MARKETING & RETAIL ANALYSIS

Milestone 1 - DSBA

By,

Pathanjali Srinivasan.

Content

- Agenda & Executive Summary of the data
- Exploratory Data Analysis and Inferences
- RFM Segmentation
- Customer Segmentation using RFM analysis
- KNIME Workflow & RFM Analysis
- Inferences from RFM Analysis and identified segments

Agenda

Agenda of this project is to find the underlying buying patterns of the customers of an automobile part manufacturer. based on the last 3 years of the Company's transaction data and recommend them customized marketing strategies for different segments of customers.

* Reference image from Google image search result

Executive Summary

We have received the 3 years data of automobile part manufacture. Consisting 2747 entries with 20 variable details regarding the demography of the product and customer information.

* Reference image from Google image search result

Problem Statement

An automobile parts manufacturing company has collected data of transactions for 3 years. They do not have any in-house data science team, thus they have hired us as their consultant. Our job is to use your magical data science skills to provide them with suitable insights about their data and their customers.

ORDERNUMBER :	Order Number	CUSTOMERNAME :	customer
QUANTITYORDERED :	Quantity ordered	PHONE :	Phone of the customer
PRICEEACH :	Price of Each item	ADDRESSLINE1 :	Address of customer
ORDERLINENUMBER :	order line	CITY :	City of customer
SALES :	Sales amount	POSTALCODE :	Postal Code of customer
ORDERDATE :	Order Date	COUNTRY :	Country customer
DAYS_SINCE_LASTORDER :	Days_ Since_Lastorder	CONTACTLASTNAME :	Contact person customer
STATUS :	Status of order like Shipped or not	CONTACTFIRSTNAME :	Contact person customer
PRODUCTLINE :	Product line – CATEGORY	DEALSIZE :	Size of the deal based on Quantity and Item Price
MSRP :	Manufacturer's Suggested Retail Price		
PRODUCTCODE :	Code of Product		

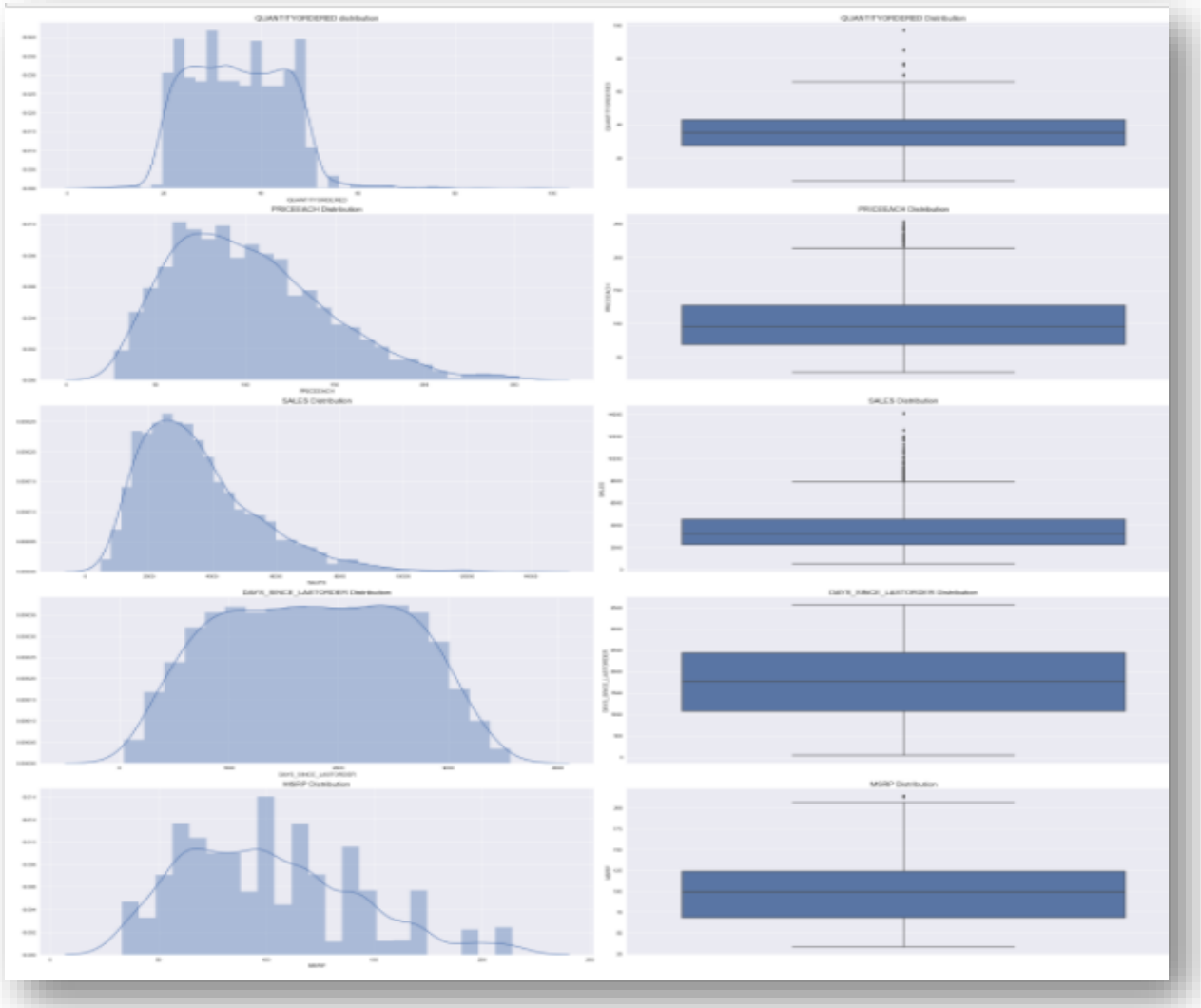
About Data

- The data is about an automobile parts manufacturing company. They have provided the data collected of transactions for last 3 years.
- The data has 2747 entries (0 To 2746) of rows and 20 columns. The data has 1 datetime64 , 2 foat64, 5 int64,and 12 Object data types.
- The dataset has no null values and no duplicate rows of data.
- This data more or less reflects the purchasing behavior of customers in different categories . The company is into automobile part manufacture, and they have different product line like Classic car , Motorcycle, plane, train, ship, Bus truck, vintage cars etc.
- The data maintained each transactions entry as order number and for each order number maintained all required information like customer identity details , and product details like price , quantity , product code, and sales for each customer.
- We noticed that one order number has many different entries with different product codes.
- Manufacturer's Suggested Retail Price(MSRP) for each product code is decided but we found that this is not matching with Price of Each item & is inconsistent with MSRP

