MRA Project ML 1

ADITYA JOSHI

G-1 DSBA AUG B 20

adityajoshi463@gmail.com

Table of Content

- Problem Statement and Business context
- Data Info
- Basic EDA: univariate, bivariate and trend analysis
- RFM analysis
- Customer segmentation and RFM insights

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 you as their consultant. Your job is to use your magical data science skills to provide them with suitable insights about their data and their customers.

CHECKING NULL VALUES: -

- Here is the basic info on the data: -
- The data consists of 2247 rows and 20 columns.
- There are no null values and no duplicate values in the data.

	Null Values
ORDER NUMBER	0
QUANTITY ORDERED	0
PRICE EACH	0
ORDER LINE	0
SALES	0
ORDER DATE	0
DAYS_SINCE_LASTORDER	0
STATUS	0
CATEGORY	0
MSRP	0
PRODUCT CODE	0
CUSTOMER NAME	0
PHONE	0
ADDRESS	0
CITY	0
POSTAL CODE	0
COUNTRY	0
CONTACT LASTNAME	0
CONTACT FIRSTNAME	0
DEAL SIZE	0

- DATA INFO: -

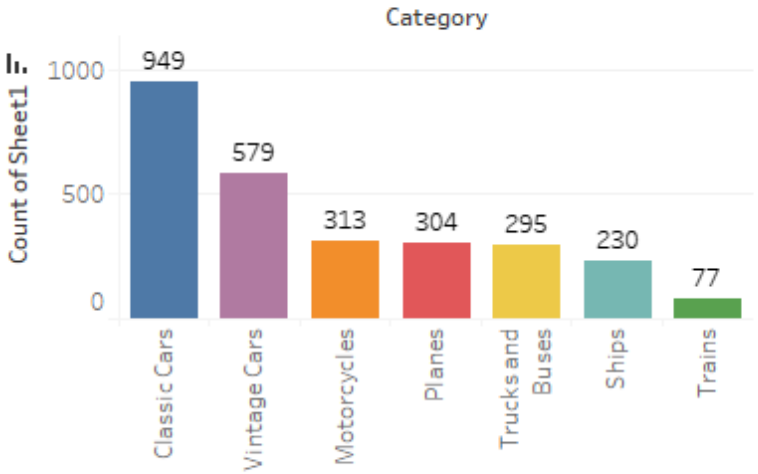
#	Column	Non-Null Count	Dtype
0	ORDER NUMBER	2747 non-null	int64
1	QUANTITY ORDERED	2747 non-null	int64
2	PRICE EACH	2747 non-null	float64
3	ORDER LINE	2747 non-null	int64
4	SALES	2747 non-null	float64
5	ORDER DATE	2747 non-null	datetime64[ns]
6	DAYS_SINCE_LASTORDER	2747 non-null	int64
7	STATUS	2747 non-null	object
8	CATEGORY	2747 non-null	object
9	MSRP	2747 non-null	int64
10	PRODUCT CODE	2747 non-null	object
11	CUSTOMER NAME	2747 non-null	object
12	PHONE	2747 non-null	object
13	ADDRESS	2747 non-null	object
14	CITY	2747 non-null	object
15	POSTAL CODE	2747 non-null	object
16	COUNTRY	2747 non-null	object
17	CONTACT LASTNAME	2747 non-null	object
18	CONTACT FIRSTNAME	2747 non-null	object
19	DEAL SIZE	2747 non-null	object

- DATA DESCRIPTION: -

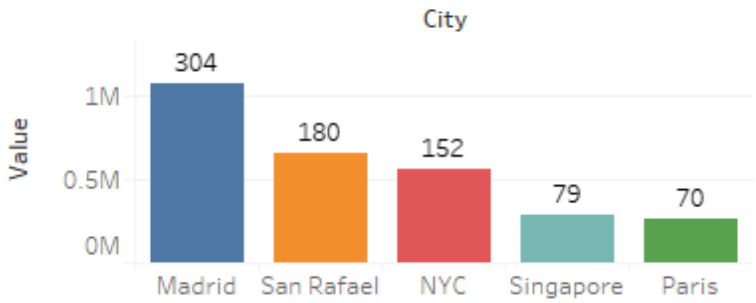
	count	mean	std	min	25%	50%	75%	max
ORDER NUMBER	2747.0	10259.761558	91.877521	10100.00	10181.000	10264.00	10334.500	10425.00
QUANTITY ORDERED	2747.0	35.103021	9.762135	6.00	27.000	35.00	43.000	97.00
PRICE EACH	2747.0	101.098951	42.042548	26.88	68.745	95.55	127.100	252.87
ORDER LINE	2747.0	6.491081	4.230544	1.00	3.000	6.00	9.000	18.00
SALES	2747.0	3553.047583	1838.953901	482.13	2204.350	3184.80	4503.095	14082.80
DAYS_SINCE_LASTORDER	2747.0	1757.085912	819.280576	42.00	1077.000	1761.00	2436.500	3562.00
MSRP	2747.0	100.691664	40.114802	33.00	68.000	99.00	124.000	214.00

