SmartBox Analytics Test

Task 1

In this section, we will take a look at the sales obtained by CraicBox company from the beginning of 2018 until mid-December 2019 as well as the forecast figures and compare them.

For the dataset, we have 4 columns 'Date', 'Product Name', 'Daily Sales' and 'Daily Forecast.

Date	Product Name	Daily Sales	Daily Forecast
01/01/2018	Product A	100,991	93,000
02/01/2018	Product A	113,184	112,000
03/01/2018	Product A	118,390	121,000
04/01/2018	Product A	152,053	153,000
05/01/2018	Product A	164,354	156,000
06/01/2018	Product A	113,929	115,000

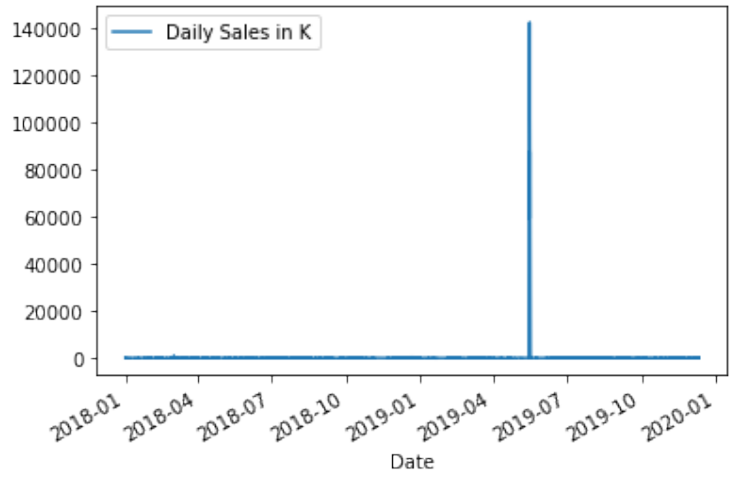
Product Types

From the dataset, we can see that the CraicBox sold 7 products in this period which are:

Product Name
Product A
Product B
Product C
Product D
Product E
Product F
Product G

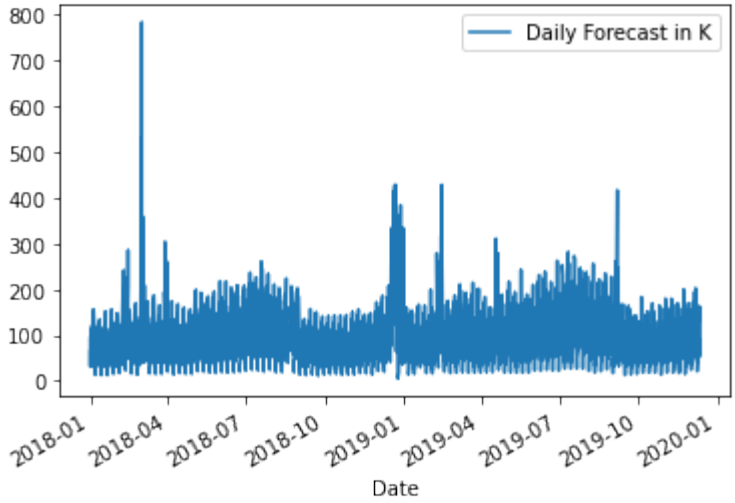
2018 and 2019 Sales

Let's now see how the sales changed over 2018 and 2019:



We can see from the figure that the sales are almost the same around 100K for all days except for a day in May 2019, there was a spike with around 140 Million.

Let's now see the Forecast amounts during the same period:



We can see that the forecast for most of the days vary between 100K and 200K except for few spikes.

