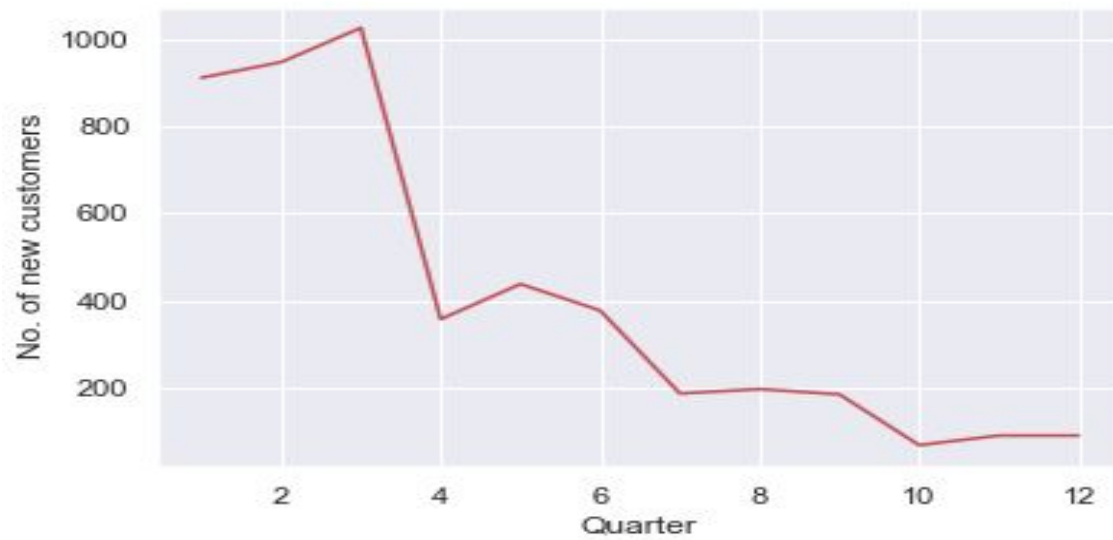


Question 1:

Trend line:



Data:

Quarter	No. of new customers
1	913
2	949
3	1028
4	357
5	438
6	377
7	186
8	196
9	184
10	67
11	89
12	89

From the trend line, we can see

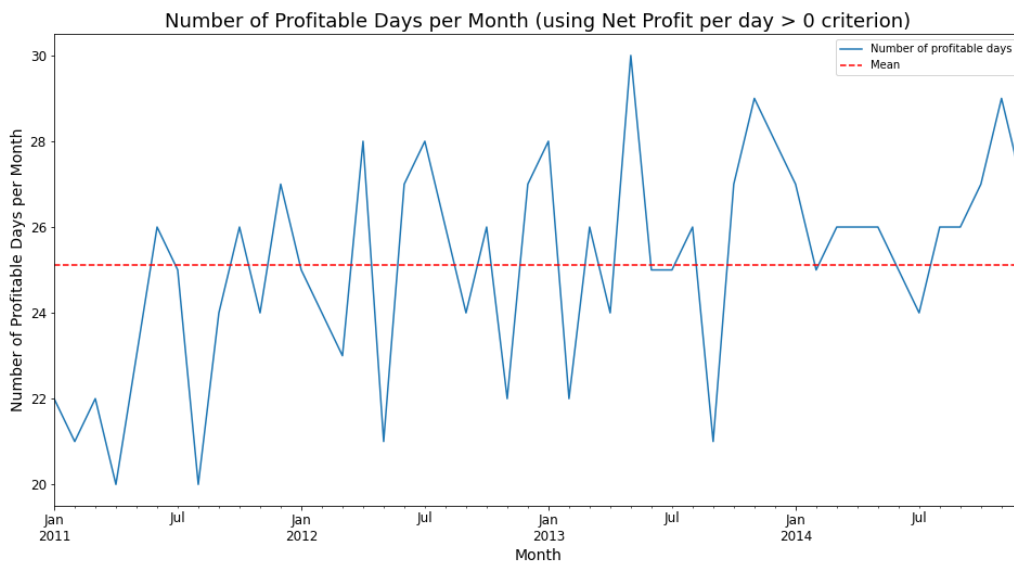
- 1) The number of new customers reached the maximum during the third quarter
- 2) There was negligible change in the number of new customers during (7 - 9) and (10 - 12) quarters. The curve is flat

