Supply Chain Data Set: Exploratory Analysis and Insights

Bharanidharan M P

AJ23SYD018

Student: MAIB23

SP Jain School of Global Management, Sydney

Australia.

**Contents:**

|  |  |
| --- | --- |
| **Topic** | **Page Number** |
| 1. Problem Statement | 2 |
| 1. Dataset | 2 |
| 1. Data Cleaning and Feature Engineering | 3 |
| 1. Exploratory Data Analysis | 3 |
| 4.1 Delivery Analysis | 5 |
| 4.2 Profit Loss Analysis | 8 |
| * 1. Market Analysis | 10 |
| * 1. Discount per Product Analysis | 12 |
| * 1. Mode of Shipment Analysis | 15 |
| 1. Conclusion and Insights | 16 |
| 1. References | 16 |

1. **Problem Statement:**

The problem is to read and analyze a dataset from a Supply Chain Industry with the details of the order and shipment of products in their Industry. The data’s might have duplicates or unwanted details which might disturb the actual accuracy of the analysis.

To ignore the errors or duplicates from the data, a neat Data Cleaning and an Exploratory Analysis is to be carried out. The project focuses on using artificial intelligence techniques to get 5 Visual Statistical summaries and insights in order to help the VP of Logistics and VP of Sales to help develop their future business.

1. **Dataset:**

Dataset Link: <https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis>

This dataset consists of **53 features and 180519 rows**. The features in the dataset include:

Type, Days for shipping (real), Days for shipment (scheduled), Benefit per order, Sales per customer, Delivery Status, Late\_delivery\_risk, Category Id, Category Name, Customer City, Customer Country, Customer Email, Customer Fname, Customer Id, Customer Lname, Customer Password, Customer Segment, Customer State, Customer Street, Customer Zipcode, Department Id, Department Name, Latitude, Longitude, Market, Order City, Order Country, Order Customer Id, order date (DateOrders), Order Id, Order Item Cardprod Id, Order Item Discount, Order Item Discount Rate, Order Item Id, Order Item Product Price, Order Item Profit Ratio, Order Item Quantity, Sales, Order Item Total, Order Profit Per Order, Order Region, Order State, Order Status, Order Zipcode, Product Card Id, Product Category Id, Product Description, Product Image, Product Name, Product Price, Product Status, shipping date (DateOrders), Shipping Mode.

Reason for selecting this dataset:

The dataset has a good collection of features. The dataset not only consists of the order details but also consists of the customer’s details and other features that help us understand the profit and loss of orders and sales (example: Profit and Loss region and country wise).

1. **Data Cleaning and Feature Engineering:**

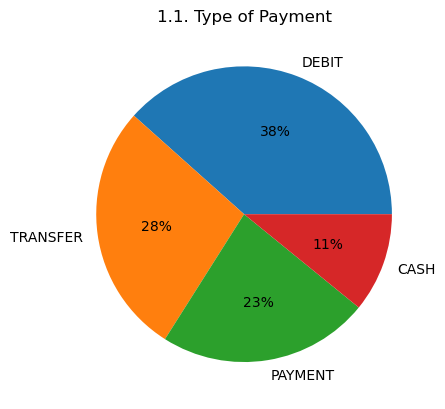
After performing initial data exploration and analyzing the features in-depth, the columns features were cleaned:

