PGP-DSBA PROJECT REPORT

MRA – MAIN PROJECT
PART 1

BY
ISHAAN SHAKTI JAYARAMAN
PGPDSBA.O.JULY24.A

AGENDA

- PROBLEM STATEMENT
- DATA CONTENTS
- DATA DESCRIPTION
- STATISTICAL SUMMARY
- EXPLORATORY DATA ANALYSIS
  - UNIVARIATE ANALYSIS
  - BIVARIATE ANALYSIS
  - MULTIVARIATE ANALYSIS
- CUSTOMER SEGMENTATION & RFM ANALYSIS
  - RFM ANALYSIS
  - KNIME WORKFLOW
  - FINAL OUTPUT
- INFERENCES FROM RFM ANALYSIS
  - TOP 5 BEST CUSTOMERS
  - TOP 5 CUSTOMERS ON THE VERGE OF CHURNING
  - TOP 5 LOST CUSTOMERS
  - TOP 5 LOYAL CUSTOMERS

PROBLEM STATEMENT

An automobile parts manufacturer, experiencing growth over three years, lacks the capability to analyze transaction data for actionable insights. The company aims to uncover patterns in customer purchasing behavior, segment its customers effectively, and optimize marketing strategies. By leveraging data science, the company seeks to implement personalized marketing approaches to improve customer satisfaction, boost sales, and enhance retention. The task is to analyze the data, derive insights, and provide recommendations for informed decision-making.

DATA CONTENTS

The dataset (Sales_Data_1.xlsx) contains information pertaining to sales data of various customers for the automobile parts manufacturing company.

• There are 2747 observations in the dataset

• There are 20 columns of which 7 are numerical variables, 12 are categorical variables and 1 is a date/time variable.

• There are no null values in the dataset.

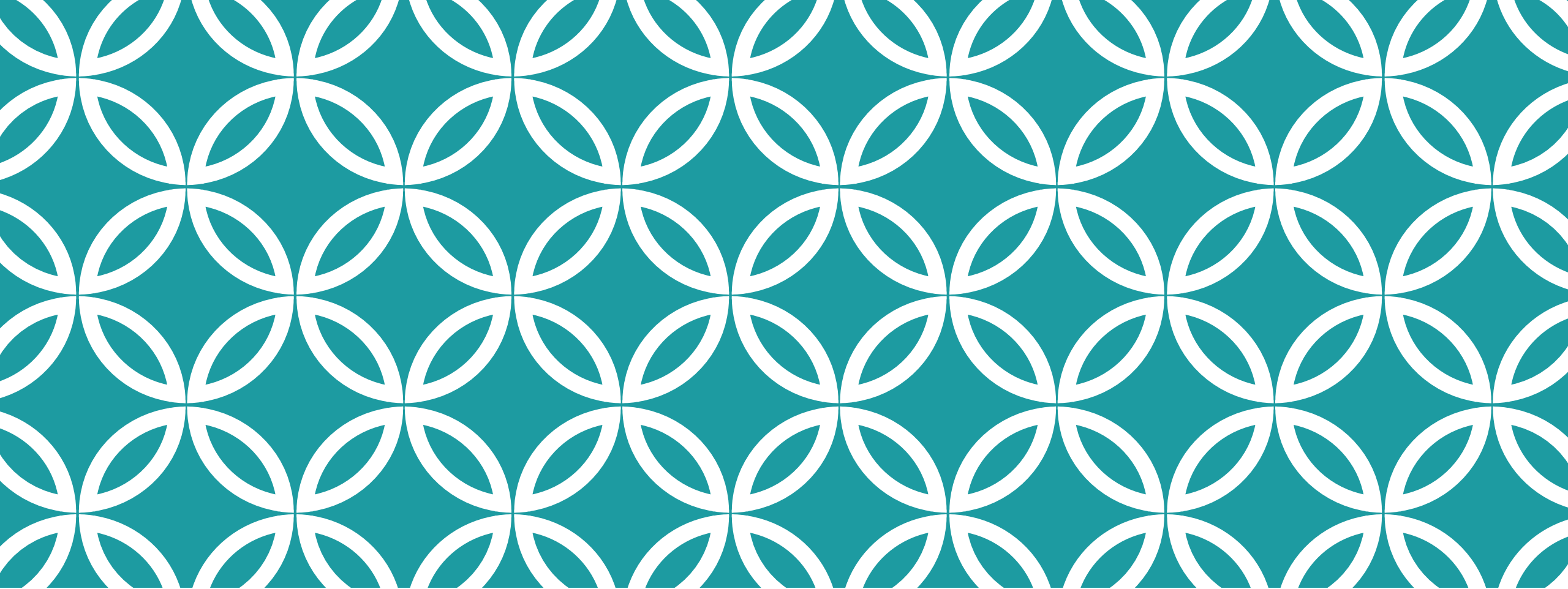• There are no duplicate entries in the dataset.

DATA DESCRIPTION

- **ORDERNUMBER**: Unique identifier for each order.

- **QUANTITYORDERED**: Number of items ordered in a specific transaction.

- **PRICEEACH**: Price per unit of the product in the order.

- **ORDERLINENUMBER**: Sequence number of the product in the order.

- **SALES**: Total sales value for the order.

- **ORDERDATE**: Date when the order was placed.

- **DAYS_SINCE_LASTORDER**: Number of days since the customer's previous order.

- **STATUS**: Current status of the order (e.g., Shipped, Disputed).

- **PRODUCTLINE**: Product category to which the item belongs (e.g., Motorcycles, Classic Cars).

- **MSRP**: Manufacturer's Suggested Retail Price for the product.

- **PRODUCTCODE**: Unique identifier for the product.

- **CUSTOMERNAME**: Name of the customer placing the order.

- **PHONE**: Customer's contact phone number.

- **ADDRESSLINE1**: Customer's primary address.

- **CITY**: City of the customer's address.

- **POSTALCODE**: Postal code of the customer's address.

- **COUNTRY**: Country of the customer's address.

- **CONTACTLASTNAME**: Last name of the customer's contact person.

- **CONTACTFIRSTNAME**: First name of the customer's contact person.

- **DEALSIZE**: Size category of the transaction (e.g., Small, Medium, Large).

STATISTICAL SUMMARY

	count	mean	min	25%	50%	75%	max	std
ORDERNUMBER	2747.000	10259.762	10100.000	10181.000	10264.000	10334.500	10425.000	91.878
QUANTITYORDERED	2747.000	35.103	6.000	27.000	35.000	43.000	97.000	9.762
PRICEEACH	2747.000	101.099	26.880	68.745	95.550	127.100	252.870	42.043
ORDERLINENUMBER	2747.000	6.491	1.000	3.000	6.000	9.000	18.000	4.231
SALES	2747.000	3553.048	482.130	2204.350	3184.800	4503.095	14082.800	1838.954
ORDERDATE	2747	2019-05-13 21:56:17.211503360	2018-01-06 00:00:00	2018-11-08 00:00:00	2019-06-24 00:00:00	2019-11-17 00:00:00	2020-05-31 00:00:00	NaN
DAYS_SINCE_LASTORDER	2747.000	1757.086	42.000	1077.000	1761.000	2436.500	3562.000	819.281
MSRP	2747.000	100.692	33.000	68.000	99.000	124.000	214.000	40.115

