MRA PROJECT - MILESTONE 1

CONTENTS:

- Project Description/Agenda
- Problem Statement
- Data Dictionary
- About Data
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

PROJECT DESCRIPTION

This project aims to find the underlying buying patterns of the customers of an automobile part manufacturer based on the past 3 years of the company's transaction data and hence recommend customized marketing strategies for different segments of customers.

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.

DATA DICTIONARY

- 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

- Info: Data consists of total 2747 entries, with zero null values.
- Shape: Data has 2747 Rows and 20 Columns.
- Data Types: It has 1 column of ‘datetime64[ns]’ data type, 2 columns of ‘float64’, 5 columns of ‘int64’ and 12 columns of ‘object’ data type.

```
RangeIndex: 2747 entries, 0 to 2746
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ORDERNUMBER           2747 non-null   int64
 1   QUANTITYORDERED       2747 non-null   int64
 2   PRICEEACH             2747 non-null   float64
 3   ORDERLINENUMBER       2747 non-null   int64
 4   SALES                 2747 non-null   float64
 5   ORDERDATE             2747 non-null   datetime64[ns]
 6   DAYS_SINCE_LASTORDER  2747 non-null   int64
 7   STATUS                2747 non-null   object
 8   PRODUCTLINE           2747 non-null   object
 9   MSRP                  2747 non-null   int64
 10  PRODUCTCODE           2747 non-null   object
 11  CUSTOMERNAME          2747 non-null   object
 12  PHONE                 2747 non-null   object
 13  ADDRESSLINE1          2747 non-null   object
 14  CITY                  2747 non-null   object
 15  POSTALCODE            2747 non-null   object
 16  COUNTRY               2747 non-null   object
 17  CONTACTLASTNAME       2747 non-null   object
 18  CONTACTFIRSTNAME      2747 non-null   object
 19  DEALSIZE              2747 non-null   object
dtypes: datetime64[ns](1), float64(2), int64(5), object(12)
```

- Summary Stats: Min Sales- 482.13, Max Sales-14082.80, Mean of Sales- 3553.04

  Max. price of item- 252.87 & Min. price of item- 26.88 with standard deviation of 42.04

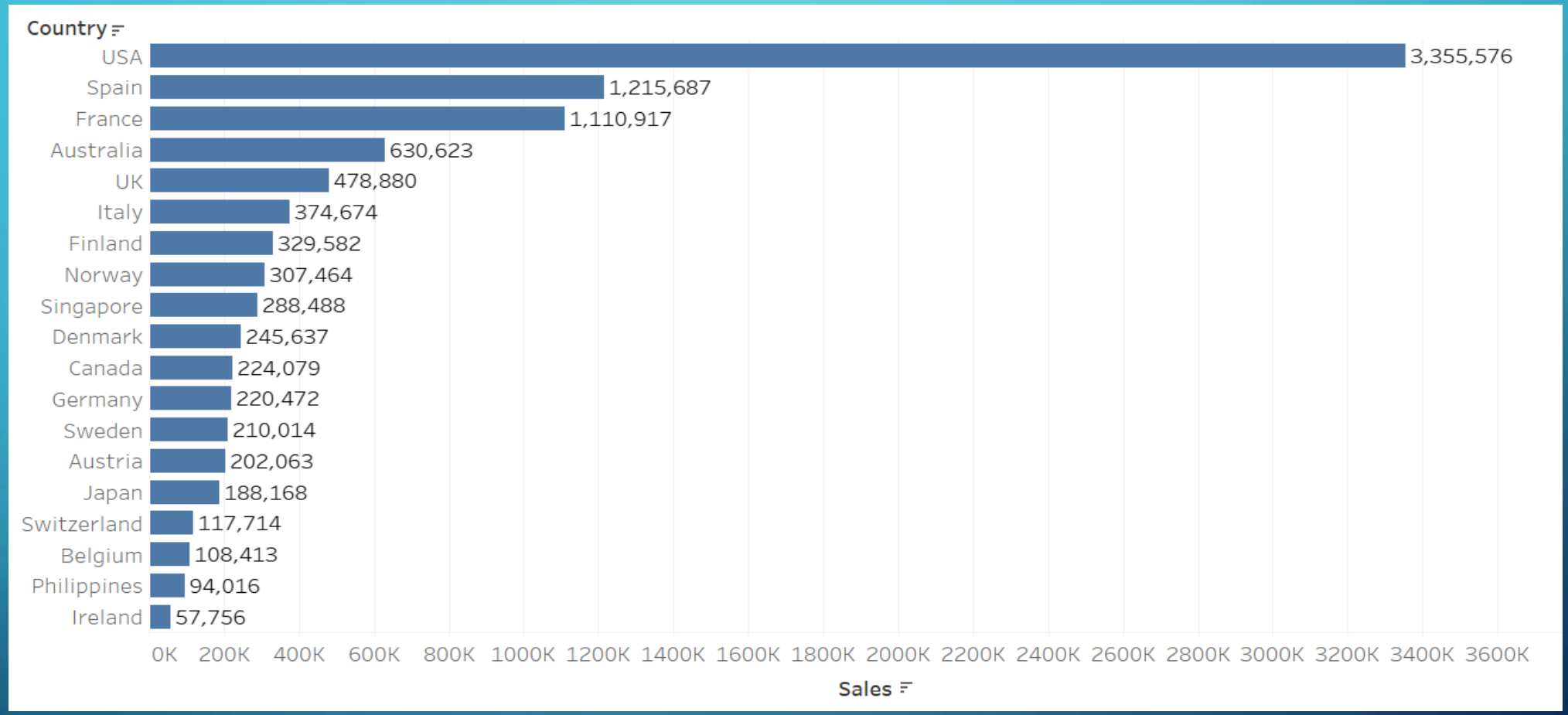
  Min. no. of days since last order- 42 days, Max. no. of days since last order-3562 days.

