



BI SOLUTION PROTOTYPE FOR LUPITA'S CAFE



Module Code: CS3DV20

Assignment report Title: CS3DV20 - Coursework

Student Number: 27001654 27020363

Date: 14/03/2022

Actual hrs spent for the assignment: 35

Assignment evaluation:

1. Helpful to learn what is needed in order to create good graphs
2. Helpful to make me try and find innovative ways for visualization
3. There are many deliverables, so we think it would be better if more information were given on what is needed on each one. For example, the information on the presentation is very little.

Contents

| | |
|--|----|
| Executive summary | 3 |
| 1. Explanatory Data Analysis | 4 |
| 1.1 Quick analysis | 4 |
| 1.2 Important facts about the data | 4 |
| 1.3 Listing more facts | 5 |
| 2. Dummy Data creation | 5 |
| 2.1 Findings on impact | 5 |
| 2.2 Overview of the findings | 6 |
| 2.3 The process of creating the 2020 data | 6 |
| Step 1: Creating entries for 2020 | 6 |
| Step 2: Preparation for changing the values of the 2020 data | 7 |
| Step 3: Perform the calculations | 8 |
| Step 4: Concat the 2020 data with the 2018 and 2019 data and export them | 9 |
| Step 5: Testing for issues | 9 |
| 2.4 Assumptions | 10 |
| 3. Layout design, storytelling reasoning, KPIs, assumptions, storyboard | 10 |
| 3.1 Identify your audience | 10 |
| 3.2 Document Goals | 10 |
| 3.3 Document Key Performance Indicators | 10 |
| 3.4 Dashboards and Story | 11 |
| 3.4.1 Section 1 – Current State | 12 |
| 3.4.2 Section 2 – Trends | 14 |
| 3.4.3 Section 3 – Forecast | 15 |
| 3.4.4 Section 4 – What if and suggestions | 16 |
| 4. Conclusion | 17 |
| 4.1 Summing up the report | 17 |
| 4.2 Lessons learned | 17 |

Executive summary

This report was commissioned to exhibit how a business intelligence solution could benefit Lupita's Café and is targeted in C-level and managerial level management.

Firstly, some explanatory data analysis was performed that revealed helpful information about the given data, for example, that all the stores are in the UK and how the product lines are clustered.

Afterwards, dummy data for 2020 were generated based on the 2019 data, taking three facts that were discovered during research into account

- *Sales were increased by 21%*
- *Marketing was decreased by 17%*
- *Inventory was decreased by 221%*

Moreover, two dashboards were created that aimed to show the current state, the business trends, forecasts, and recommendations on how to get on the right track.

Generally, the dashboards highlighted the most profitable markets and products. They discovered that the budgeting skills of the company are excellent except for a consistent trend in the summer to miscalculate the budget. Also, it was discovered that the company had the trend to do very well in marketing spending, as sales always followed the marketing spendings line.

Finally, it was explained that the business is stagnating, as its profits have not changed for the past three years and taking inflation and price gouging into account, this is not sustainable.

The following recommendations were given in order for the business to get on the right track:

- Investigate the budgeting issue after summer.
- Use the marketing skills they possess in a way that highlights and helps the key markets and the key products.

1. Explanatory Data Analysis

In this section, facts about the data will be listed. Since the requirements are only for a .knwf file and two dashboards, the analyses used can be found in the appendix folder provided in the zip.

1.1 Quick analysis

There are no missing values (see knime/workflow-1) and duplicate values (see knime/workflow-2), so no pre-processing is required.

The basic statistics can be seen in figure 1 (see knime/workflow-1)

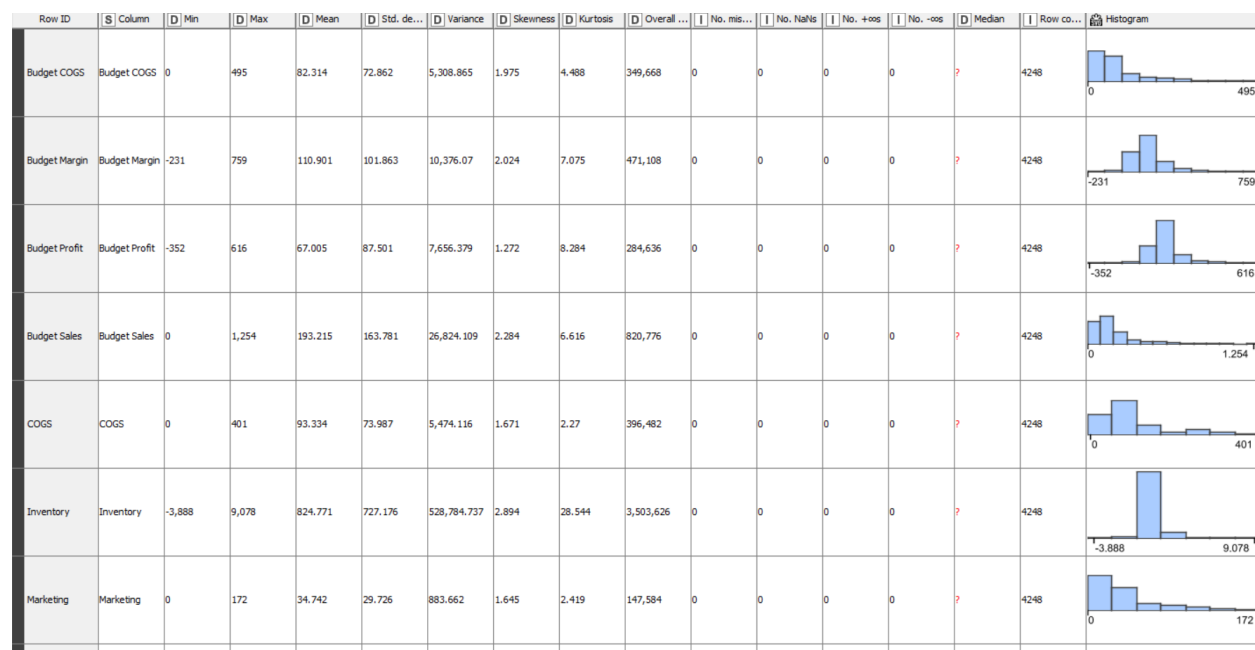


Figure 1 Basic statistics

1.2 Important facts about the data

The country field is always 'United Kingdom' (see tableau/EDA.twb), so it is clear that the company operates only in the UK.

