MINI PROJECT BY GROUP 1

CUSTOMER SEGMENTATION

USING RFM WITH DATA ORDERS

MYEDUSOLVE

Our Team

Agung Wahyu Prayogo
Informatics
Bhayangkara Jakarta Raya University

Valent Aderiandra
Information System
UPN "Veteran" East Java

Fauziya Alya Ramadhana
Statistics
Sepuluh Nopember Institute of Technology

Putri Windasari
Mathematics
Diponegoro University

Reno Darin Khalifah P
Electrical Engineering
Sepuluh Nopember Institute of Technology

WorkFlow Customer Segmentation

01 Use Case

02 Business Understanding

03 Data Understanding

04 Data Preparation

05 Data Cleansing

06 Exploratory Data Analysis

07 Modelling

08 Evaluation

09 Recommendation

01

Use Case Summary

- **Objective Statement** :
  - Get business insight about how many customers
  - Get business insight about the average, maximum and minimum amount issued by the customer
  - Get business insight about how many orders by year
  - Get business insight for each customers based on label defined such as Loyal Customer, Big Spenders, Lost Cheap Customers,Lost Customers and Best Customers,Almost Lost.
  - Build models using RFM to learn customer segmentation

02

Use Case Summary

- Challanges :
  - Large size of data, can not maintain by excel spreadsheet
  - Change the data type of column which doesn't match
  - Display data every year
  - Make a definition of each segment obtained
  - Don't know the location of retail data

03

Use Case Summary

Methodology / Analytic Technique

A

Descriptive analysis

Describe the information such as, min/max value of each column, average, and the total count of data contained in grand_total column.

B

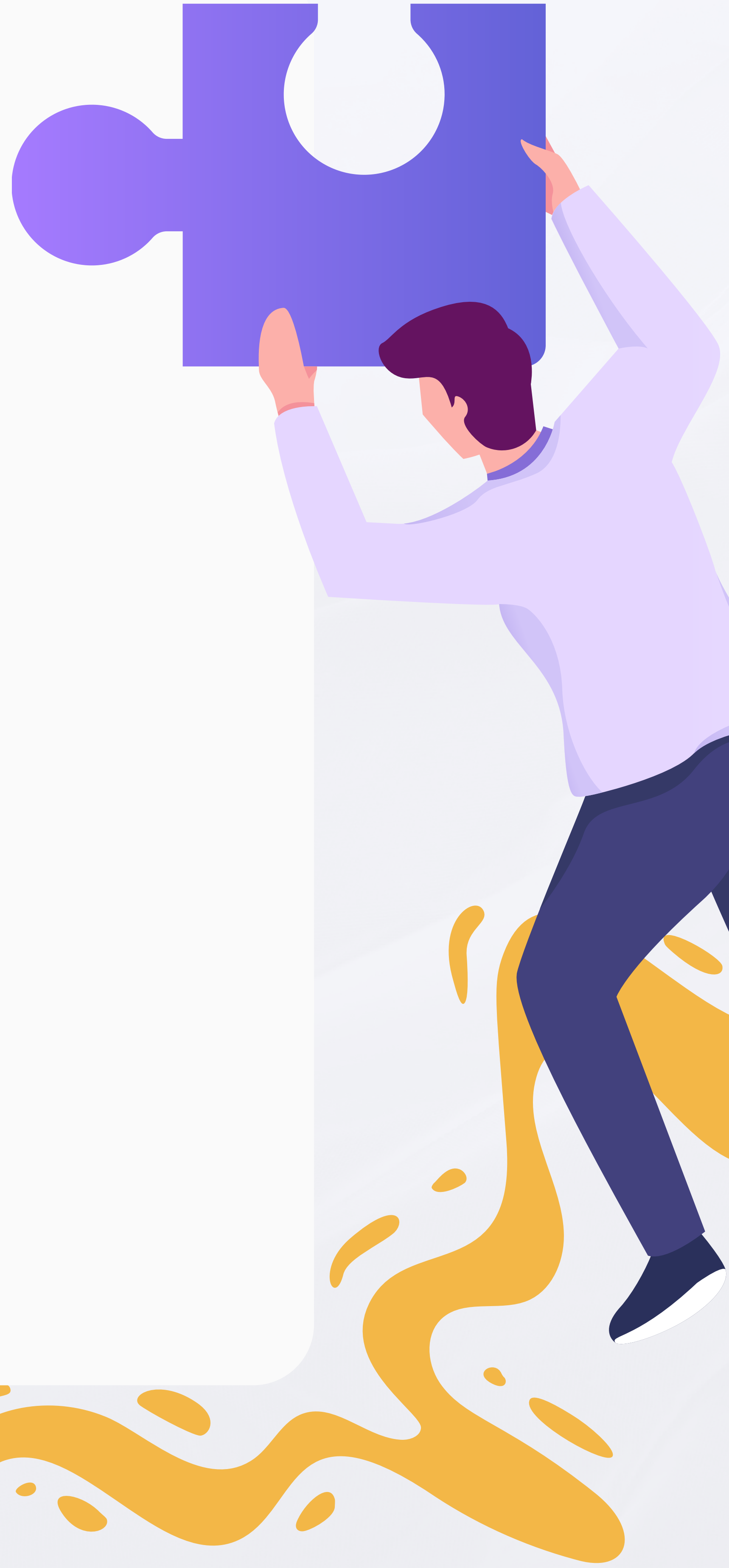
Graph analysis

View data changes by time (Year)

C

Using Modelling Segementation

RFM Model



04

Use Case Summary

Business Benefit:

- Gain insight to keep customers from churn through modification of benefits/features provided to these customers
- Gain insight to improve the quality of company services so that customers remain loyal and gain more profit for the company
- Build models using RFM to know customer segement

Expected Outcome:

- Know how many customer
- Know how about the average, maximum and minimum amount issued by the customer
- Know how many many orders by year
- Know how how behavior customers based on labeled defined such as Loyal Customer, Big Spanders, Best Customers, Lost Cheap Customers and Almost Lost
- Know how to build model using RFM

Business Understanding

- Retail business is a business that involves selling goods or services to consumers in units or retail. Consumers who buy products or services in retail are intended to consume or use them personally and do not resell them.
- This case has some business question using the data:
  - How many customer?
  - How about the average, maximum and minimum amount issued by the customer?
  - How many many orders by year?
  - How behavior customers based on labeled defined such as Loyal Customer, Big Spander, Best Customers, Lost Cheap Customers and Almost Lost?
  - How to build model using RFM modelling?

