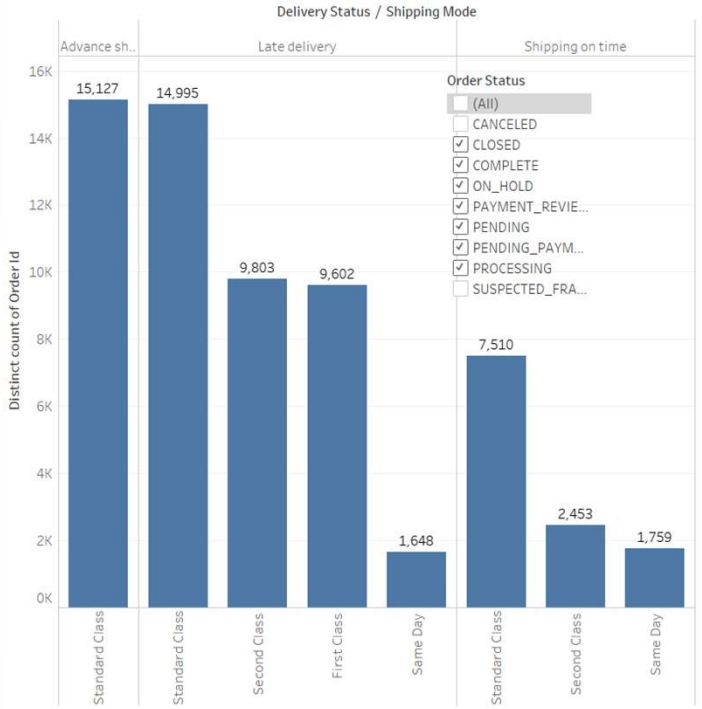
Here we are attempting to find whether the ORDER STATUS given to the Customer is correctly corelated with the ORDER STATUS of the respective order generated in the system by the company.

The common link for the Customer generated Order and the Respective ORDER generated by the company for its SCM management – is that the Customer ID and ORDER ID is the same.

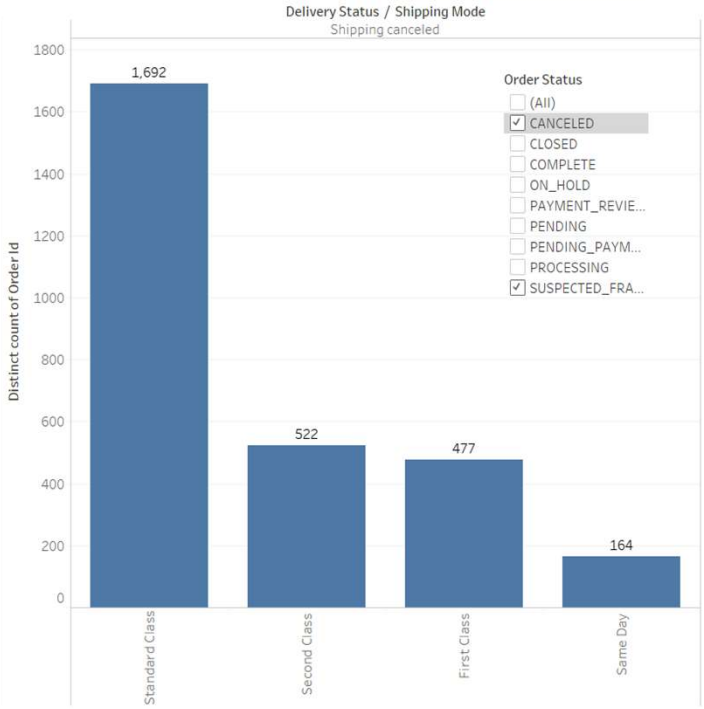
From the graphical representation , ALL SHIPMENTS THAT WERE IDENTIFIED AS 'CANCELLED' & ' SUSPECTED FRAUD' HAS BEEN INFORMED TO THE CUSTOMER AS 'SHIPMENT CANCELLED'.

Rest of all the orders which are classified under SEVEN different labels iis being processed and the customer is informed w.r.t. the timing in DELIVERY STATUS

DELIVERY Status w.r.t. ORDER Status - VALID SHIPMENTS



DELIVERY Status w.r.t. ORDER Status - CANCELLED SHIPMENTS



Click for Live Visualisation

EDA –

DEFINING VARIABLE PROPERTIES & REQUIREMENTS

- REMOVAL OF UNWANTED VARIABLE –

A Chi-square Test for 'Significance' was performed on all variables – Character as well as Numerical.