* Benefit per Order - Converted the values of this column to Profit, No Profit No Loss and Loss as the earlier values were highly imbalanced in the value counts and named it as BenefitBracket.
* Order Item Discount Rate - Converted the values of this column to less than 10%, less than 20% and more than 20% as the earlier values were highly imbalanced in the value counts and named it as Order Item Discount Bracket.
* Order Item Product Price - Converted the values of this column to less than 500, inbetween 500 – 1000, inbetween 1000 – 1500 and inbetween 1500 – 2000 as the earlier values were highly imbalanced in the value counts and named it as Product Price Bracket.
* Sales - Converted the values of this column to less than 500, inbetween 500 – 1000, inbetween 1000 – 1500 and inbetween 1500 – 2000 as the earlier values were highly imbalanced in the value counts and named it as Sales Bracket.
* Days for shipping (real) –­ This gives the number of days(real) for shipment which can be inferred from date column, hence dropped this column.
* Days for shipping (scheduled) –­ This gives the number of days(scheduled) for shipment which can be inferred from date column, hence dropped this column.
* Sales per customer – This doesn’t add value since it can be inferred by benefits, hence dropped this column.
* Category Id – This is the unique Id of the category and does not add any value hence dropped this column.
* Department Id – This is the unique Id of the department and does not add any value hence dropped this column.
* Order Customer Id – This is the unique Id of the customer and does not add any value hence dropped this column.
* Customer Fname – This is the unique detail of the customer and does not add any value hence dropped this column.
* Customer Lname – This is the unique detail of the customer and does not add any value hence dropped this column.
* Customer City – This is the unique detail of the customer and does not add any value hence dropped this column.
* Customer Email – This is the unique detail of the customer and does not add any value hence dropped this column.
* Customer Password – This is the unique detail of the customer and does not add any value hence dropped this column.
* Customer State – This is the unique detail of the customer and does not add any value hence dropped this column.
* Customer Street – This is the unique detail of the customer and does not add any value hence dropped this column.
* Customer Zipcode – This is the unique detail of the customer and does not add any value hence dropped this column.
* Latitude – This is the unique detail of the customer location and does not add any value hence dropped this column.
* Longitude – This is the unique detail of the customer location and does not add any value hence dropped this column.
* Order Item Discount – This is the unique detail of the order and does not add any value hence dropped this column.
* Order City – This is the unique detail of the order and does not add any value hence dropped this column.
* Order Customer Id – This is the unique detail of the order and does not add any value hence dropped this column.
* Order Id – This is the unique detail of the order and does not add any value hence dropped this column.
* Order Item Cardprod Id – This is the unique detail of the order and does not add any value hence dropped this column.
* Order Item Id – This is the unique detail of the order and does not add any value hence dropped this column.
* Order Item Profit Ratio – This is the unique detail of the order and does not add any value hence dropped this column.
* Order Profit Per Order – This is the unique detail of the order and does not add any value hence dropped this column.
* Order Item Quantity – This is the unique detail of the order and does not add any value hence dropped this column.
* Order Item Total – This is the unique detail of the order and does not add any value hence dropped this column.
* Order State – This is the unique detail of the order and does not add any value hence dropped this column.
* Order Zipcode – This is the unique detail of the order and does not add any value hence dropped this column.
* Product Card Id – This is the unique detail of the product and does not add any value hence dropped this column.
* Product Category Id – This is the unique detail of the product and does not add any value hence dropped this column.
* Product Description – This is the unique detail of the product and does not add any value hence dropped this column.
* Product Image – This is the unique detail of the product and does not add any value hence dropped this column.
* Product Price – This is the unique detail of the product and does not add any value hence dropped this column.
* Product Status – This is the unique detail of the product and does not add any value hence dropped this column.

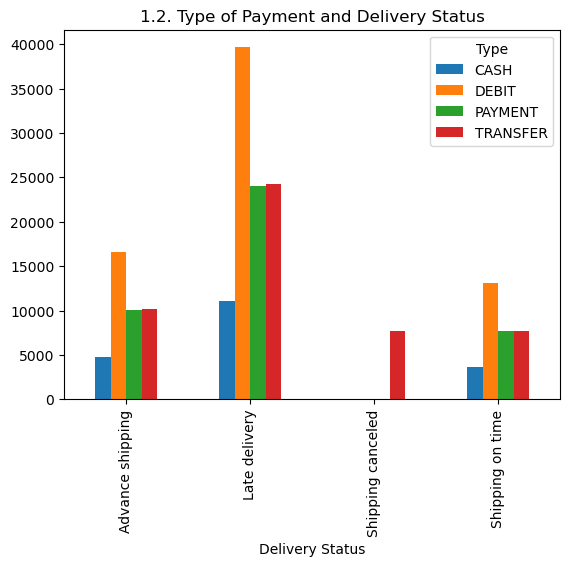
1. **Exploratory Data Analysis:**

Exploratory Data Analysis helps us understand how the features of the dataset vary for the different variables of the Supply Chain data. We first start by understanding the features alone (Univariate Analysis) and then perform bivariate and multivariate analysis to understand the data and relations between the features better.