Exploratory Data Analysis

Distplot and Boxplot inference:

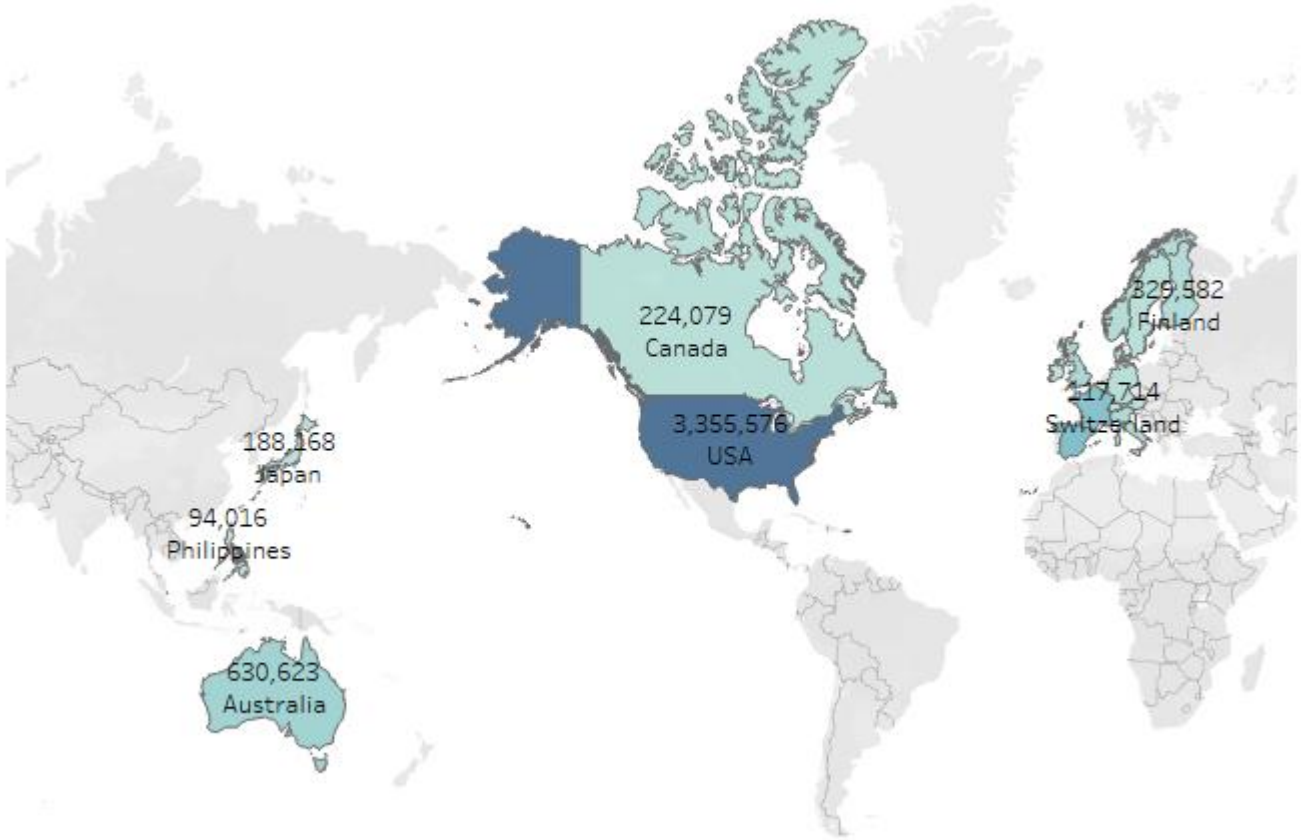
- There is presence of outliers in variables such as Quantity ordered, Price and Sales.
- Variable ‘Sales’ has highest positive skewness (0.784) and Variable ‘Days_since_lastorder’ has lowest negative skewness(-0.002)
- We can clearly see that outlier is present in each product line category and it is present in large deal size.



Exploratory Data Analysis

Bivariate Analysis:

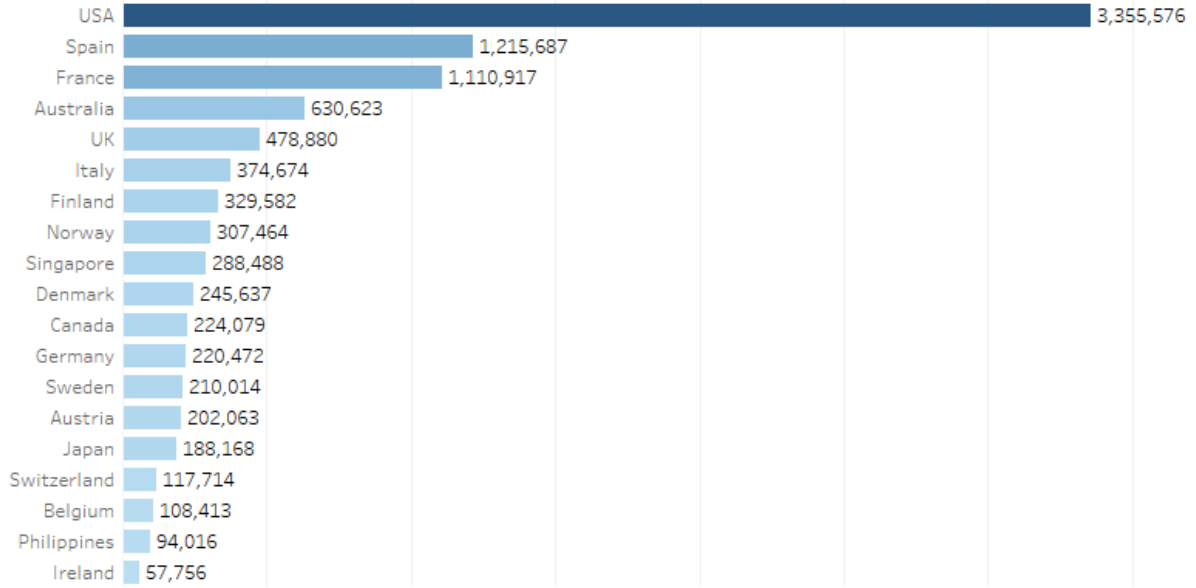
- Company is having a viable market in almost 19 countries
- This graph shows the sales across different countries
- USA is the primary market of the company contributing maximum to its turn-over
- Some European countries like Spain and Switzerland shows good number of sales following USA