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081	3553.047583	1757.085912	100.691664
std	91.877521	9.762135	42.042548	4.230544	1838.953901	819.280576	40.114802
min	10100.000000	6.000000	26.880000	1.000000	482.130000	42.000000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	1077.000000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	1761.000000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	2436.500000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	3562.000000	214.000000

- Assumptions: At first glance over the summary of the data, it seems like the items are not too costly, the average price of each item is 101.09. The difference between the max and min ordered quantity is too high, and the standard deviation(9.76) is very less, it means there are vey few customers who ordered higher quantities than others. Also, the average of MSRP is coming out to be 100.69, that isn't very high. We need to focus here on few things mainly like 'Sales' and 'Days since last order'.

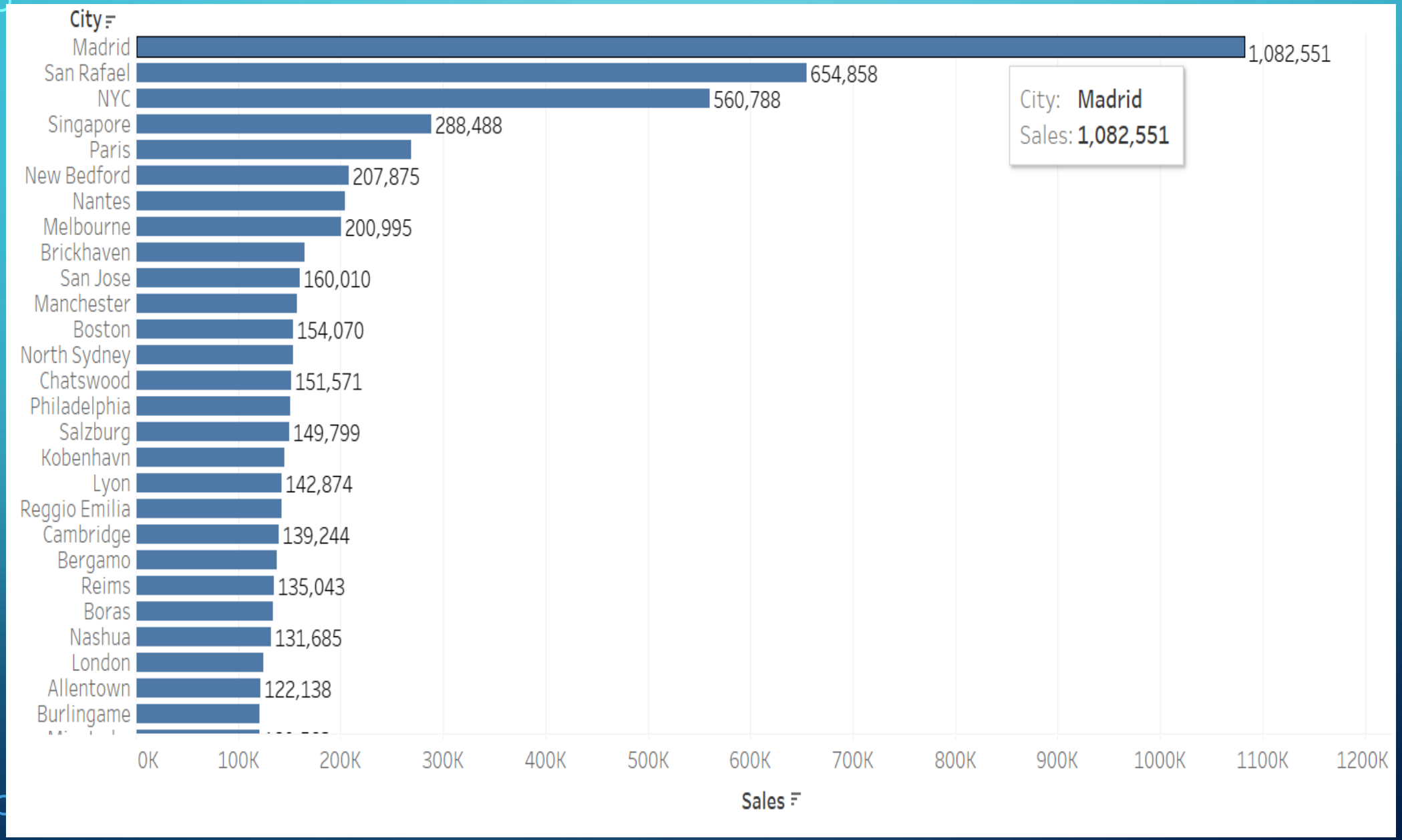
UNIVARIATE, BIVARIATE AND MULTIVARIATE ANALYSIS

1) COUNTRY: Let's first look at the count plot for the COUNTRY.

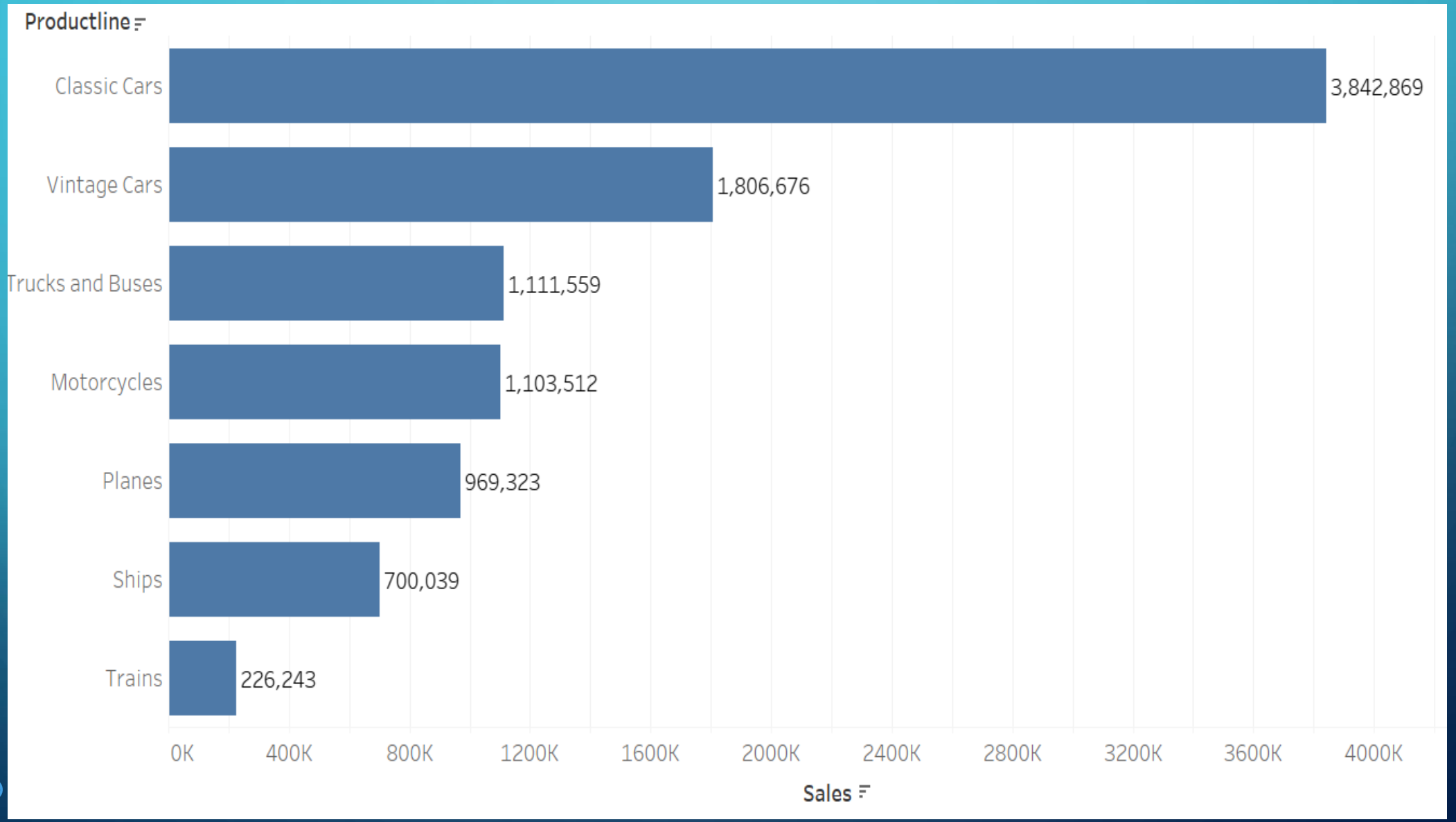


- Out of total 19 countries in the dataset, It seems like USA clearly wins the game here. The maximum number of sales is coming from USA and the minimum number of sales are here from Ireland.

