ENNOVATION 3.0

TEAM: Dracarys

Bijin Jiby

Anveshika Kamble

Arunav Chandra

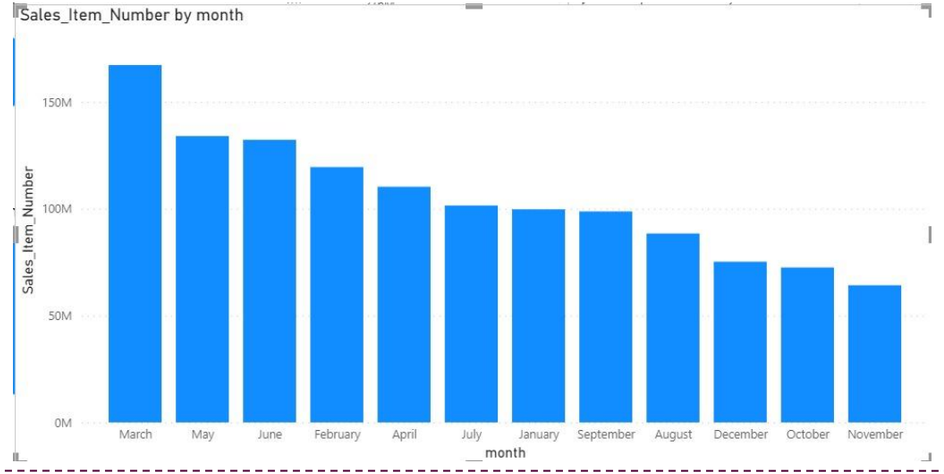
Data Analysis

Customer Insights :

Total Products : 3428

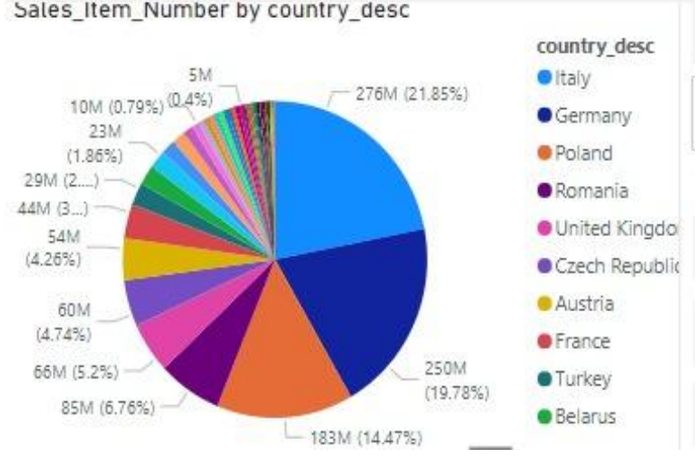
- ***March*** accounted for ***13.23%*** of Sales_Item_Number.
- ***Distributor*** accounted for ***70.25%*** of Sales_Item_Number.
- ***Italy*** accounted for ***21.85%*** of Sales_Item_Number.
- ***Finished products*** accounted for ***97.73%*** of Sales_Item_Number.
- ***Qtr 1*** had the highest Sales_Item_Number at 386531150, followed by Qtr 2, Qtr 3, and Qtr 4.
- ***Qtr 1*** accounted for 30.58% of Sales_Item_Number.

Sales by Month ::

