

PROJECT REPORT
ON
MARKETING & RETAIL ANALYTICS
(PART - A)

BY

SHAJIL FERNANDEZ

04-02-2024




Table of Contents

- Problem statement
- Information about data
- EDA and Inferences
- Customer segmentation using RFM analysis
- Inferences from RFM analysis & identified segments

Agenda & Executive Summary of the data

- To find the underlying buying patterns of the customers.
- To provide the company with suitable insights about their customers.
- Recommend customized marketing strategies for different segments of customers.

Problem Statement

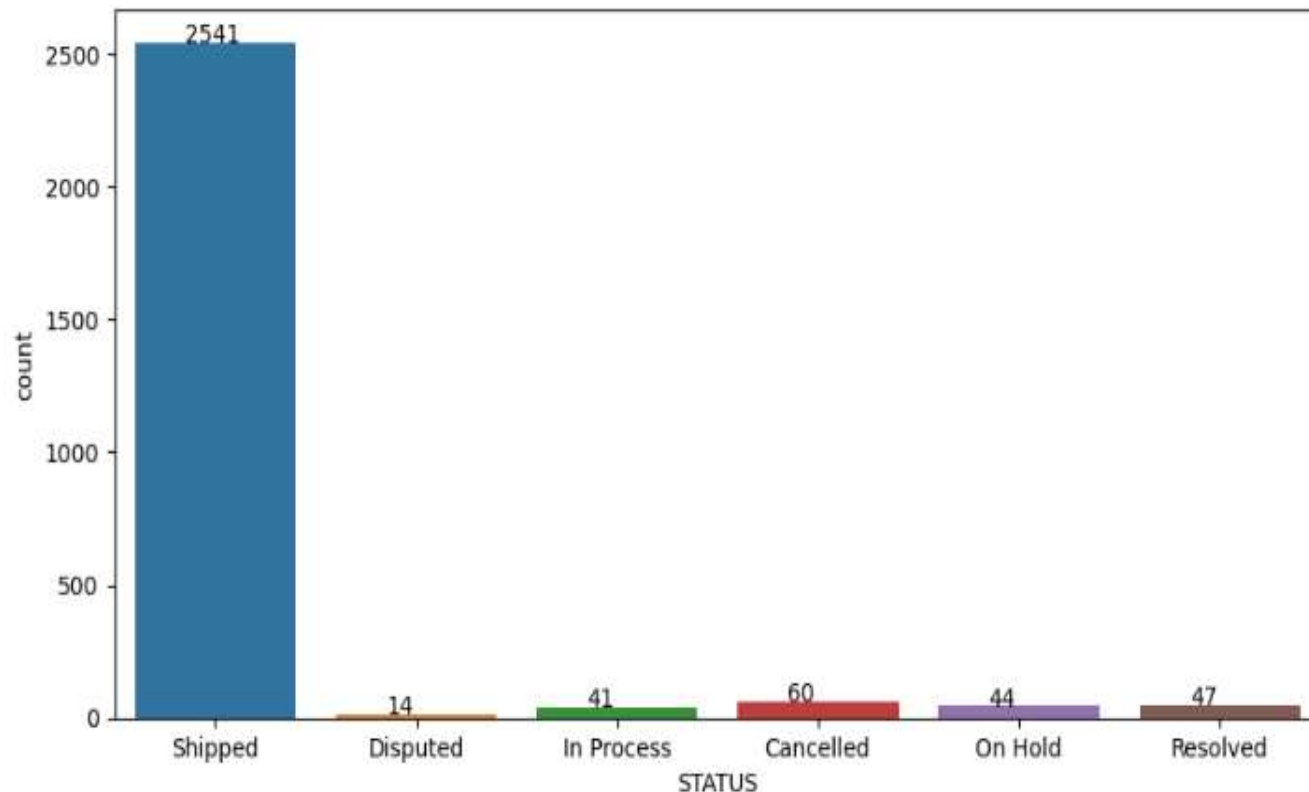
An automobile parts manufacturing company has collected data on transactions for 3 years. They do not have any in-house data science team, thus they have hired you as their consultant. Your job is to use your data science skills to find the underlying buying patterns of the customers, provide the company with suitable insights about their customers, and recommend customized marketing strategies for different segments of customers.

Information about data

- There are 2747 rows and 20 columns in the given dataset.
- There are 2 float64, 6 int64 and 12 object datatypes.
- There are no missing values in the given data.
- There are no duplicate values.
- Minimum quantity ordered is 6 and maximum is 97.
- Mean total sales amount is 3553.05 and Median is 3184.80.
- About 928 customers are from USA.