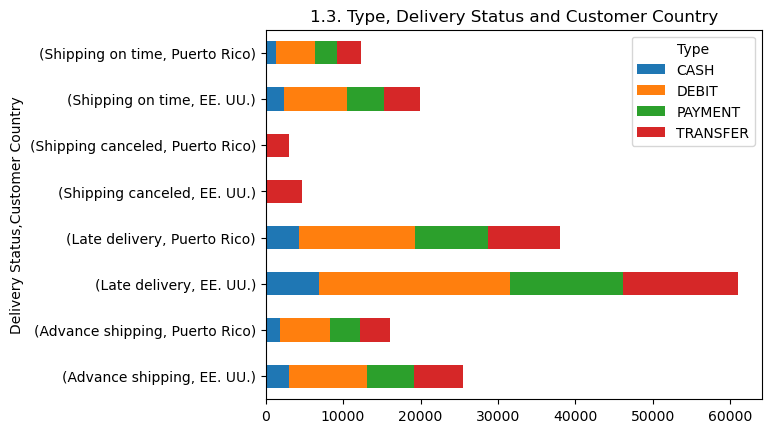
* 1. **Delivery Analysis with respect to Payment Type, Customer Country and Customer Segment:**
     1. **Distribution of Mode of Payment by Customers**

****

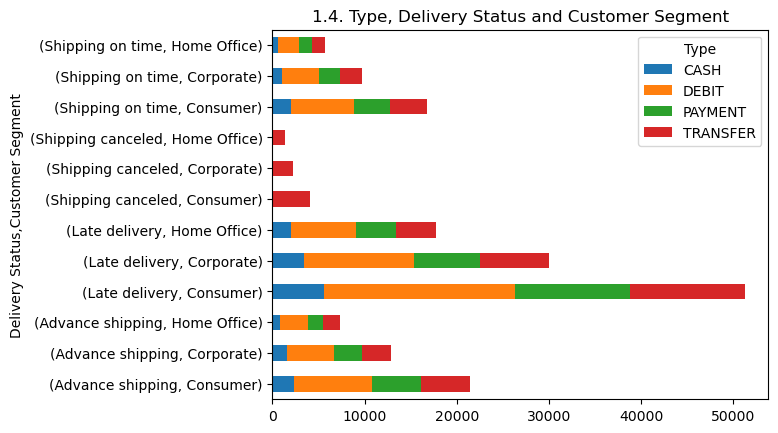
* + 1. **Distribution of Mode of Payment with respect to Delivery Status**

****

* + 1. **Distribution of Mode of Payment with respect to Delivery Status and Customer Country**

****

* + 1. **Distribution of Mode of Payment with respect to Delivery Status and Customer Segment**

****

**Observation:**

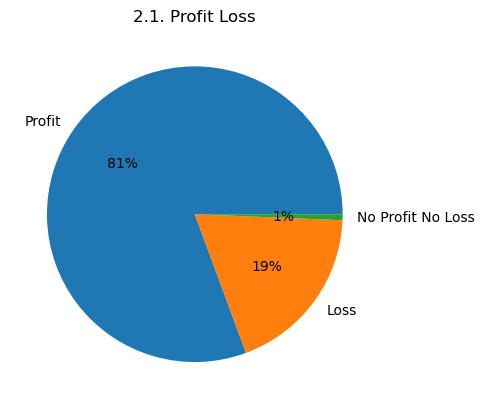
**Obs\_1:** Most of the customers have preferred to use debit cards for their purchase or order transactions.

**Obs\_2:** Shipping cancellations have happened mostly for the payment made through money transfers.

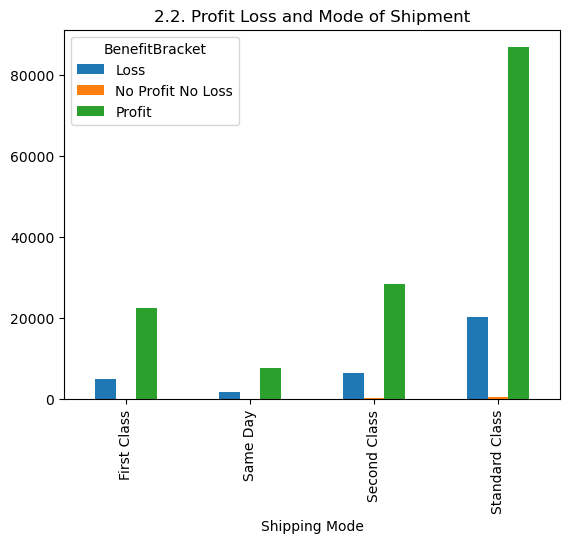
**Obs\_3:** The overall productivity or sales is more in EE.UU.