UNIQUE VALUES AND THEIR COUNT

count category

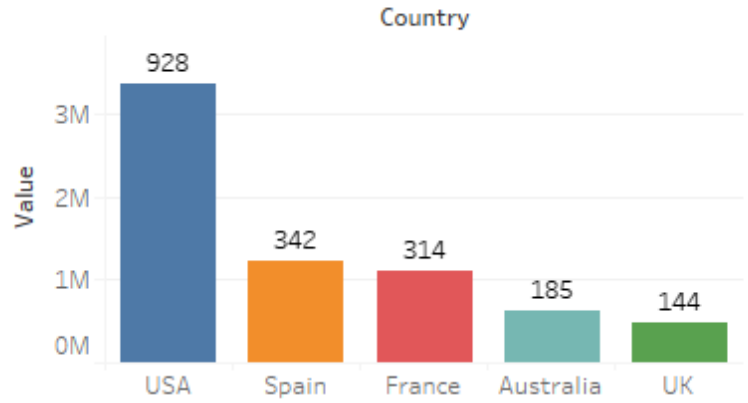


count city

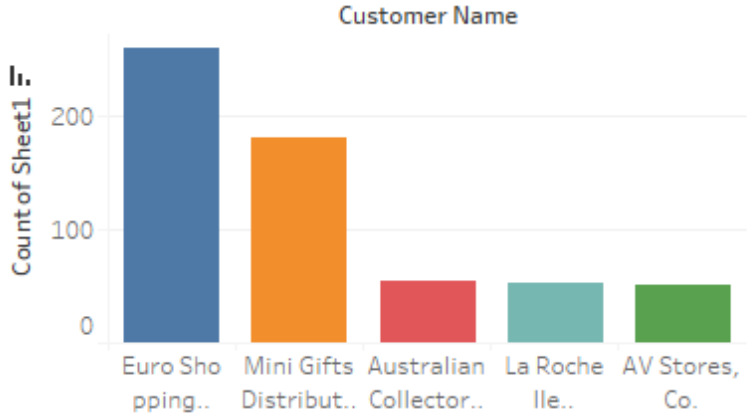


count status

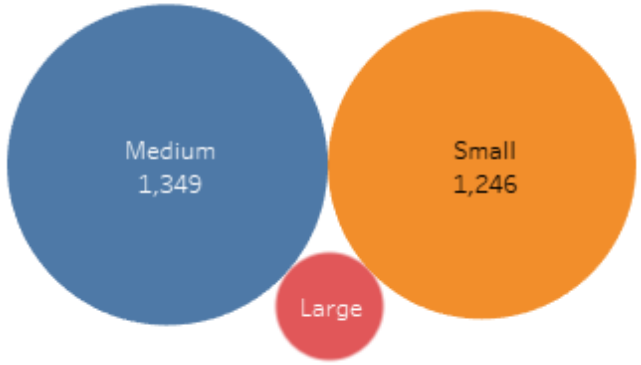
count country



count customer



count deal size



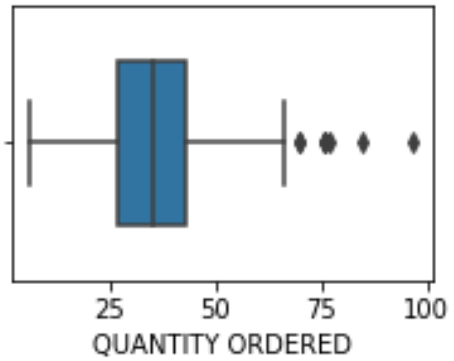
UNIVARAITE ANALYSIS

1. Quantity ordered

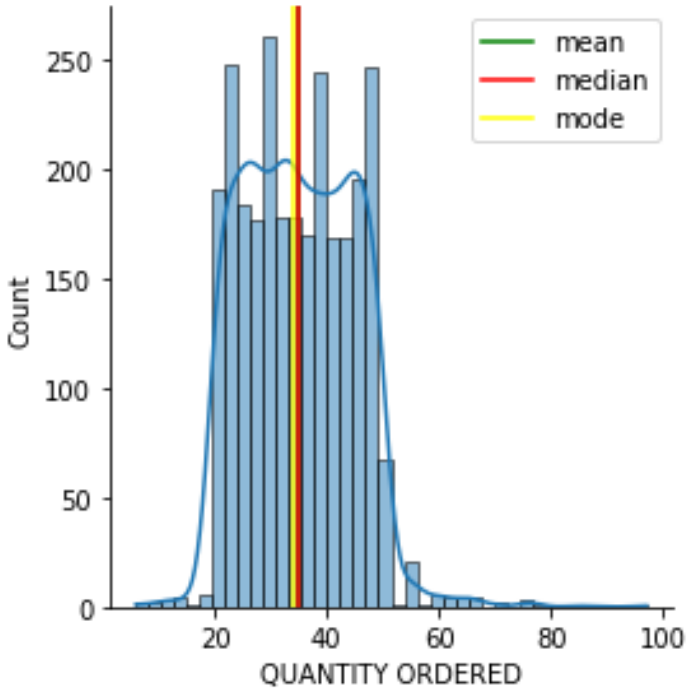
```
Description of QUANTITY ORDERED  is: -
count    2747.000000
mean       35.103021
std         9.762135
min         6.000000
25%        27.000000
50%        35.000000
75%        43.000000
max        97.000000
Name: QUANTITY ORDERED, dtype: float64

----------------------------------------
Mean is:  35.10302147797597
Median is:  35.0
Mode is:  34
```

Boxplot of QUANTITY ORDERED is: -



Distribution of QUANTITY ORDERED is: -

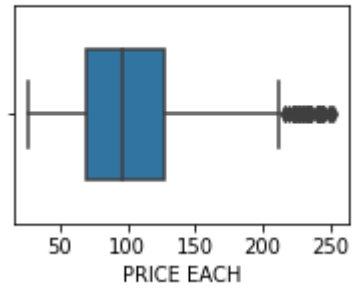


2. Price Each

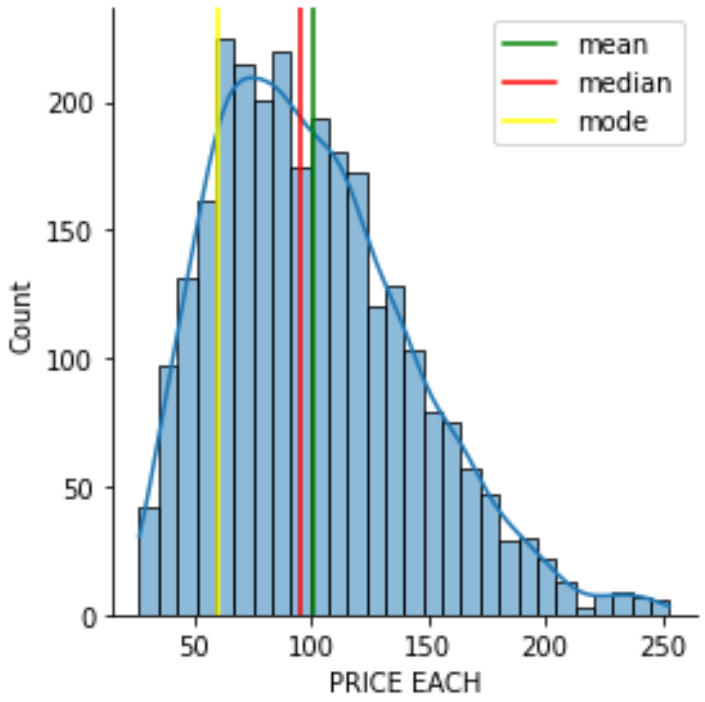
```
Description of PRICE EACH  is: -
count    2747.000000
mean      101.098951
std        42.042548
min        26.880000
25%        68.745000
50%        95.550000
75%       127.100000
max       252.870000
Name: PRICE EACH, dtype: float64

----------------------------------
Mean is:  101.09895113914965
Median is:  95.55
Mode is:  59.87
----------------------------------
```

Boxplot of PRICE EACH is: -



Distribution of PRICE EACH is: -

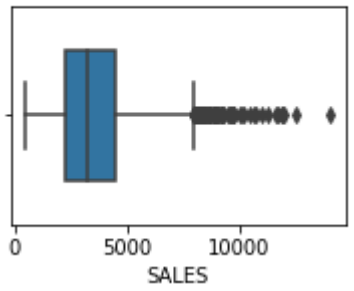


3. Sales

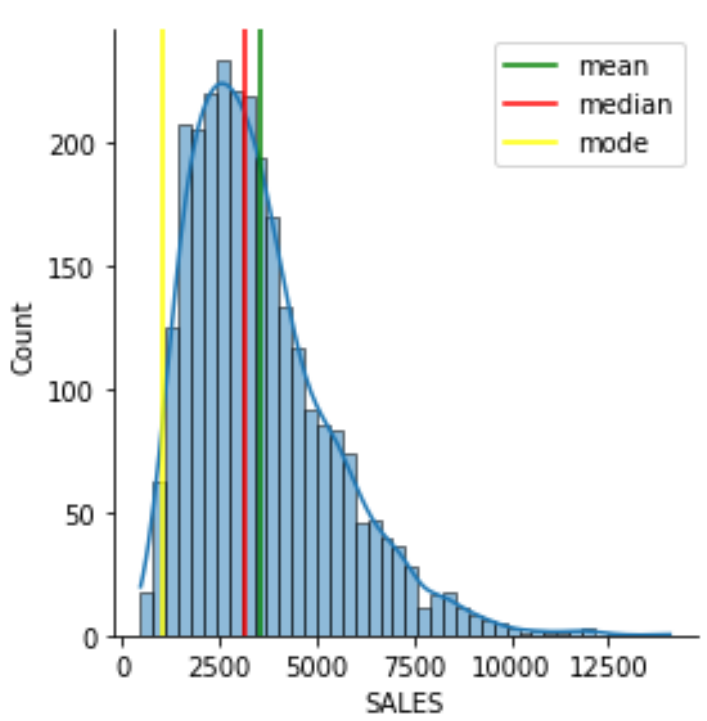
```
Description of SALES  is: -
count     2747.000000
mean      3553.047583
std       1838.953901
min        482.130000
25%       2204.350000
50%       3184.800000
75%       4503.095000
max      14082.800000
Name: SALES, dtype: float64

----------------------------
Mean is:  3553.047582817625
Median is:  3184.8
Mode is:  1030.44
```

Boxplot of SALES is: -



Distribution of SALES is: -

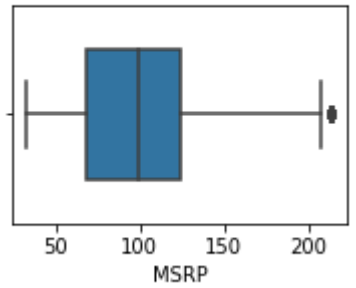


4. MSRP

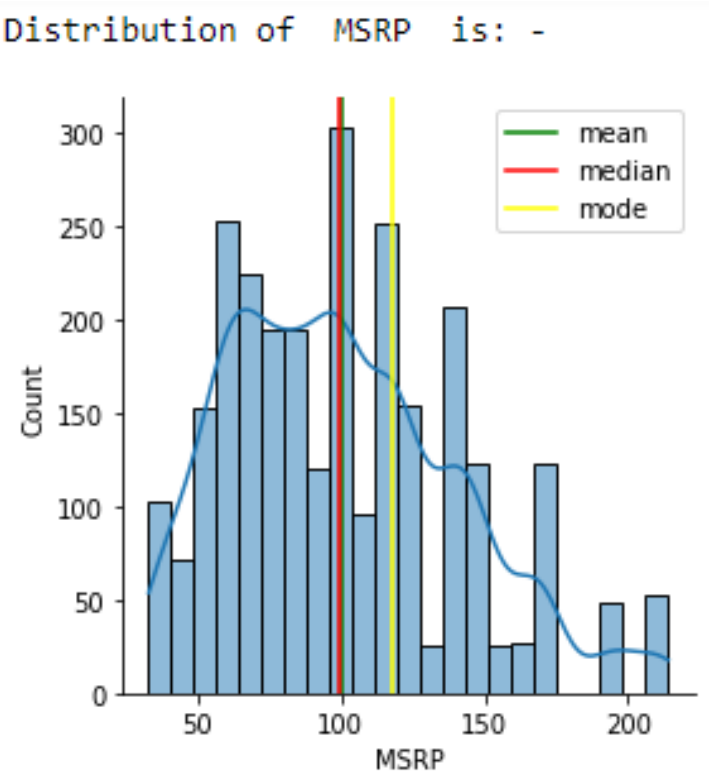
```
Description of MSRP  is: -
count    2747.000000
mean      100.691664
std        40.114802
min        33.000000
25%        68.000000
50%        99.000000
75%       124.000000
max       214.000000
Name: MSRP, dtype: float64

------------------------------
Mean is:  100.69166363305425
Median is:  99.0
Mode is:  118
------------------------------
```

