

Customer Segmentation

By Group 2

OUR TEAM MEMBER



M. Kamal
Jaza



M. Irvan
Arfandi



Risma Ashali



Rizal Maulana K.



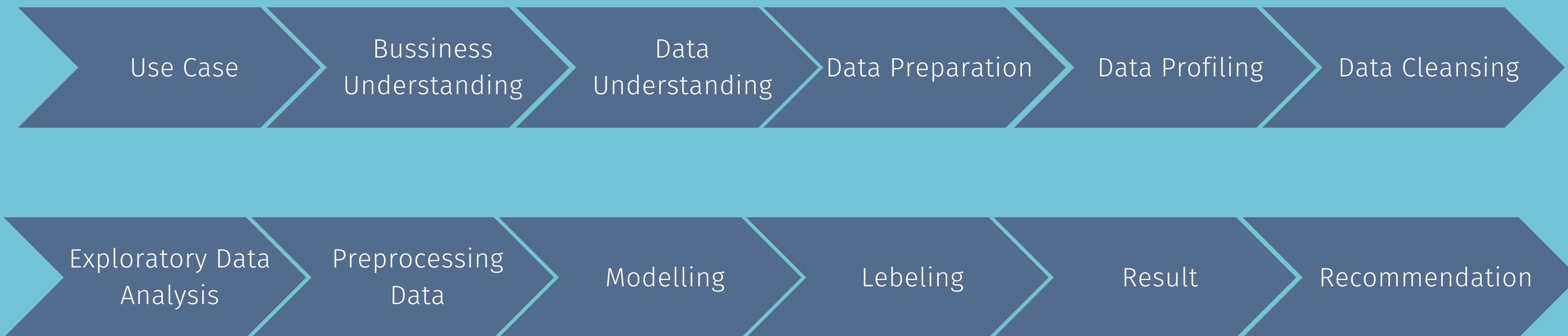
Shafira Aisyah

TABLE OF CONTENT



■ USE CASE SUMMARY	■ DATA PROFILING	■ RFM MODELLING
■ BUSINESS UNDERSTANDING	■ DATA CLEANSING	■ LABELING
■ DATA UNDERSTANDING	■ EXPLORATORY DATA ANALYSIS	■ RESULT
■ DATA PREPARATION	■ PREPROCESSING MODELLING	■ RECOMMENDATION

WORKFLOW CUSTOMER SEGMENTATION ANALYSIS



USE CASE SUMMARY

USE CASE



Objective Statement:

- To gain insight into sales fluctuations per year.
- To gain insight into the revenue earned per year and month.
- To gain insight about how often a customer orders.
- To find out how long it's been since a customer ordering.
- To find out how much money customers spend on ordering goods.
- To get insight about customer segmentation analysis.

Challenges:

- Large size of data, cannot maintain by excel spreadsheet.
- There are inappropriate data types.

Methodology/Analytics Technique:

- Descriptive Analysis.
- Graph Analysis.
- Segment Analysis using Recency Frequency Monetary (RFM) Modelling

Business Benefit

- Knowing sales performance every year so, it can be a reference for future decision making.
- Helping Business Development Team to create product differentiation based on the characteristic for each customer.
- Knowing how to treat customer with specific criteria.



Expected Outcome

- Find out how sales fluctuate per year.
- Find out how much revenue per year and per month.
- Find out how long it's been since the customer ordered.
- Find out how often customers orders.
- Find out how much the customer spends when orders.
- Know customer segmentation analysis.

BUSINESS UNDERSTANDING

BUSINESS UNDERSTANDING

- Retail is the sale of goods from a business to a consumer for their own use. A retail transaction handles small quantities of goods whereas wholesale deals with the purchasing of goods on a large scale.
- This case has some business question using the data:
 - How sales fluctuate per year?
 - How much revenue per year and per month?
 - How long it's been since the customer ordered?
 - How often customers orders?
 - How much the customer spends when orders?
 - How about customer segmentation analysis?



DATA UNDERSTANDING & DATA PREPARATION

DATA UNDERSTANDING & DATA PREPARATION

DATA UNDERSTANDING

- Data of Customer order with 4 columns and 5009 rows
- This data from 04/01/2011 to 31/12/2014
- Source Code :
[https://www.kaggle.com/datasets/siddinho/
sample-orders-dataset-retail](https://www.kaggle.com/datasets/siddinho/sample-orders-dataset-retail)
- Data dictionary :
 - `order_date` : Customer order date
 - `order_id` : Customer order unique Id
 - `customer` : Customer name
 - `grand_total` : Total cost from
customers order

DATA PREPARATION

- Code Used:
 - Python Version: 3.7.15
 - Packages: Pandas, Numpy,
Matplotlib, Seaborn,
Plotly and Warnings
Datetime,



DATA PROFILING

Import, Load, Preview and Info Dataset

Load and preview dataset

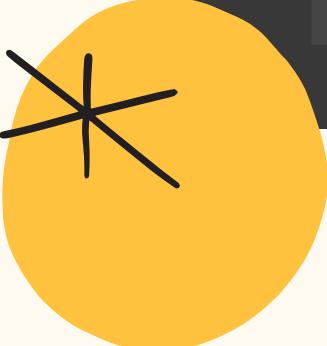
```
[4] df = pd.read_csv('data_orders.csv')
df.head()
```

	order_date	order_id	customer	grand_total
0	9/7/2011	CA-2011-100006	Dennis Kane	378
1	7/8/2011	CA-2011-100090	Ed Braxton	699
2	3/14/2011	CA-2011-100293	Neil Franzsisch	91
3	1/29/2011	CA-2011-100328	Jasper Cacioppo	4
4	4/8/2011	CA-2011-100363	Jim Mitchum	21

Info dataset

```
[5] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5009 entries, 0 to 5008
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   order_date  5009 non-null   object 
 1   order_id    5009 non-null   object 
 2   customer    5009 non-null   object 
 3   grand_total 5009 non-null   int64  
dtypes: int64(1), object(3)
memory usage: 156.7+ KB
```

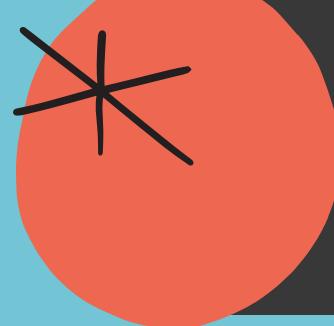


Check and Describe Dataset

Check missing value

```
[6] df.isna().sum()
```

order_date	0
order_id	0
customer	0
grand_total	0
dtype:	int64



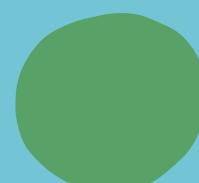
- There is no missing value

Describe dataset

```
[7] df.describe()
```

grand_total	
count	5009.000000
mean	458.626672
std	954.729307
min	1.000000
25%	38.000000
50%	152.000000
75%	512.000000
max	23661.000000

Based on the table above, the average revenue in 2011-2014 was 458.6 dollars with the smallest revenue of 1 dollar and the most at 23661 dollars. Probably, there are customers who get a discount so that there is an income of 1 dollar.