- All customers ordered an average amount of 35 items with the highest amount being 97 items.
- The average price per unit of ordered items is $101.
- On average, the days since last order is 1757 days per customers with lowest being 42 days and the highest is 3562 days
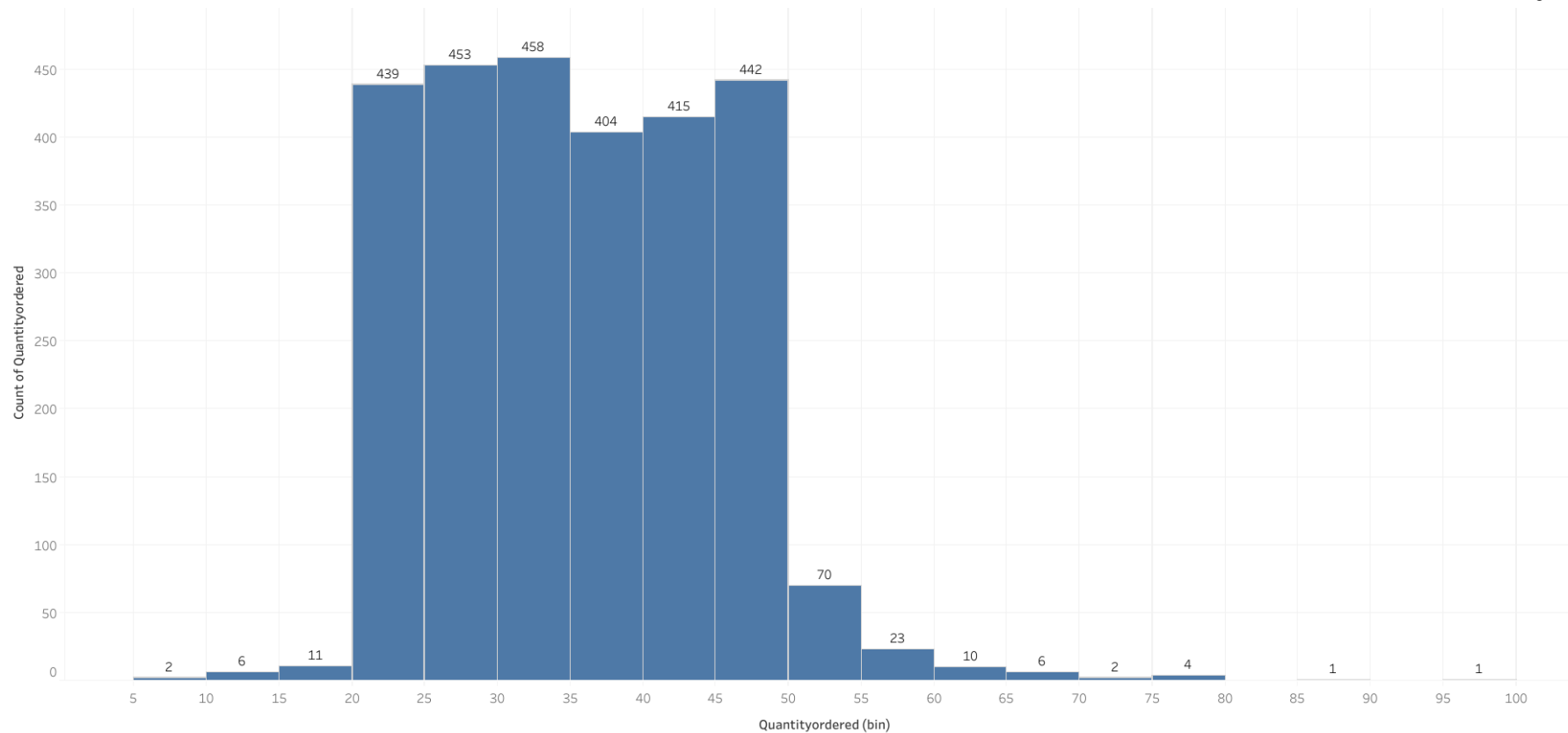- Average MSRP of the product is $100.69.

EXPLORATORY DATA ANALYSIS

Univariate analysis
Bivariate Analysis
Multivariate Analysis

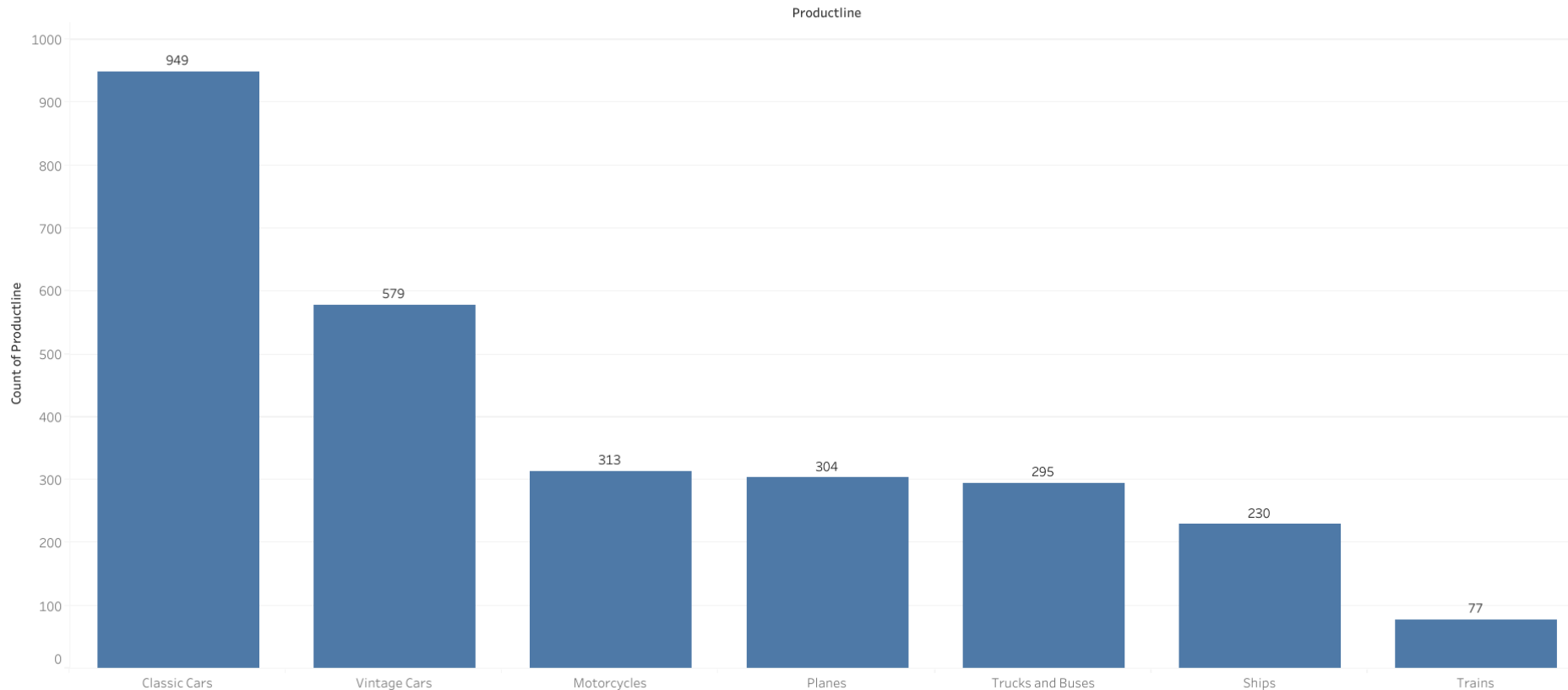# QUANTITY ORDERED

Quantity Ordered

- Majority of customers order products ranging from a quantity of 21-50 in a single order.
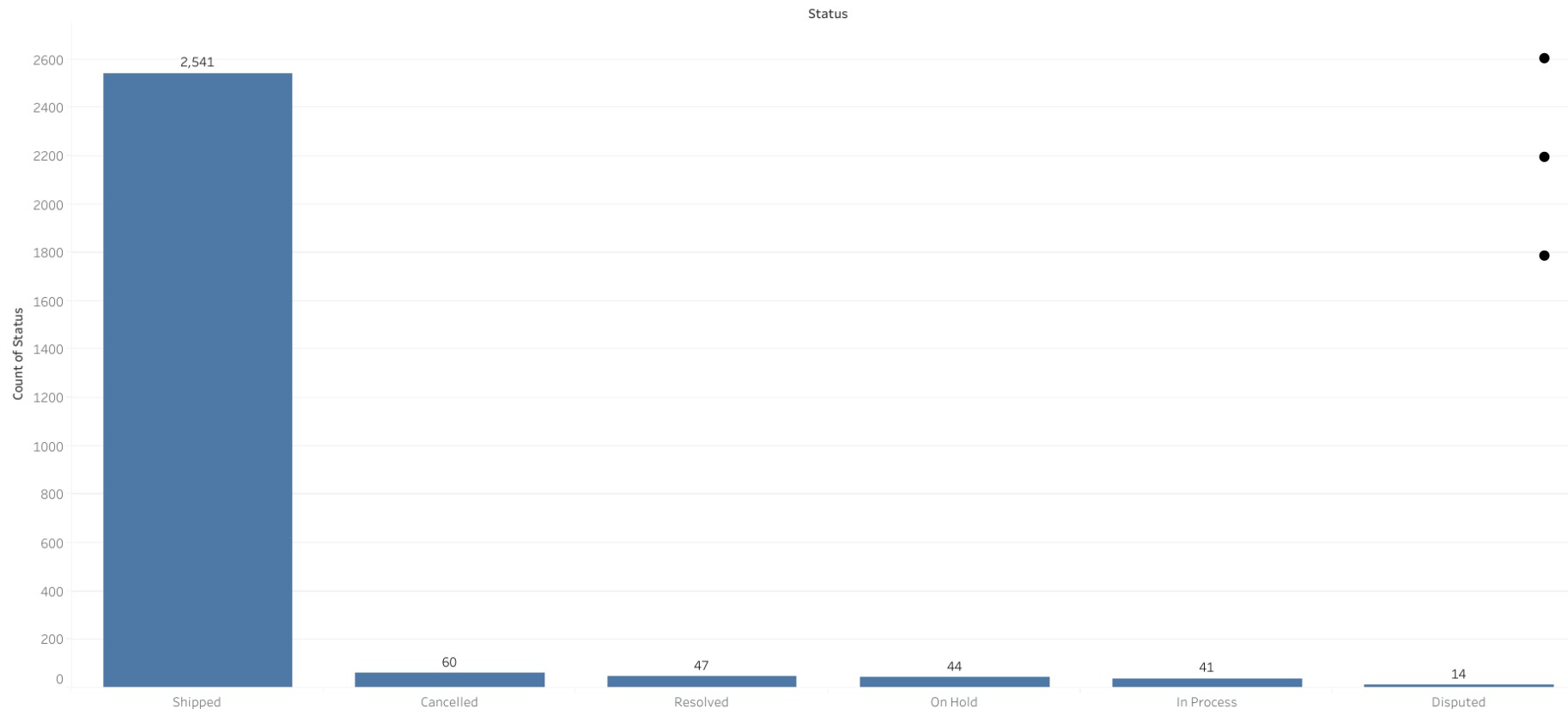
PRODUCT LINE