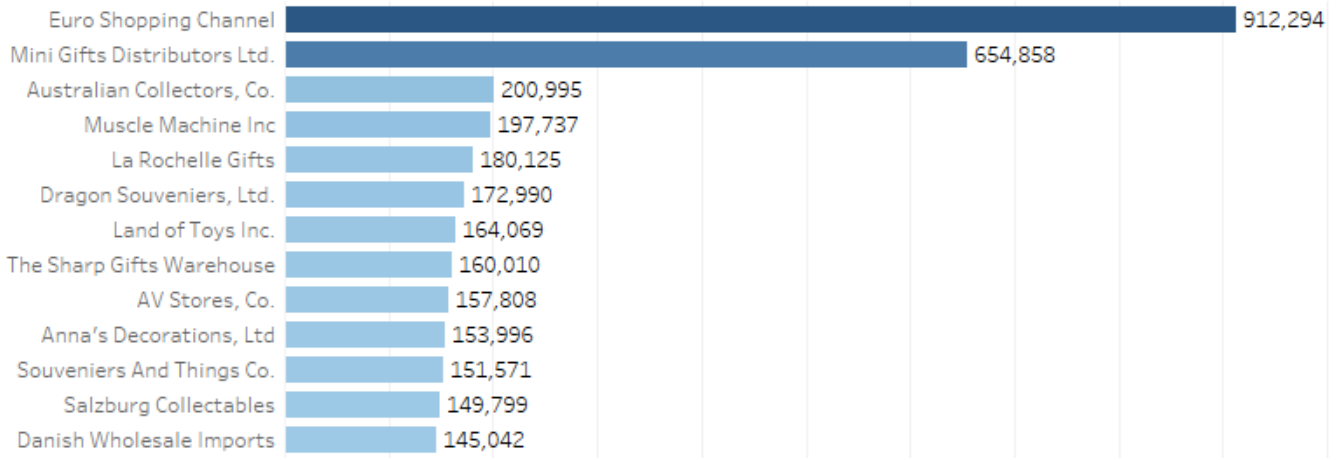


Exploratory Data Analysis

Bivariate Analysis:

- USA is having most orders shipped and high sales with 3 on hold and 1 cancelled.
- This is followed by France, and so on order of decreasing sales.
- ‘Euro Shopping channel’ is the customer having highest sales and orders shipped followed by ‘Mini Gifts Distributers ltd’.
- ‘Euro Shopping Channel’ is also having orders in cancelled, Disputed status – again an inference for company to check