Boxplot of MSRP is: -

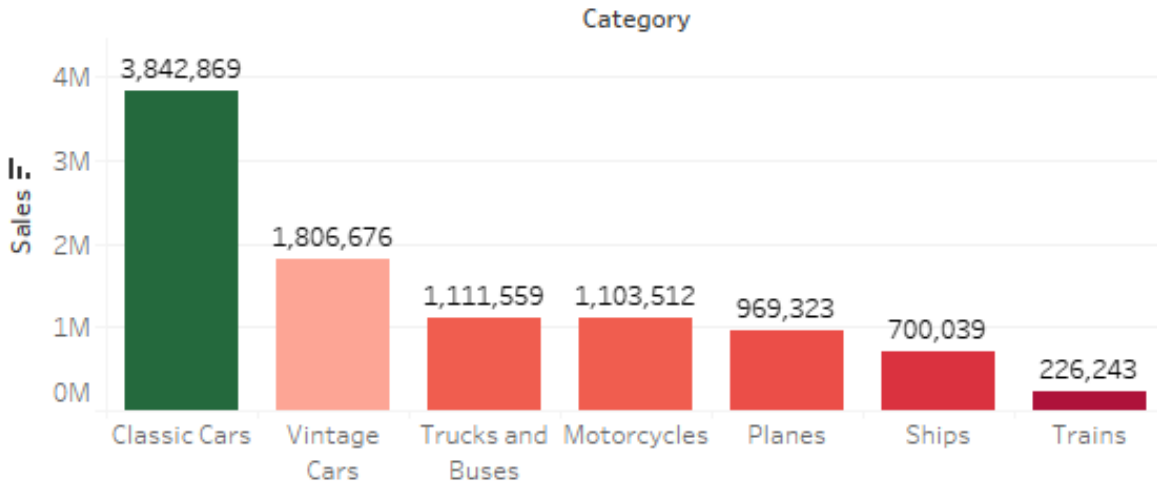


Distribution of MSRP is: -

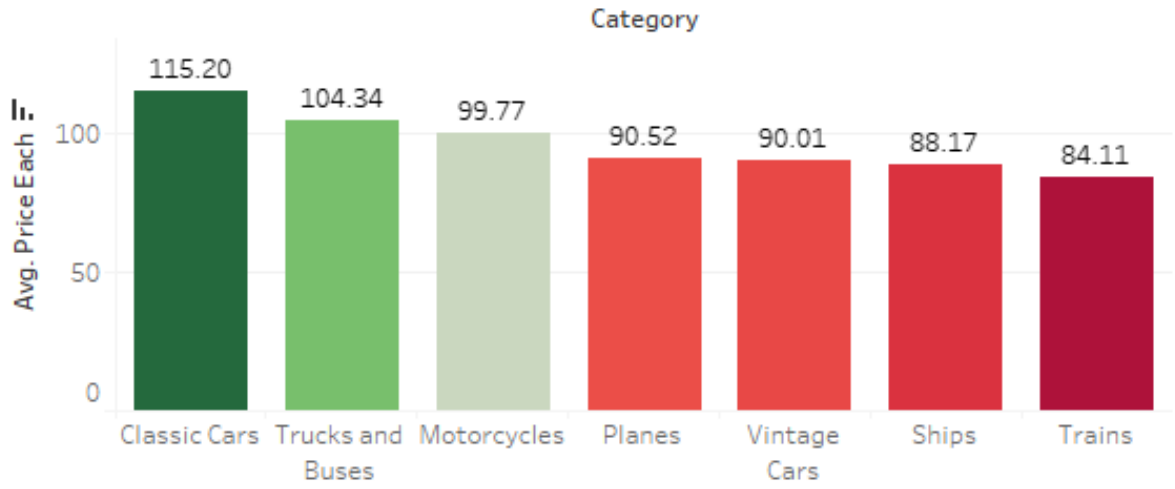


BIVARAITE ANALYSIS

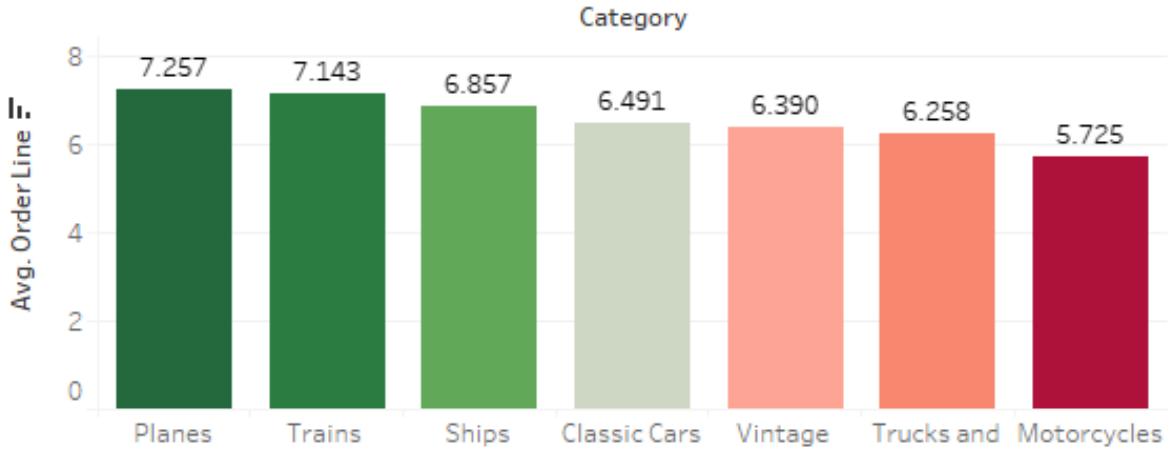
Category vs Sales



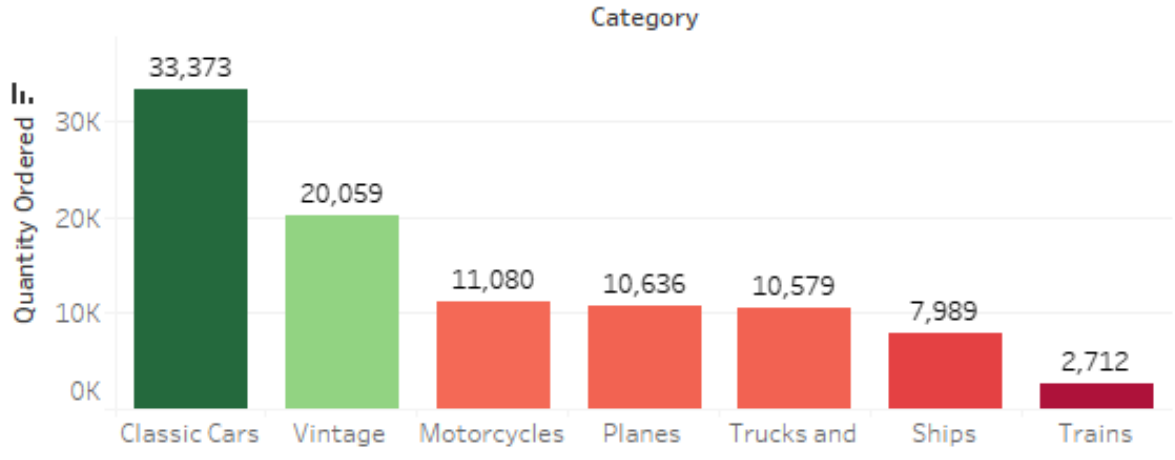
Category vs Avg Price