07

Data Understanding

Source Data:

- The dataset used is data from https://www.kaggle.com/datasets/siddinho/sample-orders-dataset-retail
- The dataset used is data from no where
- The raw data contains 5009 rows and 4 columns.

Data Dictionary

- order_date : date of ordering items by customer
- order_id : the transaction id of ordering items by the customer
- customer : the name of the customer who made the purchase
- grand_total : total price paid by customer

06

Data Preparation

- Code use:
  - Python 3.9.13
  - Package: Pandas, Numpy, Matplotlib, Seaborn, Scipy, Sklearn, and Warning

DATA
CLEANSING

Data Cleansing 'order_date'

```
#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')
```

change data type from order of object to datetime and retrieve data only month and year only

Output

```
df.head()
```

	order_date	order_id	customer	grand_total	OrderYearMonth	Date	Year
0	2011-09-07	CA-2011-100006	Dennis Kane	378	201109	2011-09	2011
1	2011-07-08	CA-2011-100090	Ed Braxton	699	201107	2011-07	2011
2	2011-03-14	CA-2011-100293	Neil Franzšsisch	91	201103	2011-03	2011
3	2011-01-29	CA-2011-100328	Jasper Cacioppo	4	201101	2011-01	2011
4	2011-04-08	CA-2011-100363	Jim Mitchum	21	201104	2011-04	2011

EXPLORARY

DATA ANALYSIS

1 How many customers?

```
df.nunique()
```

```
order_date      1238
order_id        5009
customer         793
grand_total     1358
dtype: int64
```

From here it can be seen, total customers are 793. The number of Order_id turns out to be not the same as the number of Customers, so there are some Order_ids that have the same Customers. This shows that there are customers who repeat orders or do not only shop once at the company. order_date shows that there are several transactions on the same day

Exploratory Data Analysis

2 Average, maximum and minimum amount by the customer

```
1 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 data above, it is known that on average customers make product purchases with a total price of 458.626672, the customer with the lowest purchase is rated 1 and the highest purchase is rated 23661
number 1 may get high discount for example black friday sale promo, and 23661 the possibility of buying items that are considered quite expensive

Exploratory Data Analysis

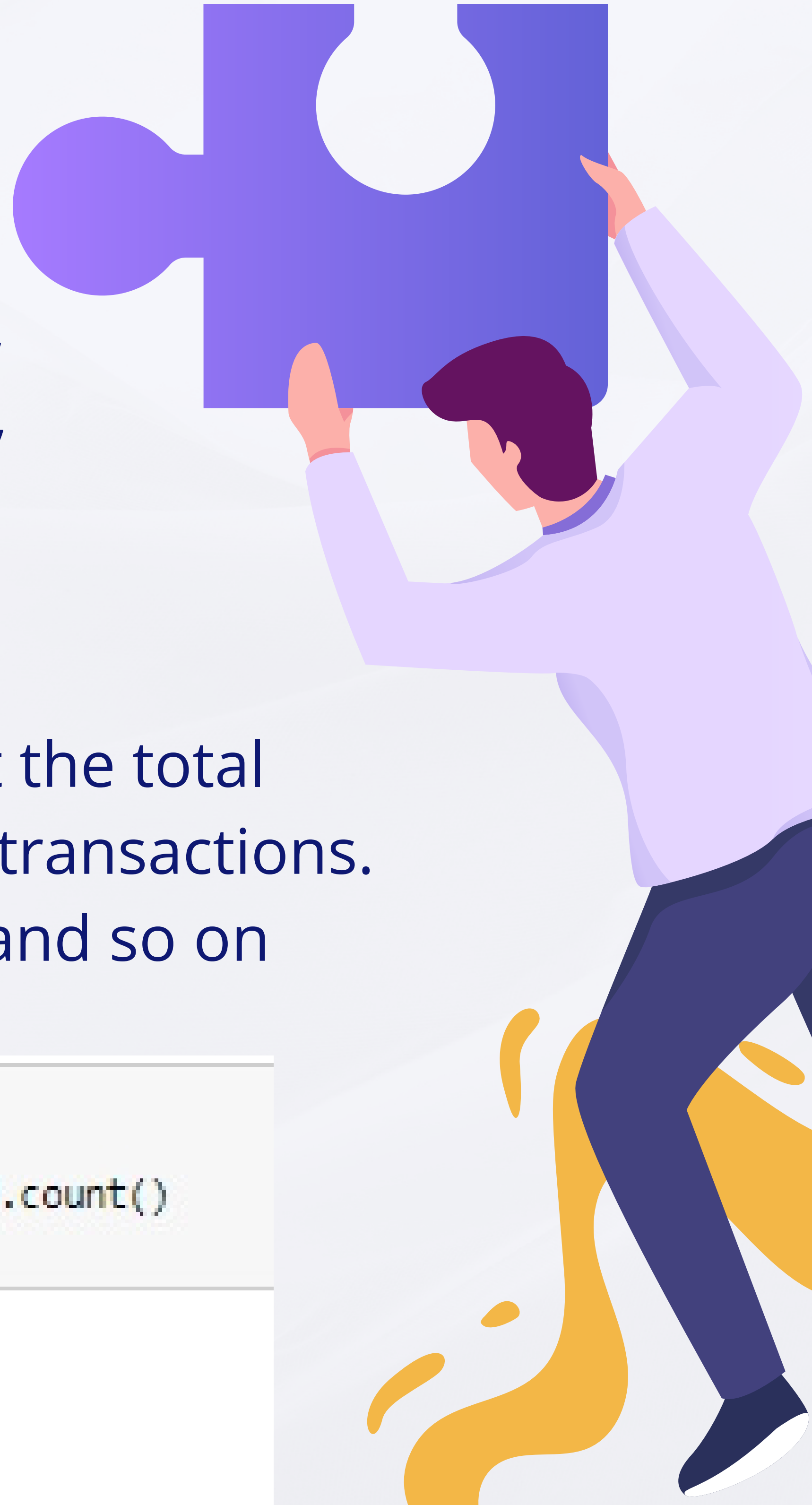
3 Timebound

```
#timebound

print(df["order_date"].min())
print(df["order_date"].max())
print("The data is from 4 year transaction")
```

```
2011-01-04 00:00:00
2014-12-31 00:00:00
The data is from 4 year transaction
```

dataset, date started from January 4, 2011. terminated until December 31, 2014.



4 Aggregating the Orders by Month