Note:

- 1) Each year is divided into four quarters (each quarter - three months). For a period of four years, we have 12 such quarters. Each instance is then classified to the respective quarter corresponding to the order date
- 2) Since each customer ID is repeated, we remove the duplicates keeping only the first occurrence of an ID. This way, we can concentrate only on the new customers

Question 2:

When it comes to measuring profitability, the main feature that comes into picture is **Profit**, ofcourse. We can sum up the profit in a day, and call a day profitable if the net profit made that day is positive. And then, we can **sum up all such days with a positive net profit within a month to get the number of profitable days** within a month.



The result is also stored in a csv file, a peek into which looks like :

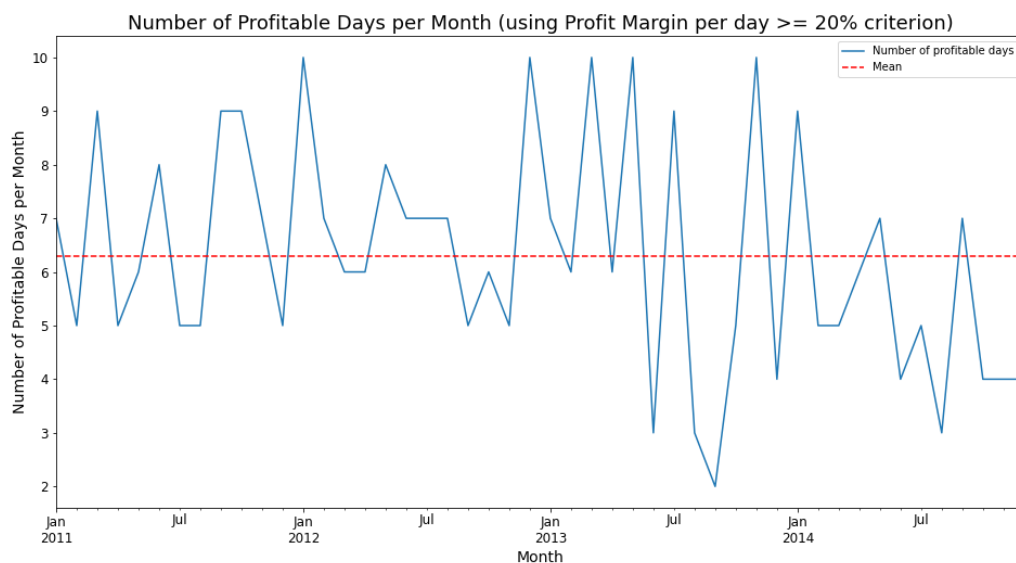
```

Order Date
2011-01-01    22
2011-02-01    21
2011-03-01    22
2011-04-01    20
2011-05-01    23
Freq: MS, Name: Number of Profitable Days per Month, dtype: int64

```

Another way of looking at profitability in more detail is via **Profit Margin**, which is Net Profit per unit total sales made. It gives a better indicator of how well we have made profit for a day from our sales irrespective of the total sales done.

Earlier we considered every day with a positive net profit as profitable, now **we consider a certain threshold above which we consider a day profitable**, as not every day even with a positive net profit is as profitable (if the threshold is zero we would get the same solution as with the earlier method). As an example, when we make a large number of sales with less profit on each, the total profit might be positive, but it may not be a great day for the business. We use this to get the number of profitable days in a month and **we use a threshold that Profit Margin \geq 20% implies a profitable day**.



This result is again stores in a csv file, a peek into which looks like :

```

Order Date
2011-01-01    7
2011-02-01    5
2011-03-01    9
2011-04-01    5
2011-05-01    6
Freq: MS, Name: Number of Profitable Days per Month, dtype: int64

```

Question 3:

	Discount	Sales	Profit	Recency	Monetary	Frequency	RFM
YOJ							
2011	0.142203	232.712907	27.894900	2.731001	2.962606	2.493366	2.728991
2012	0.145503	241.071575	20.524100	2.734151	3.084867	2.890593	2.903204
2013	0.146173	233.201070	29.503265	2.712538	2.981651	2.983180	2.892457
2014	0.144557	240.499093	31.348733	3.391410	2.960073	3.293406	3.214963
2015	0.120286	215.723810	5.235923	5.000000	3.342857	3.590476	3.977778

“YOJ” denotes the year of first order which decides the cohort a customer belongs to.