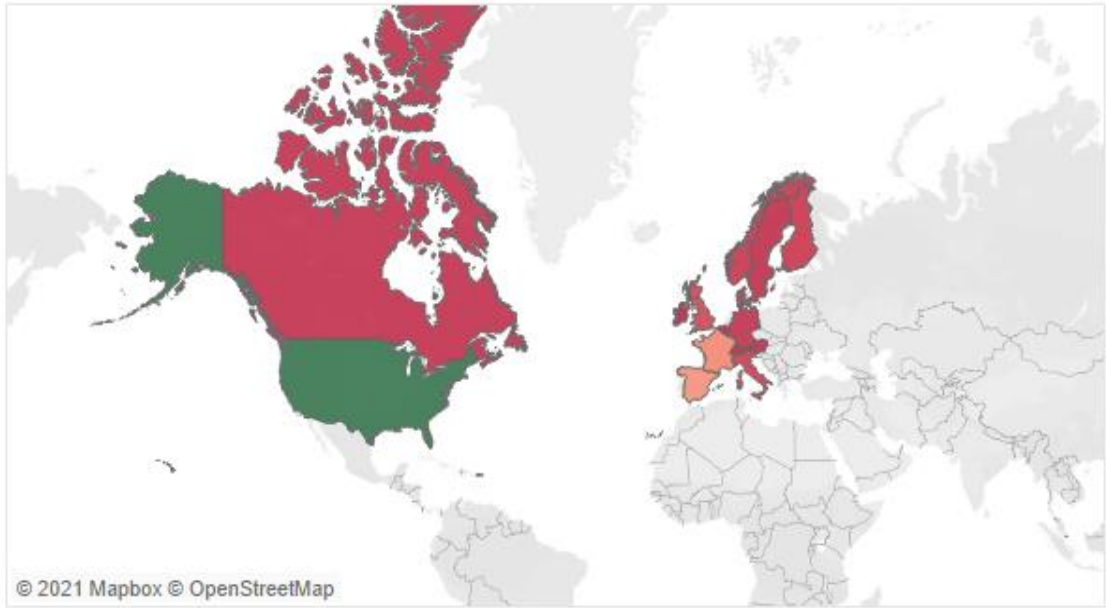
Category vs avg Order line



Category vs qty ordered

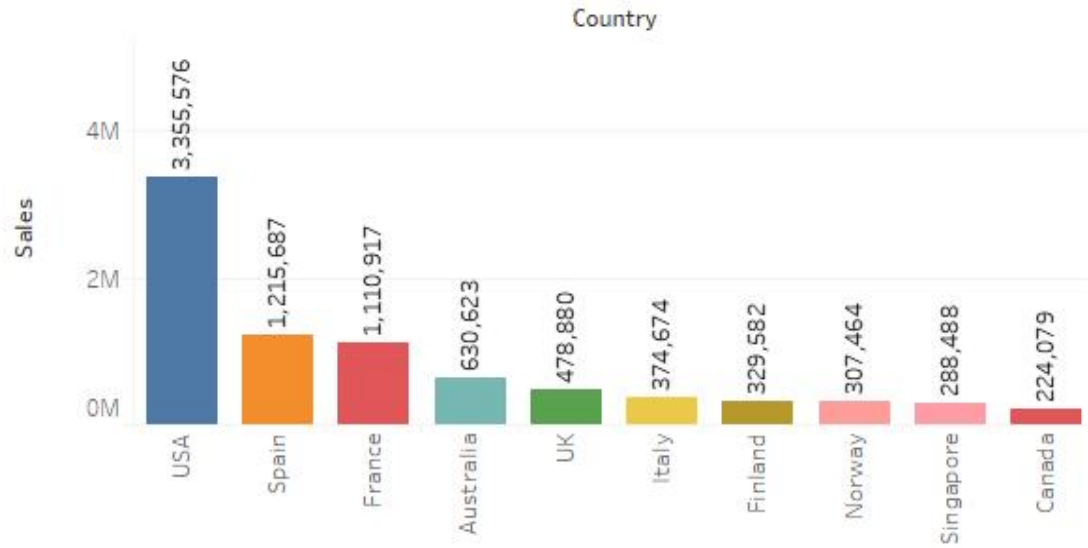


Sales on Map

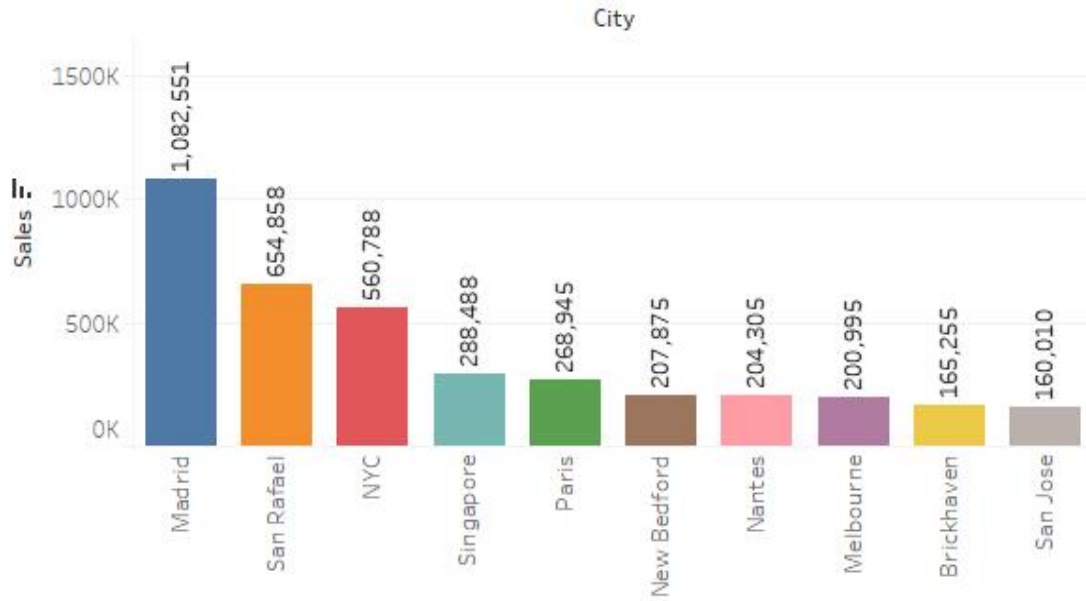


Sales by city

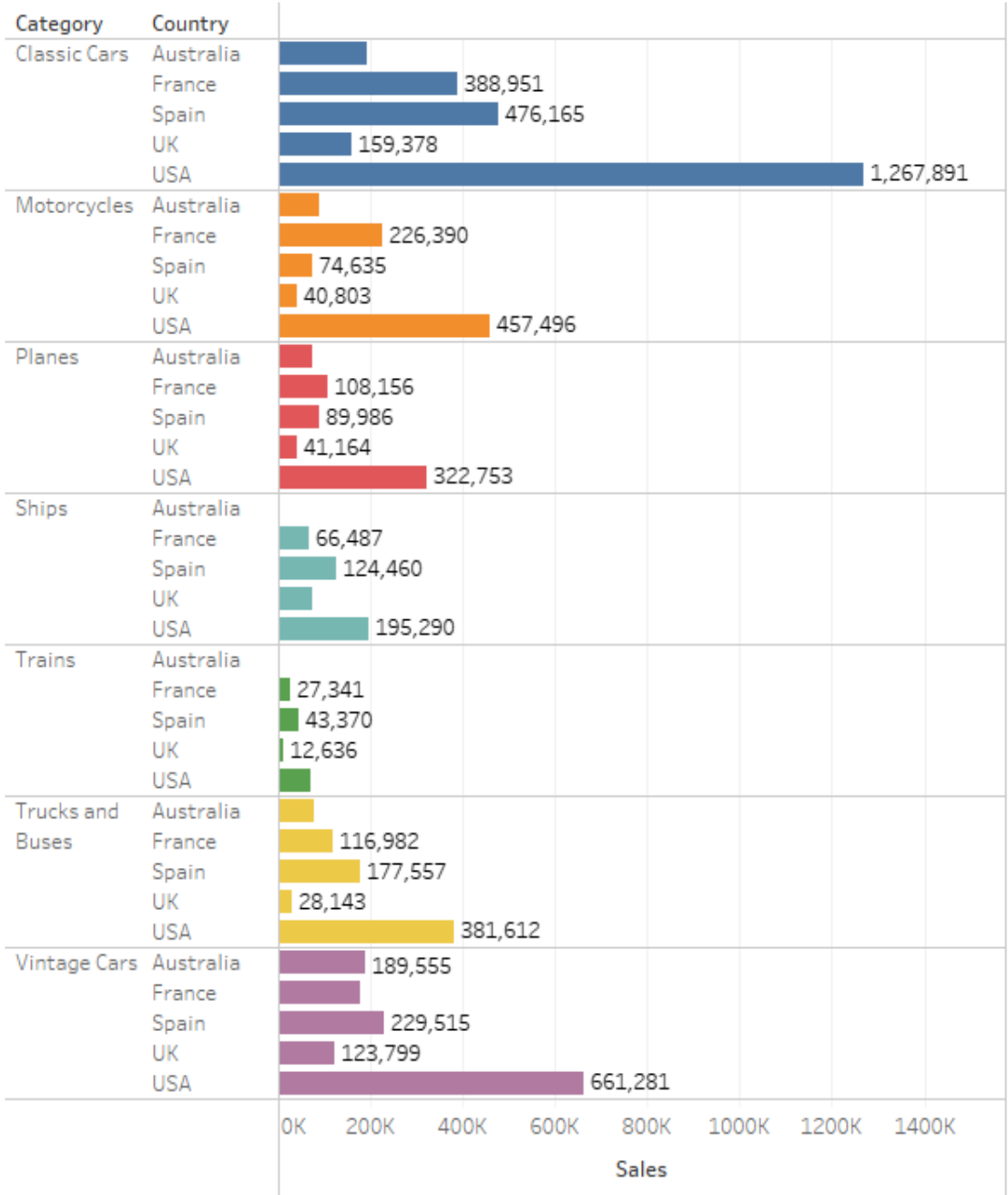
Sales vs country



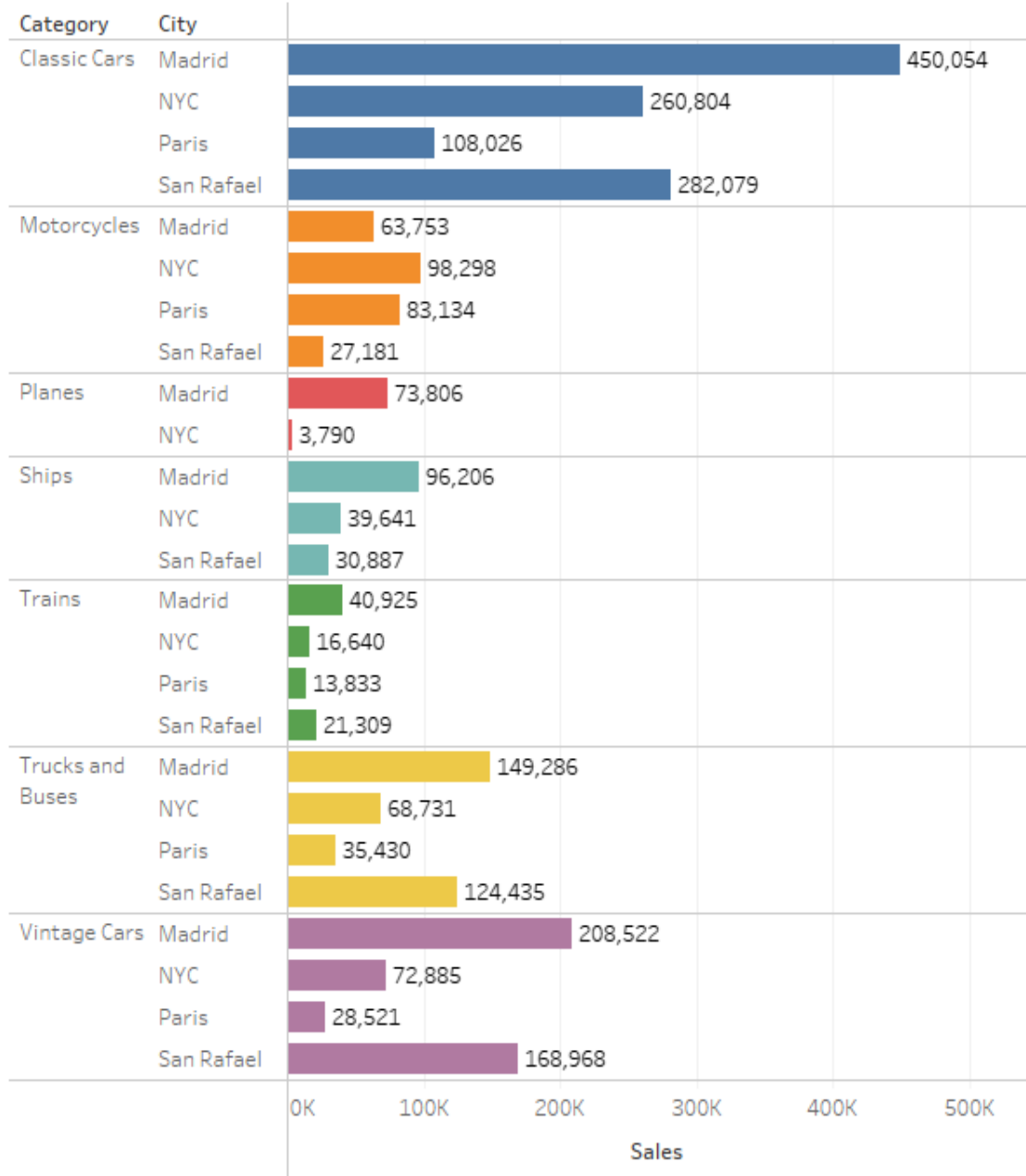
City vs Sales