```
#Aggregating the Orders by Month

df_agg = df.groupby(["Year", "Date"]).order_id.count()
df_agg.head()
```

```
Year  Date
2011  2011-01    31
      2011-02    29
      2011-03    71
      2011-04    66
      2011-05    69
Name: order_id, dtype: int64
```

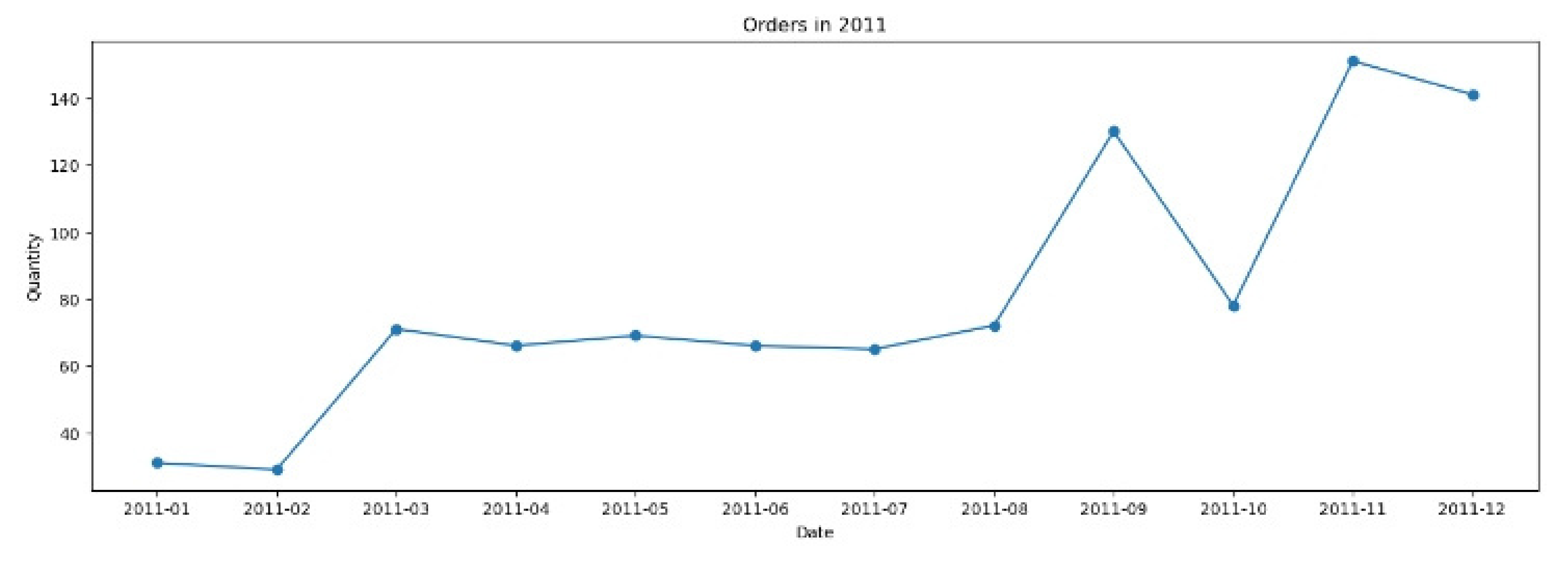
From this data, it can be seen that the total transactions in January 2011 were 31 transactions. 29 transactions in February 2011 and so on