Coffee products always have as product line 'Beans' and tea products have 'Leaves'. Moreover, all types of coffee (Coffee and espresso) can either be 'decaf' or 'regular', but in teas, the 'herbal tea' is always of type 'decaf', and the 'tea' is always 'regular' (see workflow 5-8).

Finally, it was found that not all the cities sell all the products. Figure 2 displays that. The white boxes mean that this city does not sell the specific product (see /tableau/EDA.twb. on 'typeproduct line-product name relations' Sheet for more). This can have some implications on the analysis later.

| Product.. | Product.. | Brew Type | Aberd.. | Barry | Basing.. | Bath | Birmin.. | Bristol | Briton.. | Cardiff | Clydach | Cumbe.. | Dundee | Edinbu.. | Glasgo.. | Leeds | Liverp.. | London | Manch.. | Reading | Rhond.. | Swans.. |
|-----------|------------|-----------|---------|--------|----------|-------|----------|---------|----------|---------|---------|---------|--------|----------|----------|--------|----------|--------|---------|---------|---------|---------|
| Beans | Coffee | Decaf | | 11,122 | 5,016 | 5,586 | 3,171 | 6,796 | 13,293 | 1,870 | 6,463 | 11,175 | | | | 4,740 | 7,622 | 6,674 | 3,012 | 10,522 | 5,249 | 8,723 |
| | | Regular | 12,369 | 18,413 | 4,696 | 5,723 | 5,786 | 22,609 | 18,918 | 5,179 | 7,622 | 10,190 | 23,710 | 11,067 | 32,513 | 10,843 | 11,530 | 37,370 | 3,532 | 7,448 | 10,287 | 16,229 |
| | Espresso | Decaf | 5,067 | 7,202 | 5,020 | 3,978 | 10,190 | 8,280 | 21,399 | 1,908 | 4,118 | 8,723 | | | | 13,923 | 4,830 | 26,045 | 2,254 | 7,539 | 5,188 | 3,921 |
| | | Regular | 6,255 | 8,280 | 12,478 | 7,959 | 13,667 | 18,819 | 26,045 | 2,254 | 4,830 | 10,522 | 18,940 | 7,959 | 28,956 | 10,735 | 10,692 | 35,111 | 4,075 | 9,628 | 13,923 | 7,539 |
| Leaves | Herbal T.. | Decaf | 14,085 | 24,719 | 14,196 | 5,179 | 16,229 | 10,287 | 29,535 | 39,338 | 10,580 | 17,167 | 5,188 | 3,532 | 33,873 | 8,997 | 16,653 | 40,527 | 45,923 | 16,911 | 5,663 | 11,393 |
| | | Regular | 7,704 | 16,323 | | | | | 15,482 | 47,170 | 10,475 | 9,076 | 5,663 | 4,075 | 31,019 | 23,857 | 11,937 | 27,099 | 48,570 | 17,498 | 21,365 | 11,275 |

Figure 2 City and products

1.3 Listing more facts

- Dimensions: 18
- Indexes: 9
- Measurements: 9
- General Description of the data. ‘Every store per day per product vs COGS Budget’

2. Dummy Data creation

In order to make the solution more realistic, dummy data had to be created that will reflect the 2020 values.

To create this data, first, research was performed to investigate how the coronavirus impacted the food retail industry in 2020. The research was performed using data only in the UK since the EDA showed that Lupita’s Cafe sorely operates in the UK.

2.1 Findings on the impact of the pandemic

In 2020, the highest ever number on coffee, tea, and cocoa expenditure in the UK was recorded, rising from 3,302 million in 2019 to 4,009 million in 2020. Not only that, but also the percentage increase of sales from 2019 to 2020 was double the percentage increase of 2018 to 2019, from a 10% increase to a 21%. This can be attributed to the food retail industry being one of a few industries that remained largely unaffected by the pandemic [1].

Regarding marketing, the advertising expenditure in the UK in 2020 amounted to 23.46 billion pounds, a fall of 17% compared to 2019 [2].

Finally, the inventories saw an enormous decrease of 221% across the UK in the retail sector [3]. The main reason for this remarkable decrease is that in 2020, there were negative inventories because of uncertainty, timing, supply-chain disruptions and production issues [4].

2.2 Overview of the findings

The following changes were made, in 2019's data, in order to create the 2020 dummy data :

- *Sales and Budget sales were increased by 21%*
- *Marketing was decreased by 17%*
- *Inventory was decreased by 221%*