Product Line

- Parts for Classic cars are the most popular product line followed by Vintage cars.
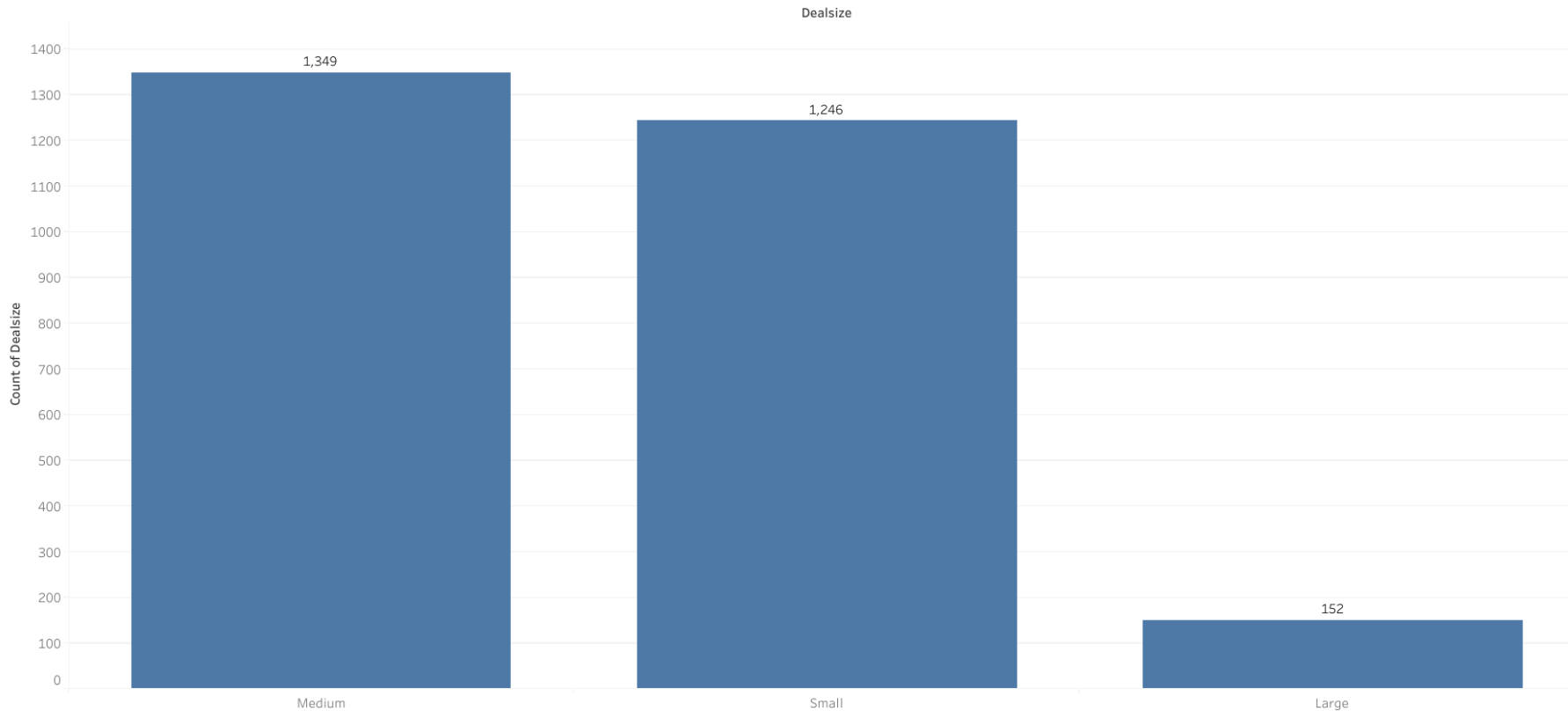- Parts for Trains are the least purchased product line.

STATUS

- Around 2500 orders have been shipped by company.
- 14 orders are currently in dispute.
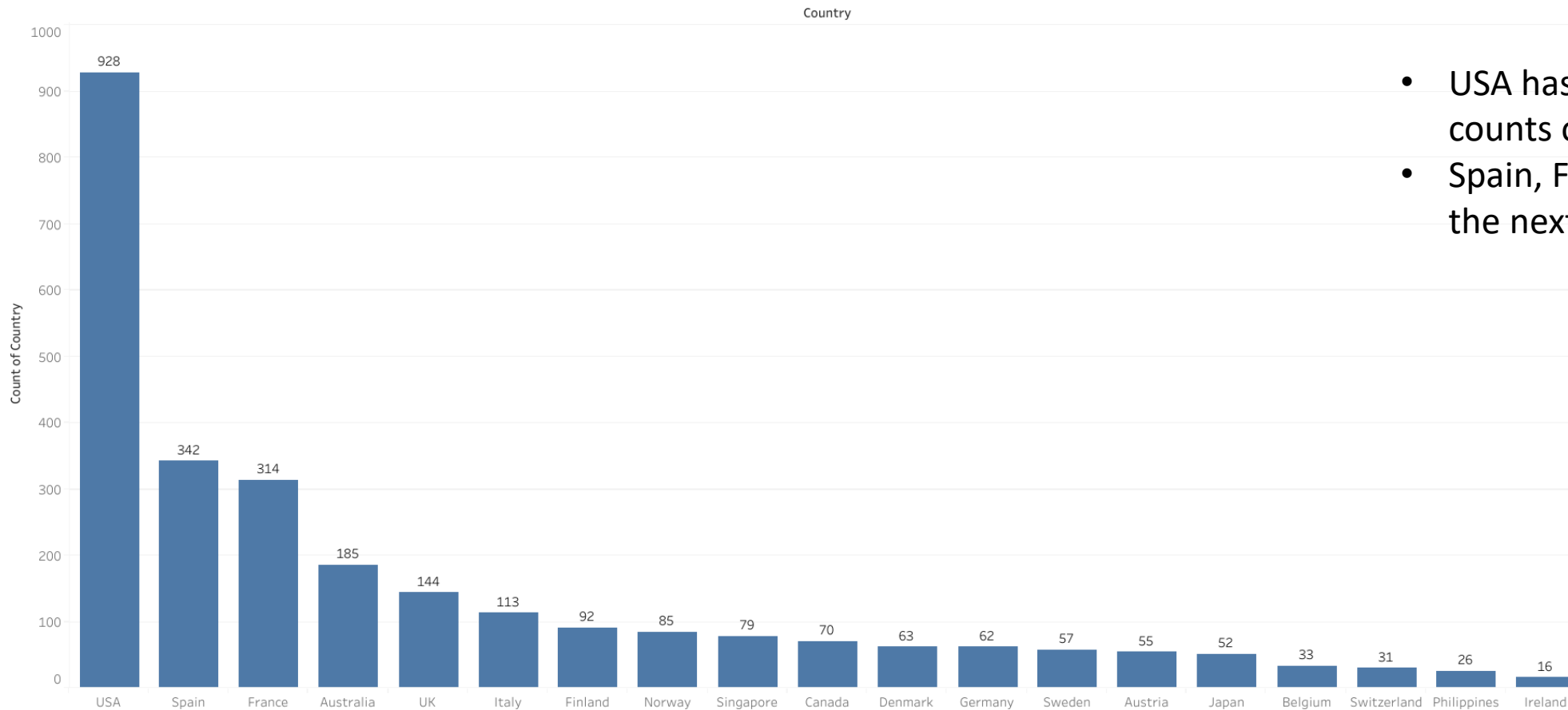- Only 60 orders have been cancelled.