EDA and Inferences

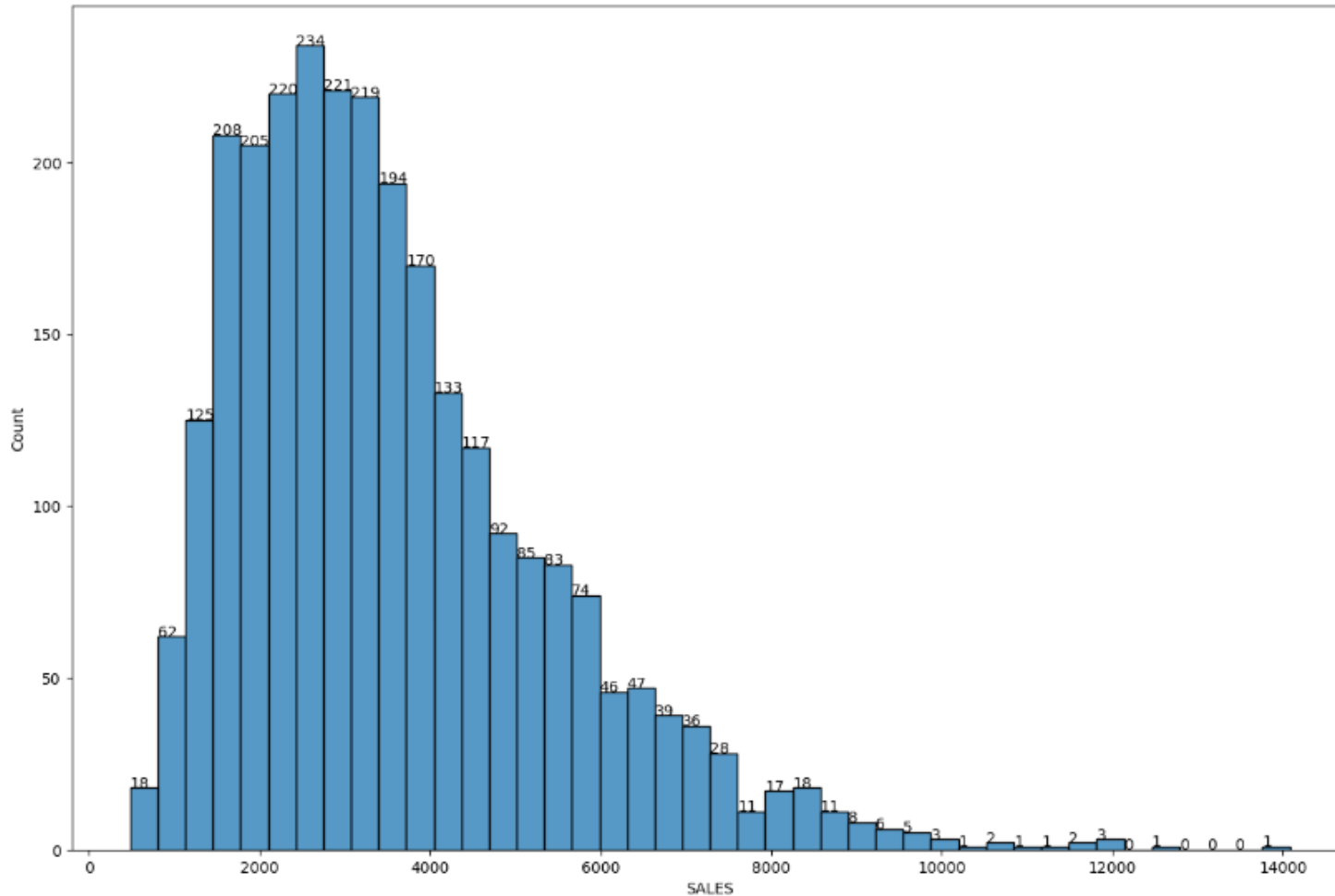
- Univariate bivariate and multivariate analysis:



Insights:

- About 60 orders are cancelled and 44 are on hold.
- 14 orders are in dispute and 47 are resolved.

Sales



Statistical summary of sales

```
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
```

Insights:

- Mean of total sales amount is 3553.08 and median is 3184.80.
- Minimum sale amount is 482.13 and maximum is 14082.80.

Yearly sales

Insights:

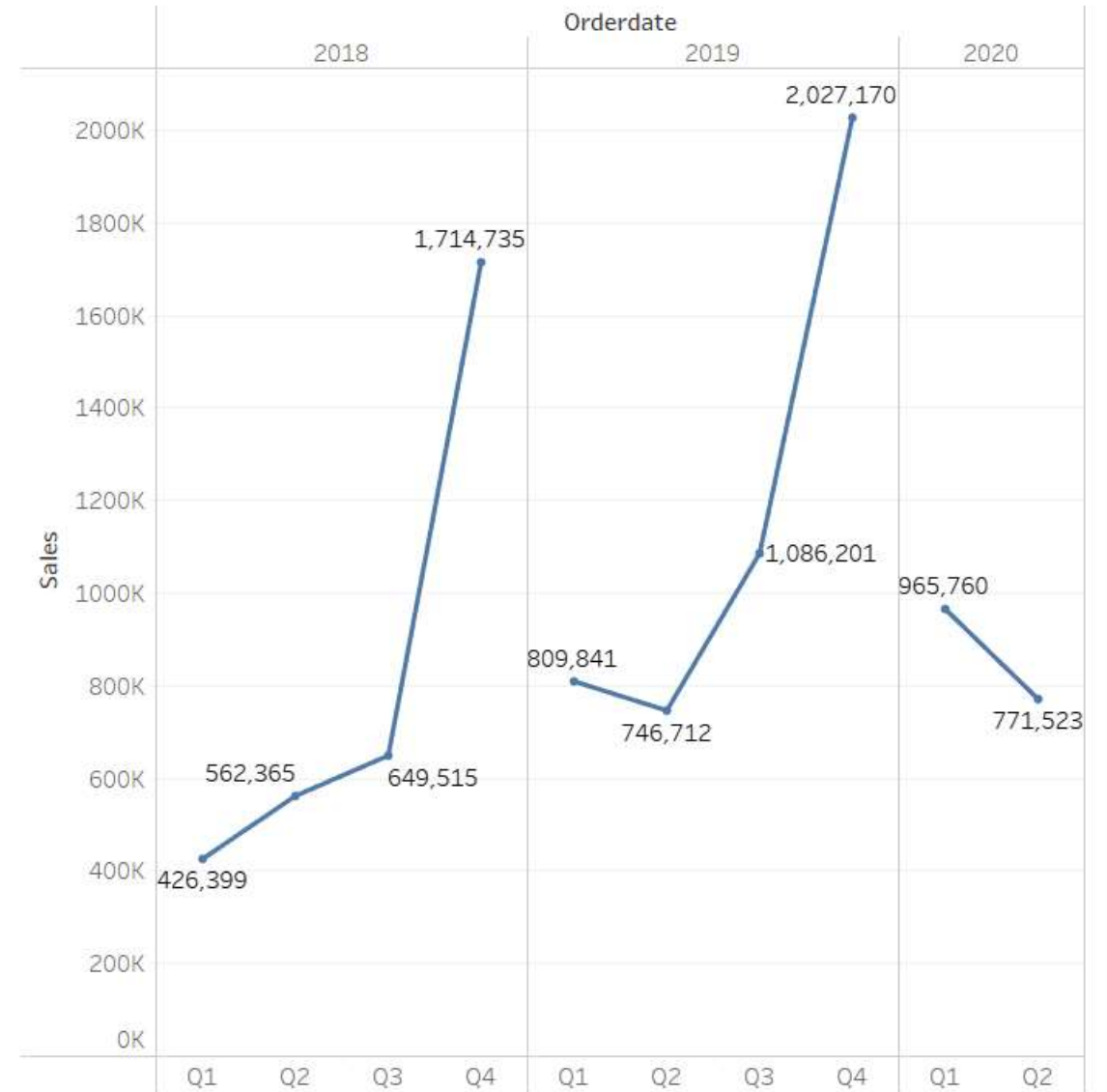
- There is an increase in yearly sales i.e., sales amount which was 3,353,014 in 2018, increased to 4,669,925 in the year 2019. However, we cannot compare with 2020 as the data is available only for 5 months.