The cohorts are compared using the following parameters: **Mean discount** offered to them (per transaction), **mean sales** (per transaction), **mean profit** (per transaction) obtained and **mean RFM score**.

We can see that the discount offered and sales is more or less similar across cohorts. Cohort 2011 and 2014 are involved in more profitable transactions, while 2015, despite having less discount are involved in less profitable transactions.

The Recency and Frequency scores of 2015 are high, but going blindly by that is misleading, since 2015 is the latest year in the dataset.

We can argue that after a few years 2015 customer's recency score will reduce, and they haven't been profitable. So we can go with 2012 or 2014 who have been users of the company's products for a longer time, and have also contributed to the company's growth (profits) more than others.

On comparing 2012 and 2014 cohorts, we conclude that **we should run the special campaign to the 2012 cohort since they have been consistently loyal to the brand for a long time.** Here are few other factors that solidifies our decision:

1. **New customers are that much harder to convince** as they have very little dealing with the business. This means not only making them aware of the brand and business, but also needs a comprehensive marketing strategy to attract them.
2. **Older customers know about the company's products and appreciate them.** They are also a lot easier to convince to try our newer products/rope them back into the fold if they leave.
3. **Loyal customers spend more** on the company's products and services, and **tend to share positive reviews** about them to their friends and family. This will save a lot of advertising costs.

Hence, we conclude that, 2012 cohort deserves the special campaign. It will result in a profitable outcome for both the company and its customers.

Question 4:

Firstly, we note that there are 3 types of products sold :

- Furniture
- Office Supplies
- Technology

It is better to **analyse and forecast the sales of each type of product separately** as they may have different kinds of demand and supply in different months.

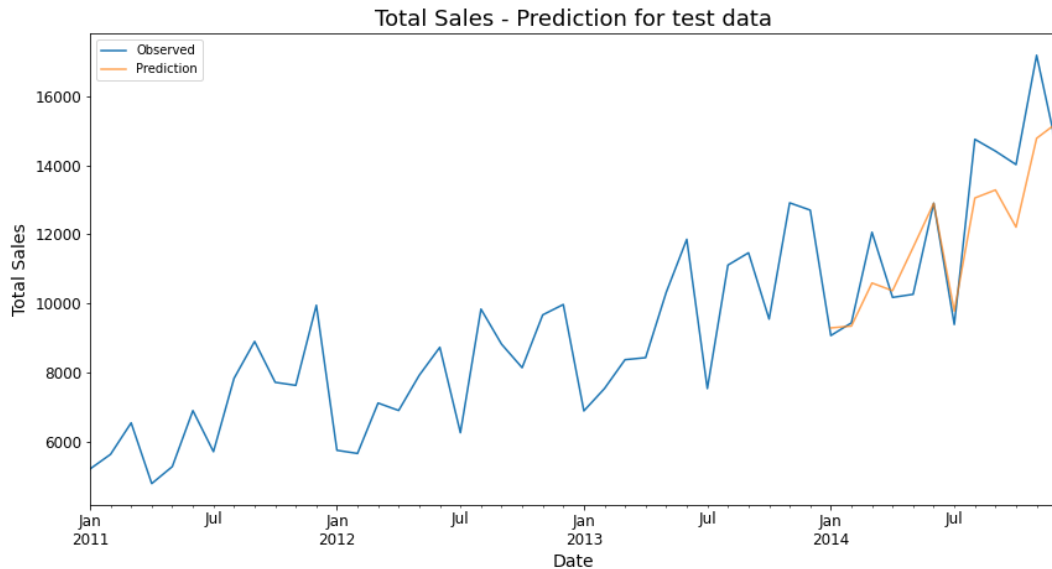
We add up the sales within a day to get day-to-day sales data for the company. However the daily sales data is random and we won't be able to make out the main smooth variations from it, to be able to predict future sales. Hence we **resample them by taking their mean over a month and get monthly sales and try to forecast the monthly sales for next two months.**