```
#Aggregating the Orders by Month

df_agg = df.groupby(["Year", "Date"]).order_id.count()
df_agg.head()
```

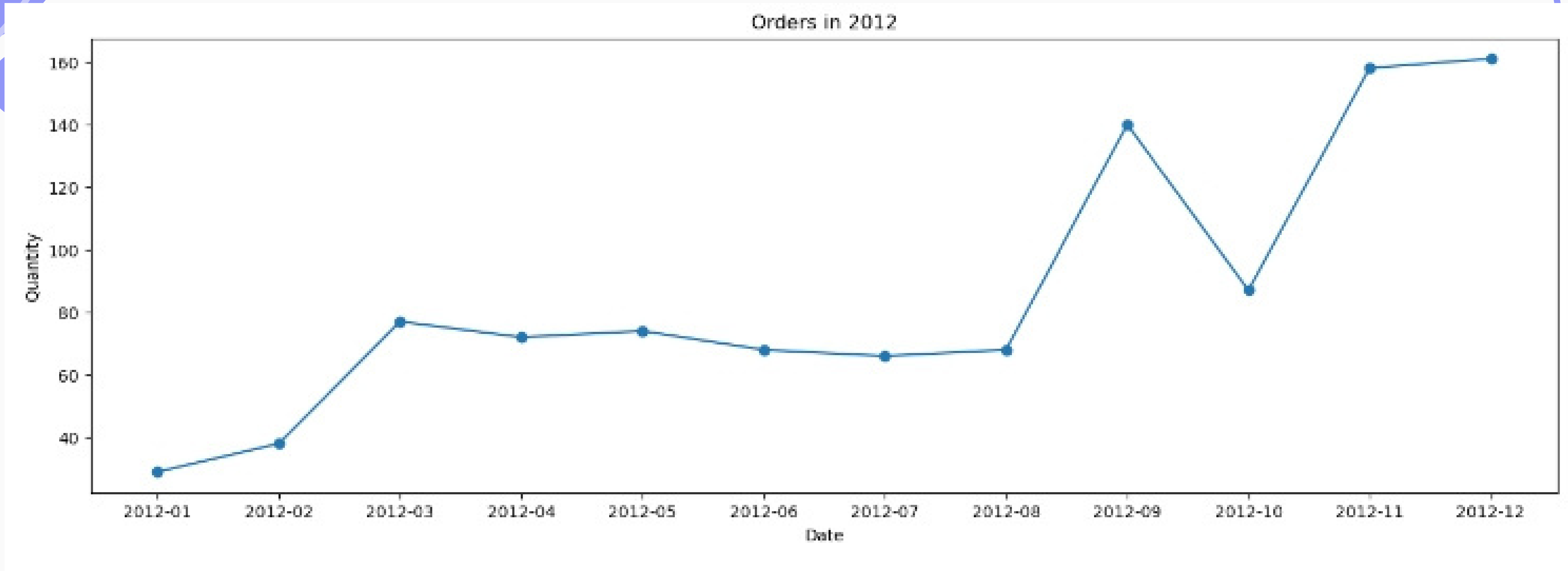
```
Year  Date
2011  2011-01    31
      2011-02    29
      2011-03    71
      2011-04    66
      2011-05    69
Name: order_id, dtype: int64
```

Order Data in 2011



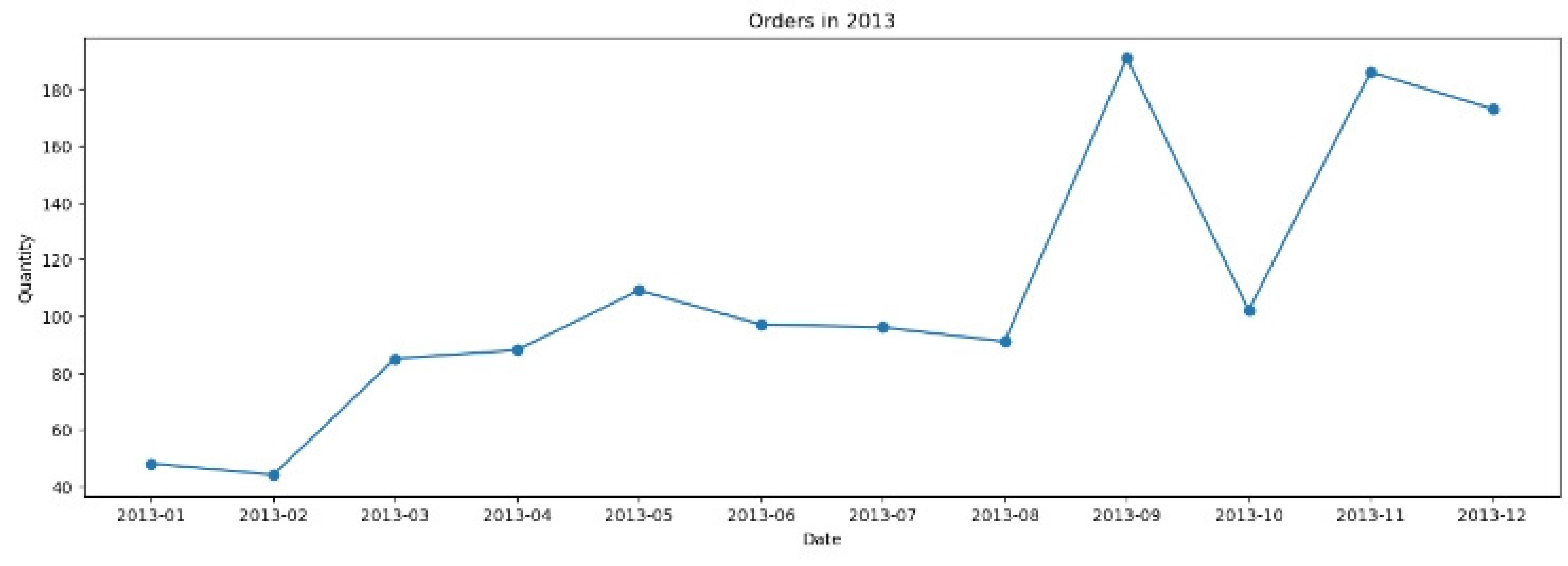
The number of items purchased tends to increase every mounth. In October sales declined sharply but increased in the following month, even exceeding sales in September because normally In most countries in the Northern Hemisphere, September is the time for students to start new school year at school. November's sales are high because there are usually lots of Black Friday promotions. Meanwhile, sales in December experienced a decline again.

Order Data in 2012



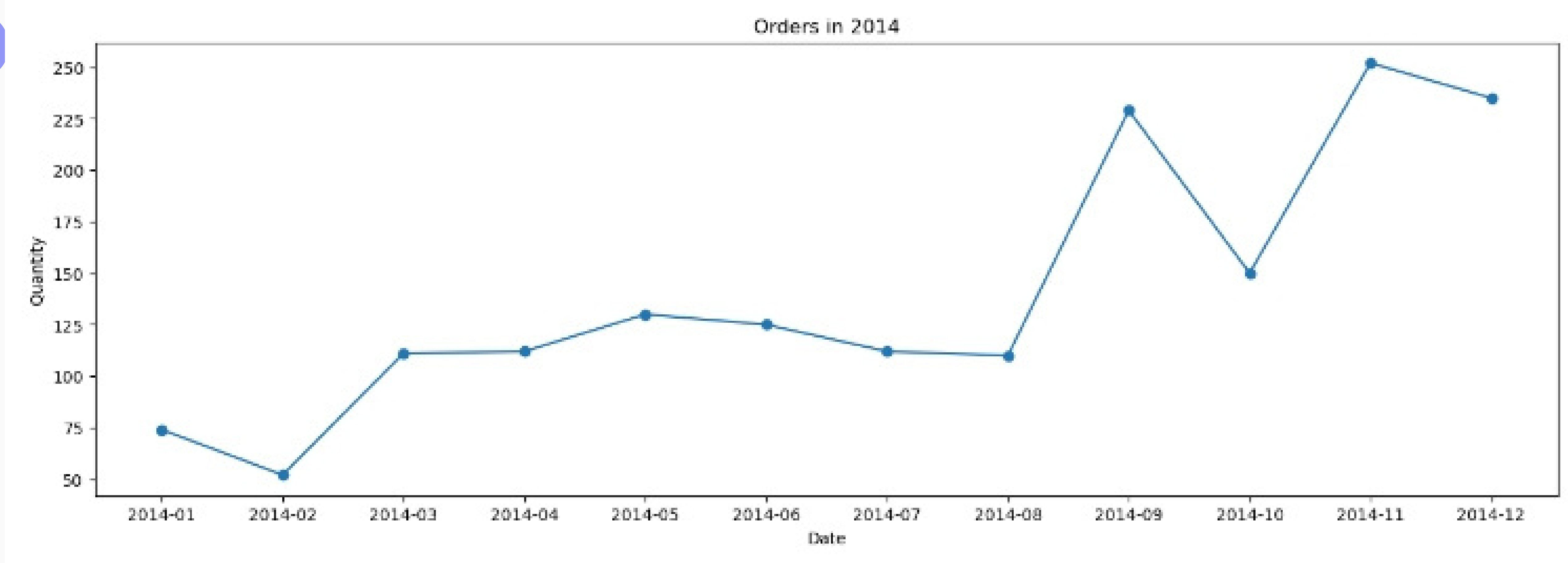
The number of items purchased tends to **increase** every month, in **February to March** there is a sharp increase of up to **2 times**. In **September** there was **a sharp increase**. However, in the following month, **October, there was a sharp decline** because in that month, **started the new school year**. in **November** there was a high increase even higher than September and again **increased** because of **Black Friday**, and sales in **December** fell again

Order Data in 2013



The number of items purchased tends to **increase** every month, in **February to March** there is a fairly sharp increase to **almost 2 times**. In the month of **September** there was **a sharp increase** because many parents bought school supplies for their children. However, in the following month, **October, there was a sharp decline** due to **new school year**. However, in **November** there was a **high** increase in the number of items purchased because **Black Fiday** promos usually occur and again **slightly decreased** sales in **December**.

Order Data in 2014



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

Show with Data Frame