The primary variable was ' Late Delivery Risk' – since this will be the Target variable for Model Building and Regression analysis. Other 52 Variables significance of presence w.r.t. 'Late Delivery Risk' was analysed.

28 variables were found to have significant correlation w.r.t. 'Late Delivery Risk'.

Removing the balance 25 Variables, a NEW Dataset 'scm2' was created for model building exercise and Regression analysis.

Variable – 'Delivery Status' was removed purposefully since presence of this variable will nullify the essence of predictability of the Model.

- MISSING VALUES TREATMENT -

The dataset 'scm2' has 08 Observations which has missing values in the column

' Customer. Lname'. These observations will be eliminated during model building.

- OUTLIER TREATMENT –

During BIVARIATE Analysis, 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'.

Hence, Outlier Treatment is kept at abeyance at this stage.

EDA –

DEFINING VARIABLE PROPERTIES & REQUIREMENTS

- VARIABLE TRANSFORMATION –

For the purpose of CLUSTERING Analysis , we will be utilising a NUMERIC VARIALE TRNSFORMATION TECHNIQUE – Normalisation using scaling.

ADDITION OF NEW VARIABLES -

BIVARIATE Analysis revealed the importance of 'ORDER PROCESSING TIME' which was to generated utilising the variables ' Order Date' and 'Shipment Date'. This variable if found to have a significant impact on analysing the Delivery Status ~ Shipment Mode.

This variable was created as a calculated field in Tableau for visualisation through analysis.

This variable will be created and added to the dataset 'scm2' in R, during Model Building exercise.

INSIGHTS FROM EDA

1. IT IS OBSERVED THAT DISCOUNT ON SALES PER ORDER DOES NOT DRIVE SALES. DISCOUNT IS EVENLY DISTRIBUTED AMONG ITEMS AND HAS A POSITIVE CORRELATION TO SALES.
2. MODE OF PAYMENT by Customer & ORDER STATUS w.r.t. the company reveals that the DELIVERY STATUS HAS NO RELATION TO THE PAYMENT STATUS OF CUSTOMER.
3. 'LATE DELIVERY' IS NOTICED WITH ORDERS THAT HAS HIGH PROFITABILITY – THE DELIVERY SATISFACTION LEVEL WITH CUSTOMERS NEED TO BE CHECKED FOR SUSTAINABLE BUSINESS.
4. THE ORDERS WHICH THE COMPANY REJECTS FOR SHIPMENT – IS CONVEYED AS SUCH TO THE CUSTOMERS. HERE, IT IS SEEN THAT COMPANY REJECTS 'SUSPECTED FRAUD' ORDERS ALSO. THIS SHOWS THE GOOD ETHICAL PRACTICES OF SUPPLY-CHAIN THE COMPANY FOLLOWS.
5. SHIPMENT MODE UTILISED FOR DEFINED 'DELIVERY STATUS' OF CUSTOMER, WHEN CORELATED WITH THE ORDER PROCESSING TIME – REVEALS A HUGE MISMATCH. E.g.- ORDERS THAT ARE PROCESSED WITHIN ONE DAY ARE BEING DISPATCHED UNDER 'STANDARD CLASS' AND END UP WITH 'LATE DELIVERY' WITH CUSTOMER. THIS CAN BE A SERIOUS THREAT TO THE SUPPLY-CHAIN MANAGEMENT AND NEEDS IMMEDIATE ATTENTION.
6. IT IS NOTICED THAT ALL THE SHIPMENT MODES ARE AVALIABLE ACROSS THE FIVE MARKET SEGMENTS OF ORDER DELIVERY – THERBY GIVING THE COMPANY HUGE POSSIBILITIES TO RECTIFY POSSIBLE ERRORS AS MENTIONED IN PT. 5 ABOVE.

INSIGHTS FROM EDA

DATA STRUCTURE – IMBALANCED OR NOT

THE DATASET AVALABLE IS A BALANCED DATA SET W.R.T. THE 'LATE DELIVERY RISK' FACTOR – WITH 54% OF THE DATA CATAGORISED AS LATE DELIVERY AND BALANCE 46% CATAGORISED AS NOT-LATE DELIVERY.

THANK
YOU!