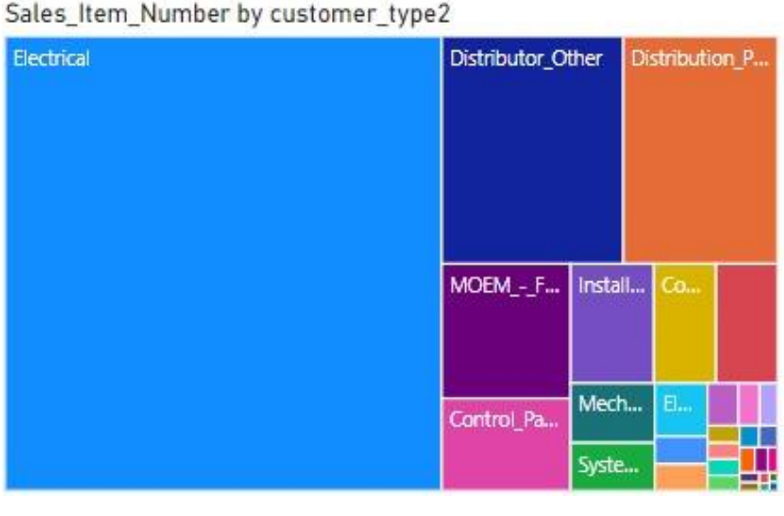


Customer Demographics

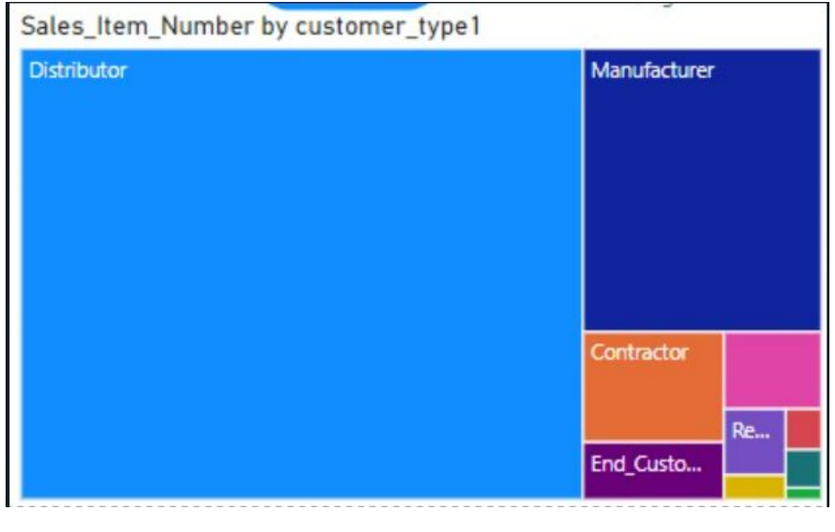
Country:



Customer Type 2:



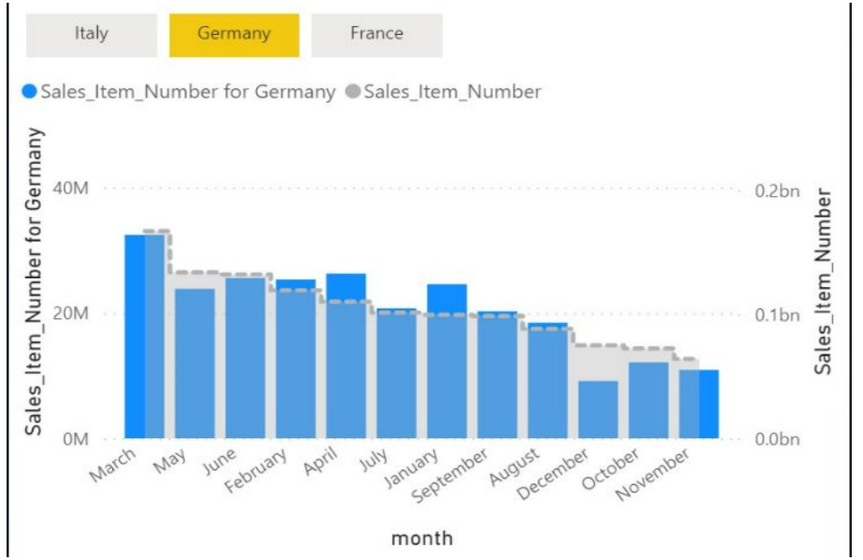
Customer Type 1:



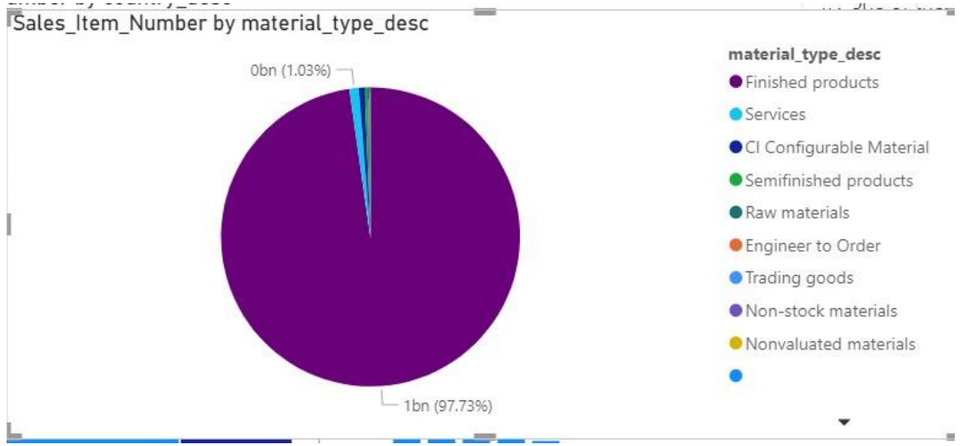
Data Analysis :

Analysis By month and Geography:

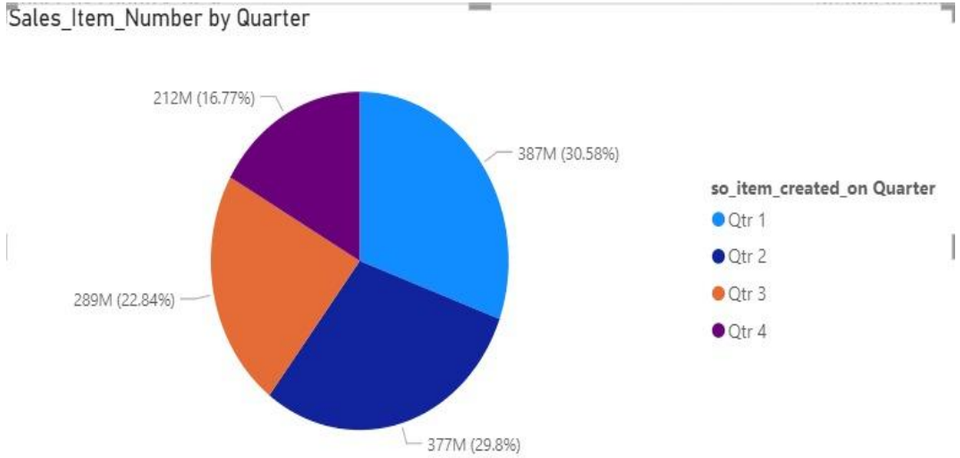
- '***Italy'*** with ***12.7*** percent of records,***'Germany'*** with ***32.8 %*** of records and ***France*** with ***6.7 %*** of records , others do not affect the distribution of sales.



Sales by Material Typet



Sales by Quarter :