```
df_year = df.groupby(df.order_date.dt.year).grand_total.sum()
df_year = pd.DataFrame(df_year)
df_year = df_year.reset_index()
df_year.head()
```

	order_date	grand_total
0	2011	484260
1	2012	470539
2	2013	608477
3	2014	733985

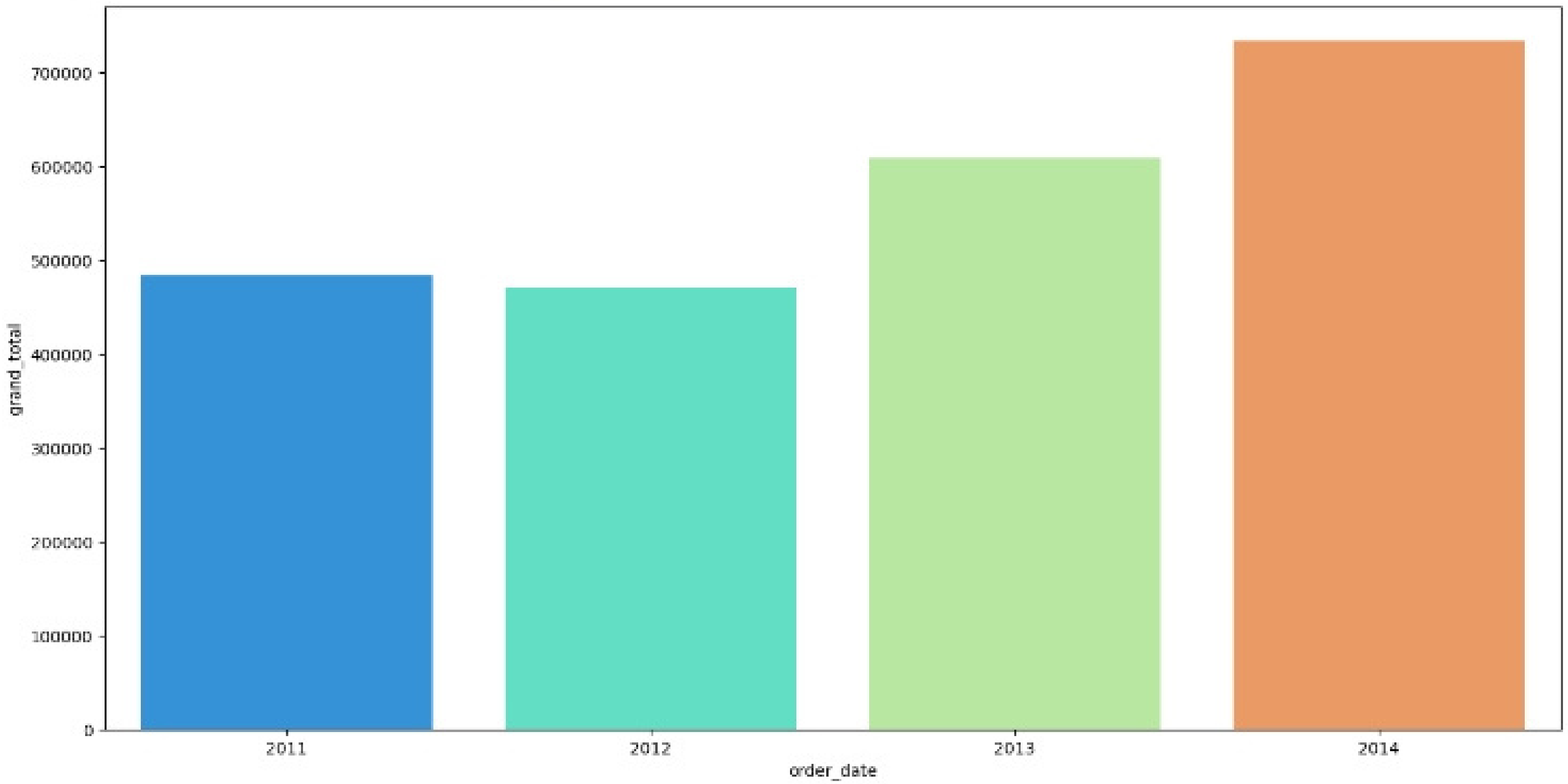
Display 'grand_total' in 2011 - 2014

Percentage grand_total

```
data = df_year
data['percent'] = round(data['grand_total']*100/data['grand_total'].sum())
print(data)
plt.figure(figsize=(16,8))
sns.barplot(data = df_year, palette='rainbow', x = 'order_date', y = 'grand_total')
```