2.3 The process of creating the 2020 data

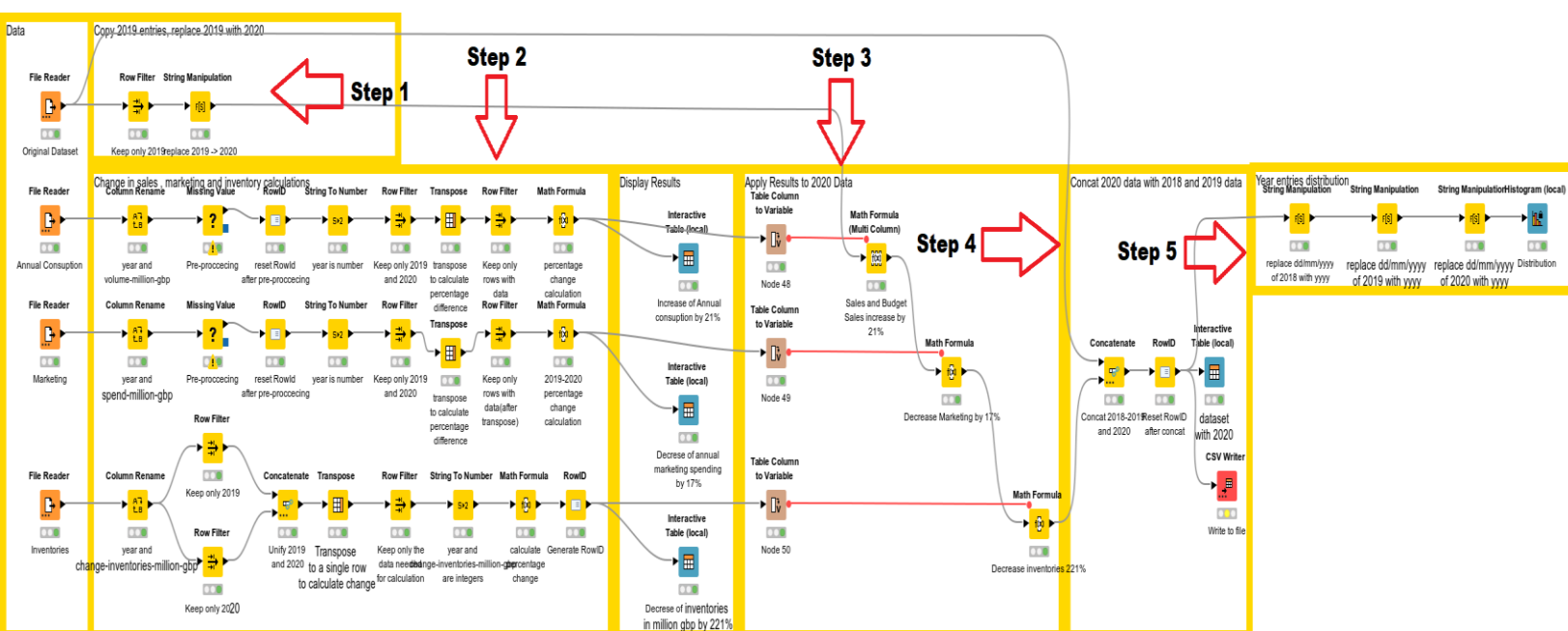


Figure 3 Complete KNIME workflow

Figure 3 shows the complete KNIME workflow used for creating the 2020 dummy data. The following sections will explain in detail what each part of the workflow does.

Step 1: Creating entries for 2020

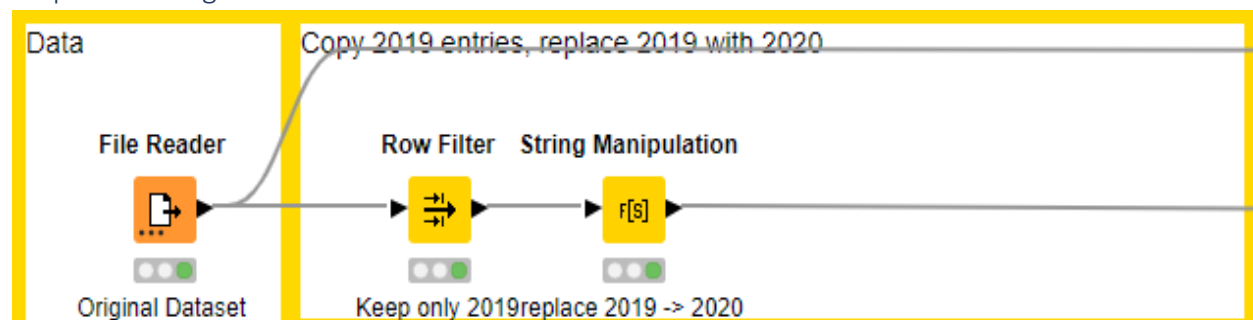


Figure 4 Creating entries

In this step, first, all the data are imported. Then the data are filtered by the year (year = 2019) in order to get only the data from 2019, and finally, the year in the date is changed from 2019 to 2020, since we are creating data for 2020.

Step 2: Preparation for changing the values of the 2020 data

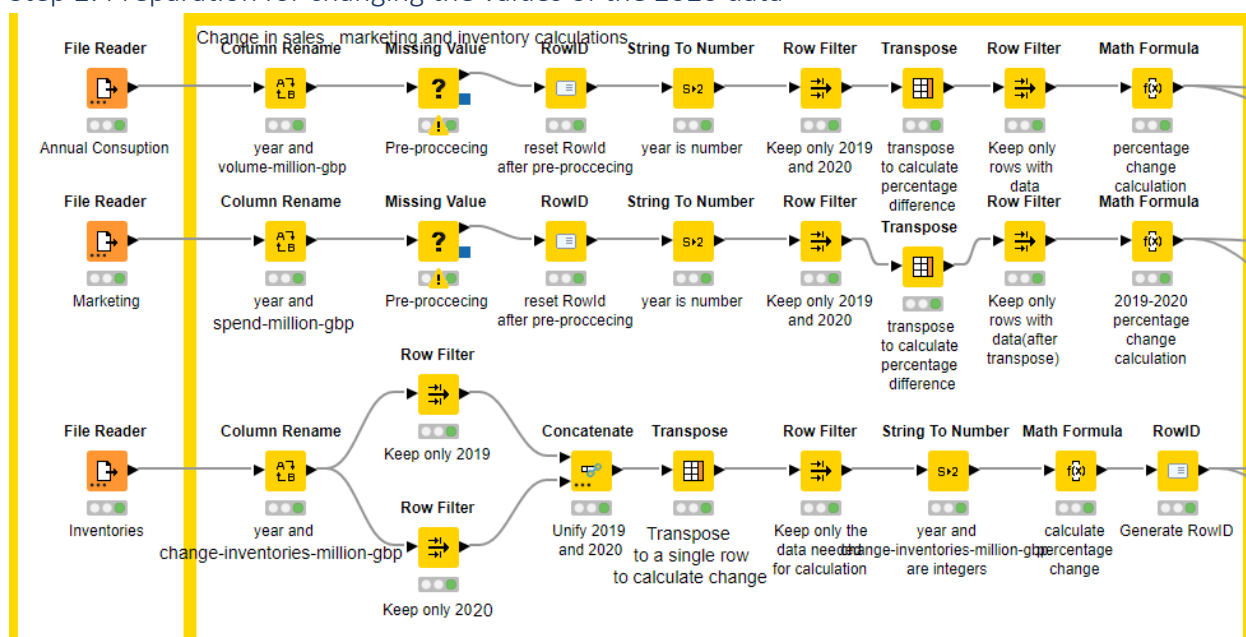


Figure 5 Preparations

In step 2, the data for annual coffee, tea and cocoa consumption, the marketing data, and the inventories data were added to the workflow. Firstly, pre-processing was applied where necessary in order to get the data compatible with the given dataset. Since these files contained data for a couple of years, the next step was to get only the data for 2019 and 2020. Finally, the percentage change for the sales and marketing were calculated and applied for the marketing and sales data. The inventory calculations are slightly more complicated due to negative percentage values involved. The formula used for the inventory calculations can be seen in figure 4. The results of the percentage change calculations are highlighted in the *Overview of the findings* section.