DEAL SIZE

Deal Size

Dealsize

- Majority of orders are Medium category size of order as well as small category.
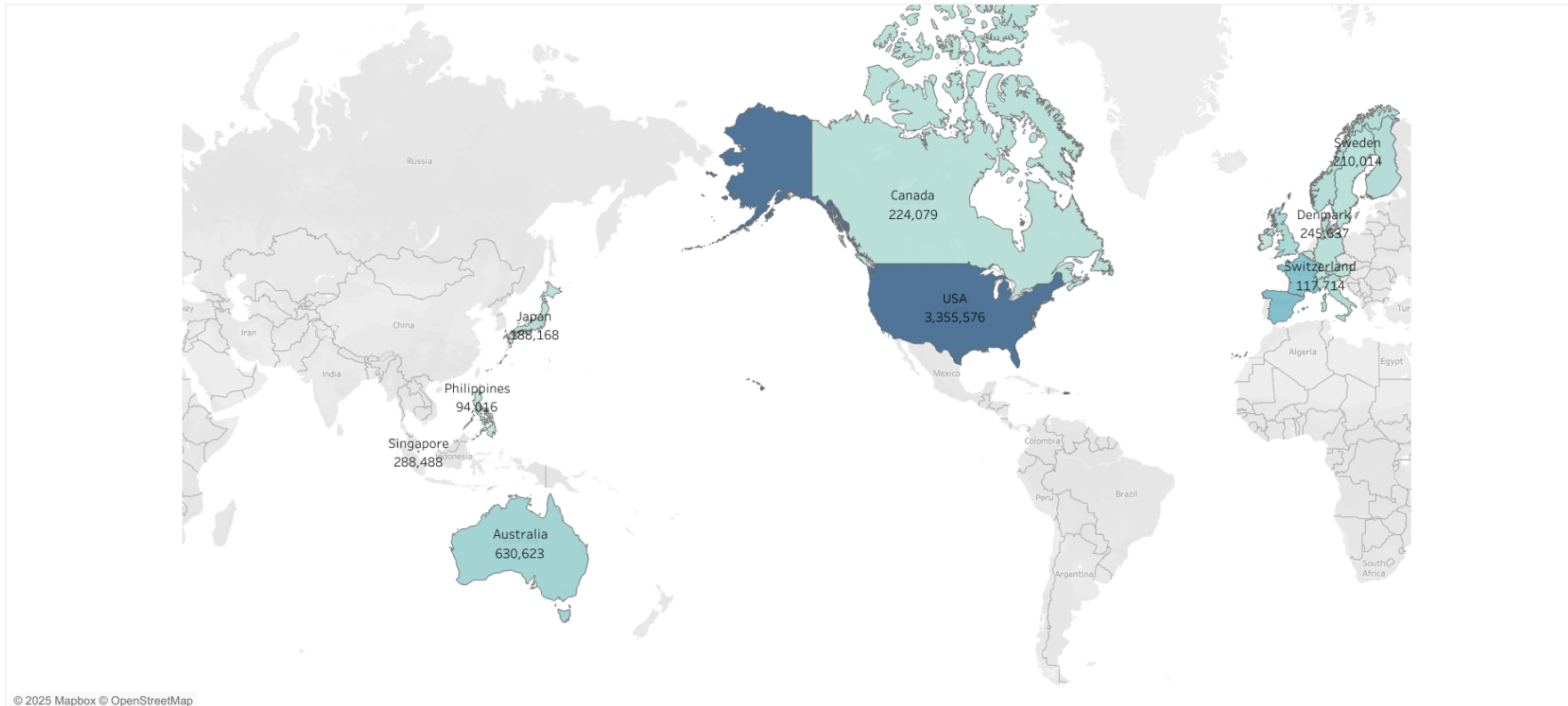- A minority of orders are large category size.

COUNTRY

Country

Country

- USA has the greatest number of counts of order at 928.
- Spain, France and Australia are the next highest countries
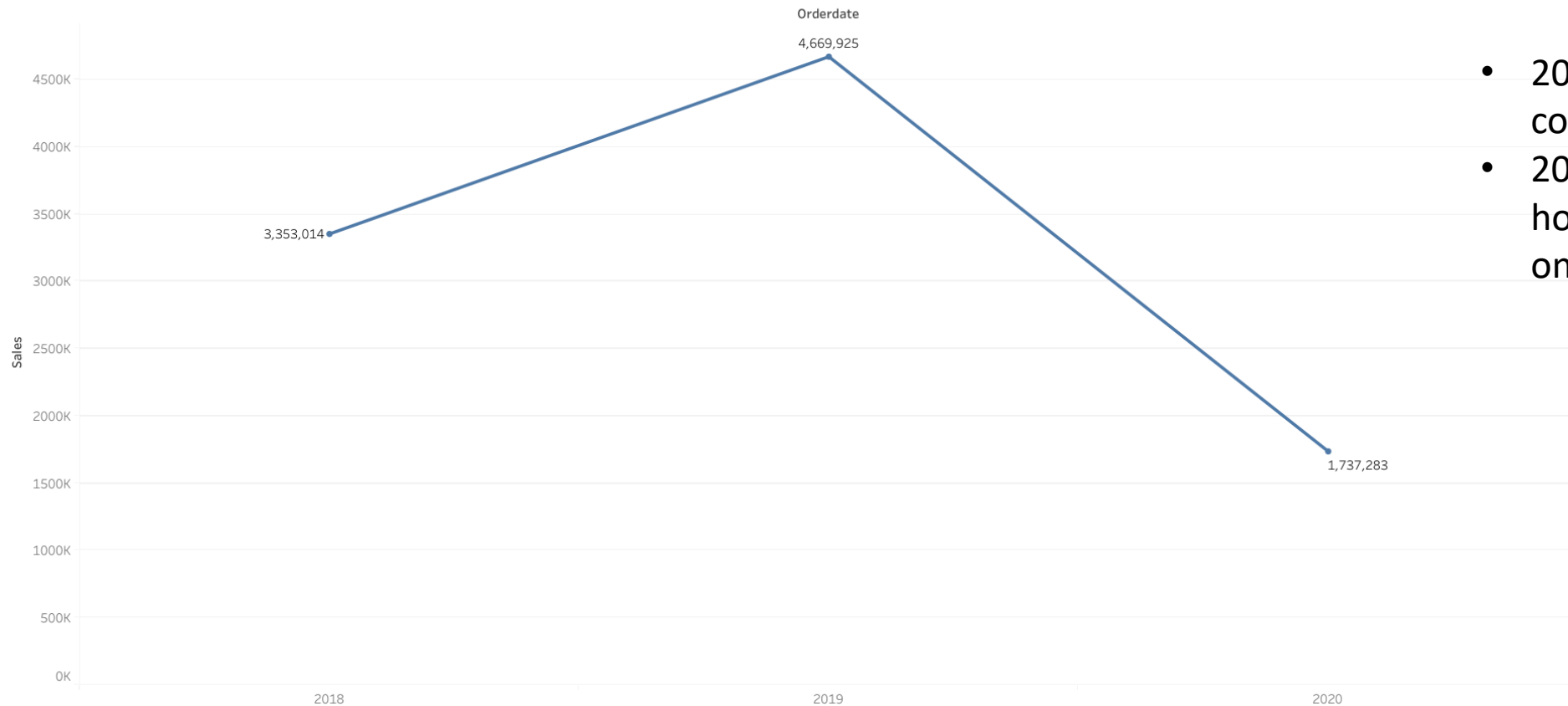
12

SALES ACROSS COUNTRIES

Sales across Countries

© 2025 Mapbox © OpenStreetMap

- USA has highest sales at around $3million.
- France and Spain have the next highest amount of sales of around $1 million.
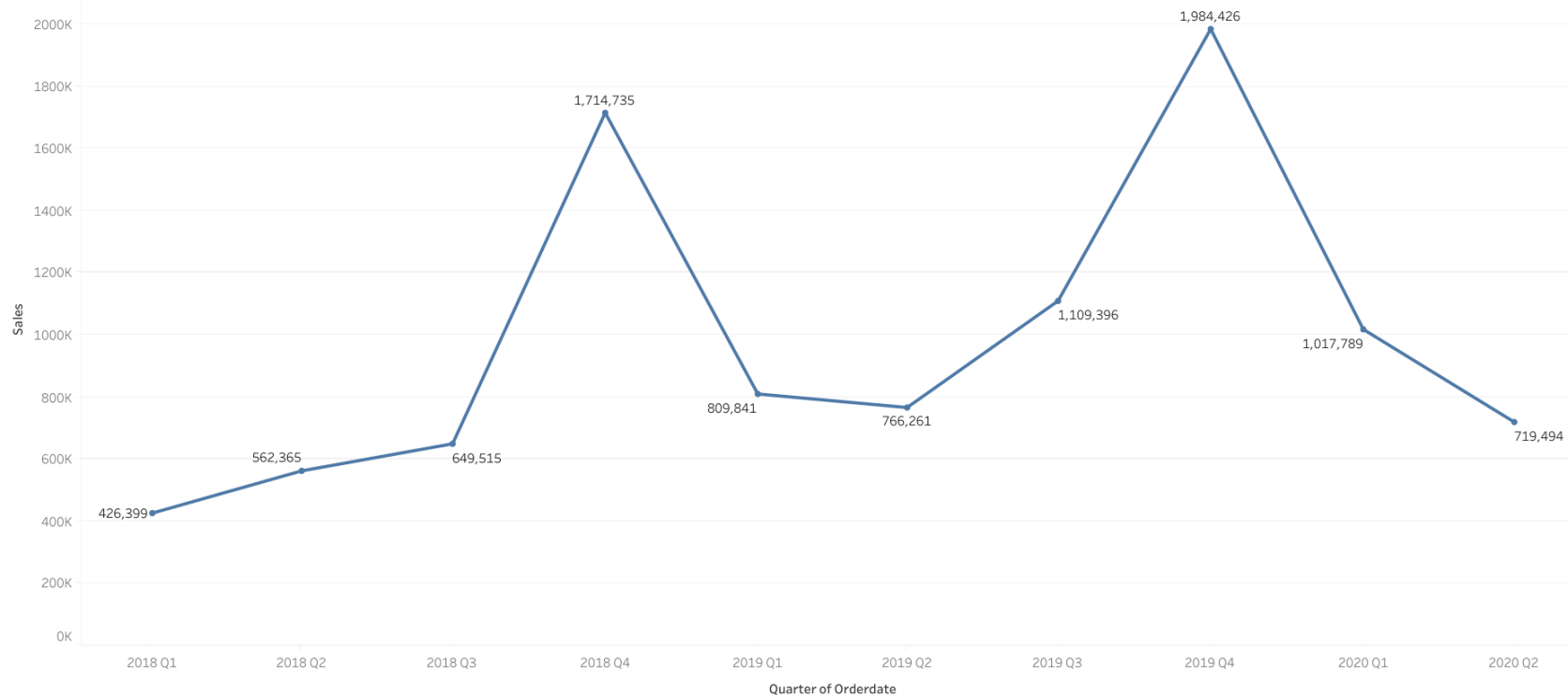
SALES ACROSS YEARS