**Obs\_4:** The customers seem to be more from Consumer segment.

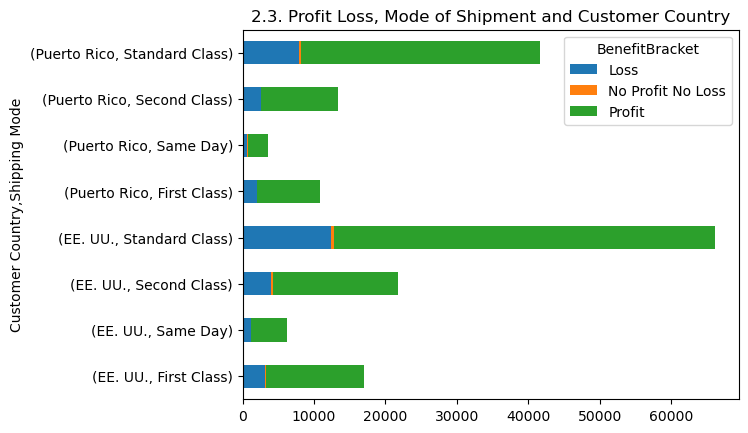
* 1. **Profit Loss Analysis with respect to Mode of Shipment, Customer Country and Customer Segment:**
     1. **Distribution of Profit and Loss per Order**

****

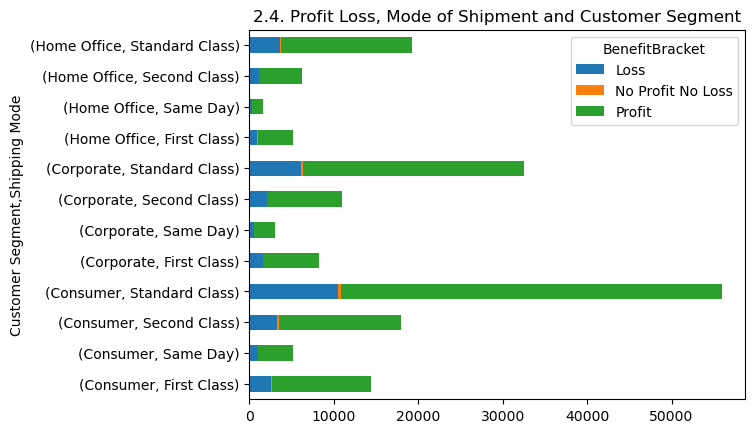
* + 1. **Distribution of Profit and Loss per Order with respect to Mode of Shipment**

****

* + 1. **Distribution of Profit and Loss per Order with respect to Mode of Shipment and Customer Country**

****

* + 1. **Distribution of Profit and Loss per Order with respect to Mode of Shipment and Customer Segment**

****

**Observation:**

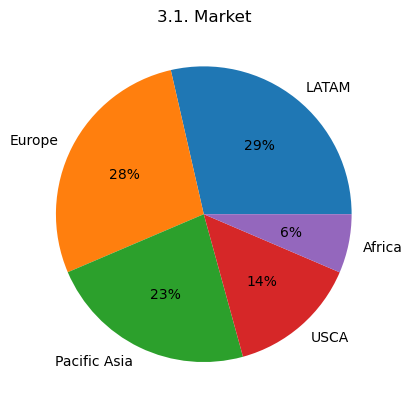
**Obs\_1:** We can see an 81% profit percentage overall in sales.

**Obs\_2:** Most customers prefer standard class rather than other mode of shipments and least of them prefer same day deliveries.