DATA CLEANSING

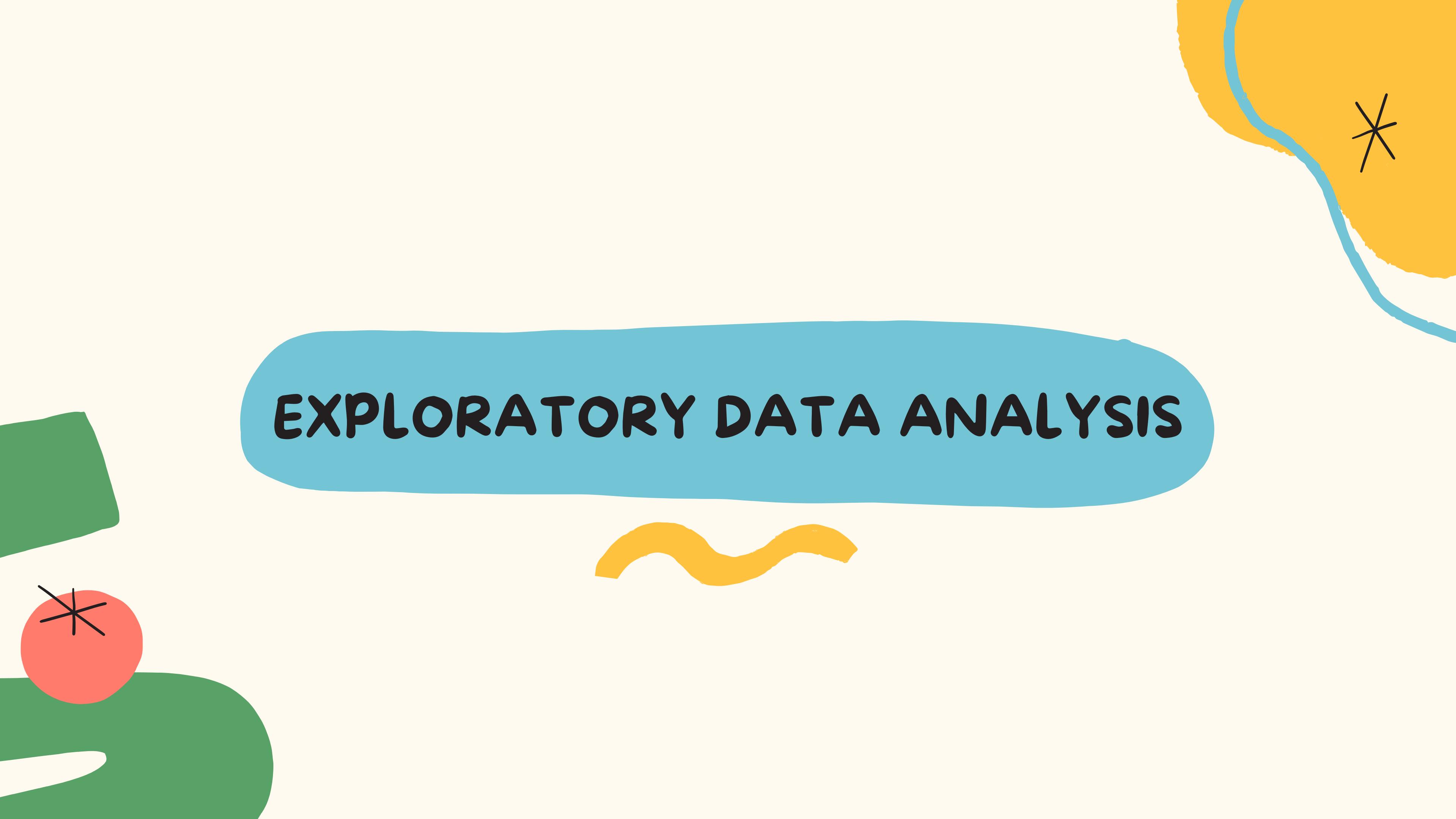
DATA CLEANSING

```
# Order date
```

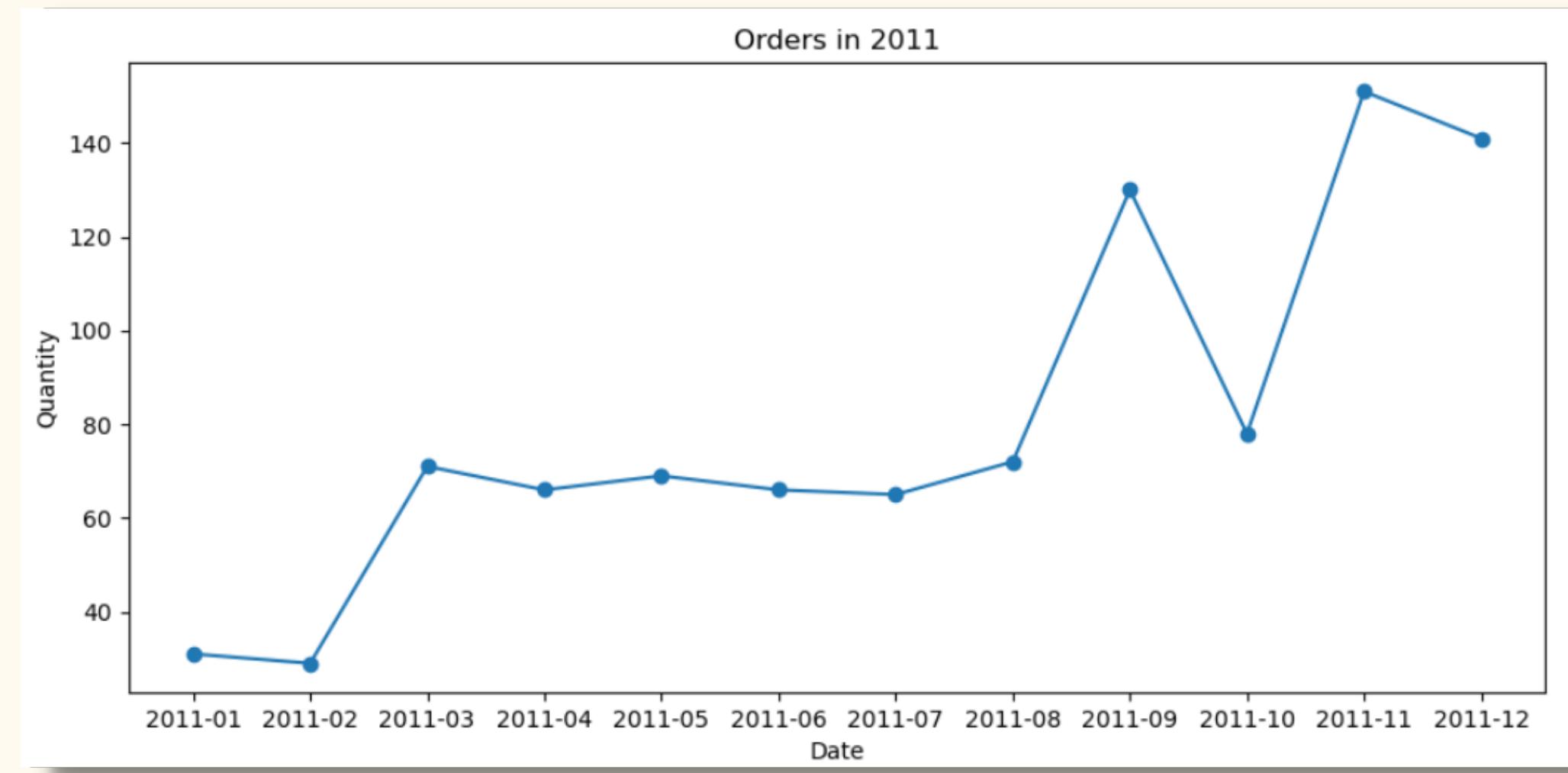
```
df['order_date'] = pd.to_datetime(df['order_date'])
df['OrderYearMonth'] = df['order_date'].map(lambda date: 100*date.year + date.month)
df['Date'] = df['order_date'].dt.strftime('%Y-%m')
df['Year'] = df['order_date'].dt.strftime('%Y')
df['Month'] = df['order_date'].dt.strftime('%m')
```