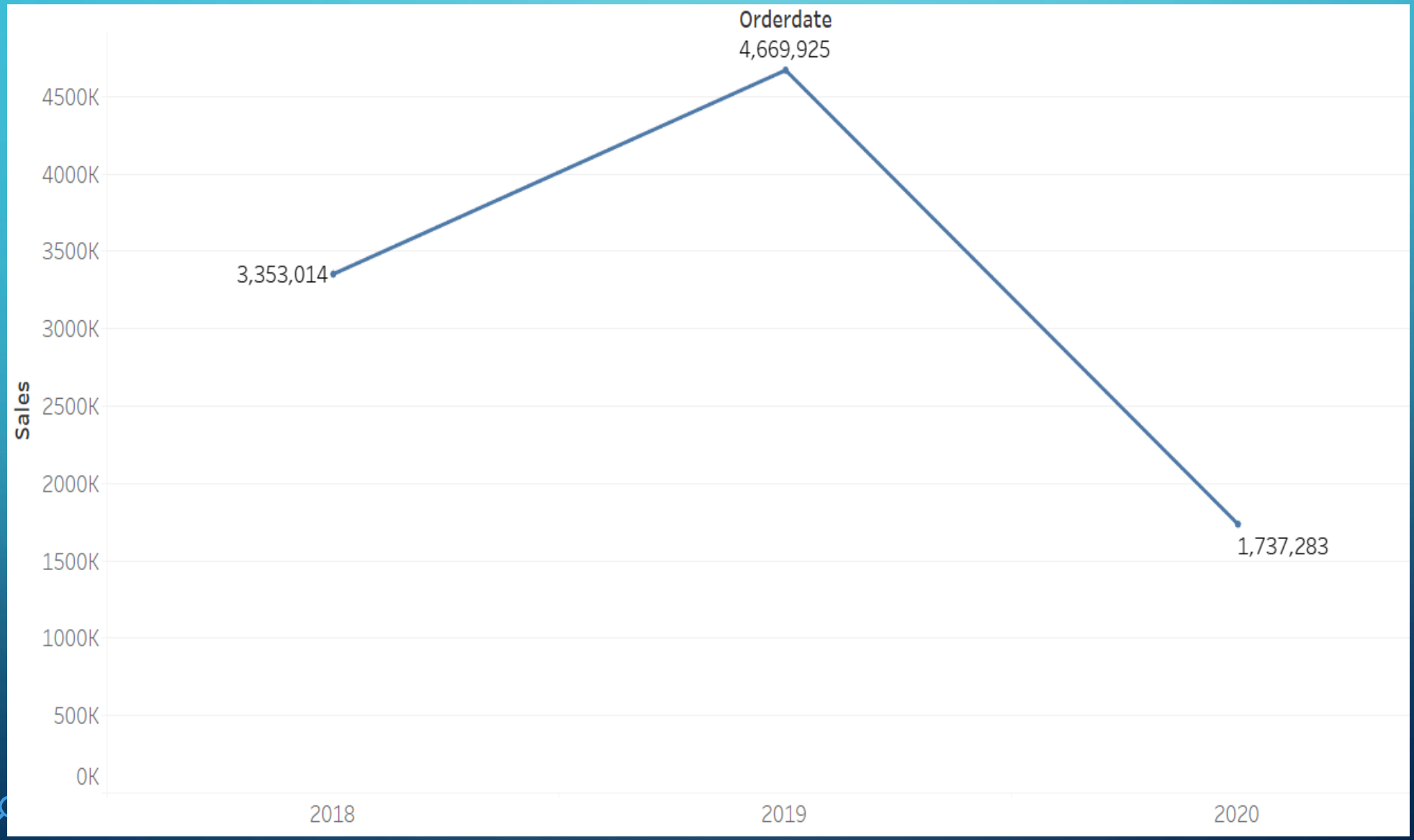
2) CITY: Here we can see that the city ‘Madrid’ has max no. of sales.