$$\frac{x - x_{ref}}{|x_{ref}|} \cdot 100 = -221\%$$

$$-\frac{x - x_{ref}}{|x_{ref}|} \cdot 1 = 2.21$$

$$-x - x_{ref} \cdot 1 = 2.21 \cdot |x_{ref}|$$

$$x = -2.21 \cdot |x_{ref}| + x_{ref}$$

$$x = \begin{cases} -2.21 \cdot x_{ref} + x_{ref}, & \text{if } x_{ref} \geq 0 \\ 2.21 \cdot x_{ref} + x_{ref}, & \text{otherwise} \end{cases}$$

Figure 6 Inventory formula

Step 3: Perform the calculations

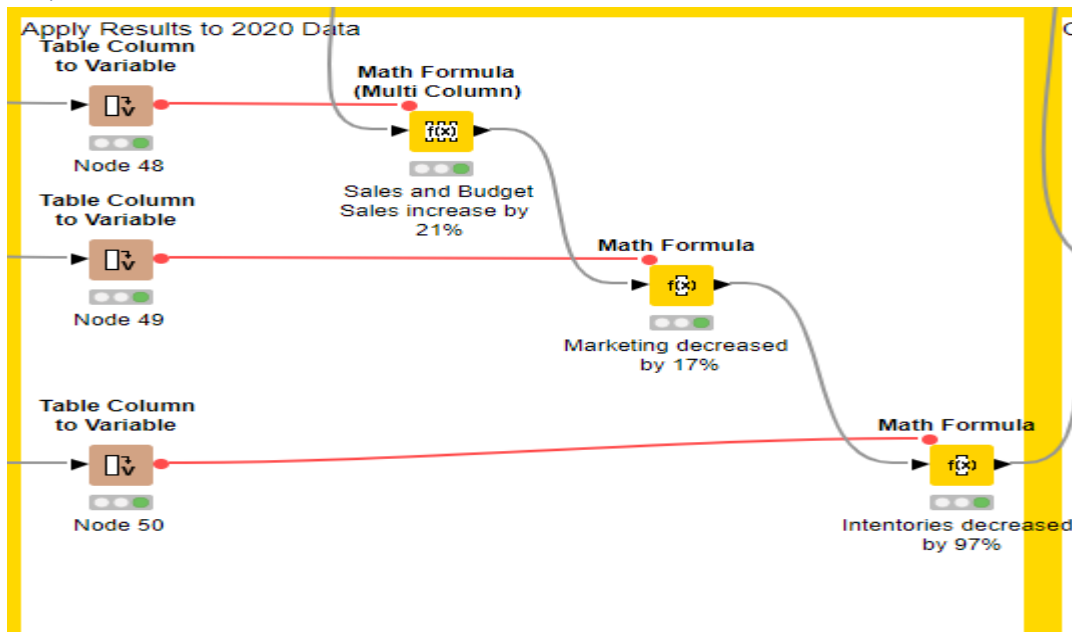
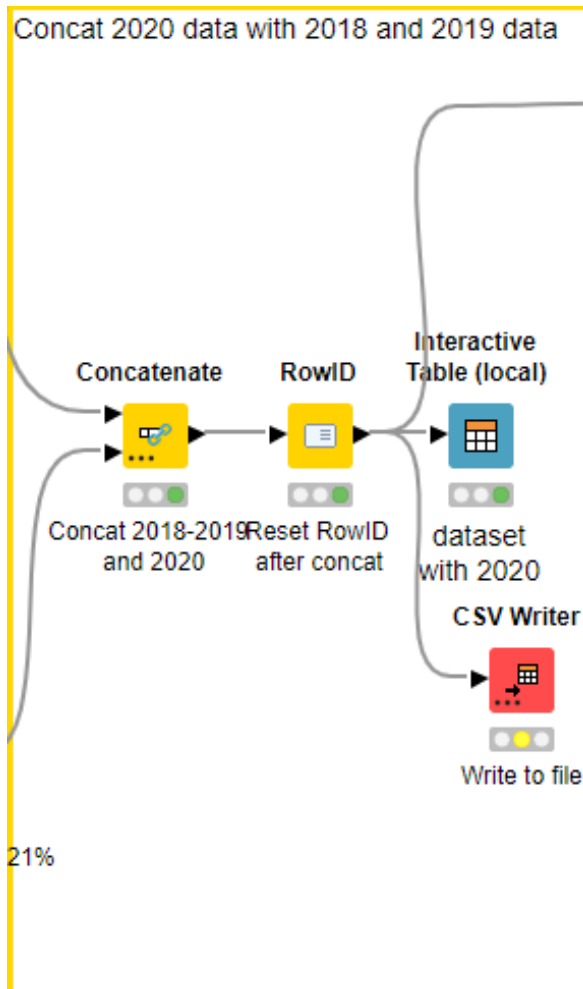


Figure 7 Calculations

In step 3, all the calculations were performed by passing the generated data from step 1 first by a math formula that changed the sales and budget sales data, then the marketing changes were applied and finally, the inventories data were changed.

Step 4: Concat the 2020 data with the 2018 and 2019 data and export them



In the penultimate step, the new 2020 dummy data are concatenated with the existing data, and all the data are exported in a new CSV file under the name “lupita-data-dummy.csv”.

Figure 8 Concat data

Step 5: Testing for issues

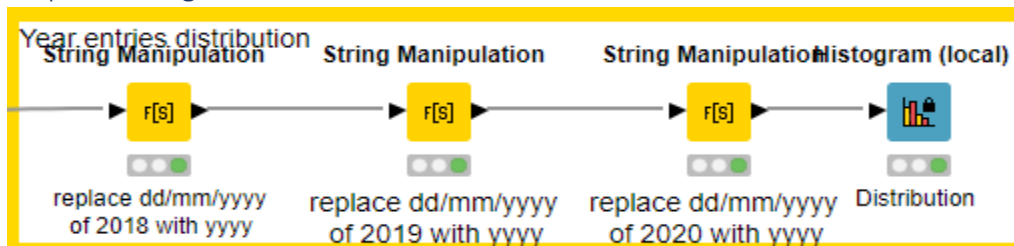


Figure 9 Testing

In the final step, it was tested that we have the exact same amount of data in each year. The reason for this is to make sure that everything in the process works as expected. Since we used 2019 data to create the data for 2020, if the volume differed, it would mean that something went wrong.