Quarterly sales

Insights:

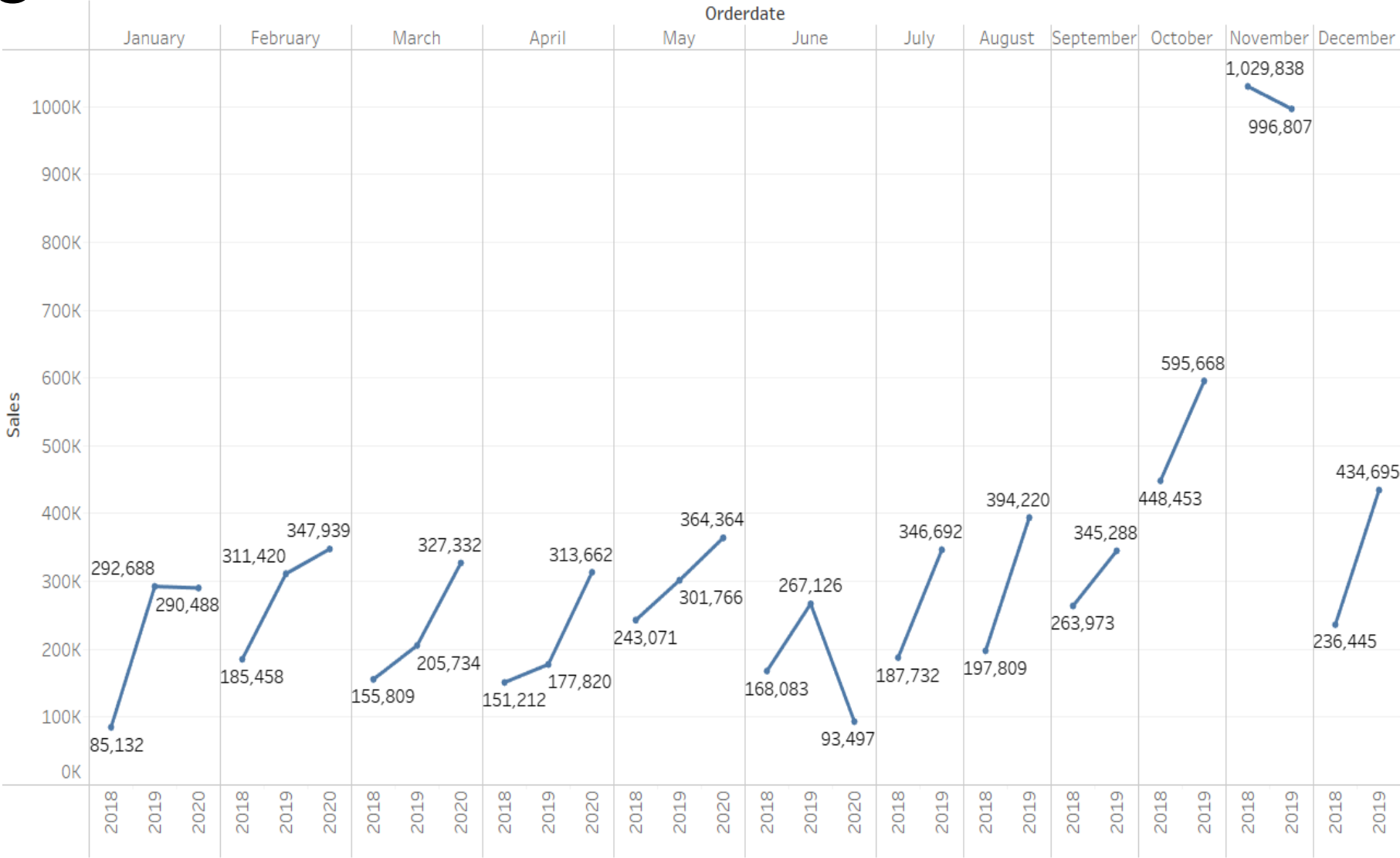
- Q4 sales is the highest, when compared with other quarters.
- In 2018, lowest sales is in Q1, whereas In 2019, Q2 has the lowest sales.



Monthly sales

Insights:

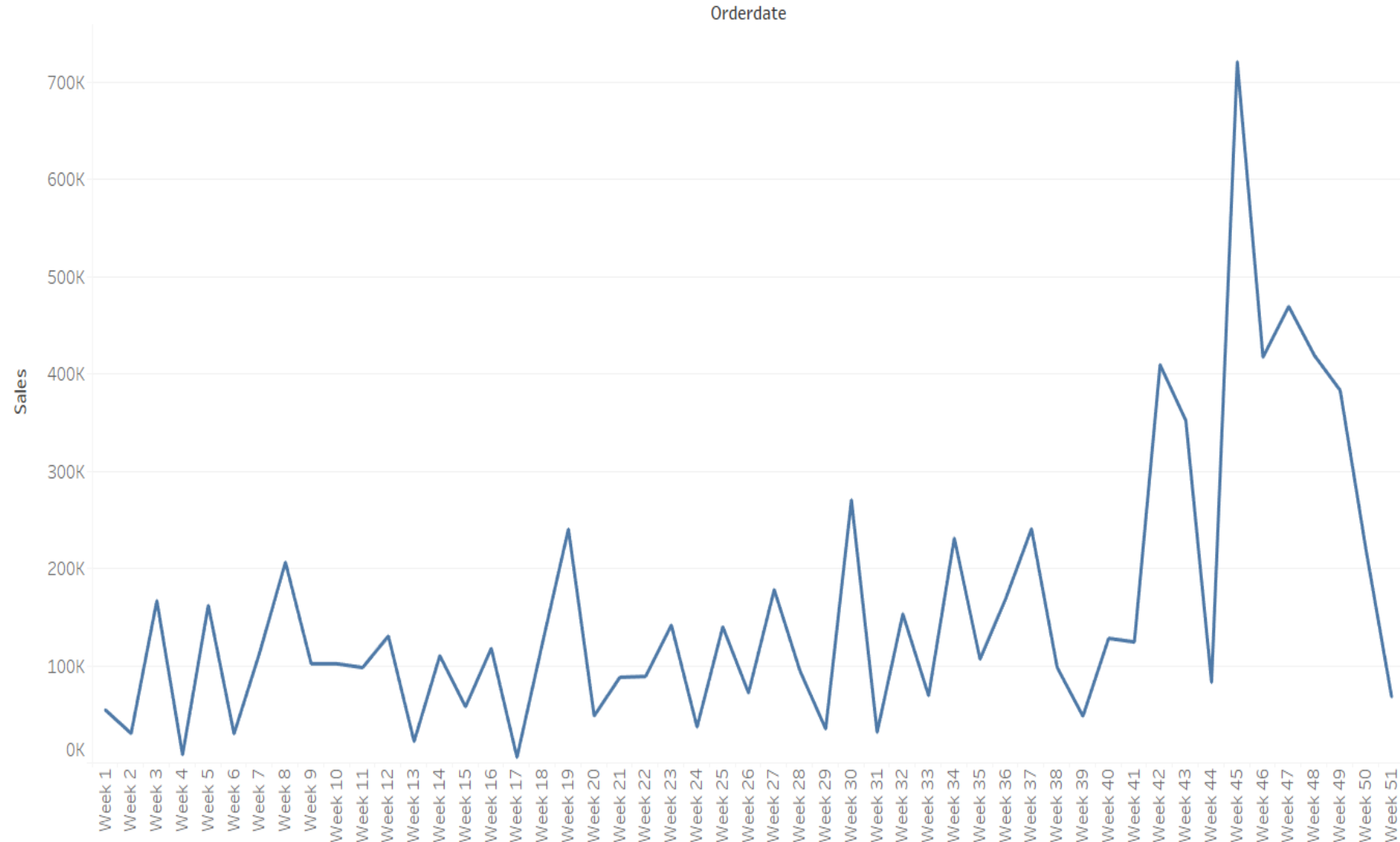
- November witnesses a drastic increase in sales.
- In 2018, January has the lowest sales, and In 2019, lowest monthly sales is in April.