Sales across Years

- 2019 sales have increased when compared to 2018.
- 2020 sales show decline however it is due to data until only Q2.
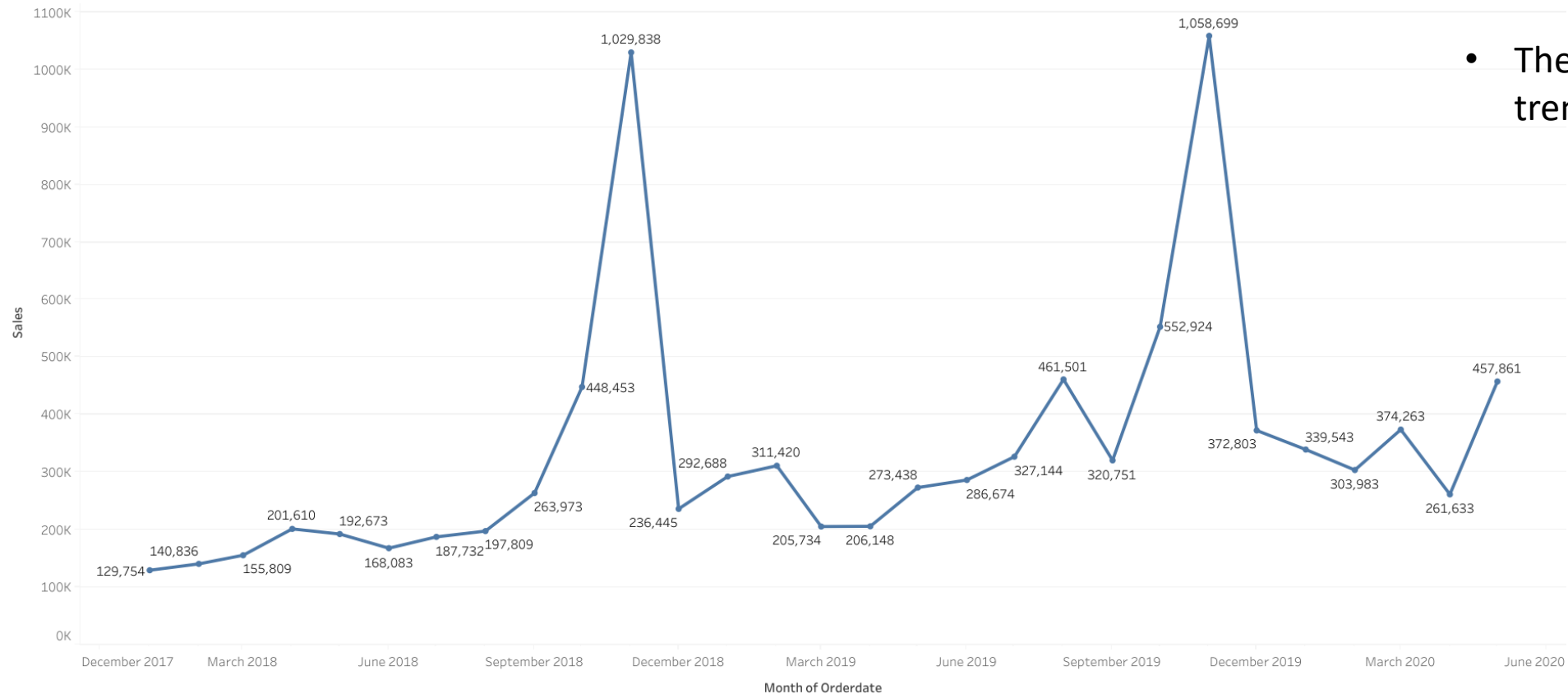
SALES ACROSS YEARLY QUARTERS

Sales across quarters

- There is an increase of sales around Q4 every year indicating seasonality.
- There is a slight decline in sales in 2020 Q2 compared to 2019 Q2.

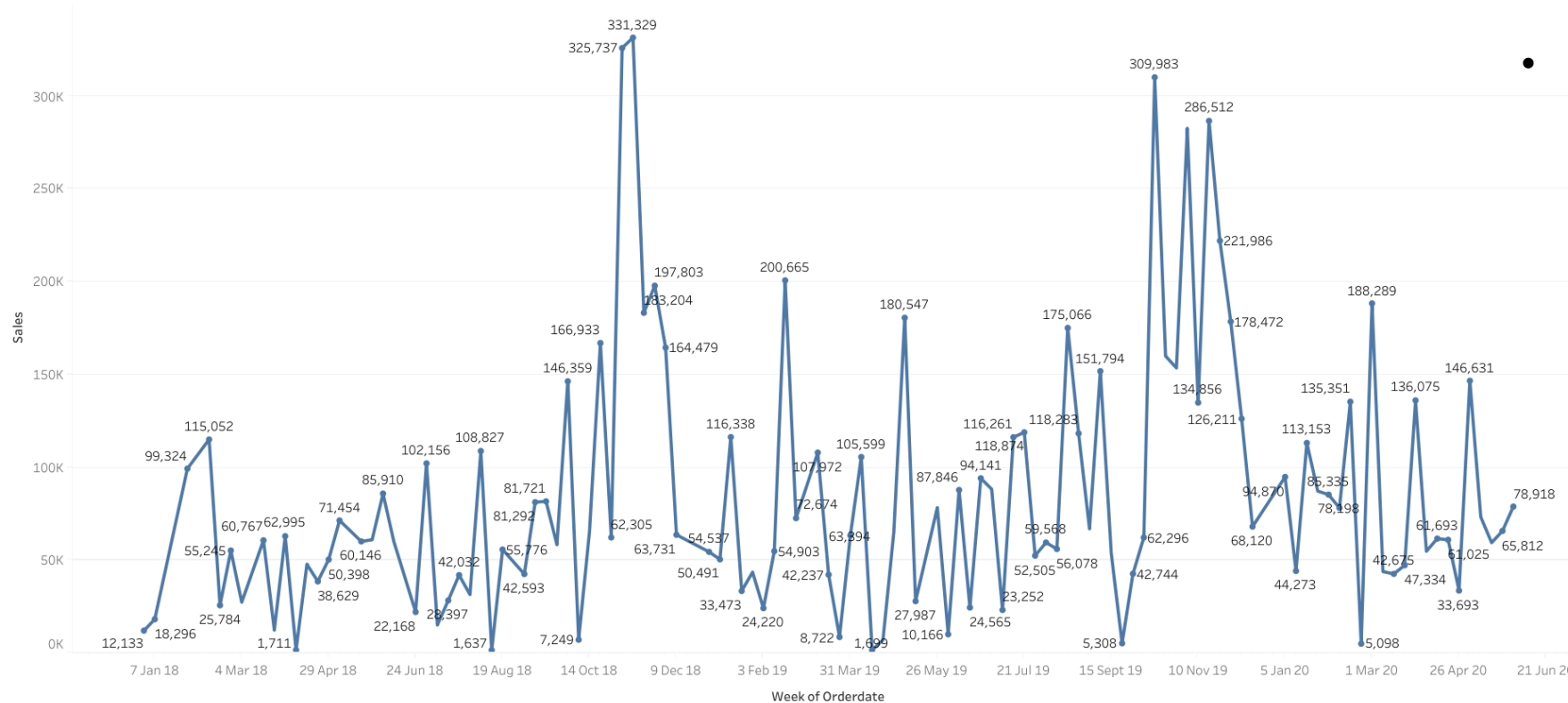# SALES ACROSS MONTHS

Sales across Months

- There is an overall increasing trend of sales over the months.

SALES ACROSS WEEKS

Sales across Weeks

- There is an overall increasing trend of sales over the weeks.

SALES BY CUSTOMERS

Sales by Customers

- Euro Shipping Channel is the top customer by sales, followed by Mini Gifts Distribution Ltd.
- Australian Collectors, Muscle Machine Inc and La Rochelle Gifts are the next best customers.