- We converting date to Month and Year, and changes the format

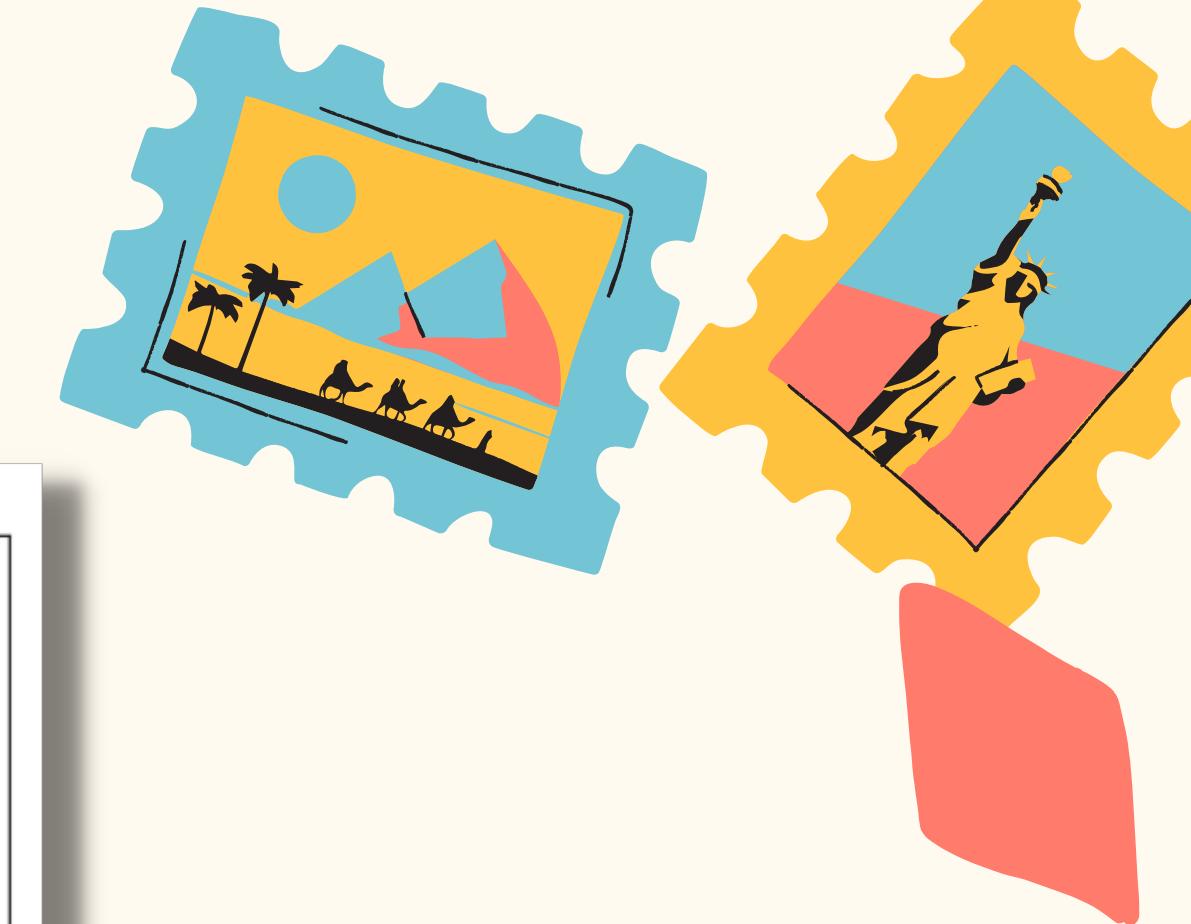
EXPLORATORY DATA ANALYSIS



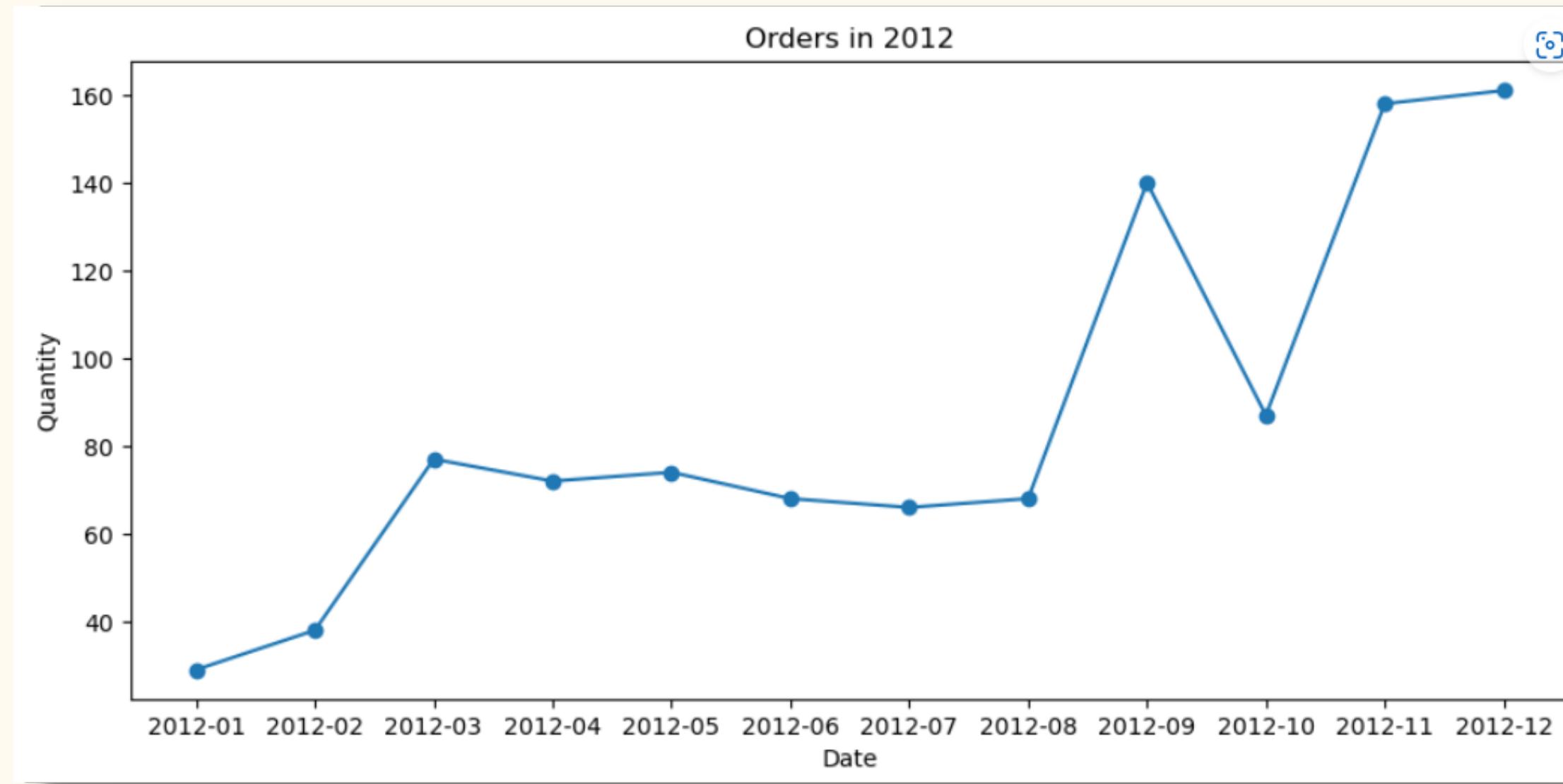
ORDERS IN 2011



The number of items purchased tends to **increase every year**. In **October sales declined sharply** but **increased again in the following month**, even exceeding sales in September. As for sales in **December experienced decline again**.