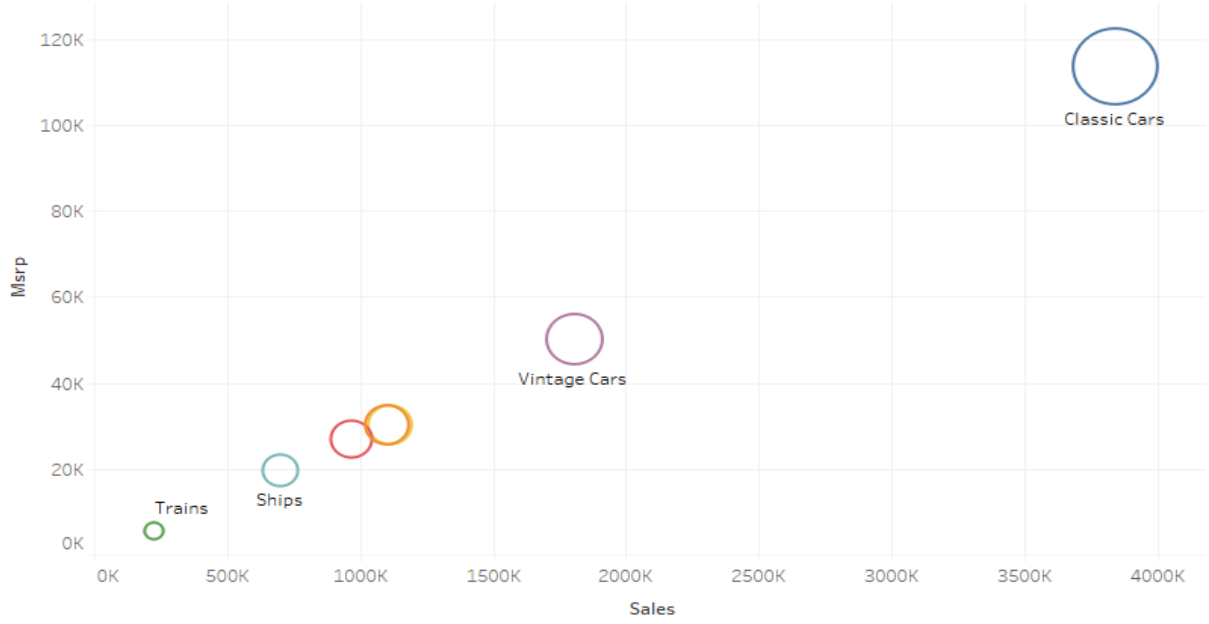


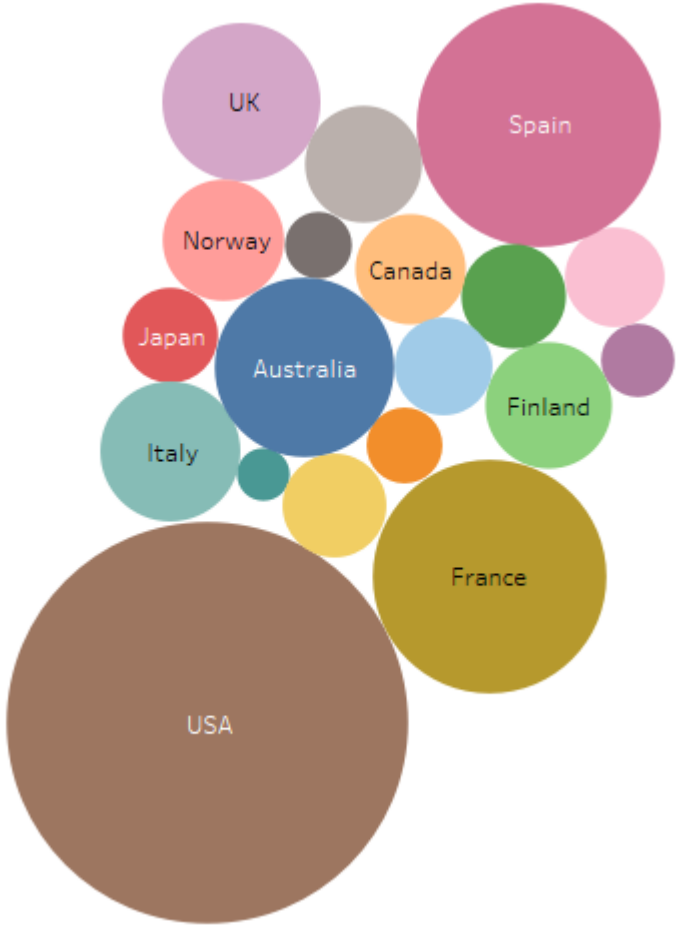


Exploratory Data Analysis

Bivariate Analysis:

- ‘Classic cars’ are visibly the most selling product line followed by ‘Vintage cars’ and so on..
- Countries that have highest orders are USA, followed by Spain and France.
- Ireland and Philippines are the least consuming countries.

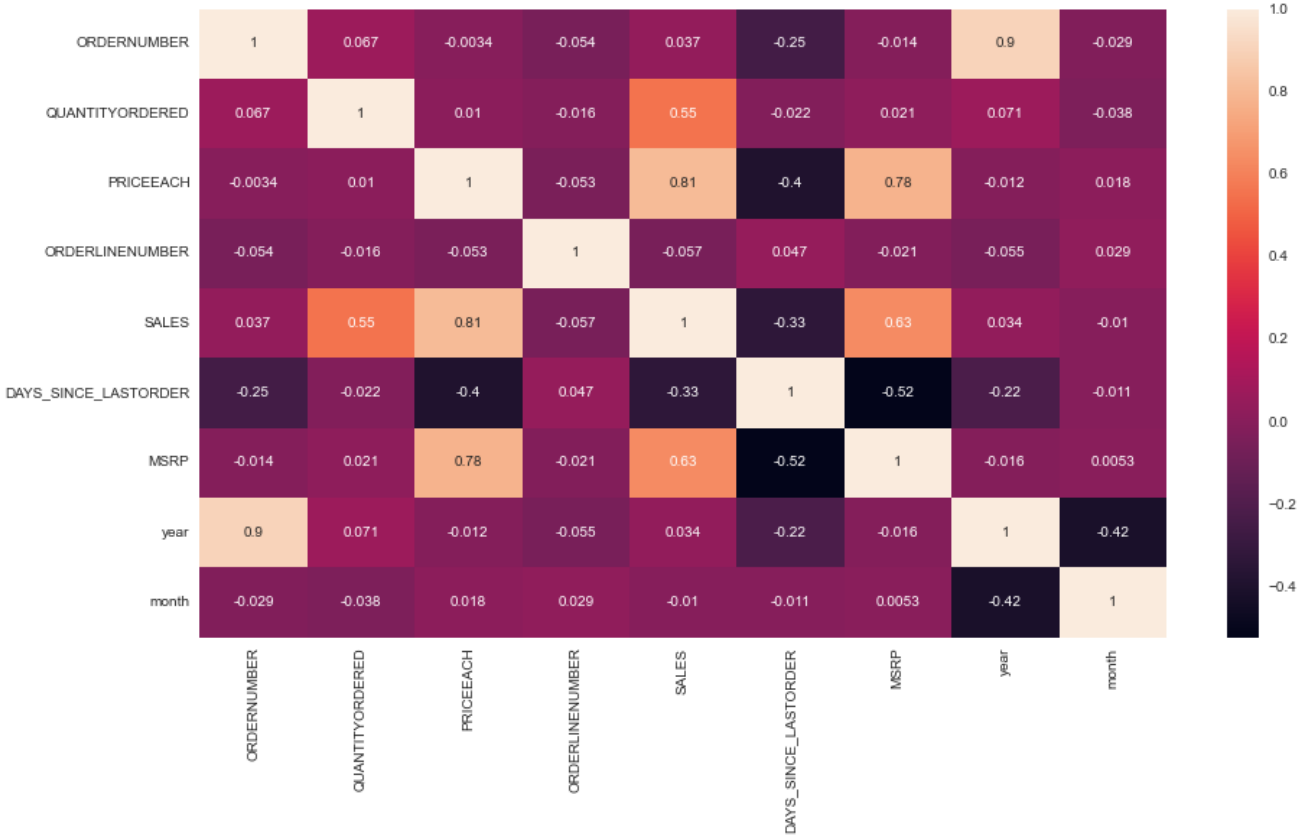




Exploratory Data Analysis

Multivariate analysis

The heatmap shows the high correlation between two variables SALES and PRICEEACH. Also variable MSRP and PRICEEACH have a high correlation shown in above heatmap