Data Cleaning

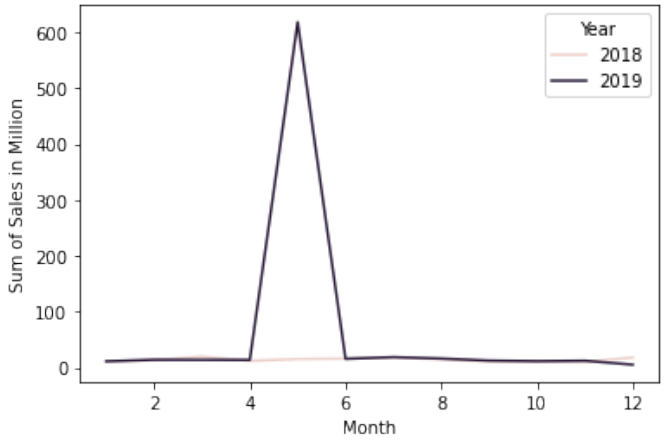
To have a better look on the data, I will check the spike details and then remove it to have a better understanding and analysis.
After checking the abnormal values in the dataset, we can see that the numbers on 16th May 2019 were huge compared to any other day in the years 2018 and 2019.

Date	Product Name	Daily Sales	Daily Forecast
2019-05-16	Product A	87696960.0	132000
2019-05-16	Product C	119558400.0	140000
2019-05-16	Product D	142506000.0	136000
2019-05-16	Product E	59529600.0	130000
2019-05-16	Product F	91728000.0	143000
2019-05-16	Product G	102648000.0	145000

I will remove those records to make better analysis on the data and return to them if needed.

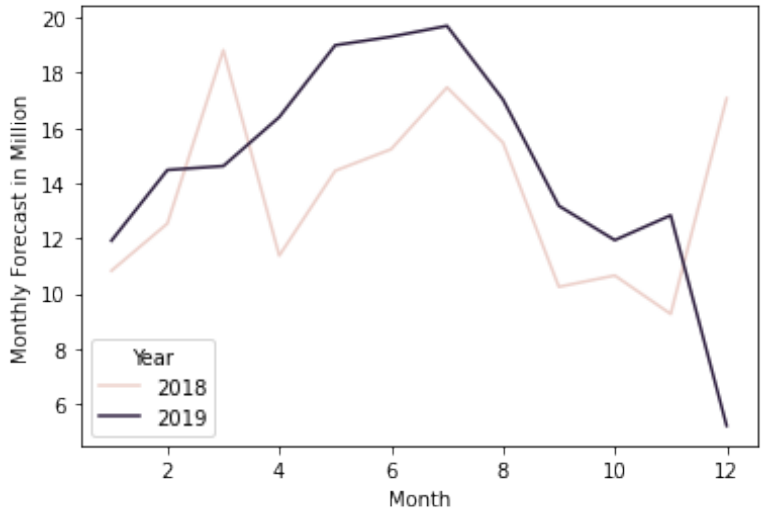
Monthly Aggregation

Let's now have a more aggregated look to check the sales per month for 2018 and 2019.