We do a decomposition of the individual sales to get the trend, seasonal and residual components and notice that all three sales do have an upward trend, and some seasonality as well. So we **use the SARIMA model to forecast the monthly sales of Furniture, Technology and Office Supplies.** We plotted the ACF and PACF plots and tried out various parameters for the SARIMA model iteratively (limited by the data available, high parameters couldn't be used) and arrived at :

- SARIMA(0,1,1)(0,1,1)₁₂ for Furniture sales which had the lowest AIC of 541.7727 . For the test data we considered sales from Jan 2014 till the last month available, and we got a RMSE of 502.6381 and the ground truth values lied within the confidence interval from predicted value.
- SARIMA(0,1,1)(0,1,1)₁₂ for Furniture sales which had the lowest AIC of 557.616 . For the test data we considered sales from Jan 2014 till the last month available, and we got a

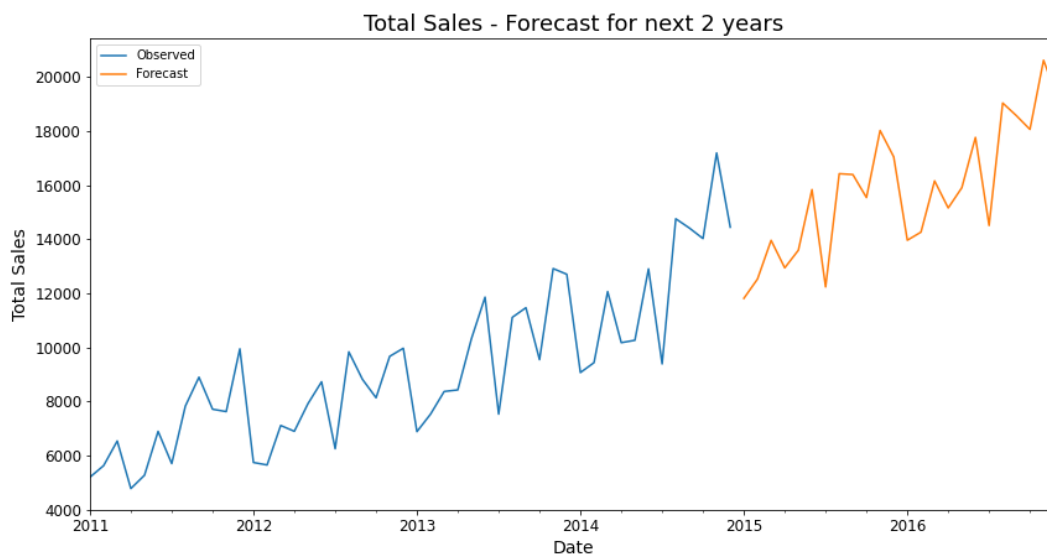
RMSE of 723.741 and the ground truth values lied within the confidence interval from predicted value.

- SARIMA(0,1,1)(1,1,0)₁₂ for Furniture sales which had the lowest AIC of 530.0924. For the test data we considered sales from Jan 2014 till the last month available, and we got a RMSE of 456.8059 and the ground truth values lied within the confidence interval from predicted value.



The RMSE between the total sales predicted from adding the predicted individual product sales and the total sales data we have is obtained to be 1228.2412 .

We then **forecast the monthly sales of the three products for the next 2 years (24 months)** and **add the individual sales of the three products to get the total sales forecast (monthly)** for the next 2 years.



The forecasted monthly total sales for next 2 years is stored in a csv file and a peek into it looks as follows :

```
Order Date
2015-01-01    11812.824130
2015-02-01    12527.413832
2015-03-01    13958.394094
2015-04-01    12940.336531
2015-05-01    13592.784110
Freq: MS, Name: Monthly Sales Forecast, dtype: float64
```

Question 5:

Top 5 countries with the highest CAGR:

Rank	Country	CAGR
1	Mauritania	1110.49
2	Slovenia	760.13
3	Uruguay	412.74
4	Croatia	356.54
5	Tajikistan	330.28

Top 5 countries with the highest CAGR (after removing the bottom ten percentile transactions):

Rank	Country	CAGR
1	Mauritania	1104.40
2	Slovenia	760.13
3	Uruguay	412.74
4	Croatia	353.71
5	Burundi	325.49