Exploratory Data Analysis

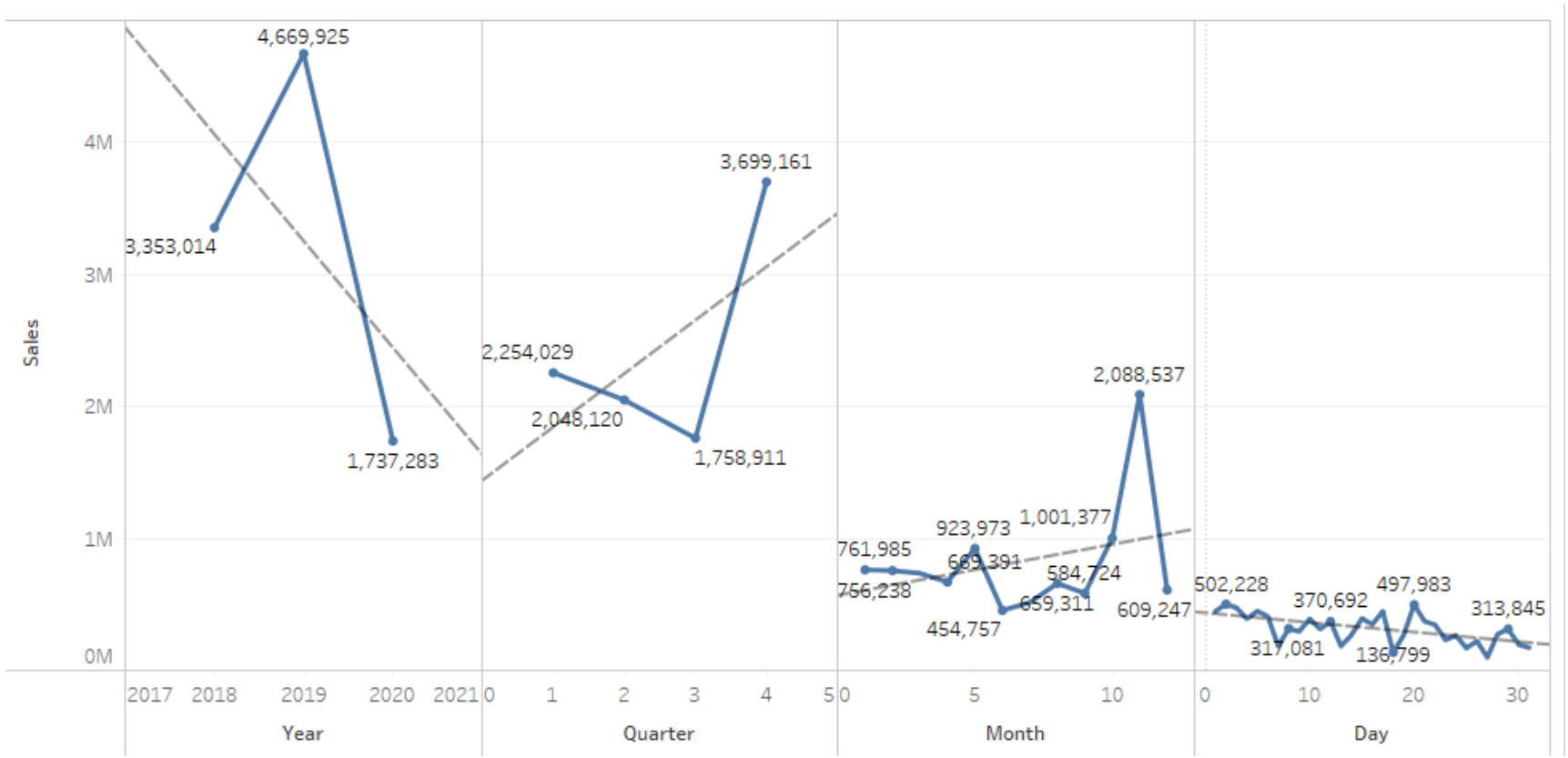
Multivariate analysis inference:

- Variables ‘Sales’ and ‘Price each’ have highest positive Correlation(0.81) and Variables ‘Days_since_lastorder’ and ‘MSRP’ have highest negative Correlation(-0.52)
- From the previous plots , we can see that none of the variables are symmetric.
- Variables ‘Sales’ and ‘Price each’ have almost linear relation ship between them.
- From the plots , we can see there is higher spread of data along the trend line for MSRP compared to Price each so we need to maximize sales by identifying respective items for which there is higher price change.
- The pair plot shows the data in most variables are normally distributed
- There’s visible correlation between Quantity Ordered, Sales, MSRP and Price each.
- Other variables don’t show any signs of correlation

Exploratory Data Analysis

Sales trend per different time period :

- Yearly trend in sales decreasing in 2020 from 2018 with 2019 having highest sales.
- Quarterly & Monthly sales having increasing trend with seasonality – indicating sales increases in Q4



Inferences & Summary

- The sales of large size deal is almost remain stagnant over the years and it can be presumed that the company should focus on getting large size chunk projects.
- Maximum number of deals are medium type of deal size
- Deals having larger deal size across sales are minimum as compared to medium and smaller deal size
- In sales across product line classic cars have higher number of sales as compared to other cars
- The sales of product line like Trucks, buses and motorcycles is in saturation stage and there is very less scope in them.
- The major chunk of sales comes from 4 – 5 customers therefore company should focus more on customer scouting in a rational way because incase there is client churn's it will impact the sales of the company greatly.
- Company have higher amount of Sales in country USA as compare to other country.
- Euro shopping channel is the most loyal customer.
- Company drives export revenue mainly from large size deals followed by medium size deals and small size deals therefore we can infer that large size deals is mainly from foreign customers and company should focus on domestic sales as well.

RFM Segmentation

- Recency, frequency, monetary value (RFM) is a marketing analysis tool used to identify a firm's best clients based on the nature of their spending habits.
- An RFM analysis evaluates clients and customers by scoring them in three categories: how recently they've made a purchase, how often they buy, and the size of their purchases.
- RFM analysis helps firms reasonably predict which customers are likely to purchase their products again, how much revenue comes from new (versus repeat) clients, and how to turn occasional buyers into habitual ones.
- Customer Segmentation is done by using KNIME & MS Excel by dividing the data based on Recency, Frequency and Monetary variables by grouping data by variable ‘Order Number’.