Weekly sales

Insights:

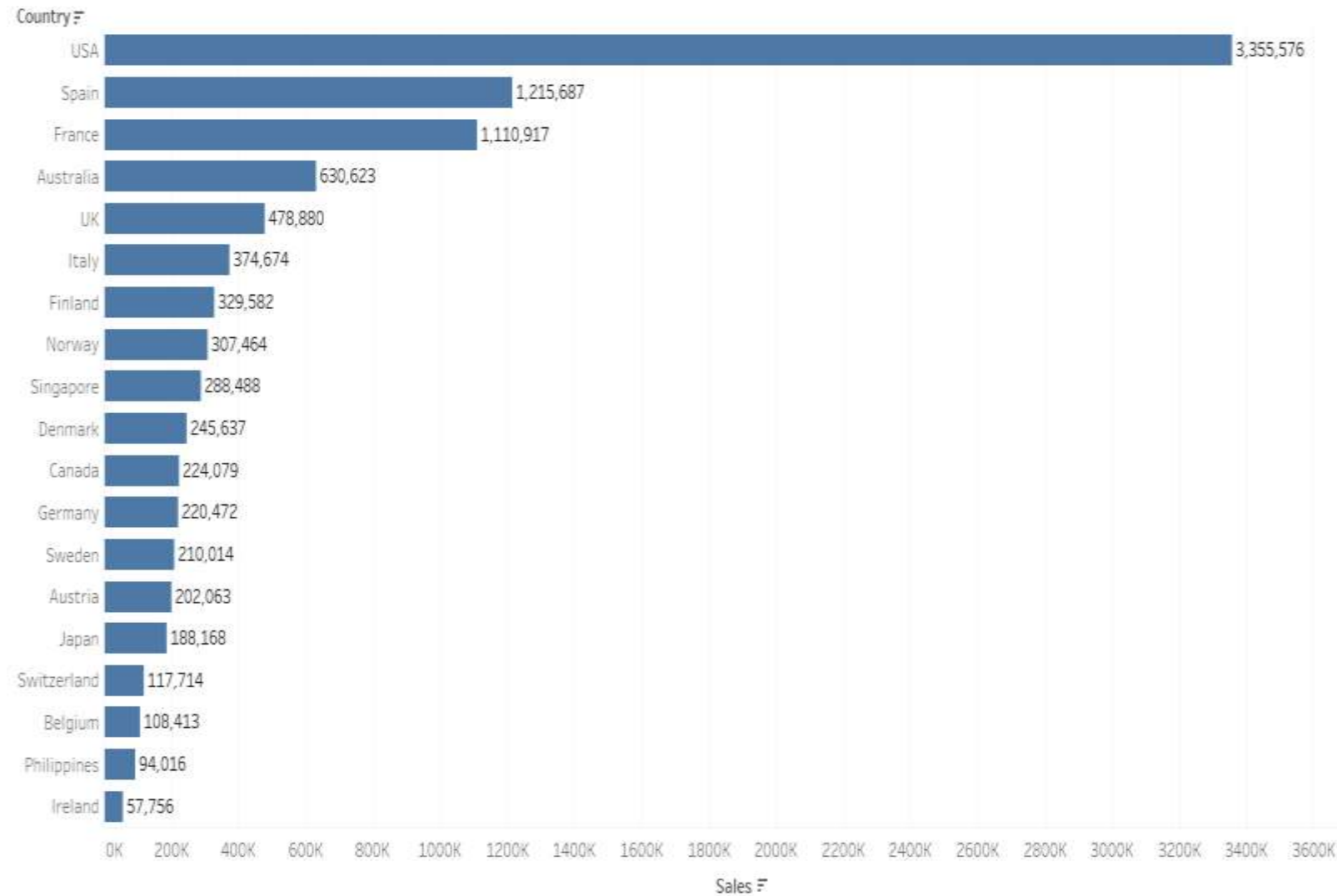
- We have compared weeks across 2018 and 2019 and excluded 2020 for consistency purpose.
- We can note that, week 45 has the highest sales and week 17 witnesses the lowest sales.