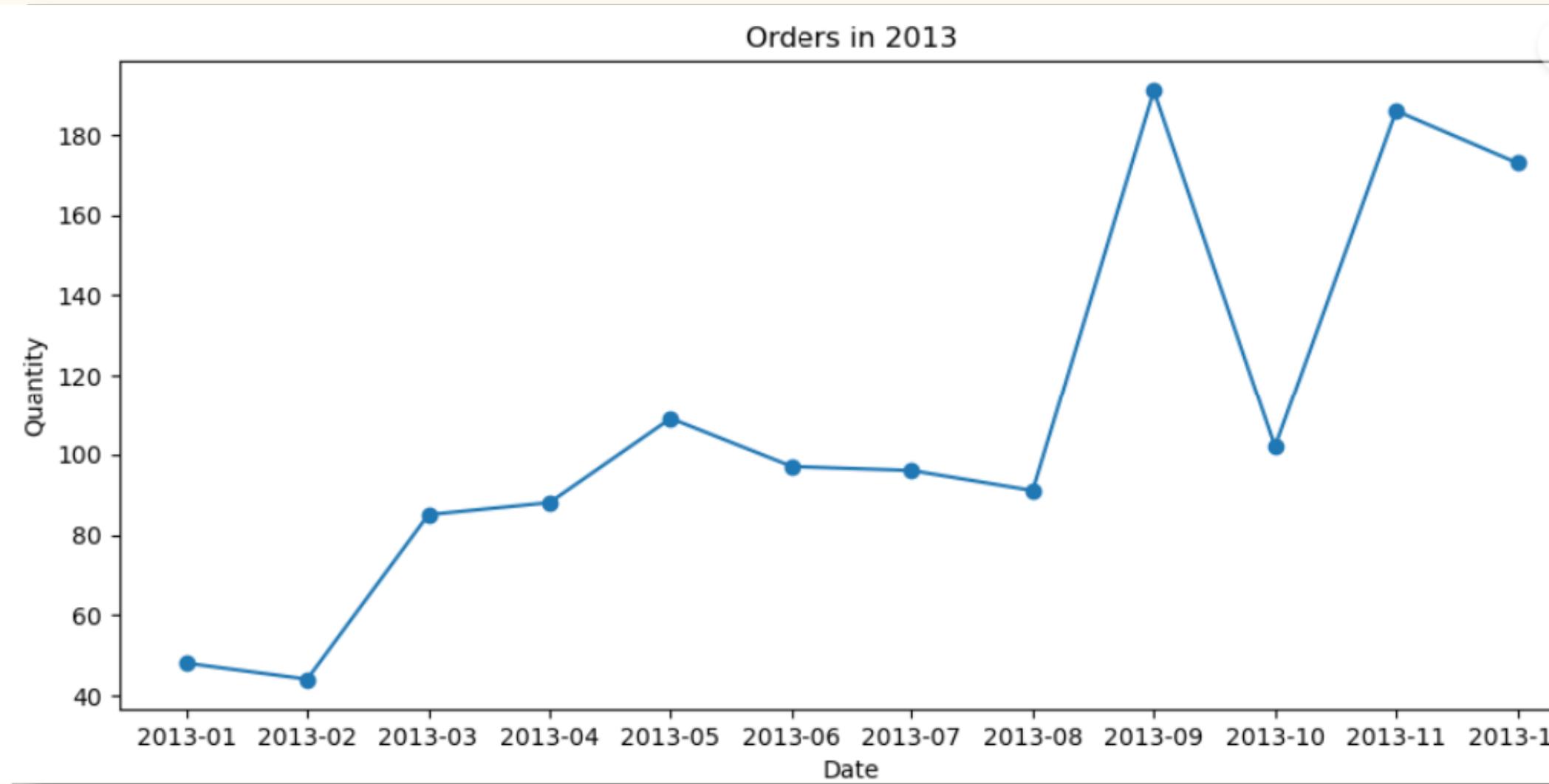
ORDERS IN 2012



The number of items purchased tends to **increase every month**, in **February to March there is a sharp increase about 2 times**. In **September there was a sharp increase**. However, in the following month, **October, there was a sharp decline**. However, in November there was a high increase even higher than in September and again **increased sales in December**.