SALES BY PRODUCT LINE

Sales by Product Line
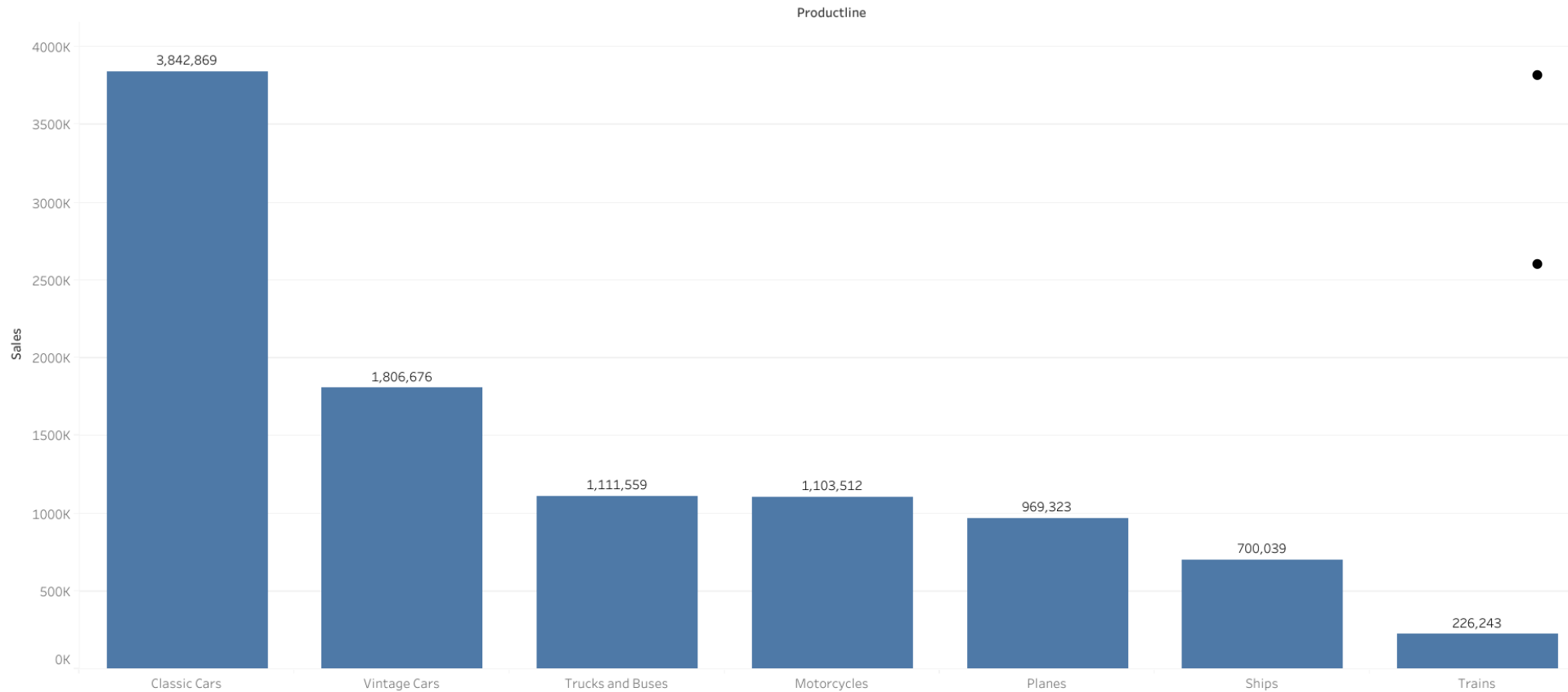
Productline

- Euro Shipping Channel is the top customer by sales, followed by Mini Gifts Distribution Ltd.
- Australian Collectors, Muscle Machine Inc and La Rochelle Gifts are the next best customers.

QUANTITY ORDERED BY PRODUCT LINE

Quantity Ordered by Product LIne
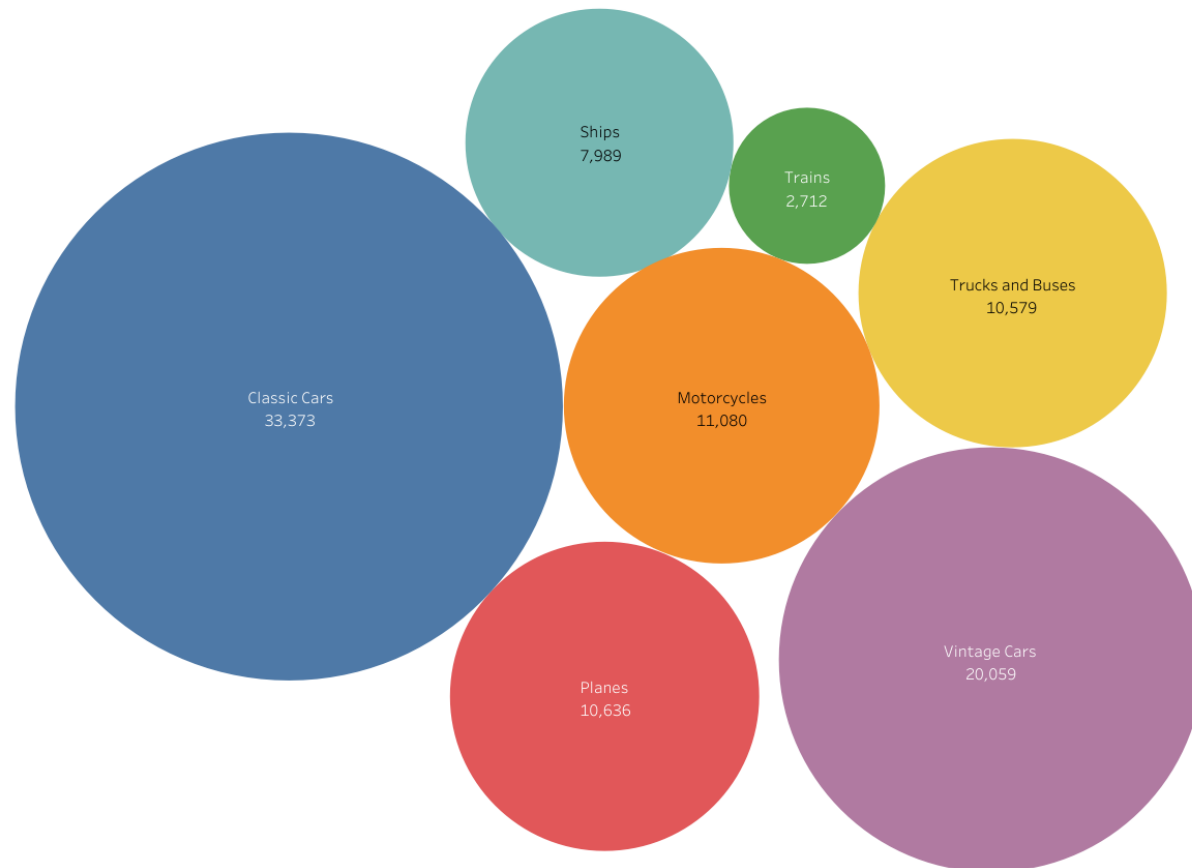
- Classic Cars and Vintage cars have the highest quantity ordered.
- Trains are the least ordered quantity of products.
- There are moderate number of quantities ordered for Planes, Motorcycles and Trucks/Buses products.

CUSTOMER SALES ACROSS YEARS

Customer Sales across Years
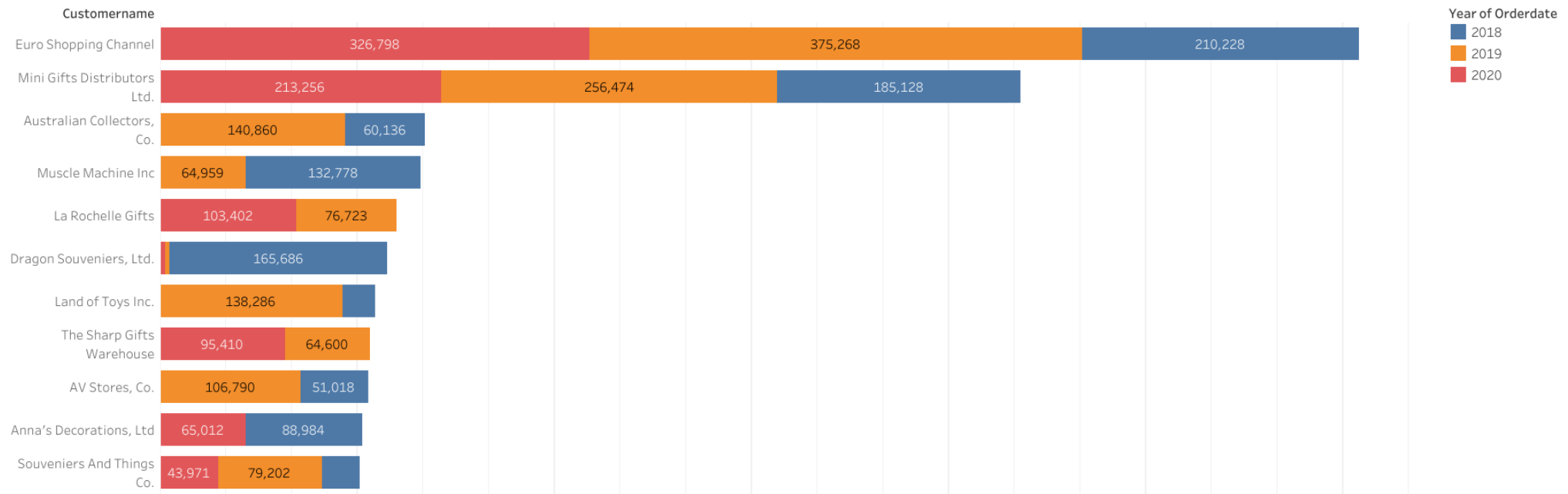
- Euro Shopping Channel and Mini Gifts Distributors Ltd. have remained high values customers throughout 3 years.
- Australian Collectors and Muscle machine Inc have not made any orders in 2020.
- Dragon Souveniers Ltd have drastically reduced their orders compared to 2018.