```
   order_date  grand_total  percent
0        2011       484260     21.0
1        2012       470539     20.0
2        2013       608477     26.0
3        2014       733985     32.0
```

Out[22]: <AxesSubplot:xlabel='order_date', ylabel='grand_total'>



In the barchart above, you can see how much the order increases each year. In 2011 with a grand total of 484260 with a value of percentage of 21%. In 2012 with a grand total of 470539 with a percentage value of 20%. In 2013 with a grand total of 608477 with a percentage value of 26%. and in 2014 with a grand total of 733985 with a percentage value of 32%.

MODELING

WITH

RECENCY, FREQUENCY, MONETARY VALUE (RFM)

RFM DEFINITION

Recency, frequency, monetary value (RFM) is a model used in marketing analysis that **segments a company's consumer** base by their purchasing patterns or habits. In particular, it evaluates customers' **recency** (how long ago they made a purchase), **frequency** (how often they make purchases), and **monetary** value (how much money they spend).

Modeling Result - RFM Quantile

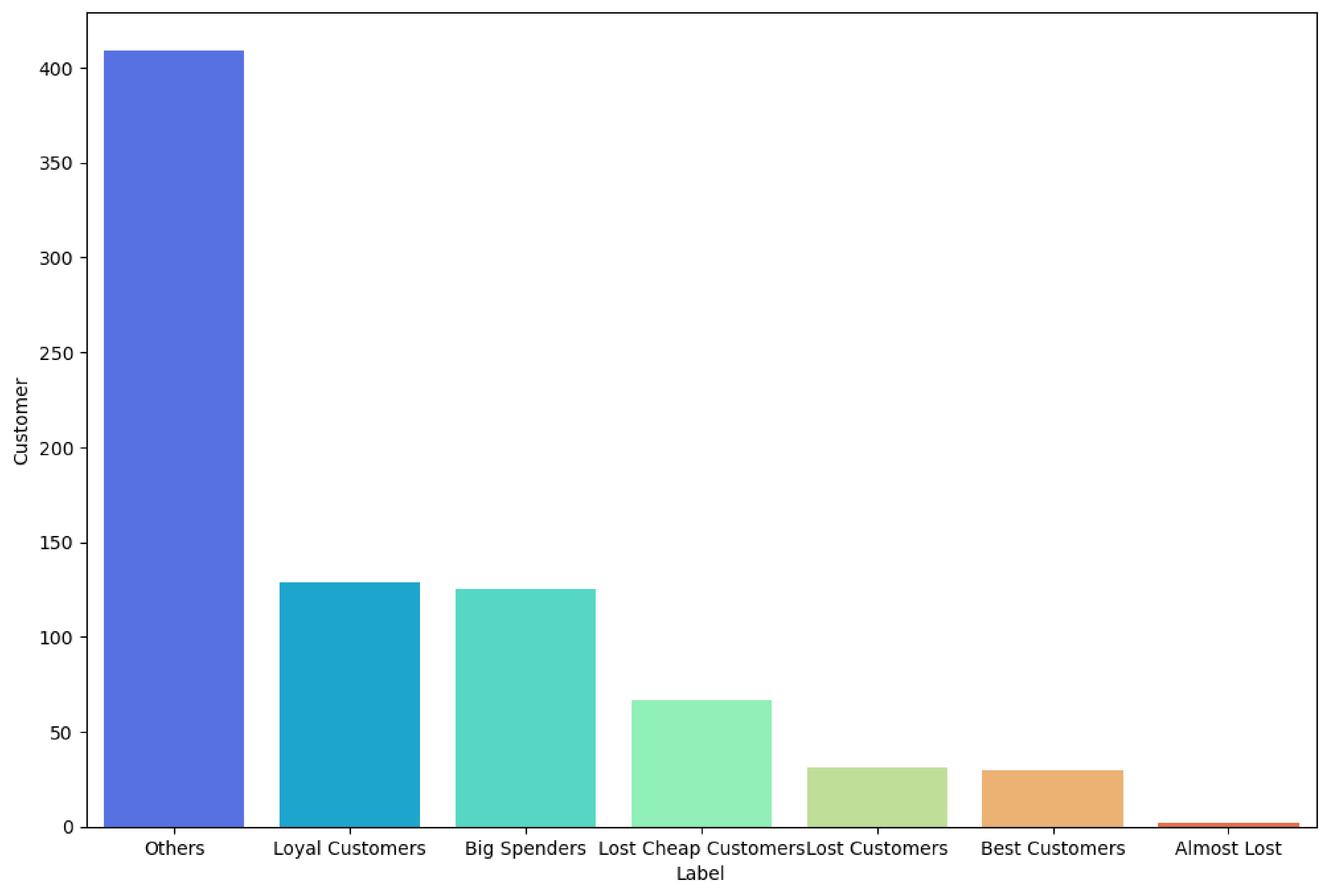
	Customer	Recency	Frequency	Monetary	R_quartile	F_quartile	M_quartile	RFM_Segment	RFM_Score
0	Aaron Bergman	415	3	887	4	4	4	444	12
1	Aaron Hawkins	12	7	1744	1	2	3	123	6
2	Aaron Smayling	88	7	3050	3	2	2	322	7
3	Adam Bellavance	54	8	7756	2	2	1	221	5
4	Adam Hart	34	10	3249	2	1	2	212	5

The **bigger** the number in **R_quartile**, then it proves that the customer is the **longer they haven't shopped**. While the **smaller the number** in the R_quartile, it proves that the customer the **more recently shopped**. From this example, it can be seen, Aaron Bergman last shopped 415 days ago, because it's been a while, he got an R_quartile 4, while Aaron Hawkins last shopped 12 days ago, he got an R_quartile 1.

The **bigger the number** in **F_quartile**, then it proves that the customer is the **less** frequency he shopped. While the **smaller the number** in F_quartile, it proves that the customer the **more frequency** he shopped.

Likewise for **M_quartile**, the **bigger the number** in M_quartile, then it proves that the customer is the **smaller** the amount spent. While the **smaller the number** in M_quartile, it proves that the customer the **more amount** he spends for shopping.

Modeling Result - RFM Quantile



There are 7 **customer segmentation** based on RFM.

- Loyal Customers
- Big Spenders
- Lost Cheap Customers
- Lost Customers
- Best Customers
- Almost Lost
- Others

Modeling Result - RFM Quantile

A total of **16.27%** of the total customers (with a total of 129 customers) are included in **Loyal Customers** where this segment's customers have the most shopping frequency than other segments.

Furthermore, **15.76%** of the total customers (with a total of 125 customers) in the company entered the **Big Spenders** segment. This segment has the highest total purchases than any other segment.

Lost Cheap Customers segment consists of 67 customers with a percentage of **8.44%** of the total customers. Customers who enter this type of segment tend to churn because they have not bought goods from the company for a long time, have the least frequency of shopping, and have the least total purchases compared to other segments.

Modeling Result - RFM Quantile

Then as much as **3.9%** of total customers (with a total of 31 customers) are classified as **Lost Customers** which is almost the same as customers in the Lost Cheap Customers segment, Lost Customers tend to churn.

A total of 30 customers with a percentage of **3.78%** of the total customers, entered the **Best Customer** segment. Customers in this segment have recently purchased goods from the company, have the most shopping frequency, and have the most total purchases than other segments.

Then **0.25%** of the total customers (with a total of 2 customers) are customers who enter the **Almost Lost** segment. Customers in the Almost Lost segment have recently purchased goods from the company, but have quite a bit of shopping frequency, and the lowest total purchases compared to other segments.

Lastly, other customers are classified into **Others segment**, where this segment is a **mixed segment** (in this case **undefined segment**).