RFM Segmentation

- Recency (R) is the most recent customer order which is calculated by taking difference of Order date & current date in Days.

  RECENCY in Days = ORDERDATE – Current Date

- Frequency (F) is how often the orders are placed by customers, from the excel sheet the variable

  DAYS_SINCE_LASTORDER.

- Monetary (M) sales can be used and we used the calculation of price & Quantity:

  Monetary = QUANTITYORDERED * PRICEEACH

Following assumptions are made for RFM segmentation:

- Assumption 1 – All the prices & sales figure are in same currency.
- Assumption 2 – Sales & (Quantity * Price) may or may not be same.

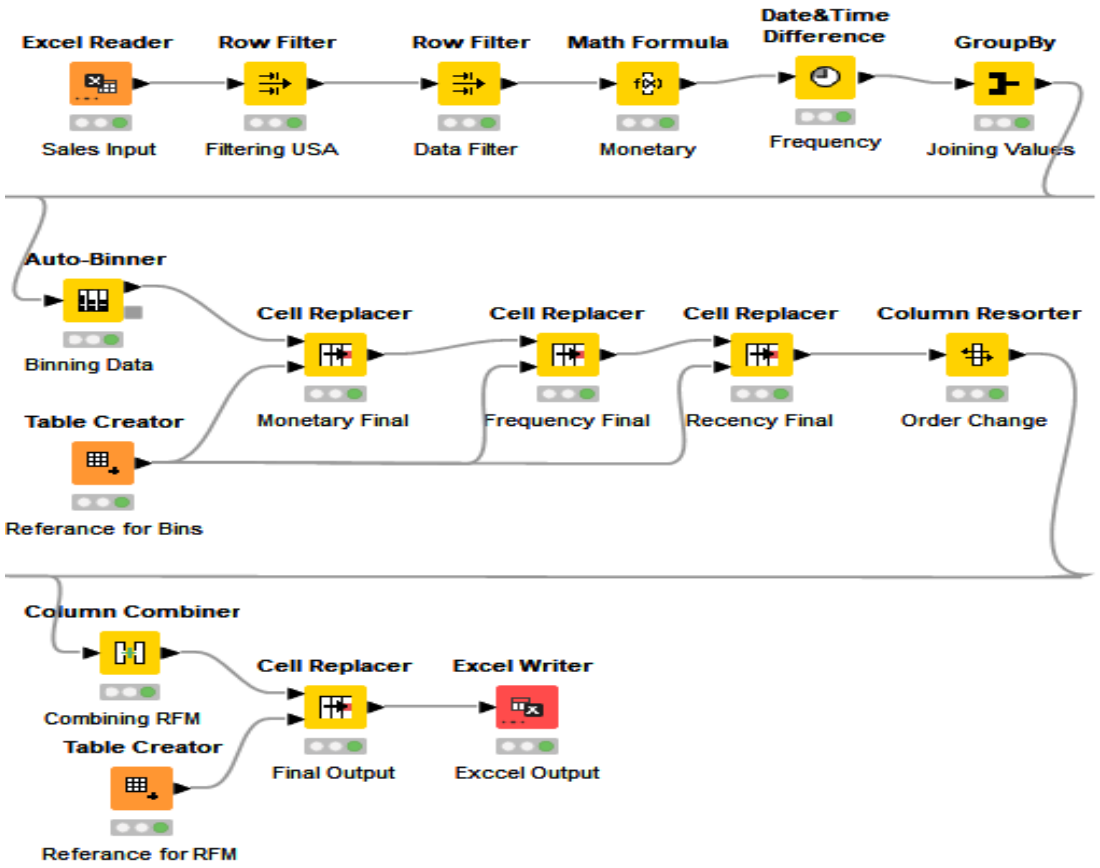
Customer Segmentation using RFM analysis

- A new column is created called Recency based on the formula mentioned previously.
- In sales column we get sales amount for each transaction. We can use sales parameter and using an assumption of sum of aggregation which we created as a new column called Monetary
- Then we created 3 different bin for each Recency, frequency & Monetary using percentile range (0.0, 0.25, 0.75, 1.0). Based on 3 bin assumption we have considered segments like Low (L), Medium (M) & High (H).
- Finally based on the combination assumption we have considered 4 segments like Loyal , Best , Lost and the customers on the verge of Churn.

	column1	column2
Row0	HHH	Best Customer
Row1	HMH	Best Customer
Row2	HML	Best Customer
Row3	HMM	Best Customer
Row4	MHH	Best Customer
Row5	MHM	Loyal Customer
Row6	MLM	Loyal Customer
Row7	MMH	Loyal Customer
Row8	MMM	Loyal Customer
Row9	MML	Loyal Customer
Row10	HLM	Loyal Customer
Row11	LMM	Customer on ...
Row12	LHH	Customer on ...
Row13	LMH	Customer on ...
Row14	MLL	Lost Customer
Row15	HLL	Lost Customer
Row16	LLL	Lost Customer
Row17	LLM	Lost Customer

KNIME Workflow

KNIME has been used in areas such as CRM customer data analysis, business intelligence, text mining and financial data analysis or modeling, data analysis and visualization without, or with only minimal, programming.