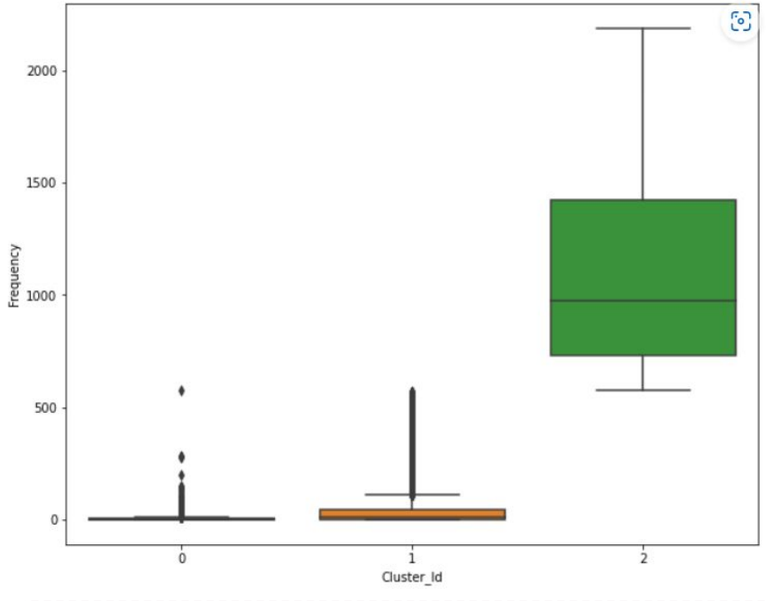


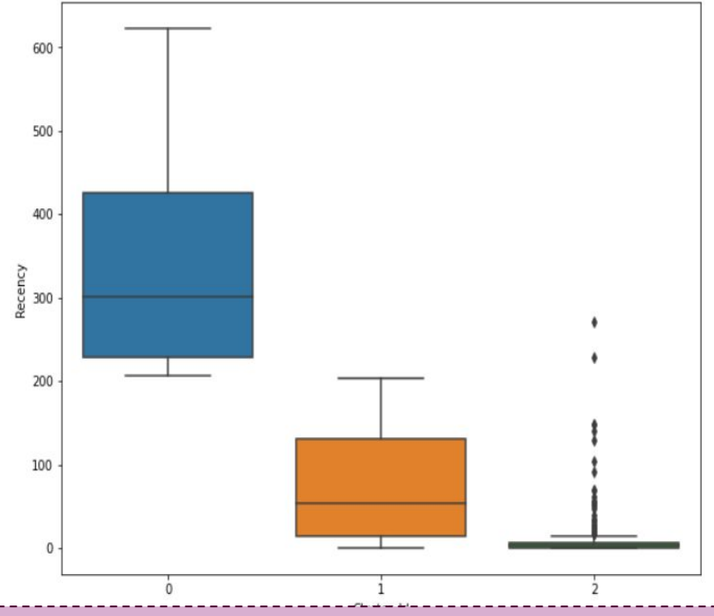
Customer segmentation

Customer segmentation was using k-means and hierarchical clustering performed on the group of customers w.r.t to the frequency as well as recency. (reduced variable attributes)

Findings through K-means:

	Customer_Group	Frequency	Recency	Cluster_Id
1	2	6	89	1
2	3	23	68	1
3	4	2	253	0
4	5	5	105	1
6	7	290	4	1





- **3 different clusters with 3 different customer Ids are formed (0,1,2)**
- **Each cluster represents behavioural pattern of different customer_groups which fall under them**

➔ Customer groups(2,3,5,7…) having **cluster Id as 1 and 0** are **very less frequent** in placing a sales order.On contrary,Customer groups having **cluster Id-2** are **more frequent** in placing orders.

➔ Customer groups having **cluster Id 1 and 2** are **recent buyers**.

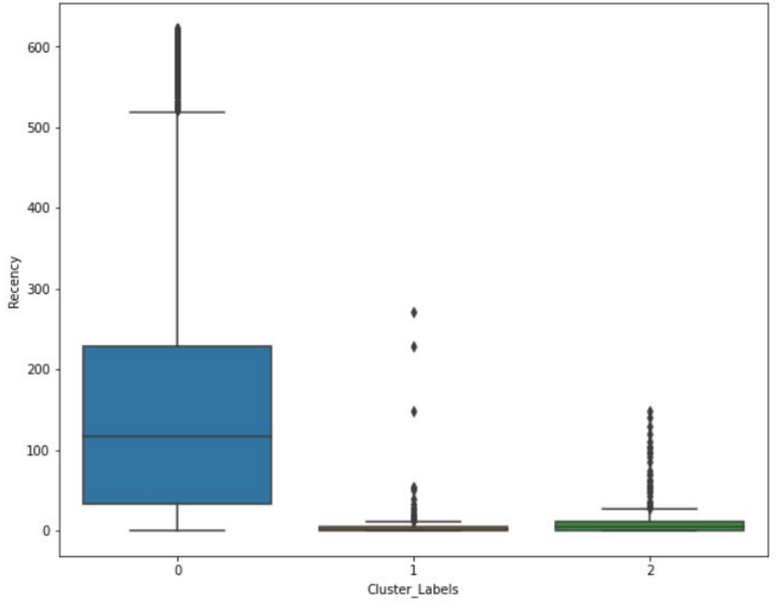
➔ Customers groups with Cluster Id 0 are not recent buyers and hence least of importance from a business point of view.