2.4 Assumptions

A few assumptions for simplicity had to be made in the process of creating the dummy data.

1. Inflation was not taken into account.
2. 2020 is a leap year, and 2019 was not.
3. Because of lockdowns over 2020 and the uncertainty covid brought, there may have been a vast variance in sales between each data point. The dummy data just used the 21% increase in sales over the whole year and applied it to each data point.

3. Layout design, storytelling reasoning, KPIs, assumptions, storyboard.

3.1 Identify your audience

One of the dashboards is created for a C level which means that the goal was to have a high level of visualization and just the company overview since C levels do not usually have much time. The second one is targeted at a person at a managerial level, so a more detailed visualization was used.

The information that will be used are:

- C level
 - Per State (Wales, Scotland, England)
 - Per Product Type
- Management
 - Per City
 - Per Line, Brew, Name

3.2 Document Goals

The goals that the dashboards tried to help accomplish were:

- Bring Costs Down
- Increase Sales and Profits
- Increase Planning Performance

3.3 Document Key Performance Indicators

The KPIs tie directly to the goals mentioned above. The KPIs are the following:

- Bring Costs Down
 - KPI: Cost of goods sold and 'Other Expenses'
- Increase Sales and Profits
 - KPI: Sales and Profit variables
- Increase Planning Performance
 - KPI: Budget values and Actual values

3.4 Dashboards and Story

Figure 10 shows the C level dashboard, and figure 11 shows the executive dashboard. The following section will explain the graphs in the managerial dashboard. The C-level dashboard graphs are the same, with fewer details. Hence, the same explanation can be applied.