Conclusion:

Even after removing the bottom ten percentile transactions from the original data, the top 5 countries ranking does not change much

Note:

- 1) **CAGR** = $((F/I)^{(1/\text{years})} - 1) * 100$ (Compound Annual Growth Rate)
where, F - Total sales in final year, I - Total sales in initial year
- 2) Most countries had sales in all the four years (2011 - end of 2014). But some countries did not have sales in some of the years and hence, the CAGR is calculated accordingly
- 3) A CAGR value of zero shows that the particular country performed sales only in the final year

Question 6:

A customer is identified using his/her customer ID.

Recency:

This is determined by grouping the data on the basis of customer ID and determining the 'maximum' date of purchase. This gives the most recent date at which the customer has bought a product of a company.

Frequency:

Frequency is determined as (number of purchases by the customer) divided by the total number of days since the first purchase (minimum order date for a given customer). First purchase is determined as the purchase with the latest (maximum) order date.

Monetary:

Monetary value of a given customer is simply the sum of sales of the customer through all transactions.

Score of a customer for each of these verticals depends on his position in the table. Top 20% get a full score of 5, customers in the 20-40% range get 4 and so on, with the bottom 20% getting a score of 1.

Overall score of each customer is simply calculated as $(R+F+M)/3$

The RFM score of each customer is given in the csv file 'q6_RFM_ANALYSIS.csv' and the code is in q3_and_q6.ipynb

$152/4873 = 3.2\%$ of the customers have a perfect RFM score of 5.

26.28% of customers have an RFM score of at least 4

Question 7

Recommendations:

01. Cohort 2012 should be rewarded with the special campaign to cement their loyalty and at the same time, to use them to grow the sales.
02. From the sales graph as well, we can see an upward trend in the sales over the years. The past sales data are an indication of the increasing demand for Furniture, Office supplies and Technology products as years progress. Hence, the company should ramp up production capacities to meet the increasing demand, as is predicted by the sales forecast as well. Also there is a clear seasonality in sales within a year, where the sales are high in the beginning of the year, reduces in the middle months around July, and again rises during the end of the year. This observation can be used to increase sales and profit, by changing the production accordingly and reducing inventory and operating costs.
- 03.

Quarter	Total number of transactions
1	2419
2	2849
3	3730
4	2797
5	3729
6	4436
7	3347
8	4923
9	5529
10	4713
11	5615
12	7603

From this table, we can see that the total number of transactions the customers make steadily increases (Each year is divided into four quarters (each quarter - three months). For a period of four years, we have 12 such quarters. Each instance is then classified to the respective quarter corresponding to the order date). But from the trend line plotted in Q1, it is intuitive to infer that after an initial spike, the company does not seem to attract new customers. Eventually, the trend line becomes flat at the bottom. Hence, it is easy to note that though the company's new customer acquisition strategy needs refinement, most of the acquired customers are very loyal which results in the increase in the number of transactions. A strong suggestion would be to use the customer's loyalty to acquire more new customers via a '**Referral marketing strategy**'. This will definitely expand the company's customer base, thereby increasing the profits.

04. Though Mauritania, Slovenia and Tajikistan have higher CAGR, the company is relatively new to these countries with fewer transactions. Especially, Slovenia just had three transactions and they were made only in the final three years. On the other hand, Uruguay and Croatia seem to be very promising as there are more transactions and it also enjoys an increasing trend in each passing year. The risk associated with Mauritania, Slovenia and Tajikistan is high as the sales in each year fluctuate high and low. Hence, the company should leverage this advantage and look to expand its market in Uruguay and Croatia.
05. To maintain the increasing sales and demand, the company needs to ensure they keep up the quality of products produced and the services offered to customers always (services can range from continuing to offer discounts as we see in most sales from the data to customer campaigns being organised).
06. The number of days in a month with total profit > 0 has been quite high all the years, but we can observe they have been particularly high in recent years. But the number of days in a month with Profit Margin $\geq 20\%$ has been becoming lesser in the later years (recent years). This means that most of the high profits recently come from high sales, and not because of better sales or marketing strategy. Hence they need to improve on their marketing strategy and sales technique to properly capitalise on improving demand for their products.