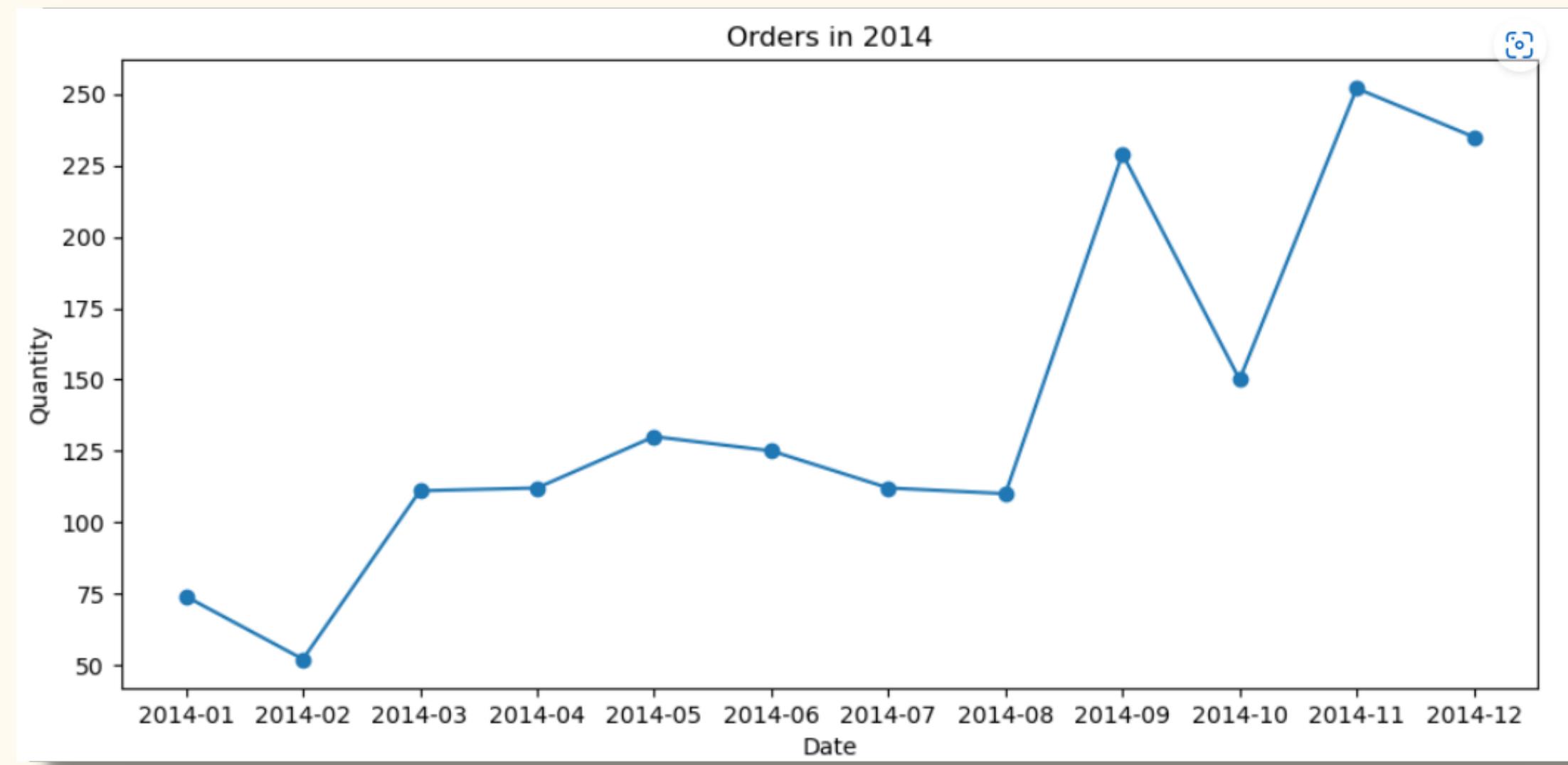


ORDERS IN 2013



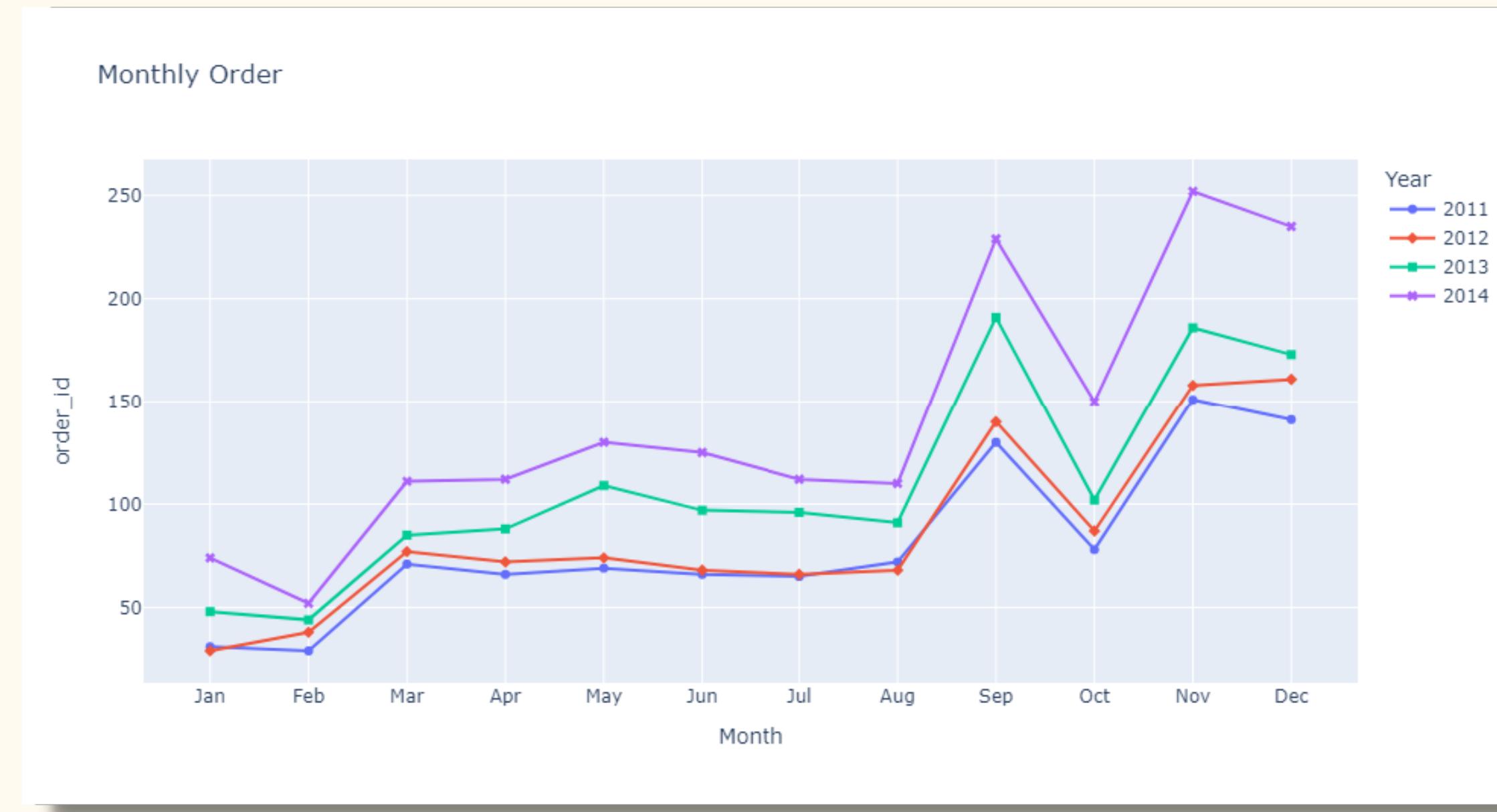
The number of items purchased tends to **increase every month**, in **February to March there is a sharp increase to almost 2 times**. In **September there was a sharp increase**. However, in the following month, **October, there was a sharp decline**. However, in November there was a high increase in the number of goods purchased and returned decreased slightly sales in December.

ORDERS IN 2014



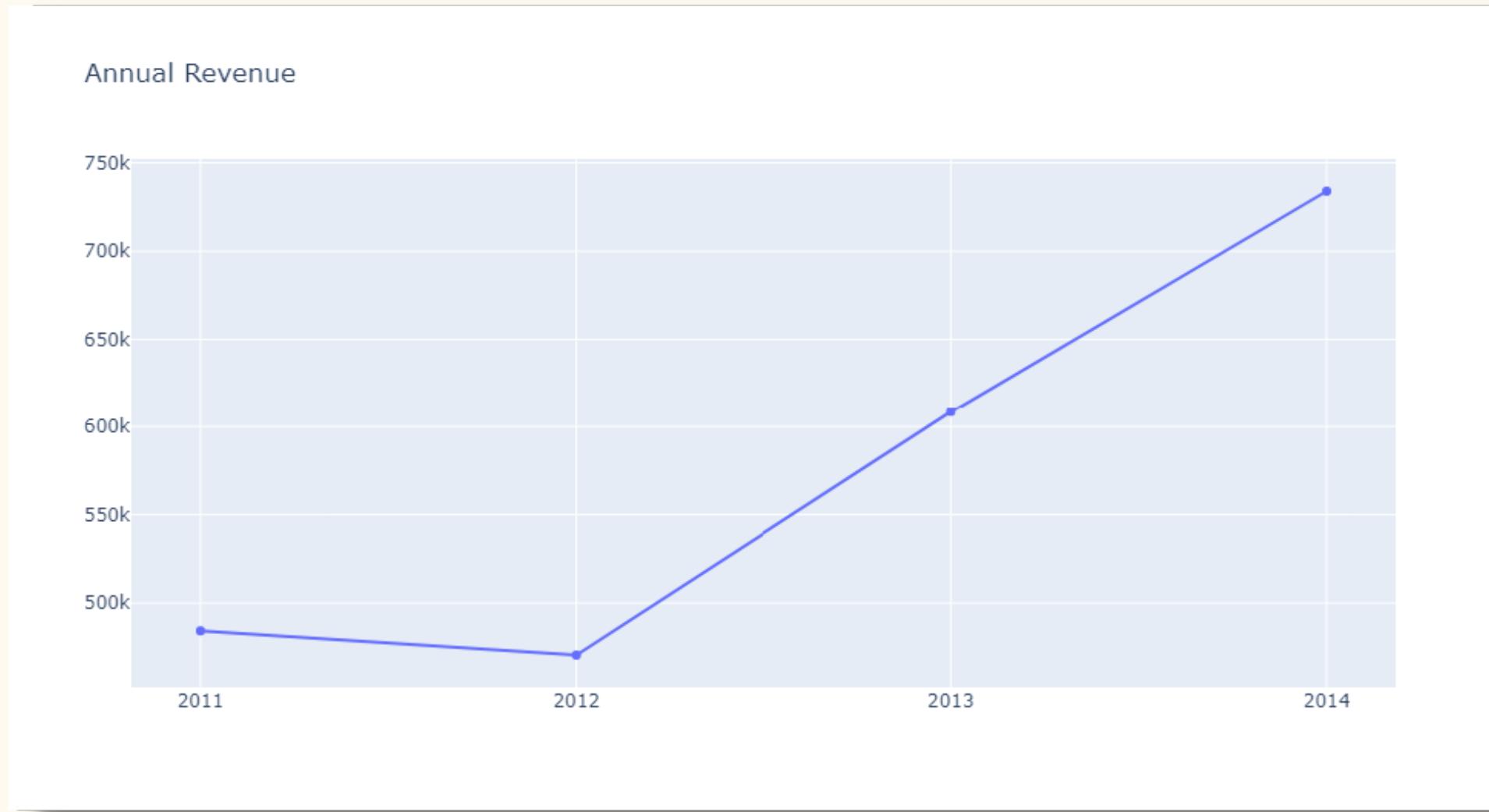
The number of items purchased tends to **increase every month**, in **February to March there is a sharp increase to almost 2 times**. In **September there was a sharp increase**. However, in the following month, **October, there was a sharp decline**. However, in **November, there was a high increase** in the number of goods purchased even higher than in September and again decreased slightly sales in December.