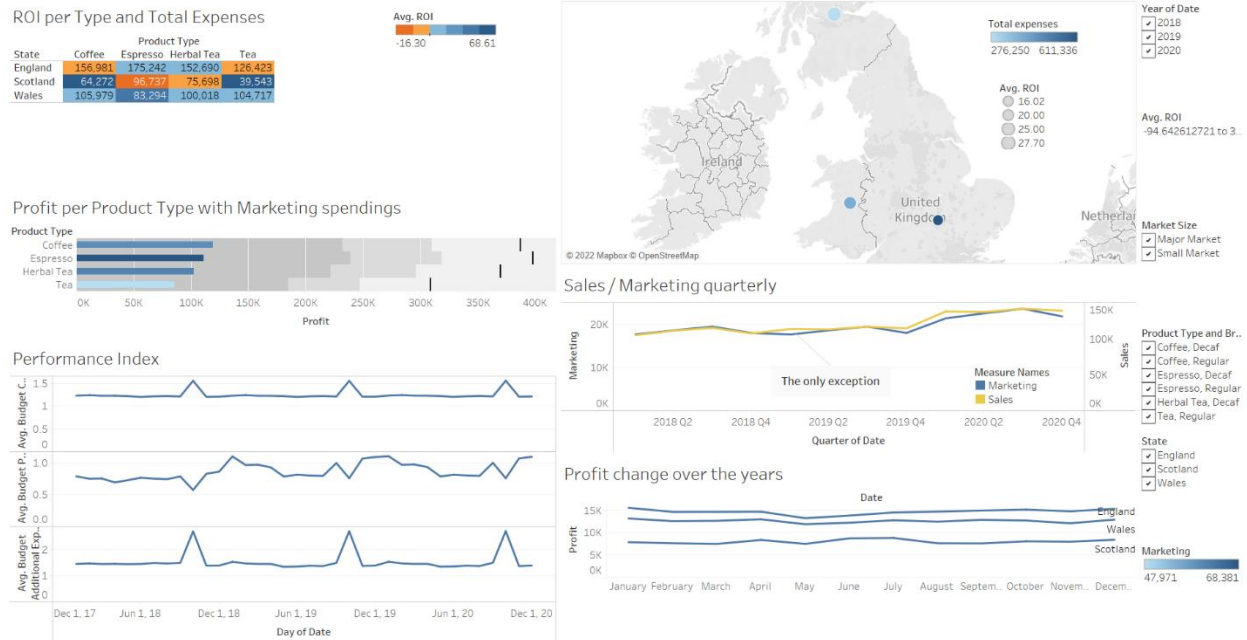


Figure 10 C-level dashboard

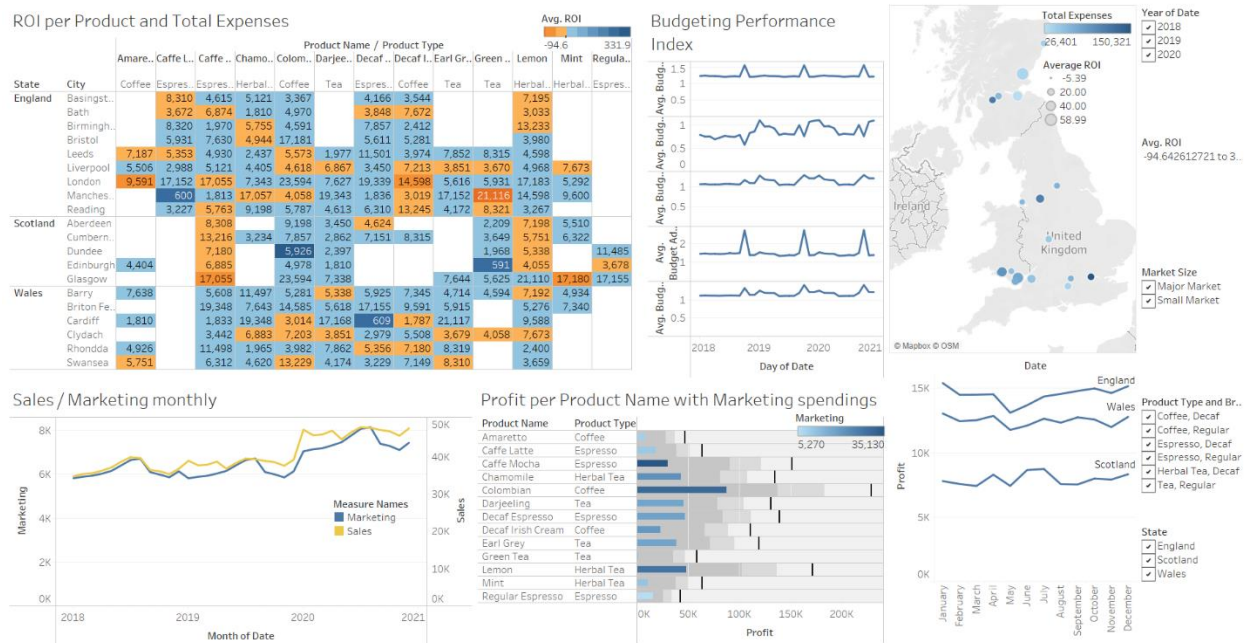


Figure 11 Managerial dashboard

3.4.1 Section 1 – Current State

| State | City | Product Name / Product Type | | | | | | | | | | | | | Avg. ROI |
|----------|--------------|-----------------------------|-----------|----------|----------|---------|----------|----------|-----------|-----------|----------|----------|----------|----------|----------|
| | | Amare.. | Caffe L.. | Caffe .. | Chamo.. | Colom.. | Darjee.. | Decaf .. | Decaf I.. | Earl Gr.. | Green .. | Lemon | Mint | Regula.. | |
| | | Coffee | Espres.. | Espres.. | Herbal.. | Coffee | Tea | Espres.. | Coffee | Tea | Tea | Herbal.. | Herbal.. | Espres.. | |
| England | Basingstoke | | 8,310 | 4,615 | 5,121 | 3,367 | | 4,166 | 3,544 | | | 7,195 | | | |
| | Bath | | 3,672 | 6,874 | 1,810 | 4,970 | | 3,848 | 7,672 | | | 3,033 | | | |
| | Birmingham | | 8,320 | 1,970 | 5,755 | 4,591 | | 7,857 | 2,412 | | | 13,233 | | | |
| | Bristol | | 5,931 | 7,630 | 4,944 | 17,181 | | 5,611 | 5,281 | | | 3,980 | | | |
| | Leeds | 7,187 | 5,353 | 4,930 | 2,437 | 5,573 | 1,977 | 11,501 | 3,974 | 7,852 | 8,315 | 4,598 | | | |
| | Liverpool | 5,506 | 2,988 | 5,121 | 4,405 | 4,618 | 6,867 | 3,450 | 7,213 | 3,851 | 3,670 | 4,968 | 7,673 | | |
| | London | 9,591 | 17,152 | 17,055 | 7,343 | 23,594 | 7,627 | 19,339 | 14,598 | 5,616 | 5,931 | 17,183 | 5,292 | | |
| | Manchester | | 600 | 1,813 | 17,057 | 4,058 | 19,343 | 1,836 | 3,019 | 17,152 | 21,116 | 14,598 | 9,600 | | |
| Scotland | Reading | | 3,227 | 5,763 | 9,198 | 5,787 | 4,613 | 6,310 | 13,245 | 4,172 | 8,321 | 3,267 | | | |
| | Aberdeen | | 8,308 | | | 9,198 | 3,450 | 4,624 | | | 2,209 | 7,198 | 5,510 | | |
| | Cumbernauld | | | 13,216 | 3,234 | 7,857 | 2,862 | 7,151 | 8,315 | | | 3,649 | 5,751 | 6,322 | |
| | Dundee | | | 7,180 | | 5,926 | 2,397 | | | | | 1,968 | 5,338 | 11,485 | |
| | Edinburgh | 4,404 | | 6,885 | | 4,978 | 1,810 | | | | | 591 | 4,055 | 3,678 | |
| Wales | Glasgow | | | 17,055 | | 23,594 | 7,338 | | | 7,644 | 5,625 | 21,110 | 17,180 | 17,155 | |
| | Barry | 7,638 | | 5,608 | 11,497 | 5,281 | 5,338 | 5,925 | 7,345 | 4,714 | 4,594 | 7,192 | 4,934 | | |
| | Briton Ferry | | | 19,348 | 7,643 | 14,585 | 5,618 | 17,155 | 9,591 | 5,915 | | 5,276 | 7,340 | | |
| | Cardiff | 1,810 | | 1,833 | 19,348 | 3,014 | 17,168 | 609 | 1,787 | 21,117 | | 9,588 | | | |
| | Clydach | | | 3,442 | 6,883 | 7,203 | 3,851 | 2,979 | 5,508 | 3,679 | 4,058 | 7,673 | | | |
| | Rhondda | 4,926 | | 11,498 | 1,965 | 3,982 | 7,862 | 5,356 | 7,180 | 8,319 | | 2,400 | | | |
| | Swansea | 5,751 | | 6,312 | 4,620 | 13,229 | 4,174 | 3,229 | 7,149 | 8,310 | | 3,659 | | | |

Figure 12 ROI per product and total expenses

In order to investigate the costs and the sales of each city and on each product, first two new values had to be calculated. The issue with calculating without the new values was that it was found that some places had very high costs, but also very high sales too.

The new values:

- *Sale Expenses* = COGS + Other Expenses
- *Return on Investment (ROI)* = $(([\text{Sales}] - [\text{Sale Expenses}] - [\text{Marketing}] / ([\text{Sale Expenses}] + [\text{Marketing}])) * 100$

After calculating the new values, figure 12 was created. Figure 12 shows a highlight table since the exact values of sale expenses had to be examined. Each value has a colour defined by its ROI, the more intense the orange, the worst ROI is since orange helps convey warning signals, and the more intense the blue, the better ROI is since blue is perceived as positive [5]. The baseline that the colours change is 12% because it was found that 12% is the average ROI for coffee shops [6].

Overall, figure 12 highlights which products have high sale expenses and small ROI and the opposite, basically showing which products are the most and the least profitable. It can be noticed that Scotland has issues with the ‘café late’ in all cities.