THE RESULT

01

Result 1

- The total of customers is 793. The number of Order_id turns out to be not the same as the number of Customers, so there are some Order_ids that have the same Customers.

- This shows that there are customers who repeat orders or do not only shop once at the company. order_date shows that there are several transactions on the same day

02

Result 2

A

Average Purchase

The average purchasing amount from the customers is 458.626672 bucks.

B

Largest Purchase

The highest purchase was rated 23661 bucks.

C

Lowest Purchase

The lowest purchasing amount is 1 bucks.

Result 3

3.1

We start the analysis **in 2011**, **the number of items purchased tends to increase every month**. In **October sales declined sharply**, **then increased in the following month**, even **exceeding sales in September** because normally In most countries in the Northern Hemisphere, September is the time for students to start the new school year at school. **November's sales are high** because there are usually lots of Black Friday promotions. Meanwhile, sales in December experienced a decline again.

Result 3

3.2

Next to **2012**, same as 2011 the transaction **tends to increase every month**, from the EDA **in February** to March there is a fairly sharp **increase of customer purchases up to 2 times**. After that, in September there was a sharp increase. However, in the following month, **October, there was a sharp decline** because, in that month, the consumer price index (CPI) was raised based on the Bureau of Labor Statistics (BLS). in November there was a high **increase even higher than in September** and again probably increased because of Black Friday, and sales in December fell again

Result 3

3.3

Similarly **in 2013**, the number of items purchased **tends to increase every month**, for the February to March there is a fairly sharp increase to almost 2 times. In the month of **September, there was a sharp increase** probably because many parents bought school supplies for their children. However, in the following month, **October, there was a sharp decline** probably due to the raise of CPI based on the BLS. However, in November there was a significant increase in the number of items purchased because Black Friday promos usually occur and again slightly decreased sales in December.

Result 3

3.4

Lastly for **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 is a sharp increase** in the new school year. In the following month, based on BLS the CPI was an increase, so that **October, there was a sharp decline**. However, in **November there was a high increase** in the number of items purchased **even higher than in September** due to Black Friday and again slightly decreased sales in December

Result 4

4.1

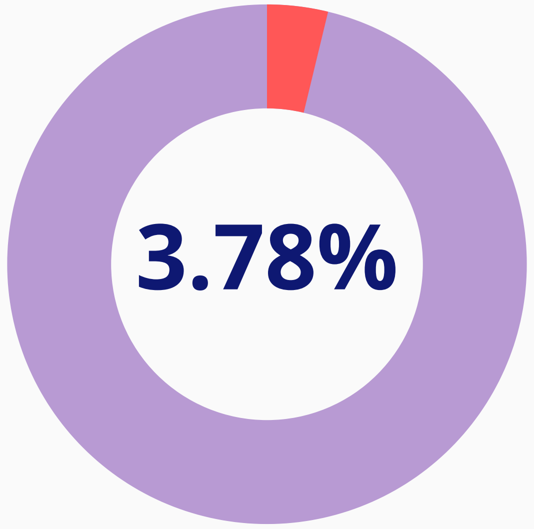
Based on results of modeling customer segmentation using RFM, the company's customers are **divided into 7 segments**

- **Best Customers**
- **Loyal Customers**
- **Big Spenders**
- **Lost Cheap Customers**
- **Lost Customers**
- **Almost Lost Customer**
- **Other Customers**

4.2

Result 4

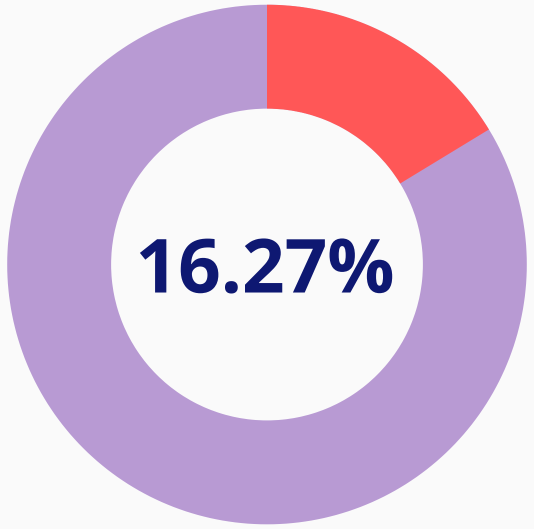
Best Customers



A total of 30 customers with a percentage of **3.78% of the total customers,** are cateogrized as **Best Customer** segment. Customers in this segment have **recently purchased goods** from the company, have the **most shopping frequency**, and have the **most total purchases** than other segments

---

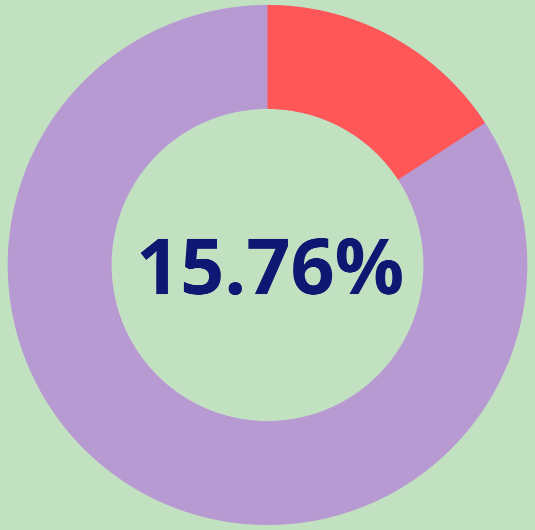
Loyal Customers



A total of **16.27% of the total customers** (with a total of 129 of 793 customers) are included in **Loyal Customers** where this segment's customers have the **most shopping frequency** than other segments.

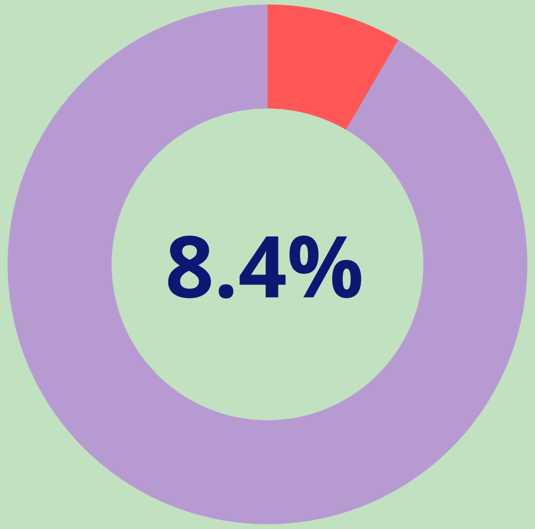
Result 4

Big Spender



Furthermore, **15.76% of the total customers** (with a total of 125 of 793 customers) in the company entered the Big Spenders segment. **This segment has the highest total purchases** than any other segment.

Lost Cheap Cust.

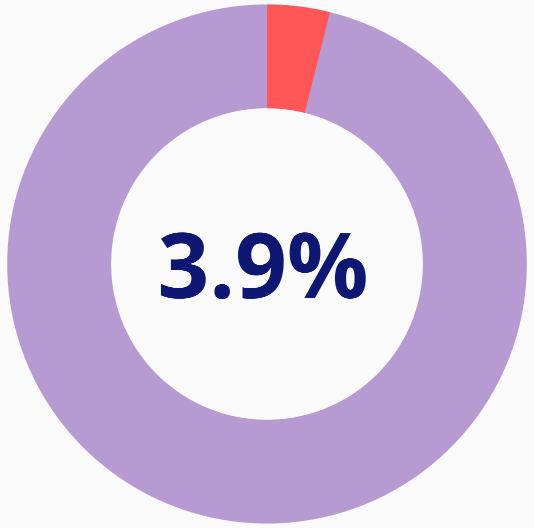


Another segment is the Lost Cheap Customers segment with a percentage of **8.44% of the total customers**. This type of customer segment highly **tends to churn** because they have **not bought goods** from the company **for a long time**, have the **least frequency of shopping,** and have **the least total purchases** compared to other segments.