SALES BY PRODUCT LINE & DEAL SIZE

Sales by Product Line and Deal size

Dealsize	Productline	Sales
Large	Classic Cars	780,091
	Vintage Cars	190,522
	Trucks and ..	56,256
	Motorcycles	151,079
	Planes	72,032
	Trains	8,977
Medium	Classic Cars	2,376,627
	Vintage Cars	1,038,236
	Trucks and ..	809,050
	Motorcycles	643,014
	Planes	535,071
	Ships	416,338
	Trains	112,895
Small	Classic Cars	686,151
	Vintage Cars	577,917
	Trucks and ..	246,253
	Motorcycles	309,419
	Planes	362,221
	Ships	283,702
	Trains	104,371

Sales

- Classic Cars and Vintage cars have the highest sales on all deal sizes.
- Trucks/buses, Motorcycles, Planes have a moderate number of sales.
- Trains have the lowest sales on all deal sizes.

22

CUSTOMER SEGMENTATION & RFM ANALYSIS

RFM ANALYSIS

RFM analysis is a marketing strategy used to evaluate customer value based on three key factors Recency, Frequency, and Monetary value.

- Recency: How recently a customer made a purchase.

- Frequency: How often a customer makes purchases.

- Monetary value: How much a customer spends.

By analyzing these dimensions, businesses can segment customers, target marketing efforts effectively, and improve customer retention strategies.

KNIME WORKFLOW

To create a workflow for RFM Analysis, KNIME Analytics Platform is used to import the excel file and through binning divide the customers into 4 segments high, medium, low and very low for each factor of the RFM metric.
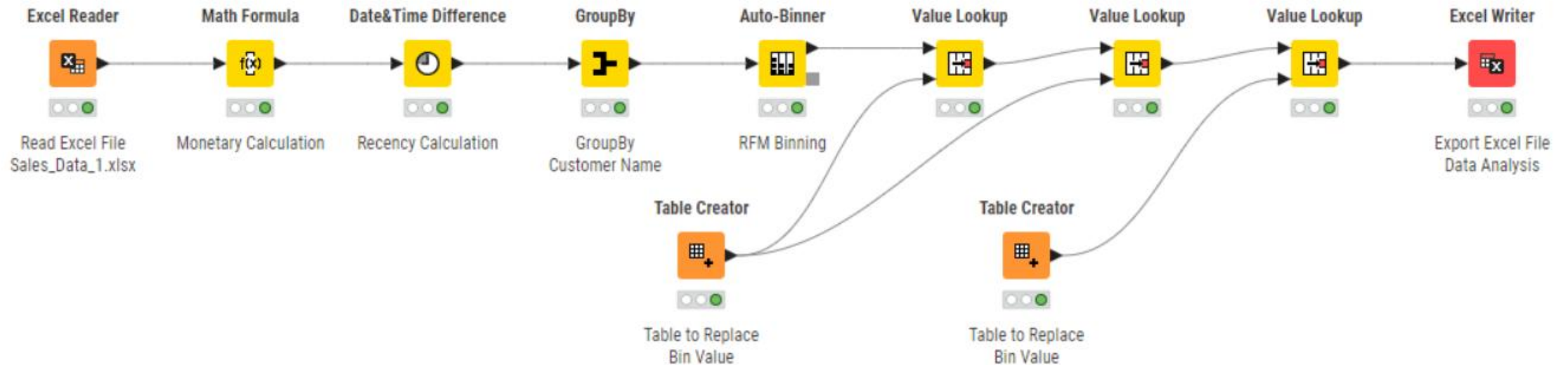
*Value Lookup Node is used to replace values as Cell Replacer node has been deprecated

FINAL OUTPUT

Rows: 89 | Columns: 25   Search

| RowID	C... St...	O... N...	Q... N...	P... N...	O... N...	S... N...	O... Lo...	DAYS_... Number ...	S... Stri...	P... St...	M... Nu...	P... St...	P... St...	A... St...	C... St...	P... St...	C... N...	C... St...	C... Stri...	D... St...	Mon... Number...	Rece... Number...	ORDERN... String	Monetary... String	Recer... String
Row0	AV Sto	51	34.863	91.085	9.02	3,094.:	2019-1	1,803.804	Shipped	Classi(	92.843	S12_1	(171) 5	Fauntle	Manch	EC2 5N	51	Ashwo	Victoria	Mediu	157,807.81	1979	High	High	Low
Row1	Alpha (	20	34.35	101.16	4.95	3,524.4	2018-1	2,236.2	Shipped	Classi(	97.15	S10_4:	61.77.(	1 rue A	Toulou	31000	20	Roulet	Annette	Mediu	70,488.44	1847	Very Low	Very Low	High
Row2	Amica	26	32.423	110.85	7.615	3,619.8	2019-0	1,318.615	Shipped	Classi(	107.654	S10_19	011-49	Via Mo	Torino	10100	26	Accort	Paolo	Large	94,117.26	2048	Low	Medium	Very Low
Row3	Anna's	46	31.935	106.42	6.435	3,347.:	2020-0	1,463.587	Shipped	Classi(	104.717	S10_19	02 993	201 Mi	North :	2060	46	O'Hara	Anna	Small	153,996.13	1866	High	High	Medium
Row4	Atelier	7	38.571	92.239	2	3,454.:	2019-0	1,424.429	Shipped	Motor(	95.571	S10_2(	40.32.:	54, rue	Nantes	44000	7	Schmit	Carine	Mediu	24,179.96	1971	Very Low	Very Low	Low
Row5	Austra	23	30.652	90.042	6.696	2,808.:	2018-1	1,689.696	Shipped	Vintag	88.13	S18_1:	61-9-3(	7 Allen	Glen W	3150	23	Conne	Sean	Mediu	64,591.46	1805	Low	Very Low	High
Row6	Austra	55	35.018	104.59	7.036	3,654.4	2019-0	1,698.782	Shipped	Motor(	103.527	S10_1(	03 952	636 St	Melbo	3004	55	Fergus	Peter	Mediu	200,995.41	1967	High	High	Medium
Row7	Austra	15	36.333	110.55	3.067	3,964.(	2018-1	1,349.133	Shipped	Classi(	111.533	S10_19	61-7-3(	31 Dun	South	4101	15	Calagh	Tony	Large	59,469.12	1902	Very Low	Very Low	Medium
Row8	Auto A	18	35.389	99.488	8.556	3,601.9	2019-1	1,266.278	Shipped	Classi(	100.389	S10_19	30.59.8	67, ave	Versail	78000	18	Tonini	Daniel	Large	64,834.32	2016	Very Low	Very Low	Very Low
Row9	Auto C	27	37.074	94.255	6.333	3,450.:	2019-0	1,763.444	Shipped	Motor(	94.852	S10_1(	(1) 47.	25, rue	Paris	75016	27	Perrier	Dominic	Mediu	93,170.66	1837	Medium	Medium	High
Row10	Auto-N	8	35.875	92.8	2	3,309.9	2018-0	2,281.75	Shipped	Ships	87.375	S18_3(	61755!	16780	Brickh:	58339	8	Taylor	Leslie	Mediu	26,479.26	1963	Very Low	Very Low	Medium