KNIME Workflow

Final Output of the Workflow

Table "default" - Rows: 32 | Spec - Columns: 30 | Properties | Flow Variables

Row ID	ZE	D Monetary	L Recency	S ORDER...	S Moneta...	S Recenc...	S Recenc...	S Freque...	S Monete...	S RFM	S Custom...
Row0		26,479.26	928	Bin 1	Bin 1	Bin 2	M	L	L	MLL	Lost Customer
Row1		9,129.35	861	Bin 1	Bin 1	Bin 1	H	L	L	HLL	Lost Customer
Row2		36,163.62	1137	Bin 1	Bin 1	Bin 3	L	L	L	LLL	Lost Customer
Row3		67,506.97	978	Bin 2	Bin 2	Bin 2	M	M	M	MMM	Loyal Custo...
Row4		77,795.2	940	Bin 2	Bin 2	Bin 2	M	M	M	MMM	Loyal Custo...
Row5		87,489.23	1208	Bin 2	Bin 2	Bin 3	L	M	M	LMM	Customer on...
Row6		81,577.98	880	Bin 2	Bin 2	Bin 2	M	M	M	MMM	Loyal Custo...
Row7		122,138.14	749	Bin 3	Bin 3	Bin 1	H	H	H	HHH	Best Customer
Row8		70,859.78	1149	Bin 2	Bin 2	Bin 3	L	M	M	LMM	Customer on...
Row9		98,923.73	837	Bin 2	Bin 2	Bin 1	H	M	M	HMM	Best Customer
Row10		101,894.79	774	Bin 2	Bin 2	Bin 1	H	M	M	HMM	Best Customer
Row11		57,294.42	927	Bin 2	Bin 1	Bin 2	M	M	L	MML	Loyal Custo...
Row12		83,209.88	773	Bin 2	Bin 2	Bin 1	H	M	M	HMM	Best Customer
Row13		164,069.44	946	Bin 3	Bin 3	Bin 2	M	H	H	MHH	Best Customer
Row14		103,080.38	979	Bin 2	Bin 2	Bin 3	L	M	M	LMM	Customer on...
Row15		33,144.93	958	Bin 1	Bin 1	Bin 2	M	L	L	MLL	Lost Customer
Row16		85,555.99	977	Bin 2	Bin 2	Bin 2	M	M	M	MMM	Loyal Custo...
Row17		108,951.13	893	Bin 3	Bin 3	Bin 2	M	H	H	MHH	Best Customer
Row18		654,858.06	750	Bin 3	Bin 3	Bin 1	H	H	H	HHH	Best Customer
Row19		83,682.16	944	Bin 2	Bin 2	Bin 2	M	M	M	MMM	Loyal Custo...
Row20		197,736.94	930	Bin 3	Bin 3	Bin 2	M	H	H	MHH	Best Customer
Row21		131,685.3	957	Bin 3	Bin 3	Bin 2	M	H	H	MHH	Best Customer
Row22		57,197.96	1012	Bin 1	Bin 1	Bin 3	L	L	L	LLL	Lost Customer
Row23		50,218.51	1224	Bin 1	Bin 1	Bin 3	L	L	L	LLL	Lost Customer

Inferences from RFM Analysis and identified segments

Best Customers

CUSTOMERNAME	ORDERI	QUANTITYORI	PRICEEA	SALES	ORDERI	DAYS_S	PRODUCTLIN	MSRP	PRODU	COUNTI	CONTAC	DEALSIZ	Moneta	Recenc	RFM	Custom	Category
Diecast Classics Inc.	31	35.83870968	108.5658	3939.94	31	31	Motorcycles	31	S10_1678	USA	Yu	Medium	122138.1	749	HHH	Best Customer	
FunGiftIdeas.com	26	34.73076923	109.5865	3804.759	26	26	Motorcycles	26	S10_1678	USA	Benitez	Medium	98923.73	837	HMM	Best Customer	
Gift Depot Inc.	25	36.12	108.9324	4075.792	25	25	Motorcycles	25	S10_1678	USA	King	Medium	101894.8	774	HMM	Best Customer	
Gifts4AllAges.com	26	35.88461538	91.56385	3200.38	26	26	Classic Cars	26	S10_4757	USA	Yoshido	Medium	83209.88	773	HMM	Best Customer	
Land of Toys Inc.	49	33.28571429	104.1206	3348.356	49	49	Motorcycles	49	S10_1678	USA	Yu	Small	164069.4	946	MHH	Best Customer	
Mini Creations Ltd.	35	32.57142857	95.12914	3112.889	35	35	Classic Cars	35	S10_4757	USA	Tam	Medium	108951.1	893	MHH	Best Customer	
Mini Gifts Distributors Ltd.	180	35.36666667	102.6963	3638.1	180	180	Classic Cars	180	S10_1949	USA	Nelson	Large	654858.1	750	HHH	Best Customer	
Muscle Machine Inc	48	36.97916667	111.1508	4119.52	48	48	Classic Cars	48	S12_1108	USA	Young	Large	197736.9	930	MHH	Best Customer	
Online Diecast Creations Co.	34	36.70588235	108.3021	3873.097	34	34	Classic Cars	34	S10_1949	USA	Young	Medium	131685.3	957	MHH	Best Customer	
Technics Stores Inc.	34	34.67647059	104.9141	3552.443	34	34	Motorcycles	34	S10_1678	USA	Hirano	Medium	120783.1	895	MHH	Best Customer	
Tekni Collectables Inc.	21	43.14285714	93.57095	3963.247	21	21	Motorcycles	21	S10_1678	USA	Brown	Medium	83228.19	806	HMM	Best Customer	
The Sharp Gifts Warehouse	40	41.4	93.37575	4000.257	40	40	Classic Cars	40	S10_4757	USA	Frick	Large	160010.3	787	HHH	Best Customer	

- On basis on Recency, frequency & monetary we have grouped our top customers.
- We have given the most significance to recency parameter as these customers has recently purchased our products.
- Also according to RFM model the most importance is given to recency. Hence we have kept it as our first parameter for selecting top customers.
- We should not loose these customers at any cost as they are the biggest contributors of the business.