4.4

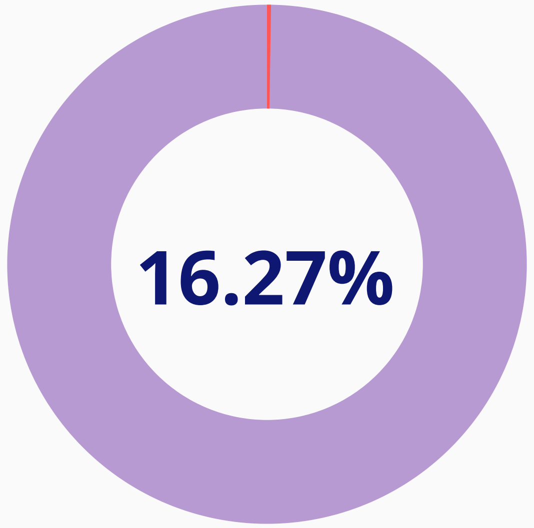
Result 4

Lost Customers



As much as **3.9% of total customers** (with a total of 31 of 793 customers) are classified as **Lost Customers** which is almost the same as customers in the Lost Cheap Customers segment, **Lost Customers tend to churn.**

Almost Lost Cust.

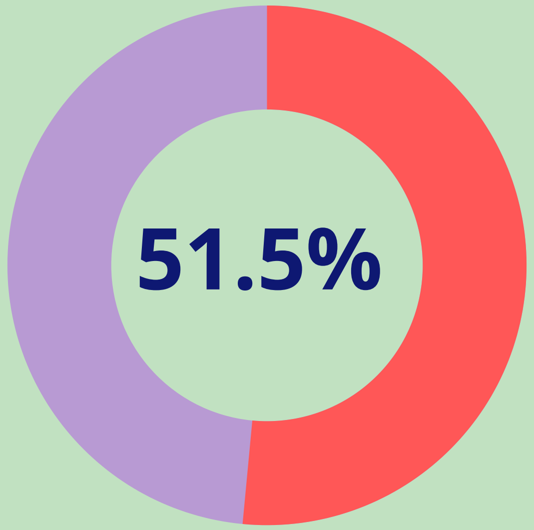


Then **0.25% of the total customers** (with a total of 2 of 793 customers) are customers which are categorized as the **Almost Lost segment**. Customers in this segment have **recently purchased goods** from the company, but have **quite a bit of shopping frequency**, and the **lowest total purchases** compared to other segments.

4.5

Result 4

Other Customers



Lastly, other customers (409 of 793) are classified into Others segment, where this segment is a mixed segment (in this case undefined segment).

THE

RECOMMENDATIONS

Action Item / Recommendation

01

Recommendations for **increased orders in a particular month** such as September and November: **focus on increasing the amount of supply** of goods to compensate the high demand. Then for a **decrease orders** in a given month such as October, the marketing team can focus on improving the company promotion strategy. The company **could also suppress the goods price as competitive** as the company can or even slightly below the market price.

Action Item / Recommendation

02

The following are the Recommendation for each customer segment.

For the Best Customers segment:

- The company **can focus on the reward for** this type of customer, such as **posting their picture/name** in the **store or social media** as the "Best Customer" and **granting the reward for them**. The reason is the **best customer tends to do repeat orders as usually** they did. In addition, **we also need to take into account** whenever the best customer **transaction behavior change** in term of a **decrease of purchase**. The company can raise their profit from implementing cross and up selling technique to the customer.

Action Item / Recommendation

03

Next to **Loyal Customers segment**, the company must **optimize services** for this customer in order to **maintain their loyalty** and increase their buying confidence. The company should **increase its engagement** with the loyal customer, such as **remembering their name** and **asking for feedback** through interview sessions and grant a reward for that.

Action Item / Recommendation

04

For the Big Spenders segment:

- the business team should focus on doing **research on which products match the big spender customer interests or products that they are buying most**.
- From that, the company can **offer special membership**, **so that they will do repeat orders** even though not so often.
- On the other hand, the company can also **offer products** with brand new **high specifications** or **top-level products** and also apply cross and up selling technique to them.
- Lastly, the big spender customers should be given good service, such as **giving a special offering** for the new and exclusive product from their **favorite media platform**.

Action Item / Recommendation

05

The next recommendation is for the **Almost Lost** segment. Since this customer segment is recently purchased, the company can **offer a high discount** or promo, **for the next one or two purchase**, based on their last purchase interest group. **This**, hopefully, **can intrigue them to purchase again** from this company. The offering could be **delivered by the cashier employee with some brochures or vouchers**.

Action Item / Recommendation

06

Lastly for the **Lost Cheap Customers** and the **Lost Customer** segment. The company should d**o some research on this customer's background**. While probably the **store distance** from their houses is **quite far**, then the company can **implement *delivery services or online shopping services**. Those strategies **will** not only **affect** the lost customer but also the **other segment of customers**. Because we are **providing an easier way** to purchasing goods.