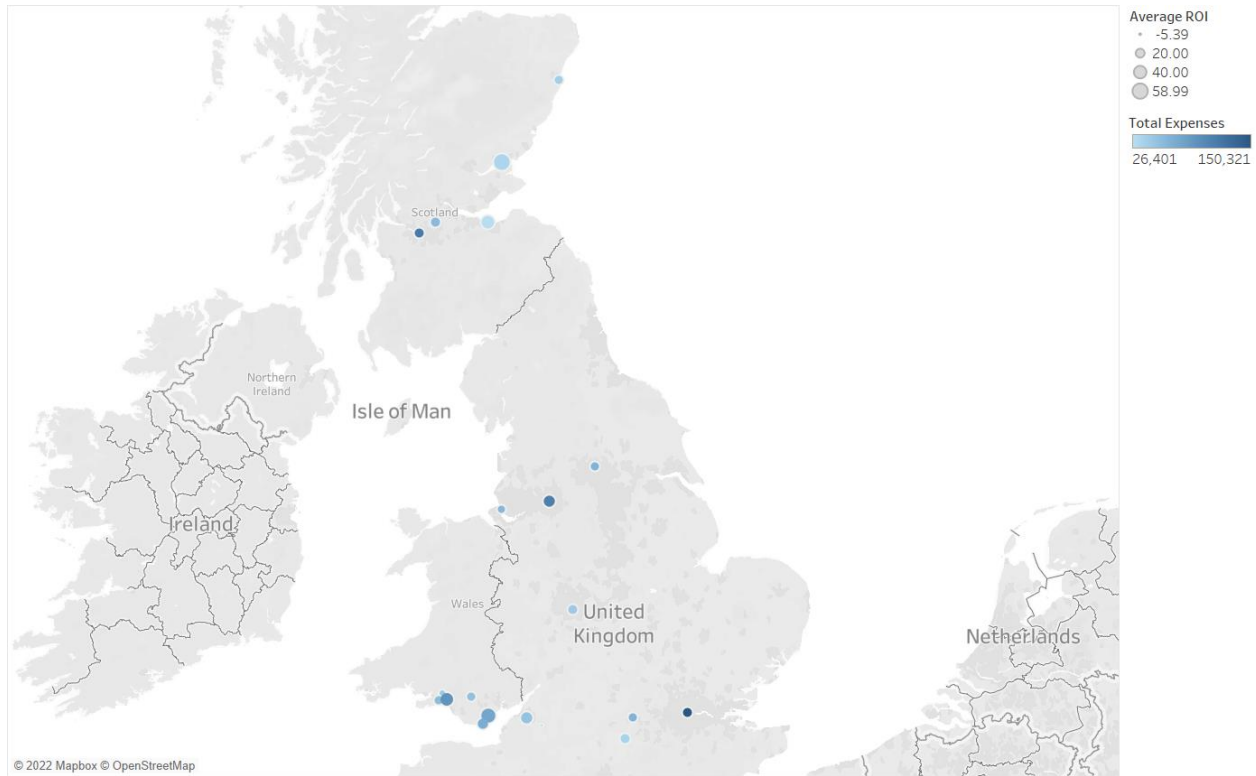


Figure 13 City analysis

Figure 13 displays similar information with figure 12 but over a symbol map. The reason a map was chosen was to display in a real environment what is happening with ROI and the total expenses. The colour blue was selected in different intensities to visualize the total expenses. Total expenses, high or low, are neither positive nor negative, so no colour that showed warnings or positives could be selected. Blue was chosen as it is one of the best liked colours [5].

Again it can be seen that usually, wherever there are high expenses, there is also a small ROI, and where the expenses are more reasonable, the ROI is more positive.

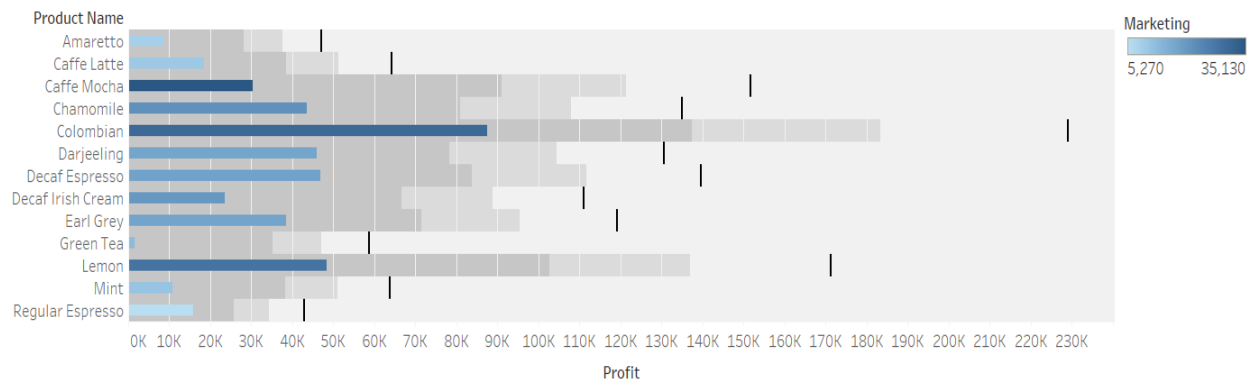


Figure 14 Profit per product with marketing spendings

Figure 14 aims to examine the most profitable products. The length of blue shows the profit of the products, grey shows the sales, and the intensity of blue shows how much were spent on sales. The reason for the colour choices is the same as figure 13 explained above. This particular format of the graph was chosen as it makes it easier to investigate the profits of its product, taking sales and marketing spendings into account.

In figure 14, it is clear that the 'Green tea' is not doing great. It has meagre profit, few sales, and the average marketing investment, so the profit should be better. Also, it is visible that the 'Chamomile' and 'Lemon' are the company's best products. It is also clear that the company is doing well in marketing. Usually, where it invests a lot, it also has a good profit. A few exceptions are the 'Caffe mocha' and 'Regular espresso' products. Marketing spending does not align with the sales and the profit.