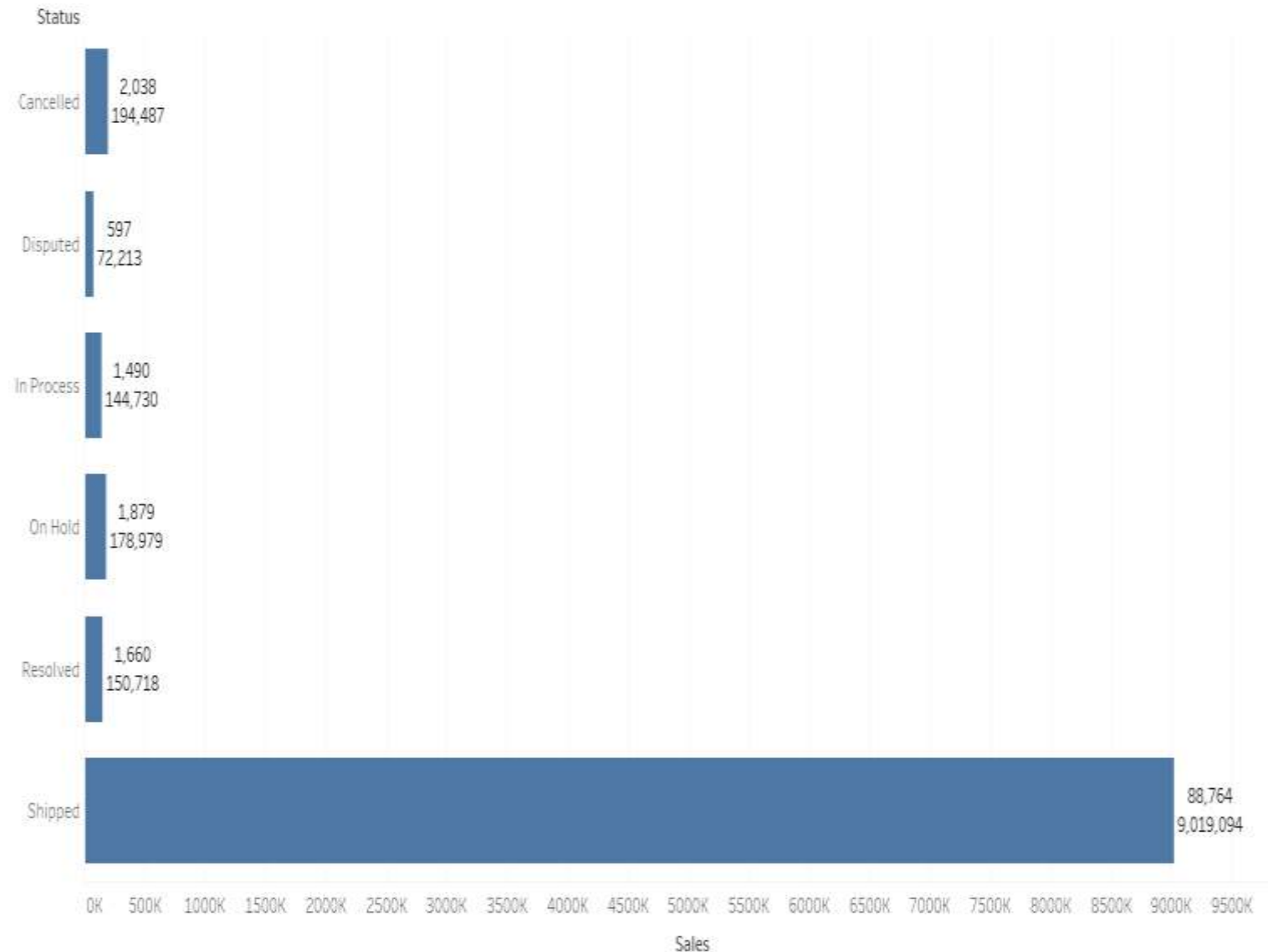
Country wise sales

Insights:

- The highest revenue is coming from the customers located in USA, followed by the Spain and France.
- Lowest revenue earned is from Ireland.



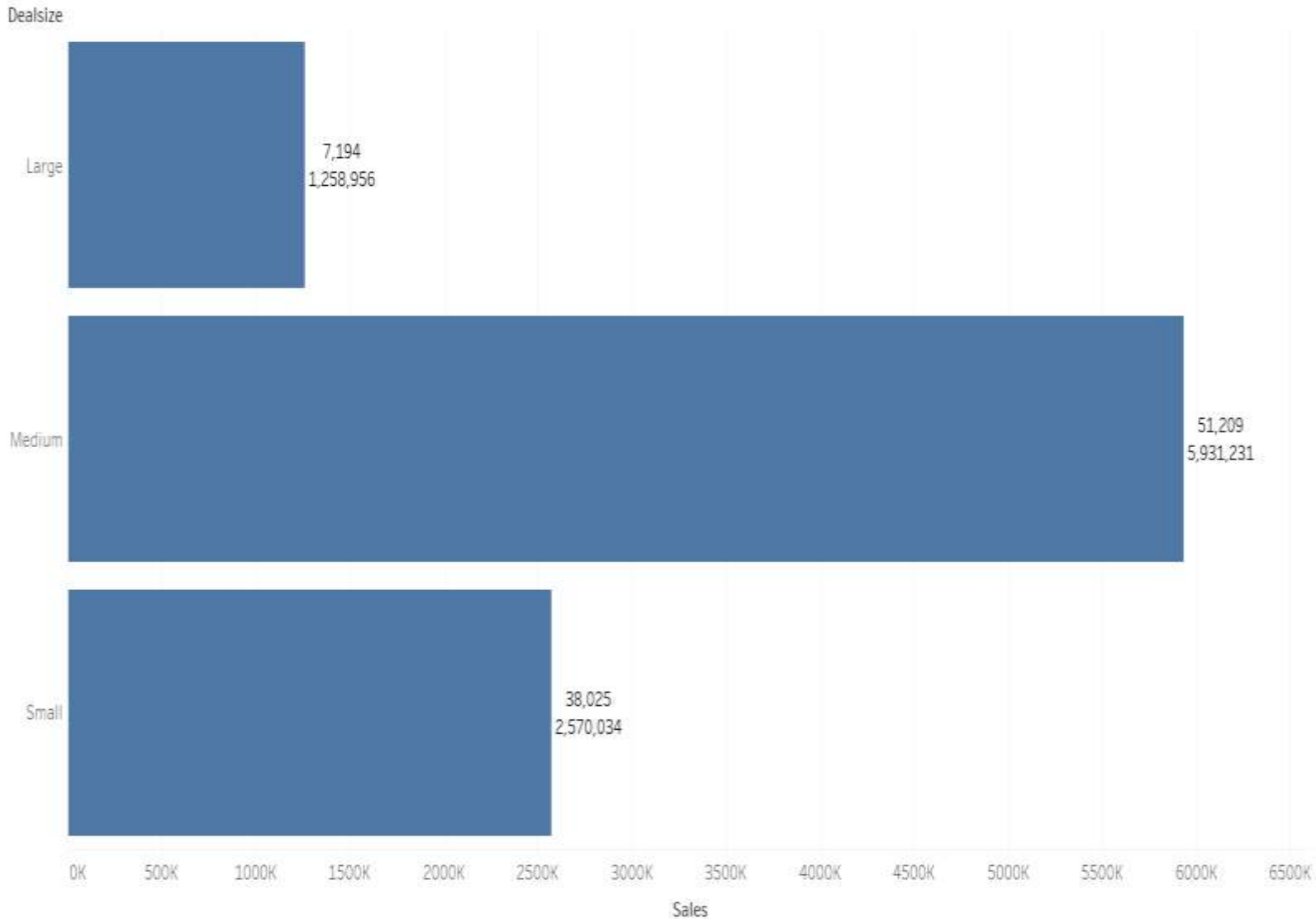
Sales based on status of the order



Insights:

- About 2038 quantities of 194487 sales amount ordered were cancelled.
- 88764 quantities ordered were shipped, generating revenue of 9,019,094.