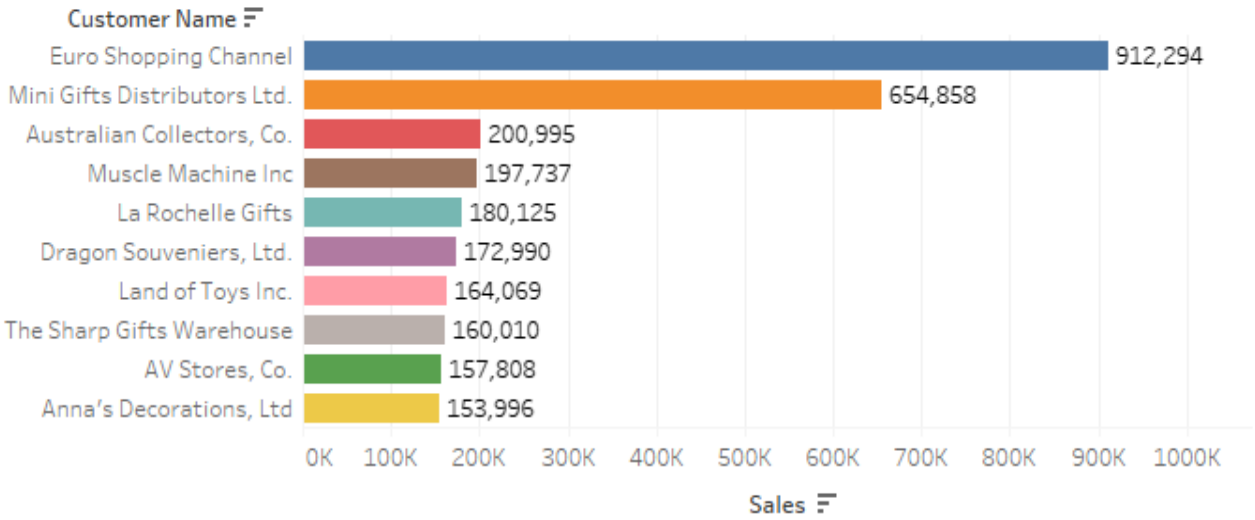
Category and Country



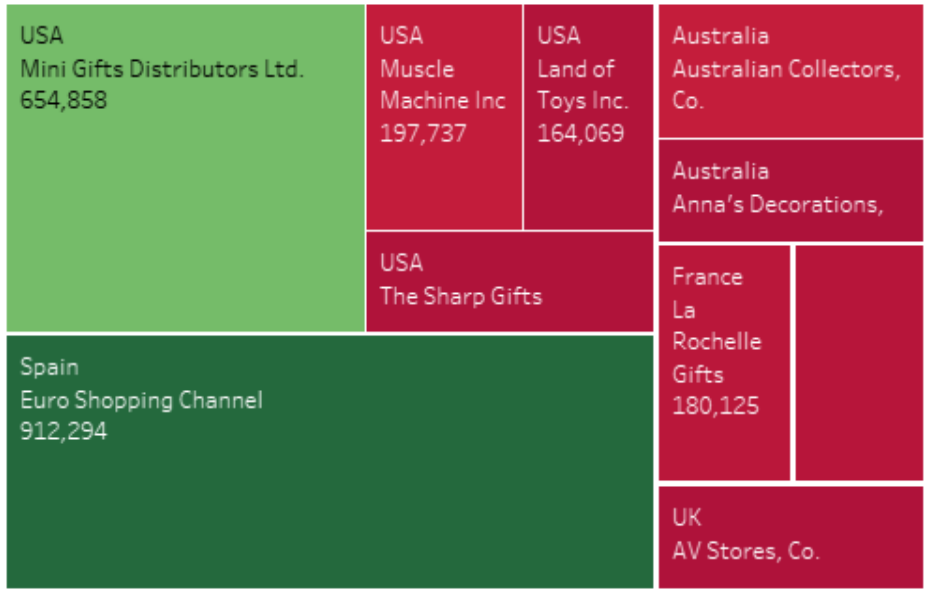
Category and City



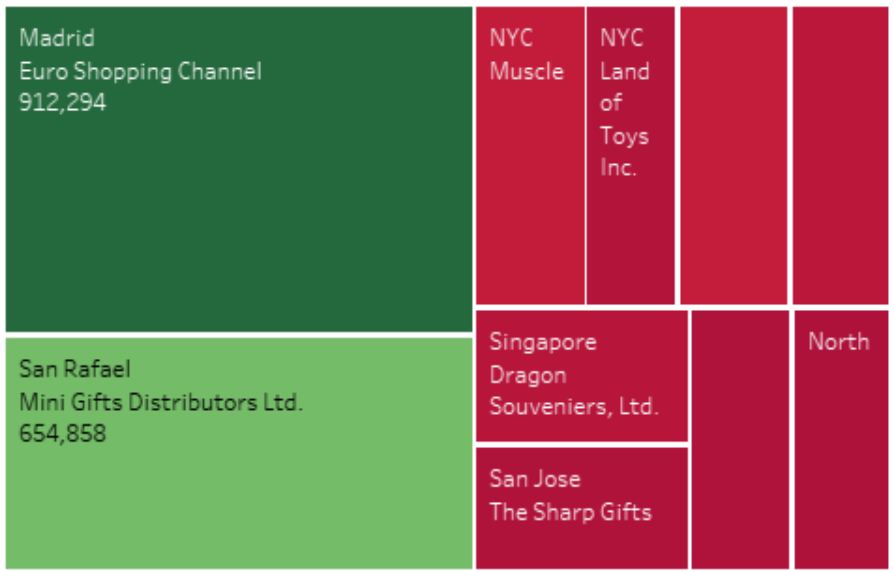
Customer and Sales



Country, Sales and Customer

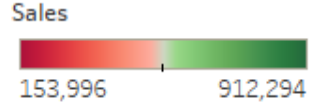


City, Sales and Customer-Table (2)