Inferences from RFM Analysis and identified segments

Loyal Customers

CUSTOMERNAME	ORDERI	QUANTITYORI	PRICEEA	SALES	ORDERI	DAYS_S	PRODUCTLIN	MSRP	PRODU	COUNT	CONTA	DEALSIZ	Moneta	Recenc	RFM	Custom	Categ
Classic Gift Ideas, Inc	21	31.80952381	103.3205	3214.618	21	21	Classic Cars	21	S10_1949	USA	Cervantes	Medium	67506.97	978	MMM	Loyal Customer	
Classic Legends Inc.	20	36	109.8035	3889.76	20	20	Classic Cars	20	S10_1949	USA	Hernandez	Medium	77795.2	940	MMM	Loyal Customer	
Collectables For Less Inc.	24	33.125	97.23708	3399.083	24	24	Classic Cars	24	S10_1949	USA	Nelson	Medium	81577.98	880	MMM	Loyal Customer	
Gift Ideas Corp.	19	35.05263158	87.6	3015.496	19	19	Planes	19	S18_1662	USA	Lewis	Small	57294.42	927	MML	Loyal Customer	
Mini Classics	26	35.73076923	95.03038	3290.615	26	26	Motorcycles	26	S10_2016	USA	Frick	Medium	85555.99	977	MMM	Loyal Customer	
Motor Mint Distributors Inc.	23	31.73913043	113.4065	3638.355	23	23	Motorcycles	23	S10_2016	USA	Hernandez	Small	83682.16	944	MMM	Loyal Customer	
Signal Gift Stores	29	32.03448276	91.42897	2853.486	29	29	Classic Cars	29	S18_1129	USA	King	Medium	82751.08	932	MMM	Loyal Customer	
Toys4GrownUps.com	30	35.33333333	97.22467	3485.399	30	30	Motorcycles	30	S10_1678	USA	Young	Medium	104562	888	MMM	Loyal Customer	
Vitachrome Inc.	25	31.48	106.1788	3521.65	25	25	Motorcycles	25	S10_1678	USA	Frick	Small	88041.26	956	MMM	Loyal Customer	

- On basis on Recency, frequency & monetary we have grouped our loyal customer's. These customers have purchased multiple times with good monetary value.
- If we focus more on this segment of customers, we can easily turn them into our top best customers too. Also, in this segment we can see the customers for product line - classic cars are many.
- These are the still valuable customers as they are with better level in all aspects Frequency and Monetary. We have to look after the needs of these customers to bring them into the pool of best.

Inferences from RFM Analysis and identified segments

Customers on the verge of churn

CUSTOMERNAME	ORDERI	QUANTITYORI	PRICEEA	SALES	ORDERI	DAYS_S	PRODUCTLIN	MSRP	PRODU	COUNTI	CONTAC	DEALSIZ	Moneta	Recenc	RFM	Custom Category
Collectable Mini Designs Co.	25	38.16	91.5348	3499.569	25	25	Classic Cars	25	S10_4757	USA	Thompson	Medium	87489.23	1208	LMM	Customer on the verg
Diecast Collectables	18	38.61111111	101.7833	3936.654	18	18	Classic Cars	18	S10_4962	USA	Franco	Medium	70859.78	1149	LMM	Customer on the verg
Marta's Replicas Co.	27	36.14814815	107.0719	3817.792	27	27	Motorcycles	27	S10_1678	USA	Hernandez	Medium	103080.4	979	LMM	Customer on the verg

- On basis on Recency, frequency & monetary we have grouped our Customers who are on verge of churning. We should definitely focus on this group before we lose them and try to convert them into our regular customers.
- If the company pays more attention and fulfill their requirement, then we can easily turn them into our regular customer and we can save them from churning out.
- We can still concentrate on them to increase business as they have got High Recency level.

Inferences from RFM Analysis and identified segments

Lost Customers

CUSTOMERNAME	ORDERI	QUANTITYORI	PRICEEA	SALES	ORDERI	DAYS_S	PRODUCTLIN	MSRP	PRODU	COUNTI	CONTA(	DEALSIZ	Moneta	Recenc	RFM	Custom	Categ
Auto-Moto Classics Inc.	8	35.875	92.8	3309.908	8	8	Ships	8	S18_3029	USA	Taylor	Medium	26479.26	928	MLL	Lost Customer	
Boards & Toys Co.	3	34	89.80667	3043.117	3	3	Classic Cars	3	S12_3380	USA	Young	Medium	9129.35	861	HLL	Lost Customer	
Cambridge Collectables Co.	11	32.45454545	101.3291	3287.602	11	11	Classic Cars	11	S10_1949	USA	Tseng	Medium	36163.62	1137	LLL	Lost Customer	
Microscale Inc.	10	38.1	88.494	3314.493	10	10	Motorcycles	10	S12_2823	USA	Kuo	Medium	33144.93	958	MLL	Lost Customer	
Online Mini Collectables	15	38.13333333	94.68067	3813.197	15	15	Classic Cars	15	S12_1099	USA	Barajas	Large	57197.96	1012	LLL	Lost Customer	
Signal Collectibles Ltd.	15	34.26666667	95.396	3347.901	15	15	Trucks and Buse	15	S12_4473	USA	Taylor	Medium	50218.51	1224	LLL	Lost Customer	
Super Scale Inc.	17	37.41176471	128.4524	4674.828	17	17	Classic Cars	17	S10_1949	USA	Murphy	Medium	79472.07	1141	LLM	Lost Customer	
West Coast Collectables Co.	13	39.30769231	88.30769	3544.972	13	13	Classic Cars	13	S10_1949	USA	Thompson	Medium	46084.64	1236	LLL	Lost Customer	

- On basis on Recency, frequency & monetary parameters we have grouped our Customers who we'd lost
- Their recency is very low and hasn't made any purchase since long so we can say these are our lost customers. If taken feedback from them and fulfill their demand we might bring them back to being a good customer.
- But suggestion is to not to invest much time on these type of customers as these are the most avoidable customers as they are with lowest level in all aspects Recency, Frequency and Monetary. There is no point in spending time and effort to maintain business with these customers.

Inferences & Recommendations

- Using Recency, frequency & monetary parameters we have grouped our Top , loyal, on the verge of churning and lose customers. Customers with good recency has been our top customers were as we also have lost customer lists.
- Customers on verge of churning can be saved and can be converted into a good buyer.
- RFM model is used for deriving the customers types like Loyal, top or best, on verge of churning & lost customers.
- Recency, frequency & monetary parameters were widely used to bifurcate the types of customer's.
- This model can be very helpful to the company to maintain its sales and customers and can focus on how the company has lost the customers & can take various actions to bring them back.
- It is vital for the company to convert the customers who are on verge of churning into a regular customer or at least maintain them and also how to increase the sales ratio can be identified.
- USA is also having most at risk customers, followed by France, Australia and Finland. If shipping issues could be handled better for these countries, it would be better for business.

Thank You!