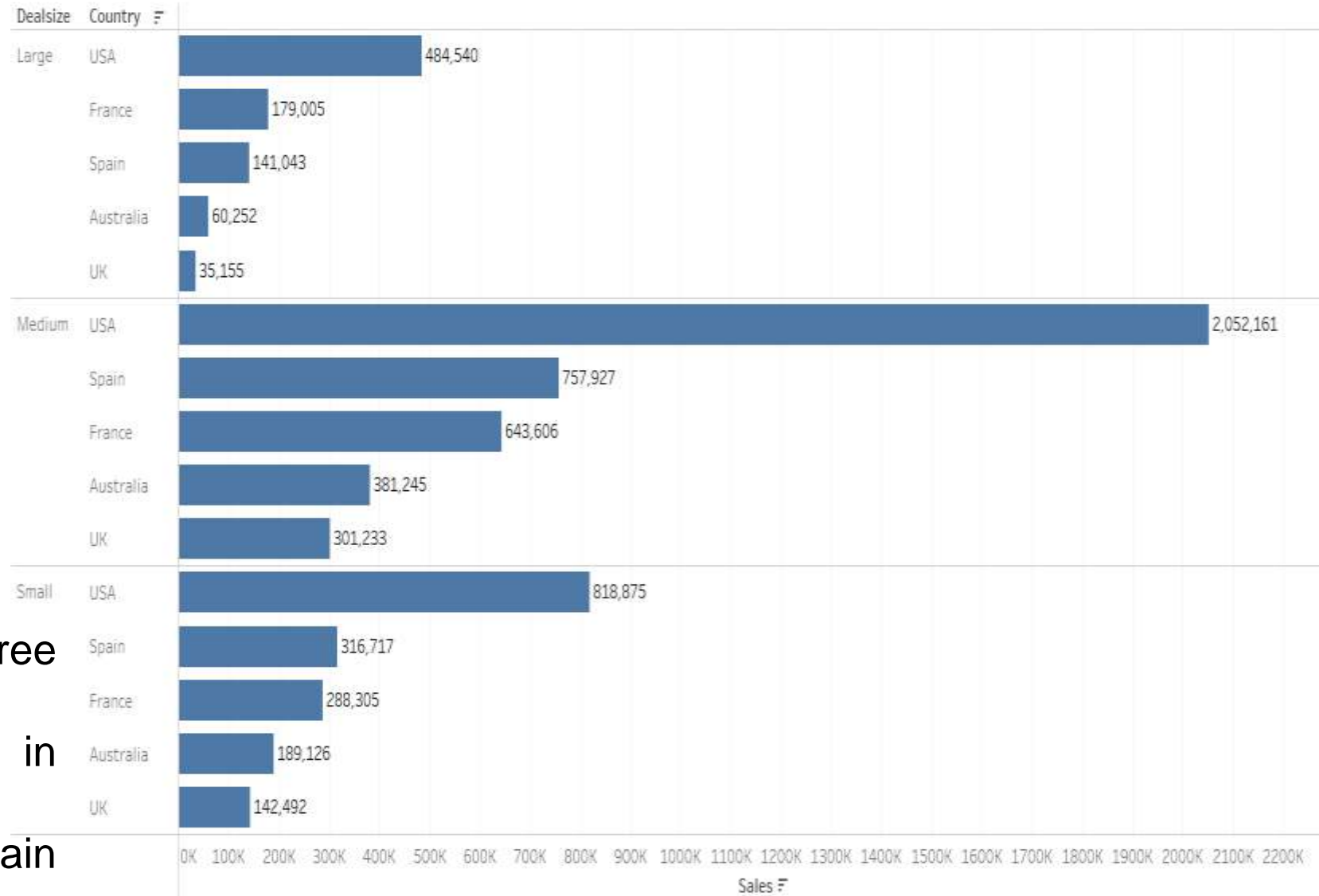
Sales based on deal size



Insights:

- About 5,931,231 revenue earned from the medium orders.
- Lowest revenue earned is from the Large orders.

Sales based on country wise deal size



Insights:

- USA is the highest among all three deal sizes.
- France is the second highest in providing Large orders.
- In Medium and small deals, Spain is the second highest.

Inferences

- Largest buyers are from USA, followed by Spain and France.
- Highest revenue earned is also from USA.
- Q4 witnesses the highest sales, of which November witnesses a drastic increase in sales.
- Maximum sale amount earned per order is 14,082.80.
- Highest revenue generated is from Medium orders.
- Cancelled orders are more than Resolved orders.
- We can observe an upward trend in yearly sales.
- Company should strategize in a way to attract countries with lower sales.

RFM analysis

RFM analysis is a strategy used by the company in order to segregate the customers and group them based on their buying behaviour and spending pattern.