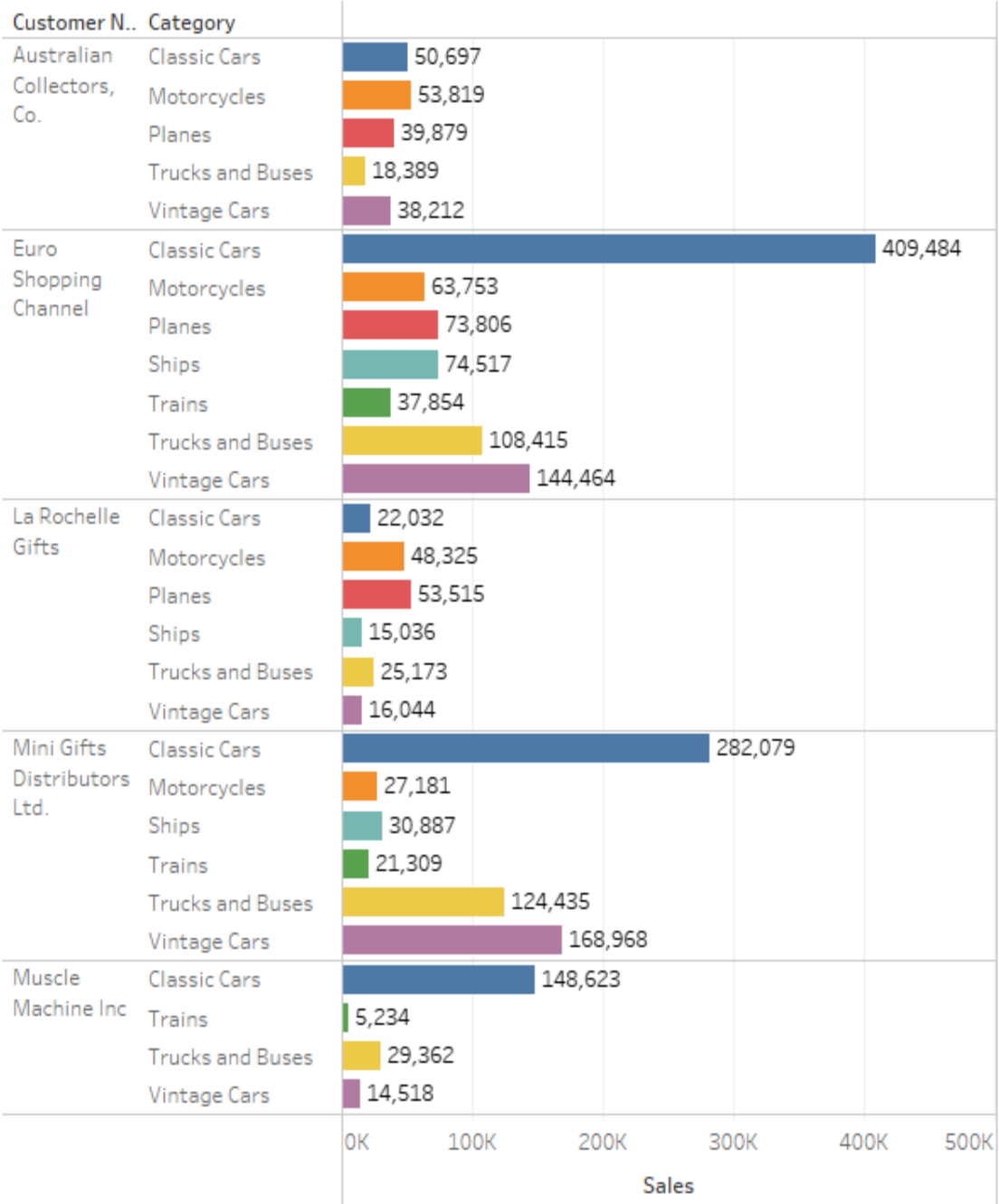


Country, Sales and Customer-Table

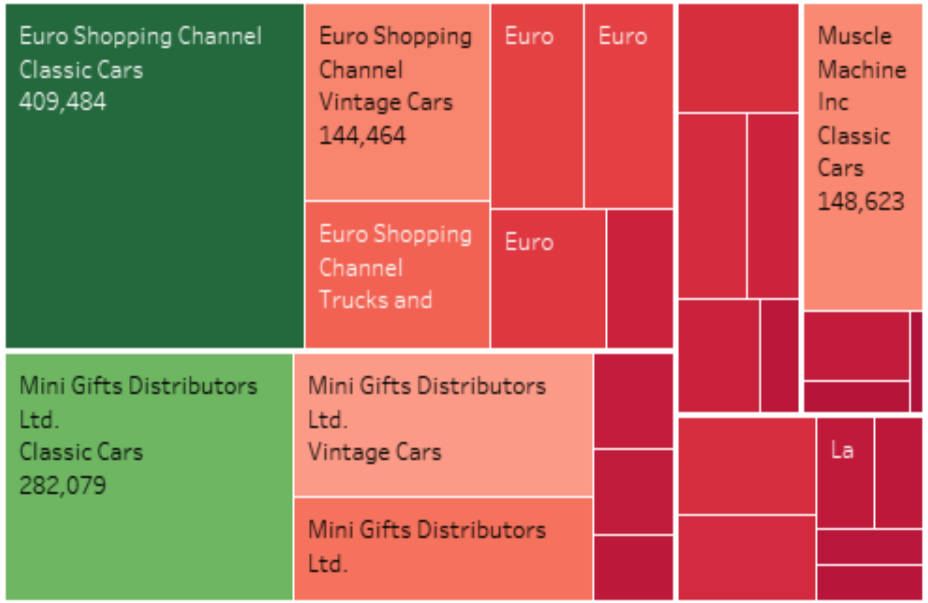
Country	City	Customer Name	
Australia	Melbourne	Australian Collectors, Co.	200,995
	North Sydney	Anna's Decorations, Ltd	153,996
France	Nantes	La Rochelle Gifts	180,125
Singapore	Singapore	Dragon Souveniers, Ltd.	172,990
Spain	Madrid	Euro Shopping Channel	912,294
UK	Manchester	AV Stores, Co.	157,808
USA	NYC	Land of Toys Inc.	164,069
		Muscle Machine Inc	197,737
	San Jose	The Sharp Gifts Warehouse	160,010
	San Rafael	Mini Gifts Distributors Ltd.	654,858



Cust,Cat andSales_bar



Customer,Cat and Sales



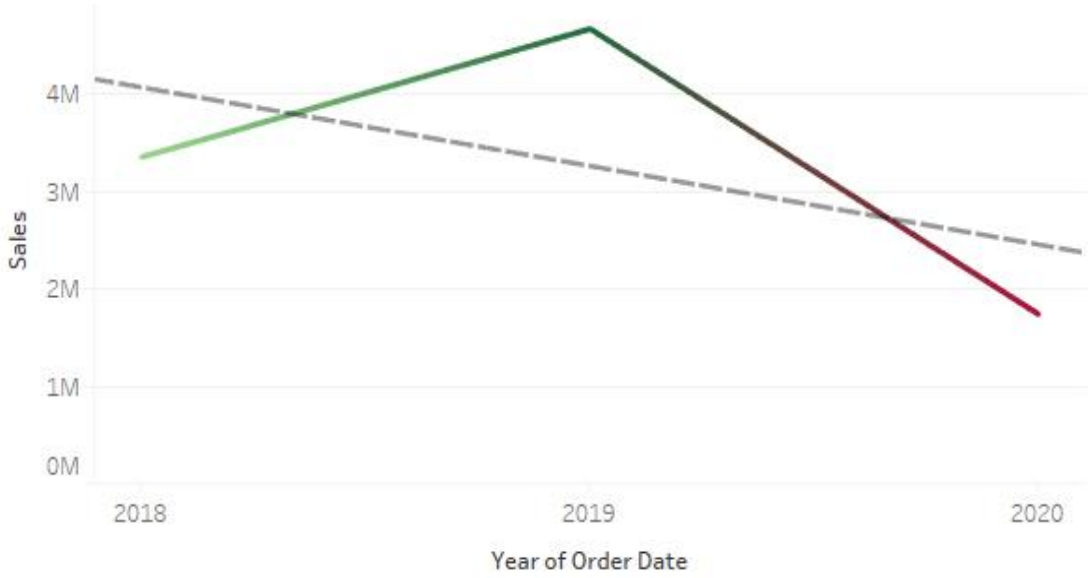
Cust,Cat andSales_bar (2)