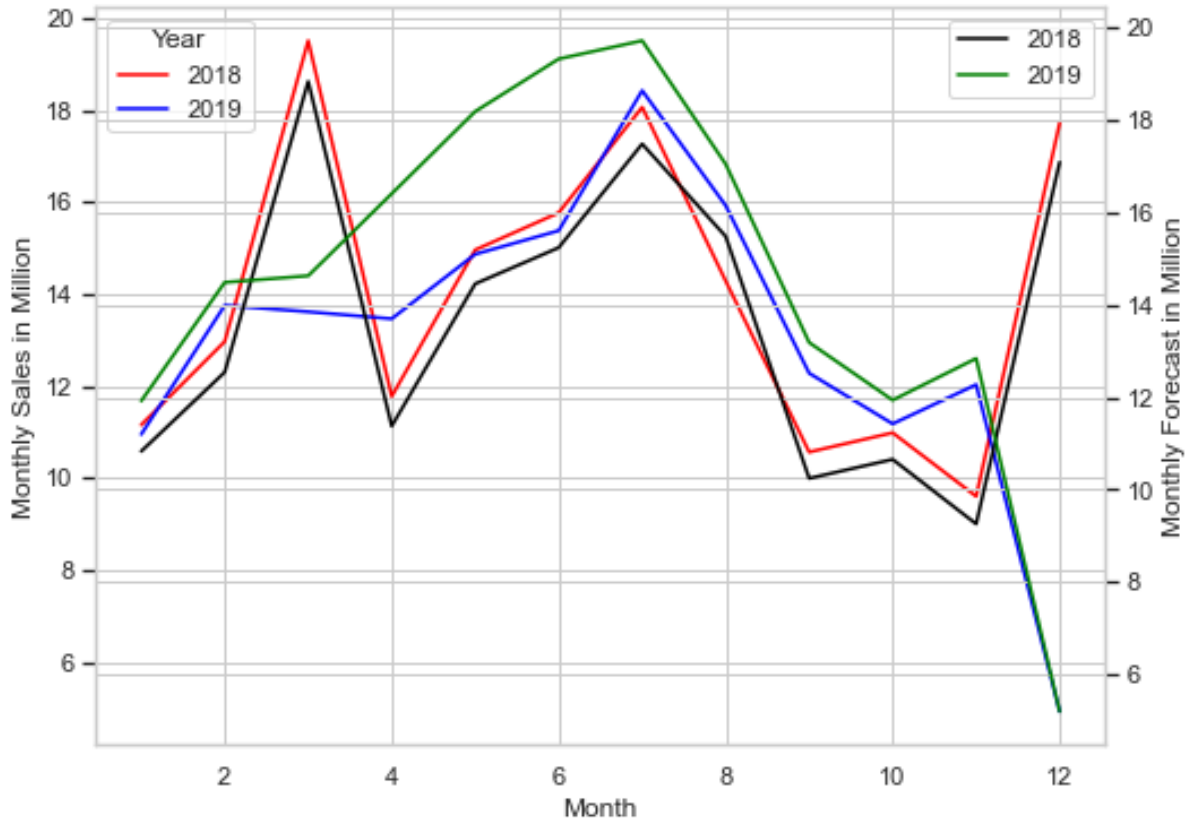
From the above figure, we can see that the sales for 2018 are between 10 and 20 Million for all months. However, in 2019 the sales got a big spike in May 2019 for around 600 Million due the huge numbers on 15thof May discussed earlier.

Let's check the Monthly forecast for 2018 and 2019:



From the above figure, we can see that there is no big difference between the forecast for each month of years 2018 and 2019 except December that there is a huge gap between the two years. We can explain thig huge gap by the data as the dataset contains only the sales and forecast for the first half of December 2019 while the whole month for December 2018.

Now, let's have a completed look on the data after removing the spike. The below figure contains the sales and forecast for 2018 and 2019.



From the above figure, we can have some conclusions:

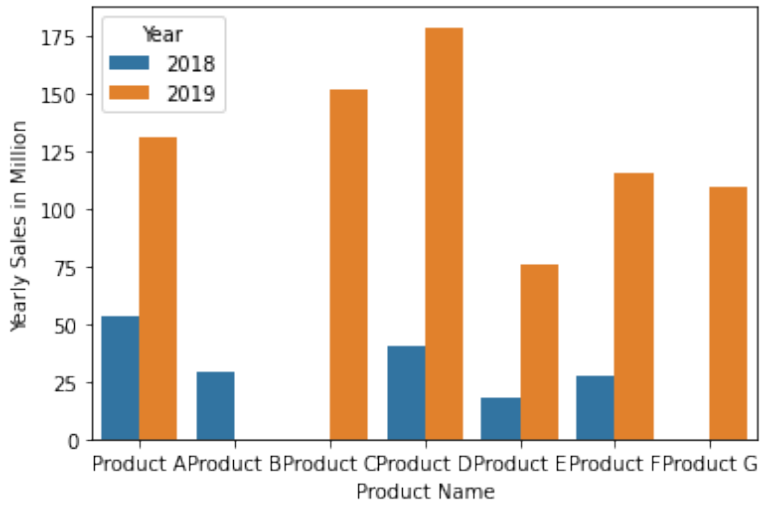
- The Sales and Forecast are higher in Summer months from June to August
- The Sales in 2018 exceeded the sales in 2019 only in March but the 2019 sales is greater for all other months.
- The Sales and Forecast in 2018 have the same trend and the sales exceeded the forecast for all months except August.
- In 2019, the Forecast is greater than the Sales for all months which means none achieved the target. If we included the spike that we removed earlier, May 2019 should achieve the target and the sales will be higher than the forecast.