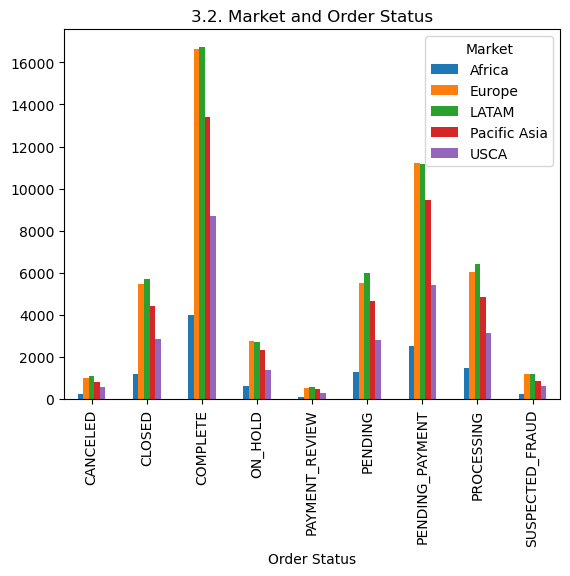
**Obs\_3:** Certain orders to both EE.UU and Puerto Rico are forced to be sold with no profit and no loss to their customers in Standard Class to meet respective sales targets.

**Obs\_4:** All same day mode of shipments is mostly managed to gain profit in all the three segments.

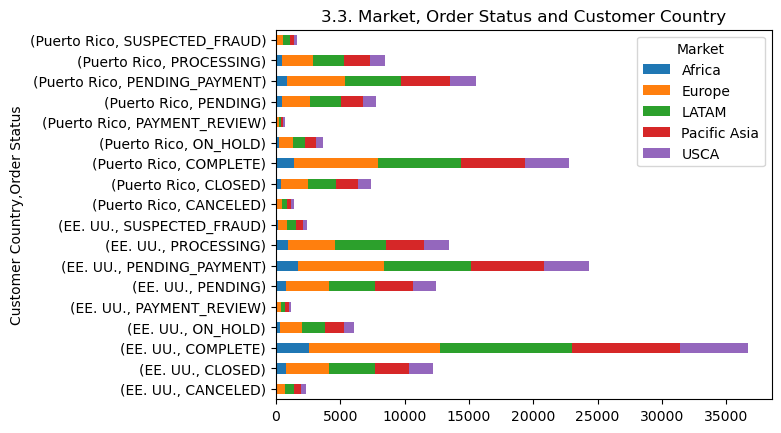
* 1. **Market Analysis with respect to Order Status, Customer Country and Customer Segment:**
     1. **Distribution of Markets**

****

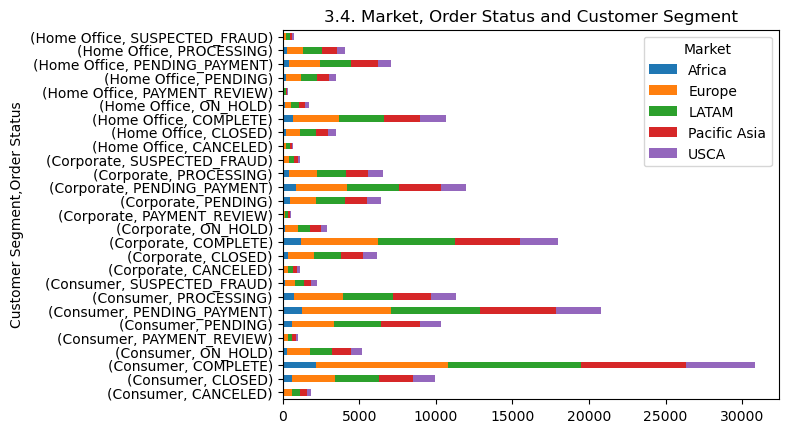
* + 1. **Distribution of Markets with respect to Order Status**

****

* + 1. **Distribution of Markets with respect to Order Status and Customer Country**

****

* + 1. **Distribution of Markets with respect to Order Status and Customer Segment**

****

**Observation:**

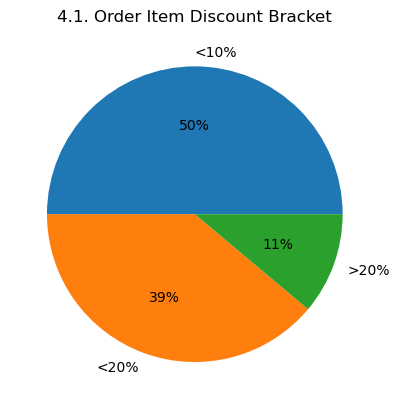
**Obs\_1:** LATAM has more market value followed by Europe with 1% behind.