Findings through hierarchical clustering:

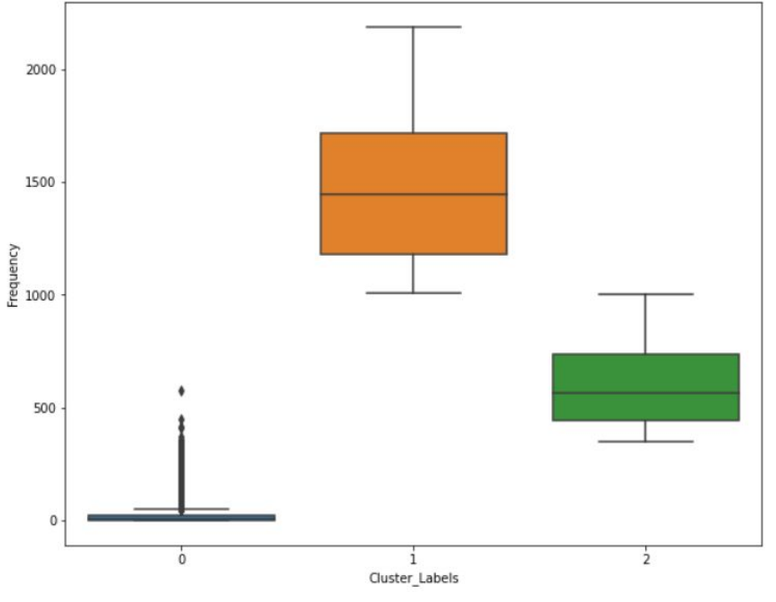
	Customer_Group	Frequency	Recency	Cluster_Id	Cluster_Labels
1	2	6	89	1	0
2	3	23	68	1	0
3	4	2	253	0	0
4	5	5	105	1	0
6	7	290	4	1	0

- 3 different clusters with 3 different customer labels are formed (0,1,2)
- Each cluster represents behavioural pattern of different customer_groups which fall under them

Customer labels w.r.t frequency



Customer labels w.r.t recency



➔ Customer groups with Cluster_Labels 1 and 2 are frequent buyers.

➔ Customers with group Cluster_Labels 0 are not recent buyers and hence least of importance from a business point of view.

Algorithm :

- ➔ We used apriori algorithm to find frequent items bought together
- ➔ To filter the best recommendations we will use the highest confidence value for each antecedent.
- ➔ We have used reduced dataset due to limited resources

length	count	mean	std	min	25%	50%	75%	max
1	4.0	29.167500	29.938902	3.82	10.5175	20.765	39.415	71.32
2	3.0	10.886667	6.493854	6.90	7.1400	7.380	12.880	18.38
3	1.0	6.230000	NaN	6.23	6.2300	6.230	6.230	6.23

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
5	(10)	(20, 30)	0.713196	0.068998	0.062321	0.087383	1.266449	0.013112	1.020145
6	(20)	(10, 30)	0.287758	0.073768	0.062321	0.216575	2.935893	0.041094	1.182285
2	(10, 20)	(30)	0.183784	0.127504	0.062321	0.339100	2.659528	0.038888	1.320164
7	(30)	(10, 20)	0.127504	0.183784	0.062321	0.488778	2.659528	0.038888	1.596599
3	(10, 30)	(20)	0.073768	0.287758	0.062321	0.844828	2.935893	0.041094	4.590002
4	(20, 30)	(10)	0.068998	0.713196	0.062321	0.903226	1.266449	0.013112	2.963646

```
antecedents
(20)        2
(30)        2
(10, 20)    1
(10, 30)    1
(20, 30)    1
(10)        1
dtype: int64
```