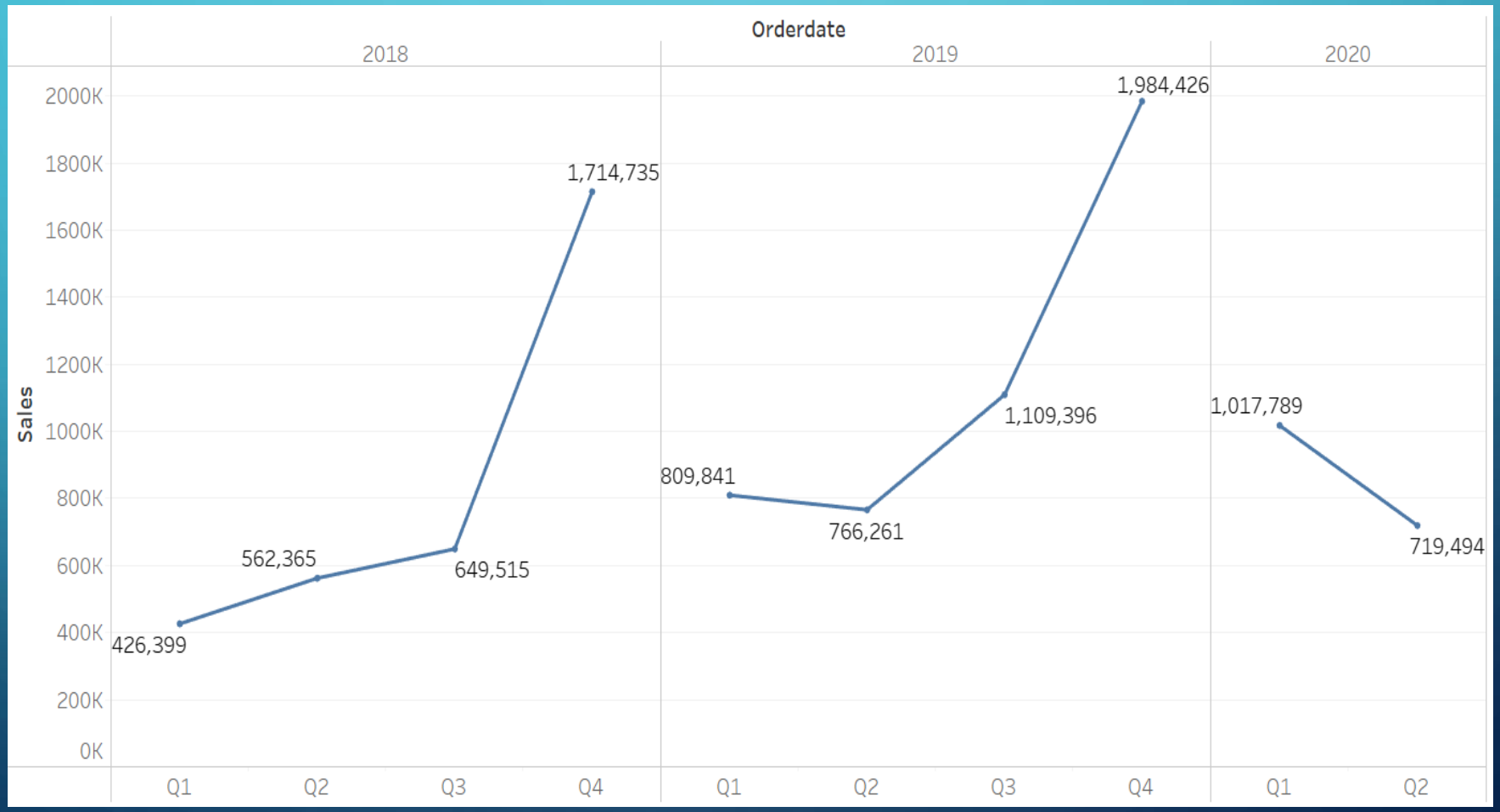
3) PRODUCT LINE- We can clearly see that Classic Cars has the max no. of sales, (above 4000) followed by Vintage Cars. While Trains has the least no. of sales.