Parameters:

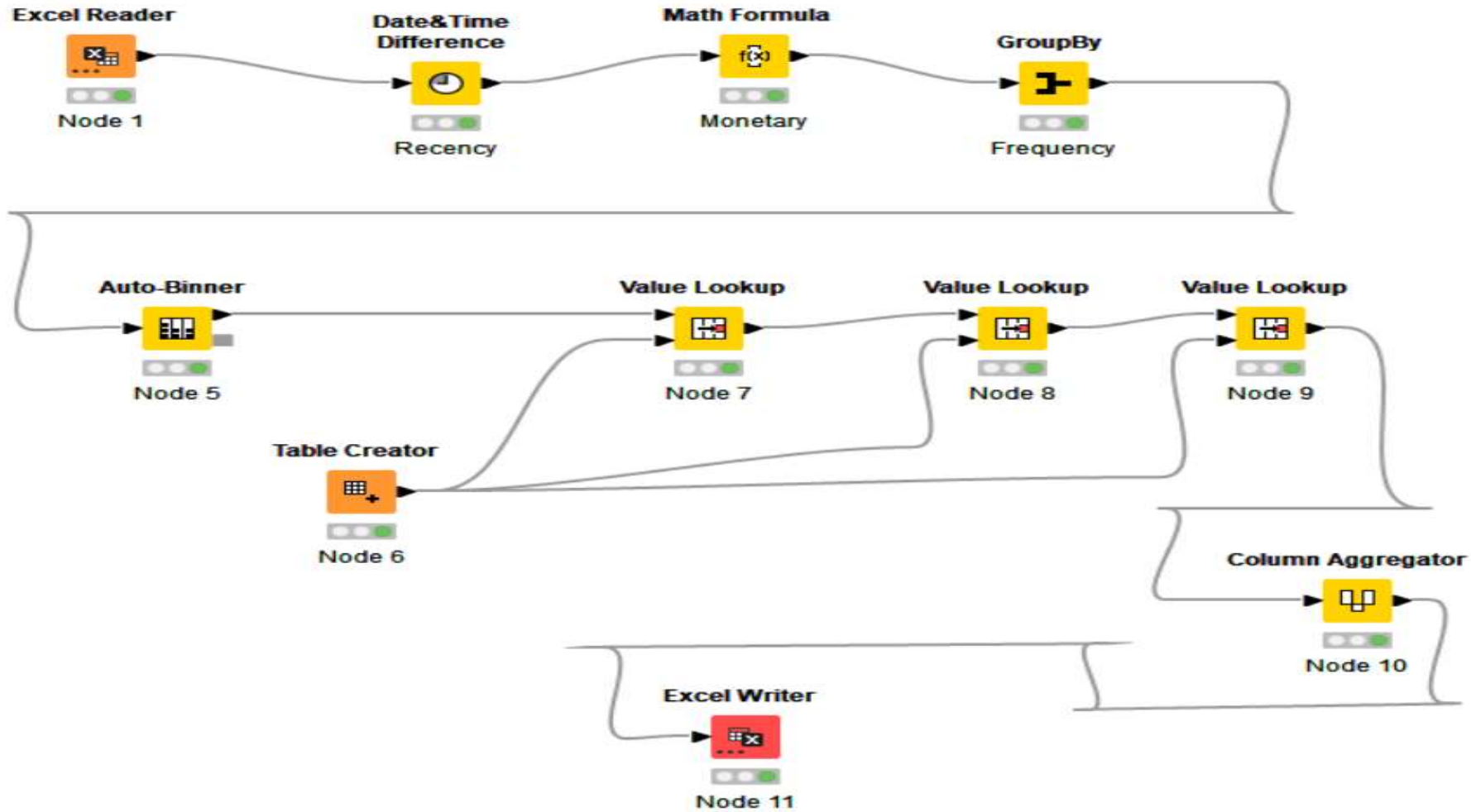
- **Excel reader:** To import and read the dataset which is in excel format.
- **Date & Time difference:** In order to identify the recency, we can subtract order date with Current execution date & time option (today's date). For Granularity we are choosing days option.
- **Math Formula:** We are choosing Sales column as Monetary.
- **GroupBy:** In order to identify the frequency, we are grouping by "ordername" and selecting the appropriate aggregation across different categories under "Manual Aggregation" option.

- **Auto-Binner:** Creating Bins for grouped 'order number', 'recency' and 'monetary' columns, i.e., by creating quantile range 0.0, 0.25, 0.5, 0.75, 1.0 based on four segments. This helps in segregating customers to their respective bins.
- **Table Creator:** Creating RFM scores to their respective bins.
- **Value Lookup:** Assigning RFM scores to their respective bins based on the keys.
- **Column Aggregator:** To concatenate the individual RFM scores.
- **Excel Writer:** To export output table head which consists of data with RFM scores.

Assumptions:

- For recency, order date is subtracted with 03-02-2024.
- The data is available only till May 2020, however, for recency, incompleteness of the data have been ignored.
- 'Cancelled', 'On hold' orders are also considered for the analysis.