MONTHLY ORDERS



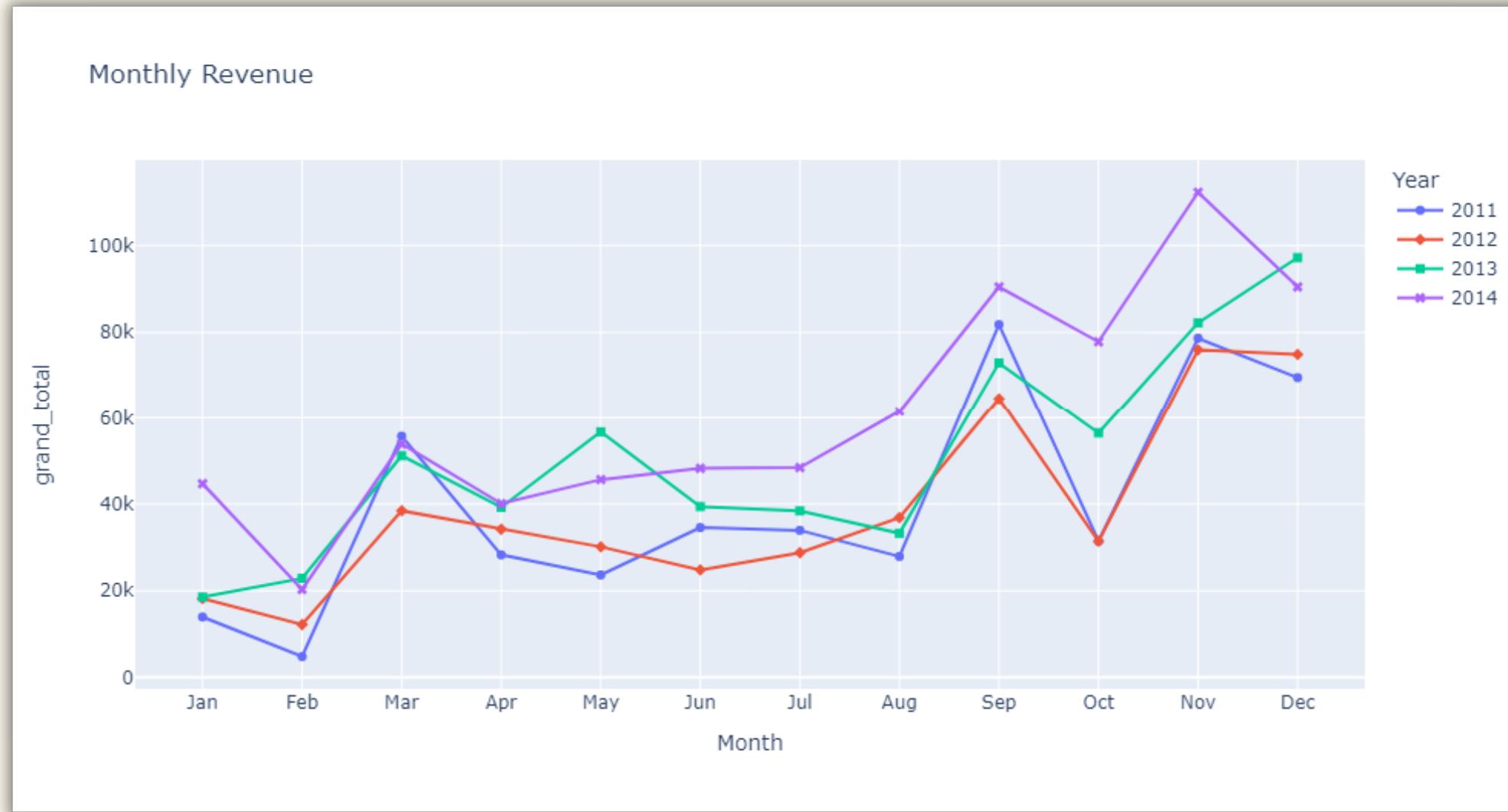
Based on the chart above, **product sales always increased in September and November in 2011-2014**. Meanwhile, there was **always a fairly sharp decreased in October in 2011-2014**.

ANNUAL REVENUE



Based on the chart above, revenue in 2011 amounted to 484.26K dollars and in 2012 it was 470,539K dollars. This means that **there was a decrease in revenue from 2011 to 2012 by 2.83%**. Meanwhile, from 2013 to 2014, revenue continued to increase. When viewed quantitatively, the increase in revenue from 2012 to 2013 was 29.3%. **From 2013 to 2014, revenue increased by 20.6%.**

MONTHLY REVENUE



Based on the chart above, monthly revenue from 2011-2014 is seasonal because in certain months there is an increase and decrease. In September and November in 2011-2014 there was **always a fairly high increase** in income, this may be due to the **high demand** for products for the Halloween and Christmas celebration preparation.