**Obs\_2:** Suspected fraud sees to be more in Europe comparatively.

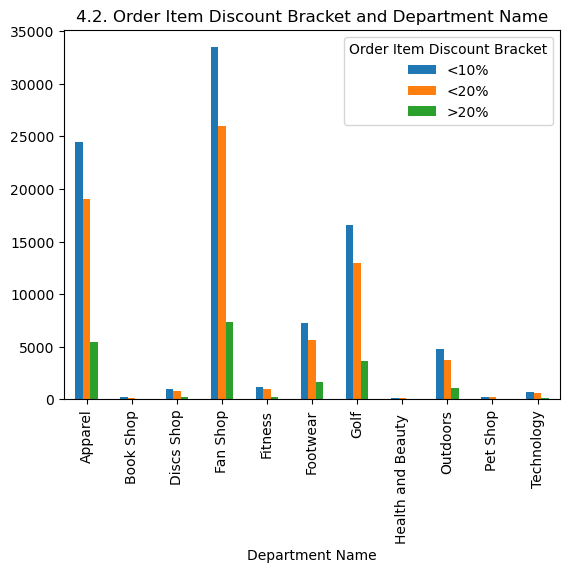
**Obs\_3:** Suspected fraud is more in EE.UU than Puerto Rico.

**Obs\_4:** USCA has a success of No Cancellation of Orders**.**

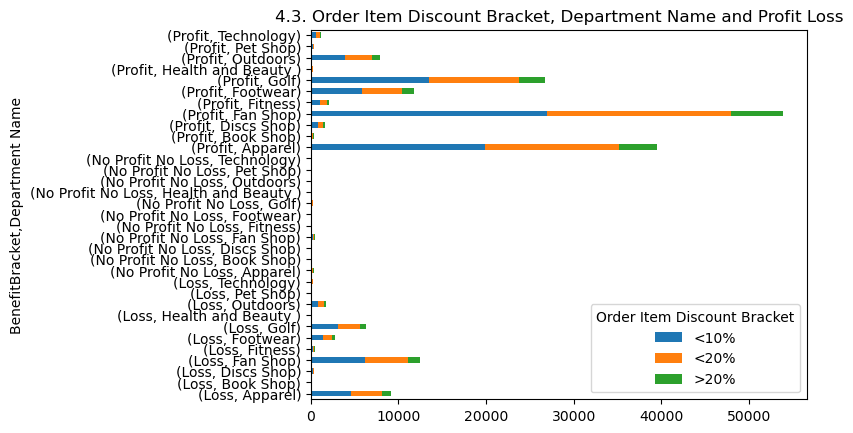
* 1. **Discount per Product Analysis with respect to Department Name and Profit Loss:**
     1. **Distribution of Product Discounts**

****

* + 1. **Distribution of Product Discounts with respect to Department Name**

****

* + 1. **Distribution of Product Discounts with respect to Department Name and Profit Loss**