KNIME workflow



Output Table head

Row ID	S ADDRESSLINE1	S CITY	S POSTA...	S COUNTRY	S CONTA...	S CONTA...	I DEAL...	L Rec...	D Monet...	S ORDER...	S Rec...	S Mon...	S Rec...	S Fre...	S Mon...	S ▼ Con
Row228	636 St Kilda Road	Melbourne	3004	Australia	Ferguson	Peter	12	1527	47,442.91	Bin 4	Bin 1	Bin 4	4	4	4	444
Row231	C/ Moralarzal, 86	Madrid	28034	Spain	Freyre	Diego	17	1524	52,269.17	Bin 4	Bin 1	Bin 4	4	4	4	444
Row238	C/ Moralarzal, 86	Madrid	28034	Spain	Freyre	Diego	14	1516	51,435.45	Bin 4	Bin 1	Bin 4	4	4	4	444
Row242	Software Engineering C...	Espoo	FIN-02271	Finland	Suominen	Kalle	15	1489	51,373.49	Bin 4	Bin 1	Bin 4	4	4	4	444
Row252	Torikatu 38	Oulu	90110	Finland	Koskitalo	Pirkko	17	1464	49,055.4	Bin 4	Bin 1	Bin 4	4	4	4	444
Row254	67, rue des Cinquante O...	Nantes	44000	France	Labrune	Janine	15	1461	59,617.4	Bin 4	Bin 1	Bin 4	4	4	4	444
Row260	5677 Strong St.	San Rafael	97562	USA	Nelson	Valarie	13	1447	55,741.02	Bin 4	Bin 1	Bin 4	4	4	4	444
Row263	C/ Moralarzal, 86	Madrid	28034	Spain	Freyre	Diego	18	1434	53,815.72	Bin 4	Bin 1	Bin 4	4	4	4	444
Row267	5677 Strong St.	San Rafael	97562	USA	Nelson	Valarie	16	1431	61,133.28	Bin 4	Bin 1	Bin 4	4	4	4	444
Row274	59 rue de l'Abbaye	Reims	51100	France	Henriot	Paul	18	1405	52,029.07	Bin 4	Bin 1	Bin 4	4	4	4	444
Row281	3086 Ingle Ln.	San Jose	94217	USA	Frick	Sue	12	1382	57,883.09	Bin 4	Bin 1	Bin 4	4	4	4	444
Row285	C/ Moralarzal, 86	Madrid	28034	Spain	Freyre	Diego	11	1371	52,287.74	Bin 4	Bin 1	Bin 4	4	4	4	444
Row287	8616 Spinnaker Dr.	Boston	51003	USA	Yoshido	Juri	14	1368	48,316.89	Bin 4	Bin 1	Bin 4	4	4	4	444
Row291	Geislweg 14	Salzburg	5020	Austria	Pipps	Georg	14	1357	59,475.1	Bin 4	Bin 1	Bin 4	4	4	4	444
Row223	636 St Kilda Road	Melbourne	3004	Australia	Ferguson	Peter	11	1532	43,779.09	Bin 4	Bin 1	Bin 3	4	4	3	443
Row230	4092 Furth Cirde	NYC	10022	USA	Young	Jeff	10	1525	41,706.52	Bin 4	Bin 1	Bin 3	4	4	3	443
Row237	5677 Strong St.	San Rafael	97562	USA	Nelson	Valarie	10	1516	42,152.11	Bin 4	Bin 1	Bin 3	4	4	3	443
Row240	Monitor Money Building, ...	Chatswood	2067	Australia	Huxley	Adrian	14	1509	37,905.15	Bin 4	Bin 1	Bin 3	4	4	3	443
Row246	78934 Hillside Dr.	Pasadena	90003	USA	Young	Julie	13	1483	44,273.36	Bin 4	Bin 1	Bin 3	4	4	3	443
Row250	5677 Strong St.	San Rafael	97562	USA	Nelson	Valarie	12	1472	36,124.27	Bin 4	Bin 1	Bin 3	4	4	3	443
Row251	2-2-8 Roppongi	Minato-ku	106-0032	Japan	Shimamura	Akiko	9	1469	38,191.39	Bin 4	Bin 1	Bin 3	4	4	3	443
Row259	C/ Moralarzal, 86	Madrid	28034	Spain	Freyre	Diego	13	1448	38,360.4	Bin 4	Bin 1	Bin 3	4	4	3	443
Row261	C/ Moralarzal, 86	Madrid	28034	Spain	Freyre	Diego	13	1442	41,249.74	Bin 4	Bin 1	Bin 3	4	4	3	443
Row265	1785 First Street	New Bedford	50553	USA	Benitez	Violeta	8	1432	37,557.66	Bin 4	Bin 1	Bin 3	4	4	3	443
Row268	201 Miller Street	North Sydney	2060	Australia	O'Hara	Anna	10	1426	35,160.25	Bin 4	Bin 1	Bin 3	4	4	3	443
Row275	3086 Ingle Ln.	San Jose	94217	USA	Frick	Sue	9	1403	37,526.84	Bin 4	Bin 1	Bin 3	4	4	3	443
Row276	7476 Moss Rd.	Newark	94019	USA	Brown	William	12	1401	46,519	Bin 4	Bin 1	Bin 3	4	4	3	443
Row278	Berkeley Gardens 12 Br...	Liverpool	WX1 6LT	UK	Devon	Elizabeth	9	1396	40,802.81	Bin 4	Bin 1	Bin 3	4	4	3	443
Row279	24, place Kluber	Strasbourg	67000	France	Citeaux	Frederique	5	1390	35,680.35	Bin 4	Bin 1	Bin 3	4	4	3	443
Row289	Strada Provinciale 124	Reggio Emilia	42100	Italy	Moroni	Maurizio	14	1364	41,509.94	Bin 4	Bin 1	Bin 3	4	4	3	443
Row292	Monitor Money Building, ...	Chatswood	2067	Australia	Huxley	Adrian	13	1345	43,971.43	Bin 4	Bin 1	Bin 3	4	4	3	443
Row296	C/ Moralarzal, 86	Madrid	28034	Spain	Freyre	Diego	6	1343	35,133.34	Bin 4	Bin 1	Bin 3	4	4	3	443
Row297	67, rue des Cinquante O...	Nantes	44000	France	Labrune	Janine	13	1343	43,784.69	Bin 4	Bin 1	Bin 3	4	4	3	443
Row224	59 rue de l'Abbaye	Reims	51100	France	Henriot	Paul	6	1532	18,681.17	Bin 4	Bin 1	Bin 2	4	4	2	442
Row225	12, rue des Bouchers	Marseille	13008	France	Lebihan	Laurence	7	1531	20,136.86	Bin 4	Bin 1	Bin 2	4	4	2	442
Row234	2440 Pompton St.	Glendale	97561	USA	Lewis	Dan	9	1522	29,343.35	Bin 4	Bin 1	Bin 2	4	4	2	442
Row235	C/ Moralarzal, 86	Madrid	28034	Spain	Freyre	Diego	10	1519	29,884.13	Bin 4	Bin 1	Bin 2	4	4	2	442

Inferences from RFM Analysis and identified segments

- RFM scores from RFM analysis would ease the process of segmenting customers.
- **Best customers** are chose based on the RFM score of 444, i.e., these customers are frequent, visit regularly and spend effectively.
- **Verge of Churning:** segmented based on RFM score of 222. Company need to strategize in order to retain these kind of customers before they fall in the category of lost customers.
- **Lost customers** are chose based on the RFM score of 111, these customer groups are inactive i.e., they no longer make a purchase or make purchase from a different company.
- **Loyal customers** are chose based on the recency score of 4 which is the highest. This group may spend less or more but visits or make a purchase regularly.
- RFM analysis helps the company in order to make decisions on retaining customers, providing offers, discounts etc.

Best Customers

CUSTOMERNAME	Concatenate
Euro Shopping Channel	444
Mini Gifts Distributors Ltd.	444
Gifts4AllAges.com	444
Oulu Toy Supplies, Inc.	444
Suominen Souvenirs	444

Verge of Churning

CUSTOMERNAME	Concatenate
Auto Canal Petit	222
Cambridge Collectables Co.	222
Vitachrome Inc.	222
Toms Spezialitten, Ltd	222
Tokyo Collectables, Ltd	222

Lost Customers

Alpha Cognac	111
Atelier graphique	111
Australian Collectors, Co.	111
Online Diecast Creations Co.	111
Reims Collectables	111

Loyal Customers

CUSTOMERNAME	Recency (#1)
Anna's Decorations, Ltd	4
La Rochelle Gifts	4
Marseille Mini Autos	4
Mini Creations Ltd.	4
Suominen Souvenirs	4

THANK YOU