Customer Name	Classic Cars	Motorcy..	Planes	Ships	Trains	Trucks and Buses	Vintage Cars
Australian Collectors, Co.	50,697	53,819	39,879			18,389	38,212
Euro Shopping Channel	409,484	63,753	73,806	74,517	37,854	108,415	144,464
La Rochelle Gifts	22,032	48,325	53,515	15,036		25,173	16,044
Mini Gifts Distributors Ltd.	282,079	27,181		30,887	21,309	124,435	168,968
Muscle Machine Inc	148,623				5,234	29,362	14,518

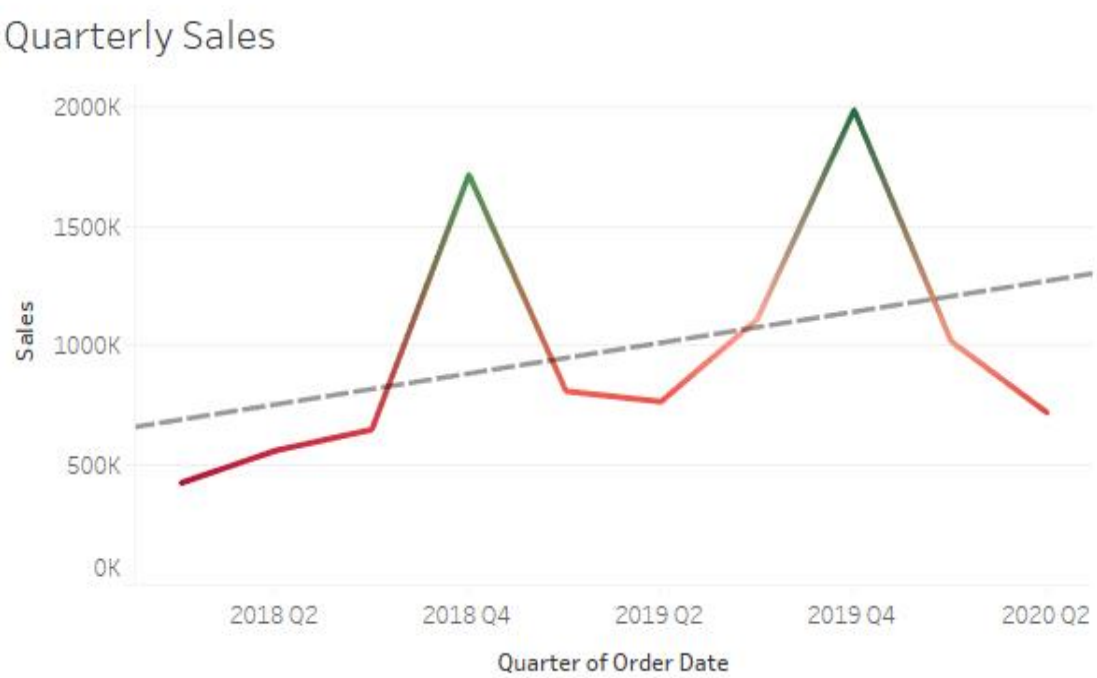
Category

Exit Presentation Mode [57]