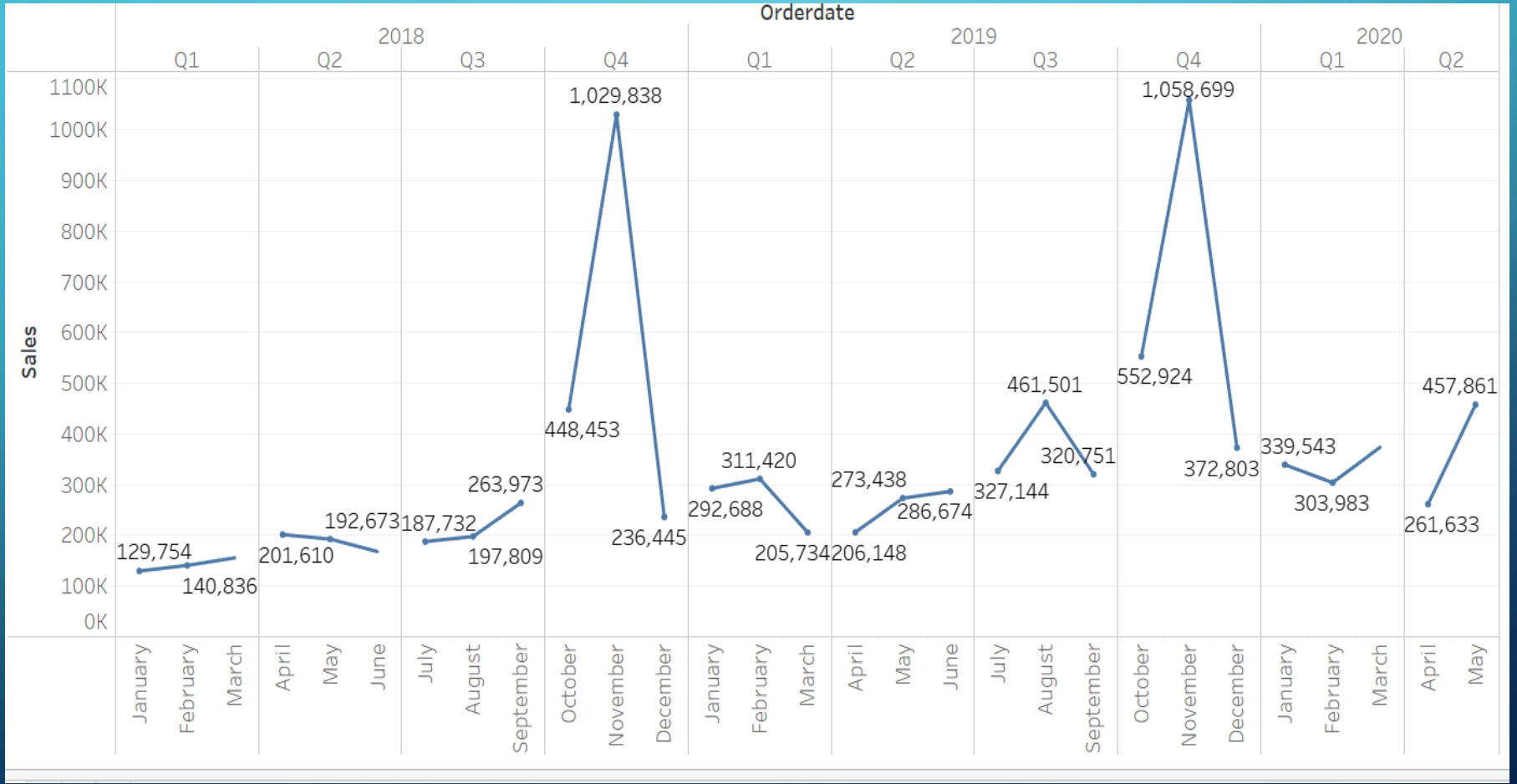
Yearly Trend in Sales



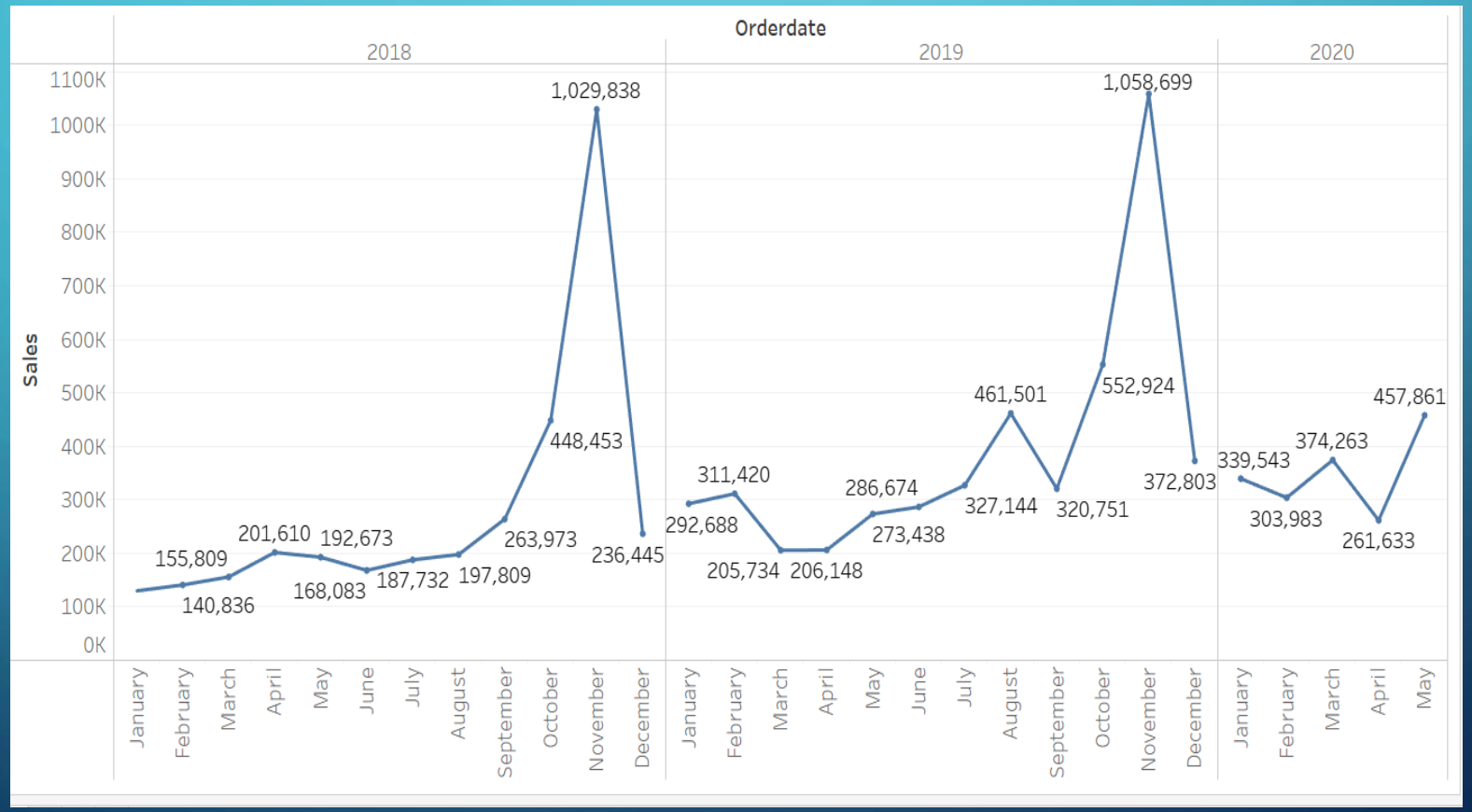