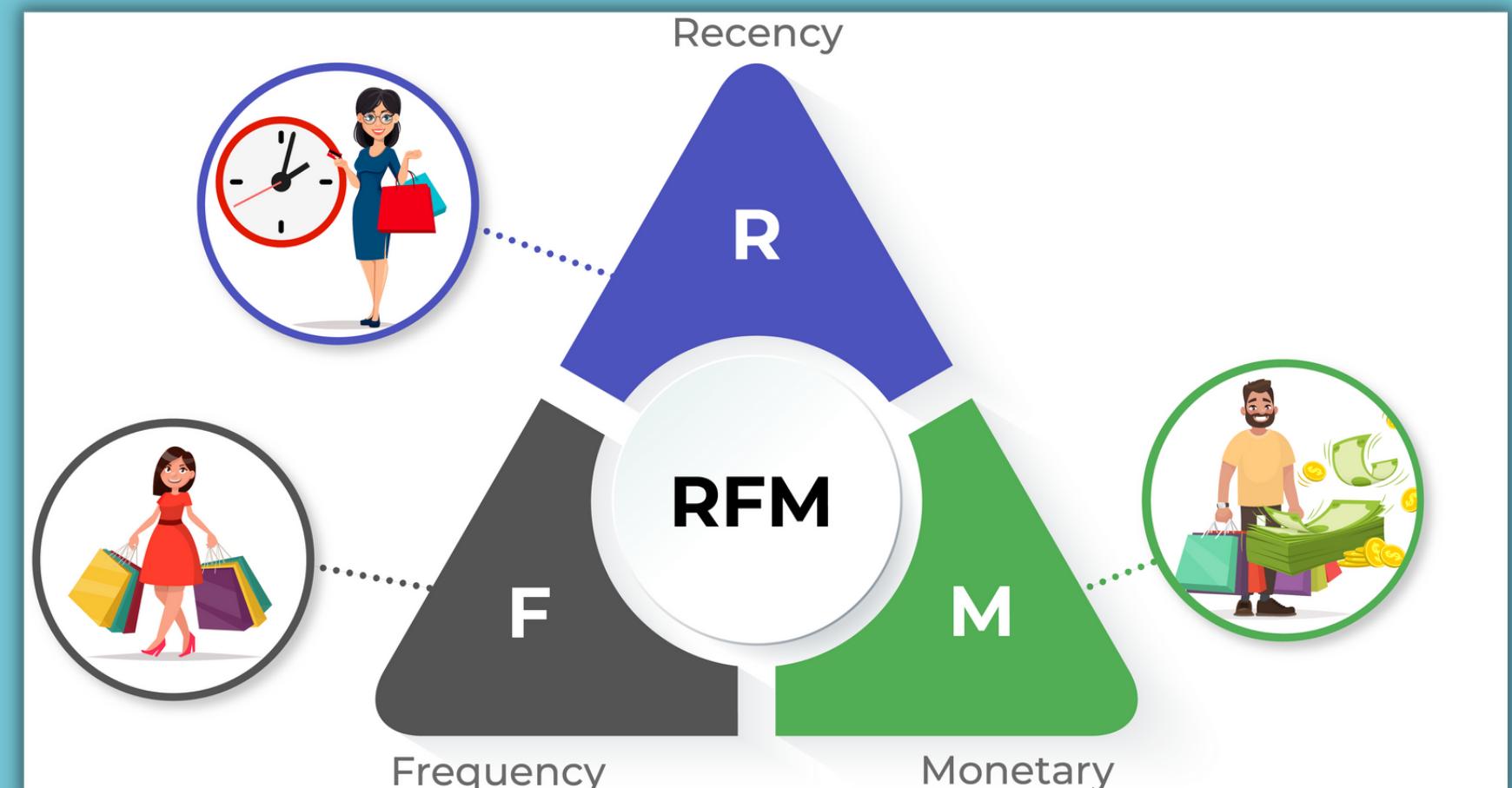
MODELING USING RFM

RFM ANALYSIS

Recency Frequency Monetary (RFM)

RFM is a model used in marketing analysis that segments a company's consumer base by their purchasing patterns or habits.

- Recency : how long it's been since a customer bought something from us.
- Frequency : how often a customer buys from us
- Monetary value : the total value of purchases a customer has made.



PREPROCESSING MODELING

- Create new data frame containing customer name column and recency column. This is done by calculating how long since customers make purchases.
- Create new data frame containing customer name column, frequency column, and monetary column. The results of frequency can be determined by calculating how often customers make a purchase. While the monetary is produced from the calculation of how much money the customer spent making purchases.
- Combine the recency table with the frequency and monetary table.

	Customer	Recency	Frequency	Monetary
0	Aaron Bergman	415	3	887
1	Aaron Hawkins	12	7	1744
2	Aaron Smayling	88	7	3050
3	Adam Bellavance	54	8	7756
4	Adam Hart	34	10	3249
...
788	Xylona Preis	43	11	2375

Modeling Data Using RFM

- Split the metrics into segments using quantiles.
- Assign a score from 1 to 4 to each Recency, Frequency and Monetary respectively.
- 1 is the highest value, and 4 is the lowest value.
- A final RFM score (Overall Value) is calculated simply by combining individual RFM score numbers.

Modeling Data Using RFM

Segments	RFM Score
BEST CUSTOMERS	III
LOYAL CUSTOMERS	F=1
BIG SPENDERS	M=1
ALMOST LOST	134
LOST CUSTOMERS	344
LOST CHEAP CUSTOMERS	444

LABELLING

LABELLING

```
print("Best Customers: ", len(segmented_rfm[segmented_rfm['RFM_Segment']=='111']))  
print("Loyal Customers: ", len(segmented_rfm[segmented_rfm['F_quartile']==1]))  
print("Big Spenders: ", len(segmented_rfm[segmented_rfm['M_quartile']==1]))  
print("Almost Lost: ", len(segmented_rfm[segmented_rfm['RFM_Segment']=='134']))  
print("Lost Customers: ", len(segmented_rfm[segmented_rfm['RFM_Segment']=='344']))  
print("Lost Cheap Customers: ", len(segmented_rfm[segmented_rfm['RFM_Segment']=='444']))
```

```
Best Customers: 30  
Loyal Customers: 159  
Big Spenders: 198  
Almost Lost: 2  
Lost Customers: 31  
Lost Cheap Customers: 67
```

We mark label and it can be seen that there are 6 types of customers

LABELLING

```
label = [0] * len(segmented_rfm)

for i in range(0,len(segmented_rfm)):
    if segmented_rfm['RFM_Segment'][i] == '111':
        label[i] = "Best Customers"
    elif segmented_rfm['RFM_Segment'][i] == '134':
        label[i] = "Almost Lost"
    elif segmented_rfm['RFM_Segment'][i] == '344':
        label[i] = "Lost Customers"
    elif segmented_rfm['RFM_Segment'][i] == '444':
        label[i] = "Lost Cheap Customers"
    elif segmented_rfm['F_quartile'][i] == 1:
        label[i] = "Loyal Customers"
    elif segmented_rfm['M_quartile'][i] == 1:
        label[i] = "Big Spenders"
    else:
        label[i] = "Others"
```

Setting up the label for each client and adding the column "Label" to the dataframe

LABELLING

```
segmented_rfm['Label'].value_counts()
```

others	409
Loyal Customers	129
Big Spenders	125
Lost Cheap Customers	67
Lost Customers	31
Best Customers	30
Almost Lost	2

Name: Label, dtype: int64

Count the frequency that a value occurs in a dataframe column for the labels

LABELLING

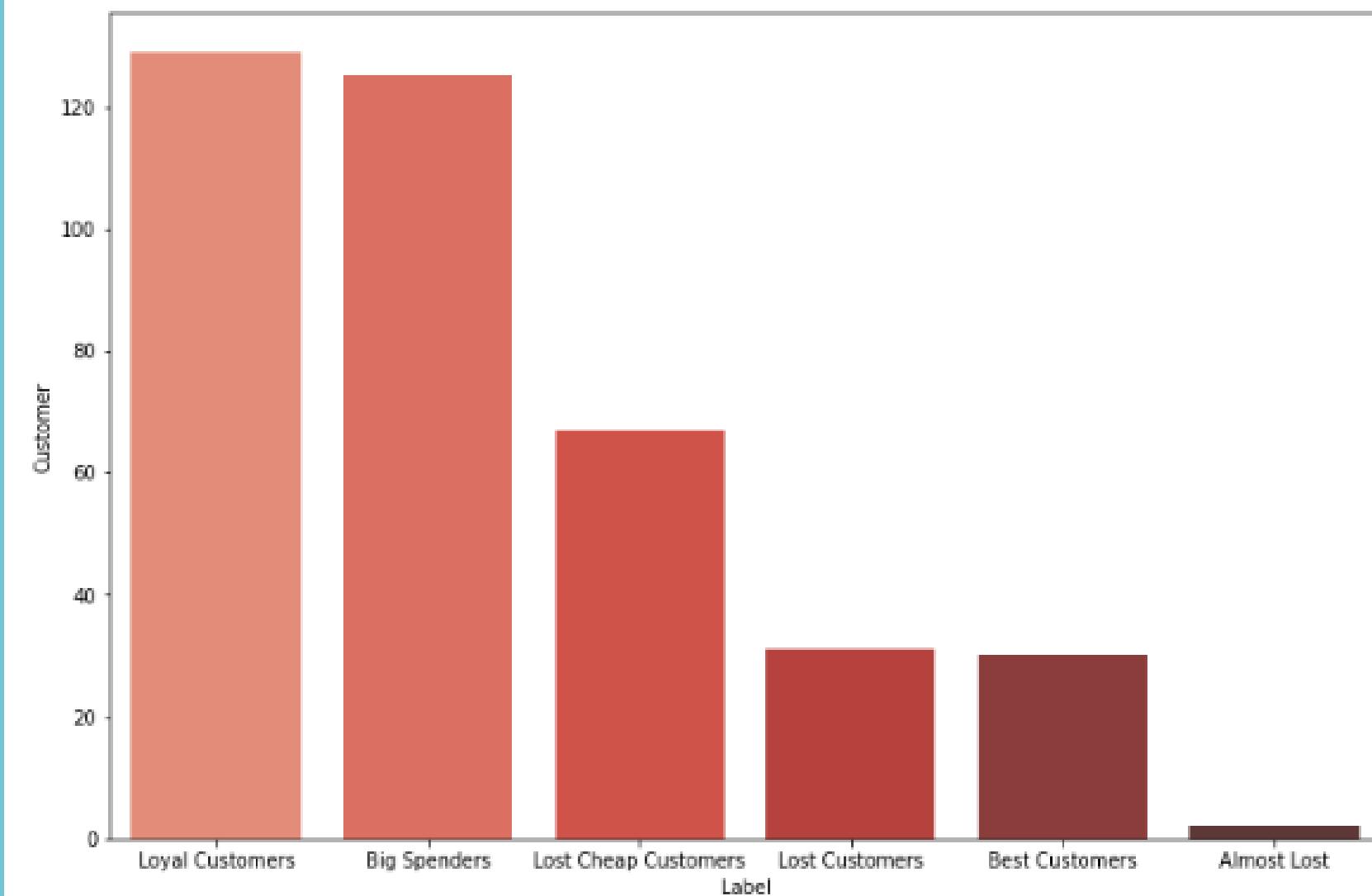
```
percent = pd.DataFrame(round(segmented_rfm.Label.value_counts()*100/segmented_rfm.Label.count(),2))
percent = percent.rename(columns={'Label':'Percent'})
percent.head(7)
```