Yearly Aggregation

The below table shows the sum of sales in 2018 and 2019 after removing the spike. If we took the spike into consideration then the sales in 2019 will be achieved.

Year	Yearly Sales in Million	Yearly Forecast in Million
2018	167.355109	163.418
2019	156.797546	174.784

Sales Per Product

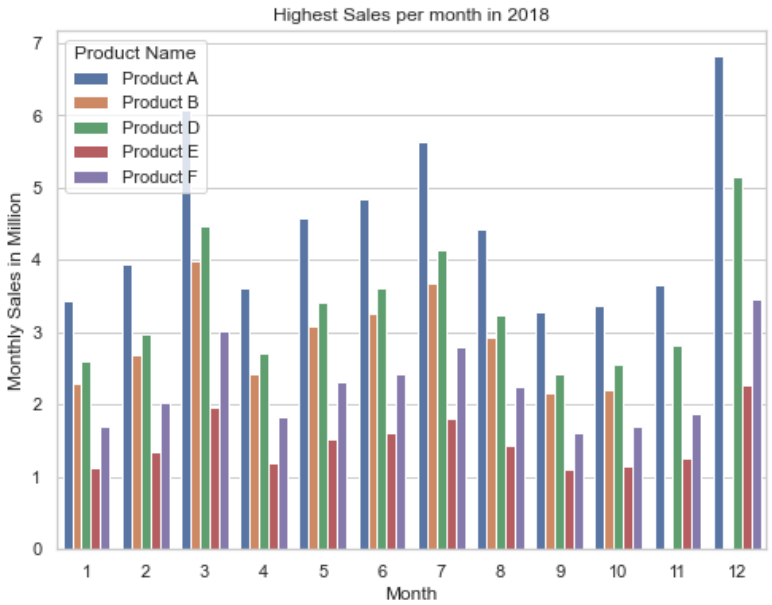


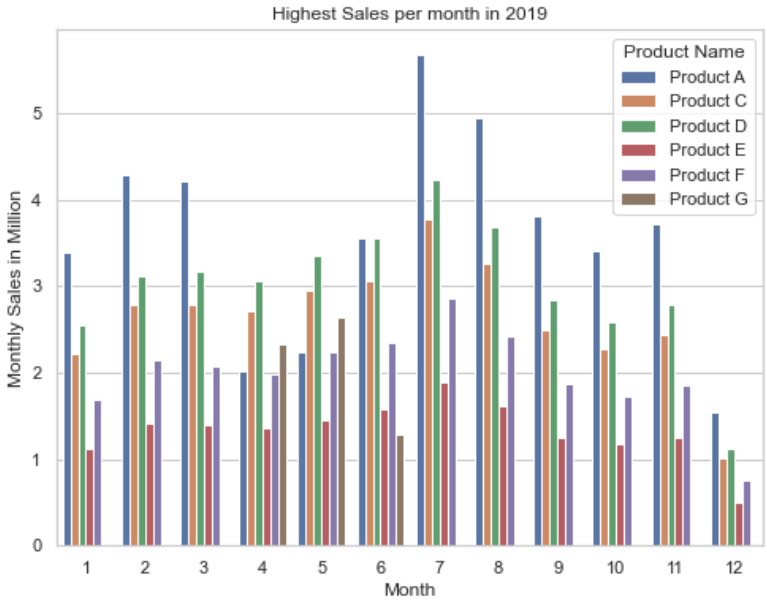
From the above figure, we can see that Product C and Product G were not sold in 2018. As well as, Product B was not sold in 2019.