QUARTERLY TREND IN SALES



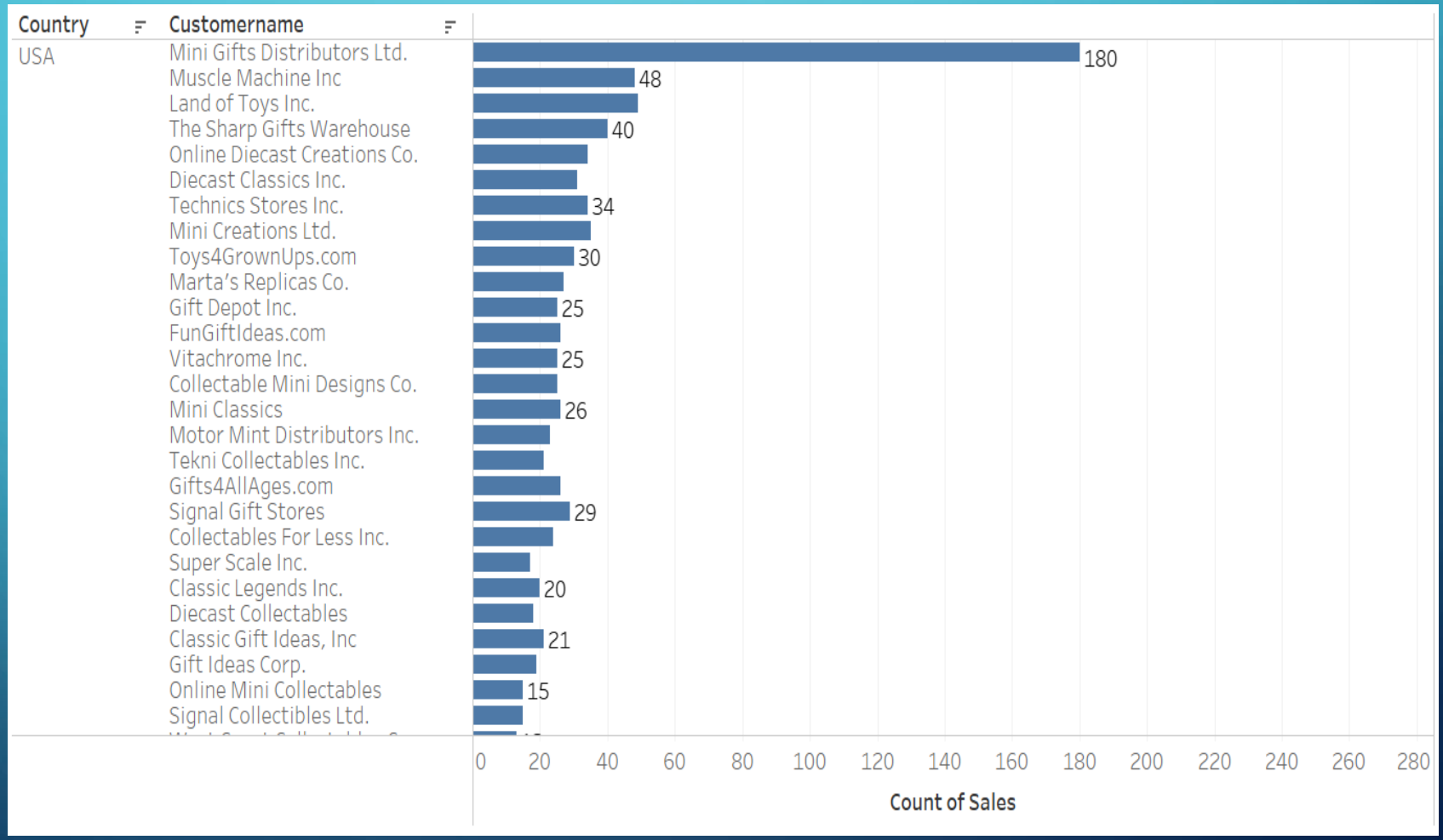
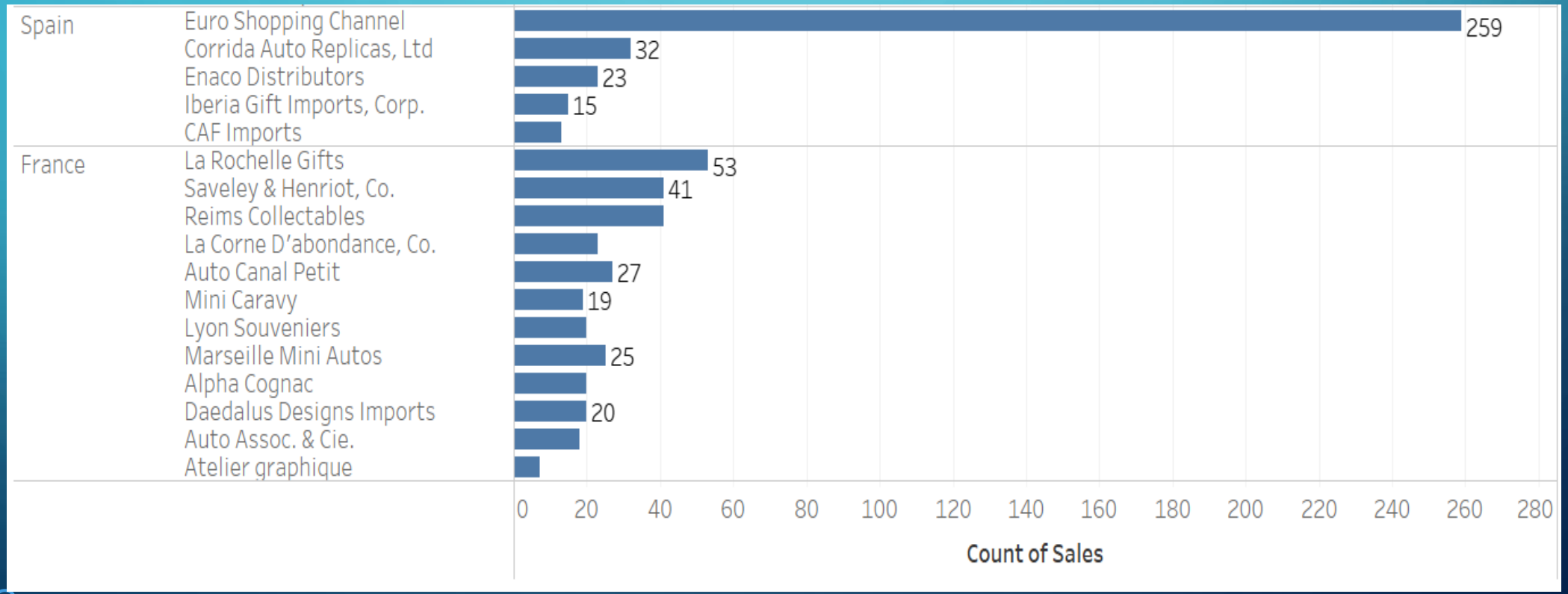
MONTHLY TREND IN SALES



YEARLY AND MONTHLY TRENDS



TOP CUSTOMERS FROM USA



TOP CUSTOMERS FROM SPAIN AND FRANCE