Yearly Sales

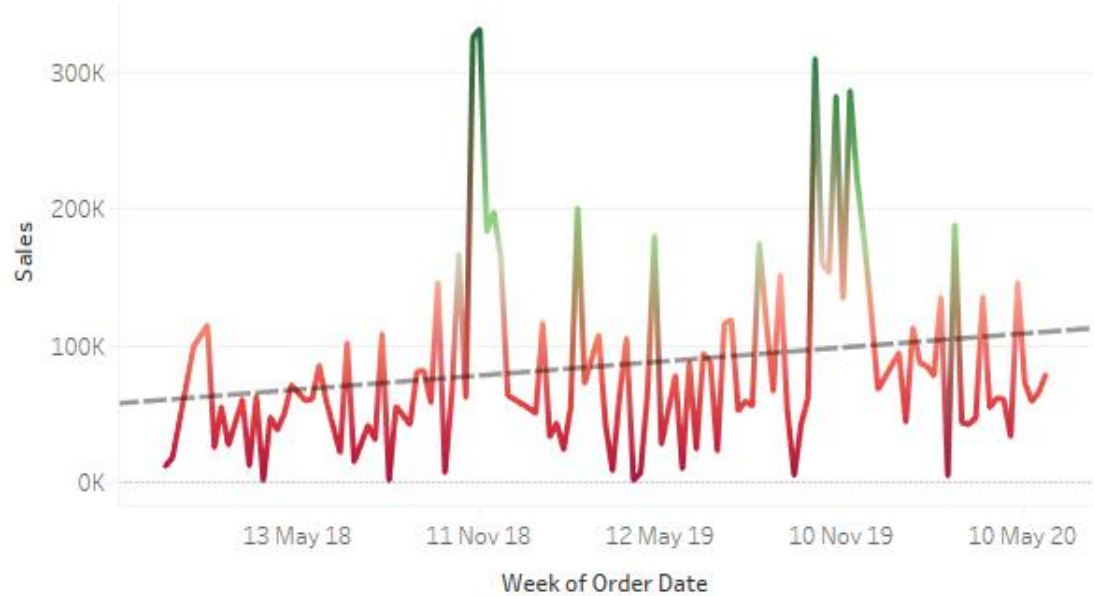


Quarterly Sales



Monthly Sales

Weekly Sales



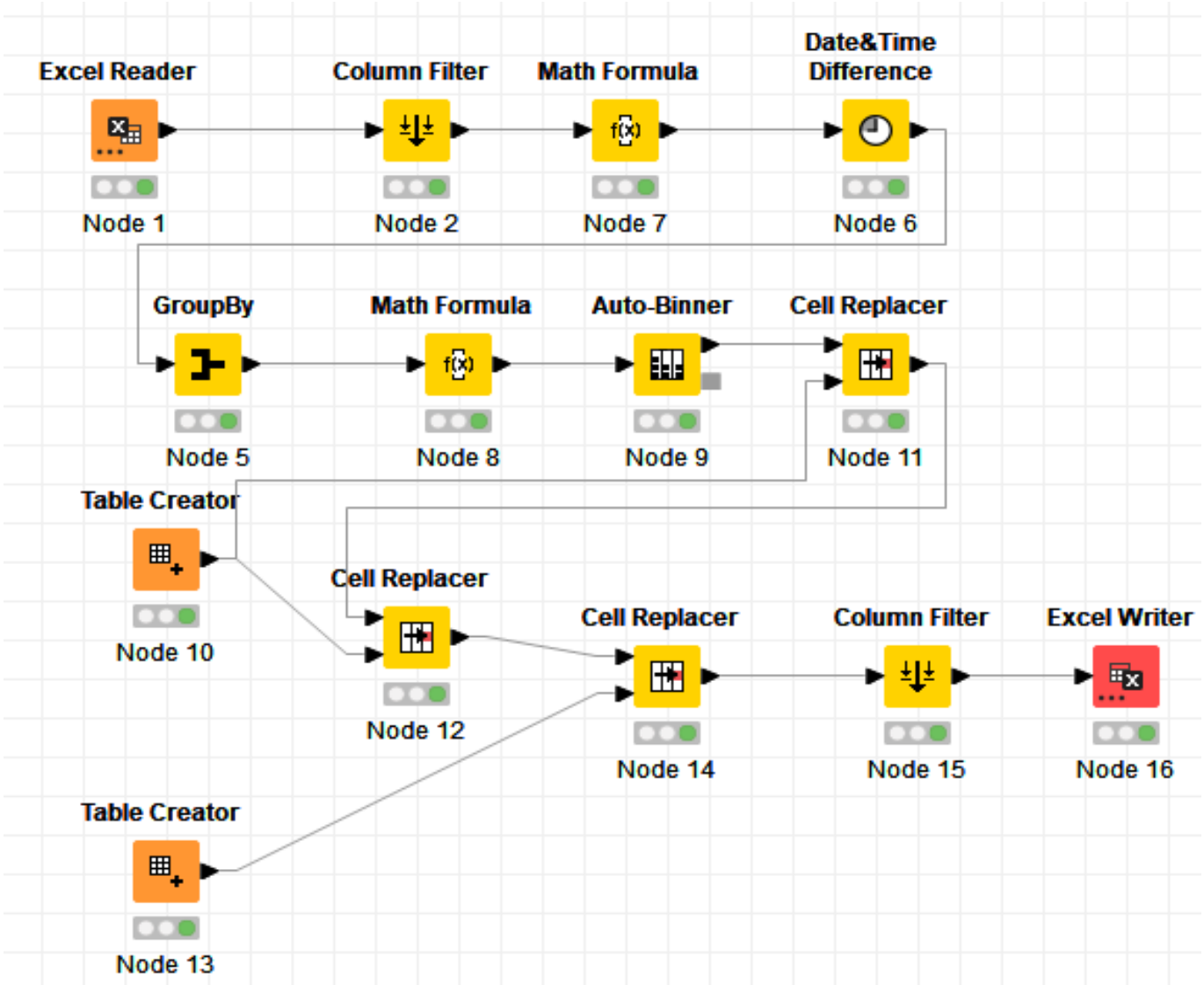
EDA SUMMARY AND INFERENCES

- Classic cars are the ones sold maximum, and comprise of the max sales revenue
- Planes Trucks and Ships have the longest order lines.
- USA and Spain are the two countries with max sales. In US San Rafael and NYC are the cities that cater to max sales, while in Spain, Madrid is the largest Market.
- Euro Shopping channel and Mini gifts distributor, are the largest customers.
- Classic cars tops the list which are catered for Euro shopping and Mini gifts.
- There is a dip in yearly sales, but when disintegrated , it shows a gradual upward trend.

RFM

- Using RFM the customers are divided into 4 segments, namely, High valued, Mid valued, At Risk/Potential and Churn. Knime workflow was used for the same.
- To start off, Recency, Frequency and Monetary columns were defined and customers segmented based upon it.
- It is assumed that the high and mid valued customers are loyal. The at risk customers might have some potential left and they could be converted to mid valued by different promotional schemes
- However, the ones with Churn could be allowed to let go.

Knime Workflow



Segmented Table

CUSTOMER NAME	ORDER NUMBER	QUANTITY ORDERED	PRICE EACH	ORDER LINE	SALES	Monetary	Recency	Frequency	Monetary_HML	Frequency_HML	Recency_HML
AV Stores, Co.	51	34.862745	91.084510	51	157807.81	157807.81	197	51	High Valued	High Valued	Risk/Potential
Alpha Cognac	20	34.350000	101.160000	20	70488.44	70488.44	65	20	Churn	Churn	High Valued
Amica Models & Co.	26	32.423077	110.852692	26	94117.26	94117.26	266	26	Mid Valued	Risk/Potential	Churn
Anna's Decorations, Ltd	46	31.934783	106.424130	46	153996.13	153996.13	84	46	High Valued	High Valued	Mid Valued
Atelier graphique	7	38.571429	92.238571	7	24179.96	24179.96	189	7	Churn	Churn	Risk/Potential

RFM Inferences

Best Customers: -

CUSTOMER NAME	SALES	CATEGORY	CITY	COUNTRY	Monetary_HML	Frequency_HML	Recency_HML
Danish Wholesale Imports	145041.6	Classic Cars	Kobenhavn	Denmark	High Valued	High Valued	High Valued
Euro Shopping Channel	912294.11	Motorcycles	Madrid	Spain	High Valued	High Valued	High Valued
L'ordine Souveniers	142601.33	Classic Cars	Reggio Emilia	Italy	High Valued	High Valued	High Valued
La Rochelle Gifts	180124.9	Motorcycles	Nantes	France	High Valued	High Valued	High Valued
Mini Gifts Distributors Ltd.	654858.06	Classic Cars	San Rafael	USA	High Valued	High Valued	High Valued
Reims Collectables	135042.94	Motorcycles	Reims	France	High Valued	High Valued	High Valued
Salzburg Collectables	149798.63	Motorcycles	Salzburg	Austria	High Valued	High Valued	High Valued
Souveniers And Things Co.	151570.98	Motorcycles	Chatswood	Australia	High Valued	High Valued	High Valued
The Sharp Gifts Warehouse	160010.27	Classic Cars	San Jose	USA	High Valued	High Valued	High Valued

RFM Inferences

Verge of Churning: -

CUSTOMER NAME	SALES	CATEGORY	CITY	COUNTRY	Monetary_HML	Frequency_HML	Recency_HML
Blauer See Auto, Co.	85171.59	Classic Cars	Frankfurt	Germany	Risk/Potential	Risk/Potential	Risk/Potential
Canadian Gift Exchange Netwo	75238.92	Classic Cars	Vancouver	Canada	Risk/Potential	Risk/Potential	Risk/Potential
Enaco Distributors	78411.86	Classic Cars	Barcelona	Spain	Risk/Potential	Risk/Potential	Risk/Potential
Mini Classics	85555.99	Motorcycles	White Plains	USA	Risk/Potential	Risk/Potential	Risk/Potential
Motor Mint Distributors Inc.	83682.16	Motorcycles	Philadelphia	USA	Risk/Potential	Risk/Potential	Risk/Potential
giftsbymail.co.uk	78240.84	Classic Cars	Cowes	UK	Risk/Potential	Risk/Potential	Risk/Potential

RFM Inferences

Lost Customers: -

CUSTOMER NAME	SALES	CATEGORY	CITY	COUNTRY	Monetary_HML	Frequency_HML	Recency_HML
Auto Assoc. & Cie.	64834.32	Classic Cars	Versailles	France	Churn	Churn	Churn
Bavarian Collectables Imports,	34993.92	Planes	Munich	Germany	Churn	Churn	Churn
CAF Imports	49642.05	Classic Cars	Madrid	Spain	Churn	Churn	Churn
Cambridge Collectables Co.	36163.62	Classic Cars	Cambridge	USA	Churn	Churn	Churn
Clover Collections, Co.	57756.43	Classic Cars	Dublin	Ireland	Churn	Churn	Churn
Daedalus Designs Imports	69052.41	Motorcycles	Lille	France	Churn	Churn	Churn
Double Decker Gift Stores, Ltd	36019.04	Classic Cars	London	UK	Churn	Churn	Churn
Iberia Gift Imports, Corp.	54723.62	Trucks and Buses	Sevilla	Spain	Churn	Churn	Churn
Online Mini Collectables	57197.96	Classic Cars	Brickhaven	USA	Churn	Churn	Churn
Osaka Souveniers Co.	67605.07	Motorcycles	Osaka	Japan	Churn	Churn	Churn
Signal Collectibles Ltd.	50218.51	Trucks and Buses	Brisbane	USA	Churn	Churn	Churn
West Coast Collectables Co.	46084.64	Classic Cars	Burbank	USA	Churn	Churn	Churn

RFM Inferences

Loyal Customers: -

Any customer with a decent recency matrix can be termed as loyal

CUSTOMER NAME	SALES	CATEGORY	CITY	COUNTRY	Monetary_HML	Frequency_HML	Recency_HML
Alpha Cognac	70488.44	Classic Cars	Toulouse	France	Churn	Churn	High Valued
Anna's Decorations, Ltd	153996.13	Classic Cars	North Sydney	Australia	High Valued	High Valued	Mid Valued
Australian Collectables, Ltd	64591.46	Vintage Cars	Glen Waverly	Australia	Churn	Risk/Potential	High Valued
Australian Collectors, Co.	200995.41	Motorcycles	Melbourne	Australia	High Valued	High Valued	Mid Valued
Australian Gift Network, Co	59469.12	Classic Cars	South Brisbane	Australia	Churn	Churn	Mid Valued
Auto Canal Petit	93170.66	Motorcycles	Paris	France	Mid Valued	Mid Valued	High Valued
Auto-Moto Classics Inc.	26479.26	Ships	Brickhaven	USA	Churn	Churn	Mid Valued
Boards & Toys Co.	9129.35	Classic Cars	Glendale	USA	Churn	Churn	Mid Valued
Collectables For Less Inc.	81577.98	Classic Cars	Brickhaven	USA	Risk/Potential	Risk/Potential	Mid Valued
Danish Wholesale Imports	145041.6	Classic Cars	Kobenhavn	Denmark	High Valued	High Valued	High Valued
Diecast Classics Inc.	122138.14	Motorcycles	Allentown	USA	High Valued	Mid Valued	High Valued
Dragon Souveniers, Ltd.	172989.68	Classic Cars	Singapore	Singapore	High Valued	High Valued	Mid Valued
Euro Shopping Channel	912294.11	Motorcycles	Madrid	Spain	High Valued	High Valued	High Valued