Percent	
Others	51.58
Loyal Customers	16.27
Big Spenders	15.76
Lost Cheap Customers	8.45
Lost Customers	3.91
Best Customers	3.78
Almost Lost	0.25

We made the precentage of each type of customers

LABELLING

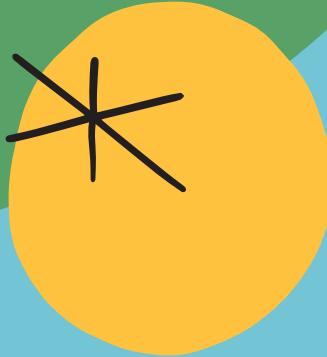
```
sq1=segmented_rfm.groupby('Label')['Customer'].nunique().sort_values(ascending=False).reset_index()  
plt.figure(figsize=(12,8))  
sq1.drop([0],inplace=True)  
sns.barplot(data=sq1, x="Label", y="Customer", palette = "Reds_d");
```



Based on the graph, there are 6 segmentations of customers using RFM modeling. The Loyal Customers segmentation has the highest percentage of 16.27% of total customers. Followed by Big Spenders with a percentage amount of 15.76%, Lost Cheap Customers with a percentage amount of 8.45%, Lost Customers with a percentage amount of 3.91%, Best Customers with a percentage amount of 3.78%, Then Almost Customers with a percentage amount of 0.25%.



RESULT



RESULT

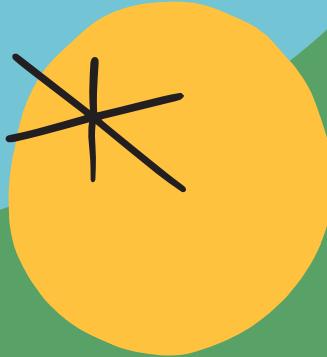


- In 2011, The number of items purchased tends to increase every year. In October sales declined sharply but increased again in the following month, even exceeding sales in September. As for sales in December experienced decline again.
- In 2012, The number of items purchased tends to increase every month, in February to March there is a sharp increase about 2 times. In September there was a sharp increase. However, in the following month, October, there was a sharp decline. However, in November there was a high increase even higher than in September and again increased sales in December.
- In 2013, The number of items purchased tends to increase every month, in February to March there is a sharp increase to almost 2 times. In September there was a sharp increase. However, in the following month, October, there was a sharp decline. However, in November there was a high increase in the number of goods purchased and returned decreased slightly sales in December.
- In 2014, The number of items purchased tends to increase every month, in February to March there is a sharp increase to almost 2 times. In September there was a sharp increase. However, in the following month, October, there was a sharp decline. However, in November, there was a high increase in the number of goods purchased even higher than in September and again decreased slightly sales in December.

RESULT

- Product sales always increased in September and November in 2011-2014. Meanwhile, there was always a fairly sharp decreased in October in 2011-2014.
- Based on the chart Annual Revenue, revenue in 2011 amounted to 484.26K dollars and in 2012 it was 470,539K dollars. This means that there was a decrease in revenue from 2011 to 2012 by 2.83%. Meanwhile, from 2013 to 2014, revenue continued to increase. When viewed quantitatively, the increase in revenue from 2012 to 2013 was 29.3%. From 2013 to 2014, revenue increased by 20.6%.
- Based on the chart Monthly Revenue, monthly revenue from 2011-2014 is seasonal because in certain months there is an increase and decrease. In September in 2011-2014 there was always a fairly high increase in income, this may be due to the high demand for products for the feast day.
- Based on the graph Labelling, there are 6 segmentations of customers using RFM modeling. The Loyal Customers segmentation has the highest percentage of 16.27% of total customers. Followed by Big Spenders with a percentage amount of 15.76%, Lost Cheap Customers with a percentage amount of 8.45%, Lost Customers with a percentage amount of 3.91%, Best Customers with a percentage amount of 3.78%, Then Almost Customers with a percentage amount of 0.25%.

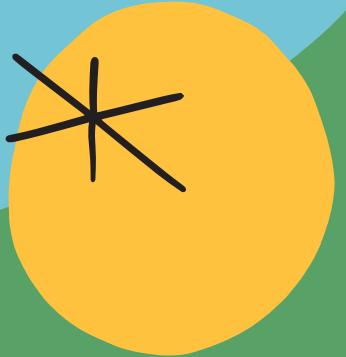
RECOMMENDATION



RECOMENDATION



- Recommendations for increased orders in a given month such as September and November: focus on increasing the amount of production of goods because in certain months such as September and November will receive high demand.
- Recommendations for a decrease orders in a given month such as October: focus on improving marketing strategy so that customers will orders in that month. It can also be by optimizing sales such as multiplying products in October with Halloween celebrations or give points to increase customer purchasing power by giving vouchers or other gifts. The company also can hold an event or sale that smells of Valentine's Day in February to increase sales for that month.
- Recommendations for Loyal Customers segment: The company must optimize the campaign for this customer segment in order to maintain their loyalty. It can also give additional points for customer loyalty.
- Recommendations for Big Spenders segment: focus on offering other products, especially products that match customer interests or products that are often sought after. Can also offer products with a high price range.



RECOMENDATION



- Recommendations for Lost Cheap Customers segment: focus on activating customers and making repurchases by forming a Reactivation Strategy. One of them is by offering products that have been purchased.
- Recommendations for Lost Customers segment: focus on activating customers and making repurchases by forming a Reactivation Strategy. One of them is by offering products that have been purchased.
- Recommendations for Best Customers segment: focus on increasing customer purchases therefore it is necessary to form a Cross Selling or Up Selling Strategy by offering other products that can be complementary to frequently purchased products.
- Recommendations for Almost Lost segment: focus on activating customers and making repurchases by forming a Reactivation Strategy. One of them is by offering products that have been purchased.

**THANK YOU FOR
WATCHING OUR
PRESENTATION!**