For the product sold in both years, we can see that the sales of 2019 exceeded the sales of 2018 for each product.

In addition, Product E has the least sales in both years. Product A has the highest sales in 2018 while Product D has the highest sales in 2019.

What is the most sold product per month?





From the above figures, we can see that Product A is the most sold product in each month of 2018. While Product A and Product D interchanges the maximum sales in months of 2019.

Target Achievement

Now, let's check if the target or forecast values have been achieved during 2018 and 2019. We checked the sales of each product aggregated monthly and we concluded that none of the products sales have achieved the forecast value in any month in 2019.
For 2018, only August sales didn't achieve the forecast values for all products. Other than August, all sales have exceeded the forecast for all months of the year.

If we included the spike that we removed earlier, May 2019 should achieve the target and the sales will be higher than the forecast for all products.

Date	Product Name	Daily Sales	Daily Forecast
2019-05-16	Product A	87696960.0	132000
2019-05-16	Product C	119558400.0	140000
2019-05-16	Product D	142506000.0	136000
2019-05-16	Product E	59529600.0	130000
2019-05-16	Product F	91728000.0	143000
2019-05-16	Product G	102648000.0	145000

Conclusion

I have created a jupyter notebook to make the data analysis and create the above visualizations. Please check it here. ()

In addition, I have created a Power BI Dashboard with all the visualizations of Monthly, Quarterly and Yearly aggregation in the two attached files (Power BI export and PDF version).

Task 2

Question

What you got in Task 1 was aggregated daily sales by Product. Assume now that you work for CraicBox and have access to company's raw data (transactions, master data etc.)
Since CraicBox is a data-driven company, it records all relevant information about its sales, products, shops and customers.
Your task is to build a forecast for next year.
How will you do it? What data (internal and/or external) would you need to produce an accurate forecast? What assumptions will you need to make?

Answer

After checking the data from Task 1, there is no explicit equation or pattern to calculate the forecast from 2018 to 2019. However, to think about building a forecast, there are main points to be taken into consideration:

- Product Types sold each month in the previous year. This can give an idea of continue producing the product or stop selling it.
- The number of products sold monthly and yearly and the sales produced from each product to calculate the forecast for each product per month. For example, if a product is sold more during a certain month then the forecast of this month for that product will be higher than the other months for the same product.
- The shops and regions that sell the products to check if a certain product is sold more in a specific region or country or shop.
- The types of customers purchasing the products. If a certain product is only purchased by a specific category of people.
- The relation between the main events or occasions in the country selling the product and the sales achieved.

The forecast then will be calculated based on the above points and should be higher than the previous year with some percentage to be decided by the company strategy. Also, the forecast might be lower for next year for the products that didn’t achieve the target or if the company would like to stop producing it.

Task 3

Question

Recently there is a surge in the number of inbound calls to CraicBox Customer Care line. Customer Care director asked you to understand the reason for the spike in the calls and present your findings and recommendations.
How will you approach this request? What things will you check? What might be your recommendations? Please describe your thought process in a structured way.

Answer

First of all, we should check the reason for that spike of received calls which might be one of the following reasons:

1- Technical issue like an issue on the company's website.
2- Complains
3- New Requests like a Request to buy
4- Respond to a Marketing campaign

Then, we should have a report for all the calls received and check if all or most of them have the same reason.

If the reason is a technical issue, we should escalate the problem to the technical team to solve the root cause in an acceptable time frame. The company should have a statement that can be seen by all the customer that the problem is addressed and is being resolved with a timeline.

If the reason is a complain, we should check the circumstances of that complain. If it is related to a certain shop, product, delivery…etc. The, we should dig deeper to know the reason for the complains and solve the problem.

If the reason is a Request to buy, we should check if that happens in a specific time of the year for example Christmas each year. If so, the company should hire more agents during that period to be able to manage that huge number of calls.

If it is related to a marketing campaign, then the marketing team should have a plan for managing the requests before releasing any new offer.