****

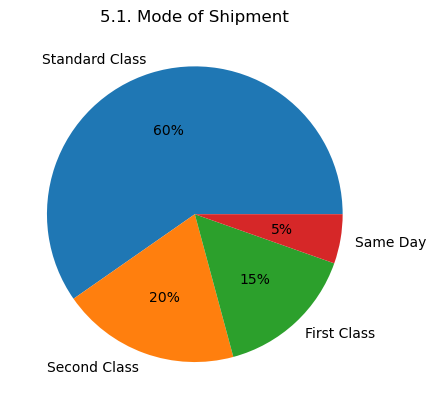
**Observation:**

**Obs\_1:** Half of the discounts range less than 10%.

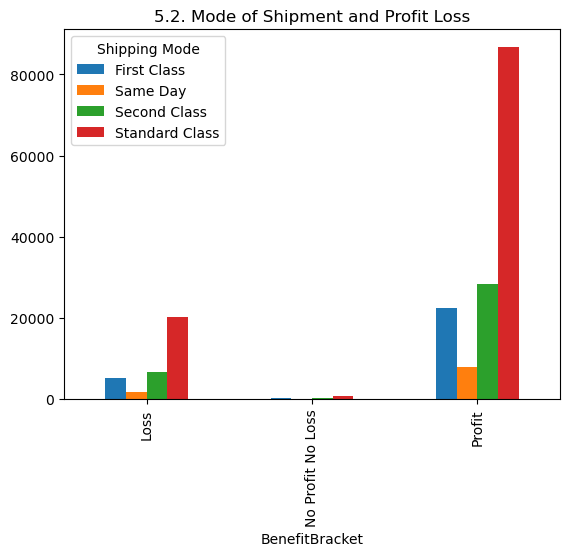
**Obs\_2:** Discounts of more than 20% is given to customers buying from Fan Shop.

**Obs\_3:** Losses with respect to Book Shop, Pet Shop, Health and Beauty, does not have any discount rates.

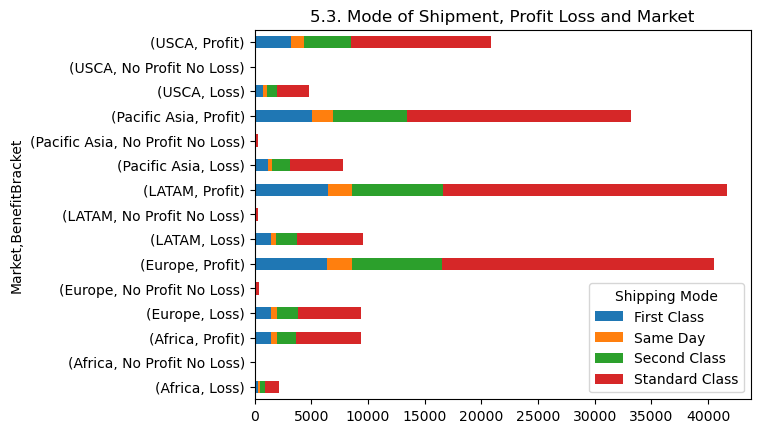
* 1. **Mode of Shipment Analysis with respect to Market and Profit Loss:**
     1. **Distribution of Mode of Shipment**

****

* + 1. **Distribution of Mode of Shipment with respect to Profit Loss**

****

* + 1. **Distribution of Mode of Shipment with respect to Profit Loss and Market**

****

**Observation:**

**Obs\_1:** Majority of sales are from Standard class customers.

**Obs\_2:** More profits are gained from Standard class customers.

**Obs\_3:** USCA and Africa has sales with either Profit or Loss but not with No Profit or Loss.

1. **Conclusion and Insights:**

The given data seem to be collected with some partiality in terms of segment and region distribution. The variables from the data are considered keeping in mind their impact on the analysis.

The suggestions to the VP of Logistics and VP of Sales is that:

1. Focus on Suspected Fraud category and follow up on the backend details regarding the fraud.
2. More offers or additional privileges can be given to Consumers as they take part in major profits to the company.
3. Offers and benefits can be implemented to Debit type of transactions.
4. Should worry about the shipment cancellation, since it happens only in situations where the mode of transaction is transfer based.
5. The same day shipment can be advertised more and can be focused on to improvise the perks.
6. **References:**

<https://medium.com/mlearning-ai/univariate-bivariate-and-multivariate-data-analysis-in-python-341493c3d173>

<https://www.tutorialspoint.com/python-display-only-non-duplicate-values-from-a-dataframe>

<https://www.digitalocean.com/community/tutorials/exploratory-data-analysis-python>

<https://towardsdatascience.com/exploratory-data-analysis-in-python-a-step-by-step-process-d0dfa6bf94ee>

<https://www.analyticsvidhya.com/blog/2022/07/step-by-step-exploratory-data-analysis-eda-using-python/>

Kaggle Code links for this dataset:

<https://www.kaggle.com/code/shashankmkadiwal/dataco-smart-supply-chain-eda>

<https://www.kaggle.com/code/divyeshardeshana/supply-chain-shipment-price-data-analysis>

<https://www.kaggle.com/code/jhurley/walmart-supplychain-eda-kk/notebook>

Top of Form

Bottom of Form