Each customer has been identified a value of either High, Medium, Low and Very Low for the 3 factors of Recency, Frequency and Monetary. Customers can be targeted based on these factors to improve retention by promoting marketing strategies.
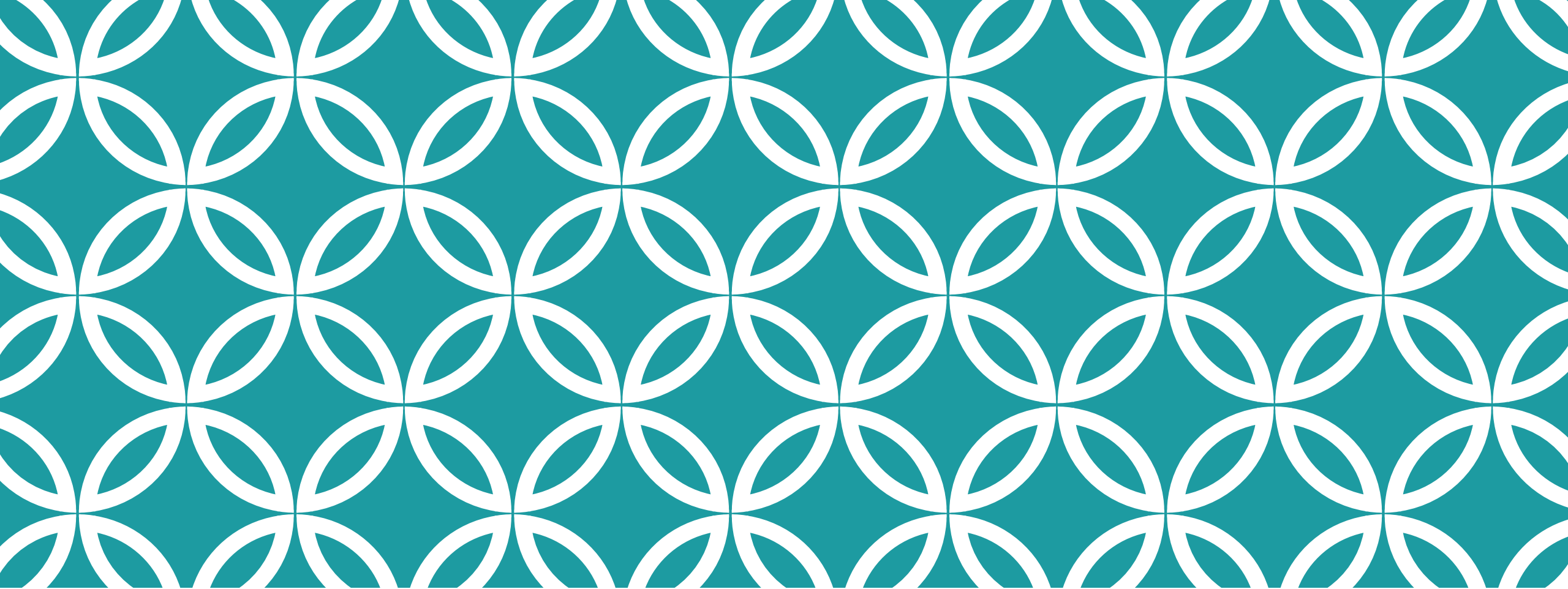
INFERENCES FROM RFM ANALYSIS

TOP 5 BEST CUSTOMERS

CUSTOMERNAME	ORDERNUMBER	QUANTITY ORDERED	SALES	Monetary	Recency	ORDERNUMBER [Binned]	Monetary [Binned]	Recency [Binned]
Euro Shopping Channel	259	36.01	3522.371081	912294.11	1783	High	High	High
Mini Gifts Distributors Ltd.	180	35.37	3638.100333	654858.06	1785	High	High	High
La Rochelle Gifts	53	34.57	3398.583019	180124.9	1783	High	High	High
The Sharp Gifts Warehouse	40	41.40	4000.25675	160010.27	1822	High	High	High
Souveniers And Things Co.	46	34.80	3295.021304	151570.98	1785	High	High	High

The Top 5 Best Customers are those who have the highest RFM scores, they are

- Euro Shopping Channel
- Mini Gifts Distributors Ltd.
- La Rochelle Gifts
- The Sharp Gifts Warehouse
- Souveniers and Things Co.

TOP 5 CUSTOMERS ON THE VERGE OF CHURNING

CUSTOMERNAME	ORDERNUMBER	QUANTITY ORDERED	SALES	Monetary	Recency	ORDERNUMBER [Binned]	Monetary [Binned]	Recency [Binned]
Heintze Collectables	27	32.66666667	3725.761111	100595.55	2005	Medium	Medium	Low
Corrida Auto Replicas, Ltd	32	36.34375	3769.2275	120615.28	1995	Medium	High	Low
Online Diecast Creations Co.	34	36.70588235	3873.097059	131685.3	1992	High	High	Low
Baane Mini Imports	32	33.8125	3643.724688	116599.19	1991	Medium	Medium	Low
Rovelli Gifts	48	34.375	2874.0775	137955.72	1984	High	High	Low

The Top 5 Customers on the verge of churning are those who have low recency score but high-medium frequency and monetary scores, they are
- Heintze Collectables
- Corrida Auto Replicas, Ltd
- Online Diecast Creations Co.
- Baane Mini Imports
- Rovelli Gifts

TOP 5 LOST CUSTOMERS

CUSTOMERNAME	ORDERNUMBER	QUANTITY ORDERED	SALES	Monetary	Recency	ORDERNUMBER [Binned]	Monetary [Binned]	Recency [Binned]
Daedalus Designs Imports	20	34.95	3452.6205	69052.41	2248	Very Low	Very Low	Very Low
Osaka Souveniers Co.	20	34.6	3380.2535	67605.07	2197	Very Low	Very Low	Very Low
Auto Assoc. & Cie.	18	35.38888889	3601.906667	64834.32	2016	Very Low	Very Low	Very Low
Clover Collections, Co.	16	30.625	3609.776875	57756.43	2041	Very Low	Very Low	Very Low
Online Mini Collectables	15	38.13333333	3813.197333	57197.96	2047	Very Low	Very Low	Very Low

The Top 5 Lost Customers are those who have very low RFM scores, they are
- Daedalus Designs Imports
- Osaka Souveniers Co.
- Auto Assoc. & Cie.
- Clover Collections, Co.
- Online Mini Collectables

TOP 5 LOYAL CUSTOMERS

CUSTOMERNAME	ORDERNUMBER	QUANTITY ORDERED	SALES	Monetary	Recency	ORDERNUMBER [Binned]	Monetary [Binned]	Recency [Binned]
The Sharp Gifts Warehouse	40	41.40	4000.25675	160010.27	1822	High	High	High
Danish Wholesale Imports	36	36.53	4028.933333	145041.6	1829	High	High	High
Salzburg Collectables	40	36.05	3744.96575	149798.63	1797	High	High	High
Euro Shopping Channel	259	36.01	3522.371081	912294.11	1783	High	High	High
Mini Gifts Distributors Ltd.	180	35.37	3638.100333	654858.06	1785	High	High	High

The Top 5 Loyal Customers are those who have high Recency and Frequency Scores, they are
- The Sharp Gifts Warehouse
- Danish Wholesale Imports
- Salzburg Collectables
- Euro Shopping Channel
- Mini Gifts Distributors Ltd.
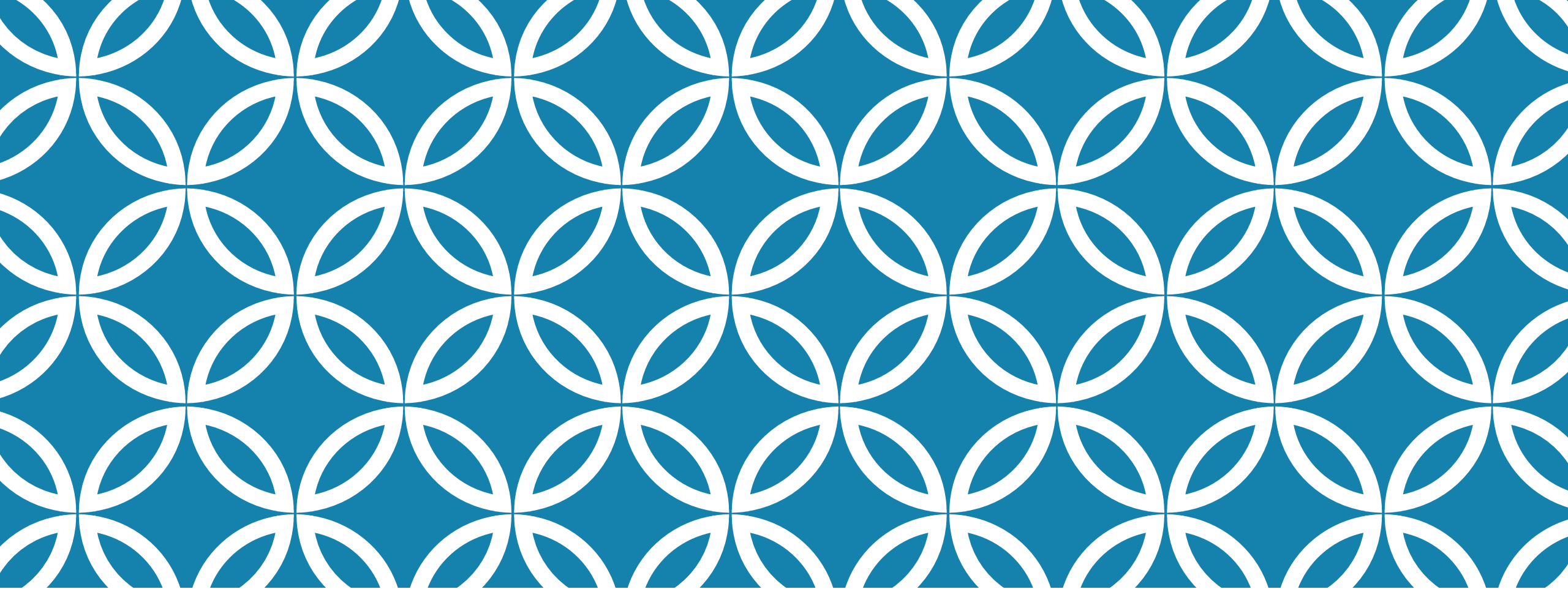
END OF PART 1

PGP-DSBA MRA Main Project
BY
ISHAAN SHAKTI JAYARAMAN
PGPDSBA.O.JULY24.A