3.4.2 Section 2 – Trends

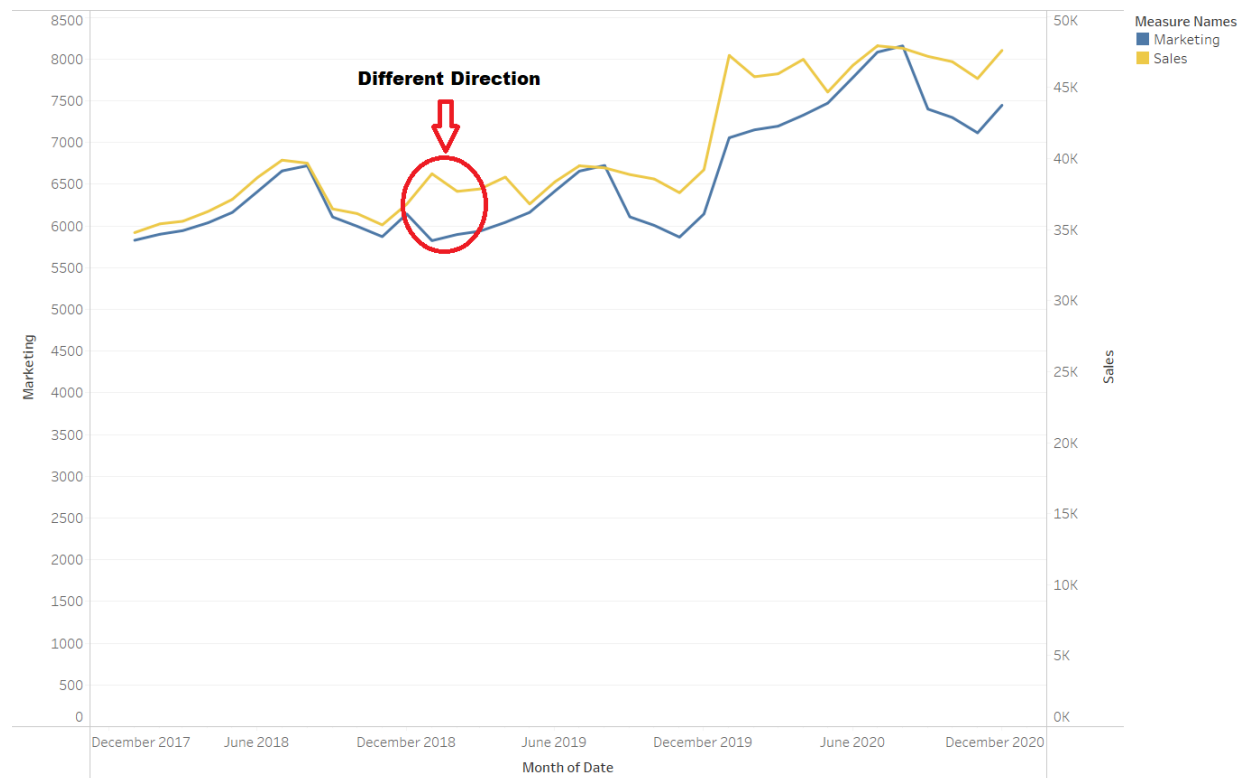


Figure 15 Sales / Marketing

Figure 15 shows the marketing spending and the sales over time. It uses a line chart, as the purpose was to show the changes over time. The colours blue and yellow were picked so the graph would be natural [5].

As it can be seen in the line chart, the sales usually respond to the marketing consistency. There is only one exception where they took different directions around December 2018. It is important to

note that whenever the sales are greater than the marketing spending, it stabilizes (look at the first quarter of 2020). Overall, the company has the trend to deliver in their marketing spendings.

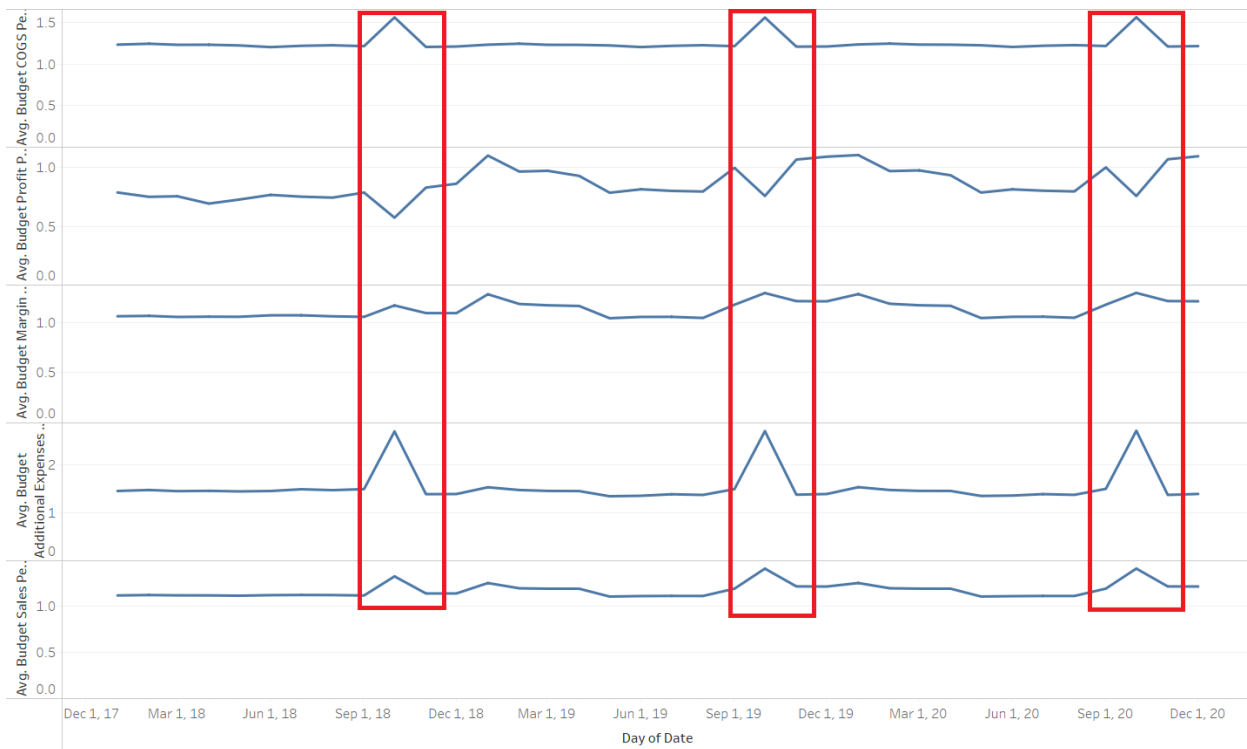


Figure 16 Budgeting performance index

Figure 16 shows the budgeting performance over time. Again a line chart was used, as the purpose was to show the changes over time.

Looking at the figure, it can be seen that the company's budgeting planning is quite good. Generally, the company hits its budget targets all over the year. However, there is a notable exception every September, October, and November. This issue is probably a trend that the business has every summer when the retail sector changes due to the end of summer and the tourism season.

3.4.3 Section 3 – Forecast

Generally, Lupita's café is doing okay. They have established some good stores and products and some bad ones and good marketing and sales trends. The business has good budgeting skills, with the exception of summer. However, doing okay can appear to be problematic in the future.

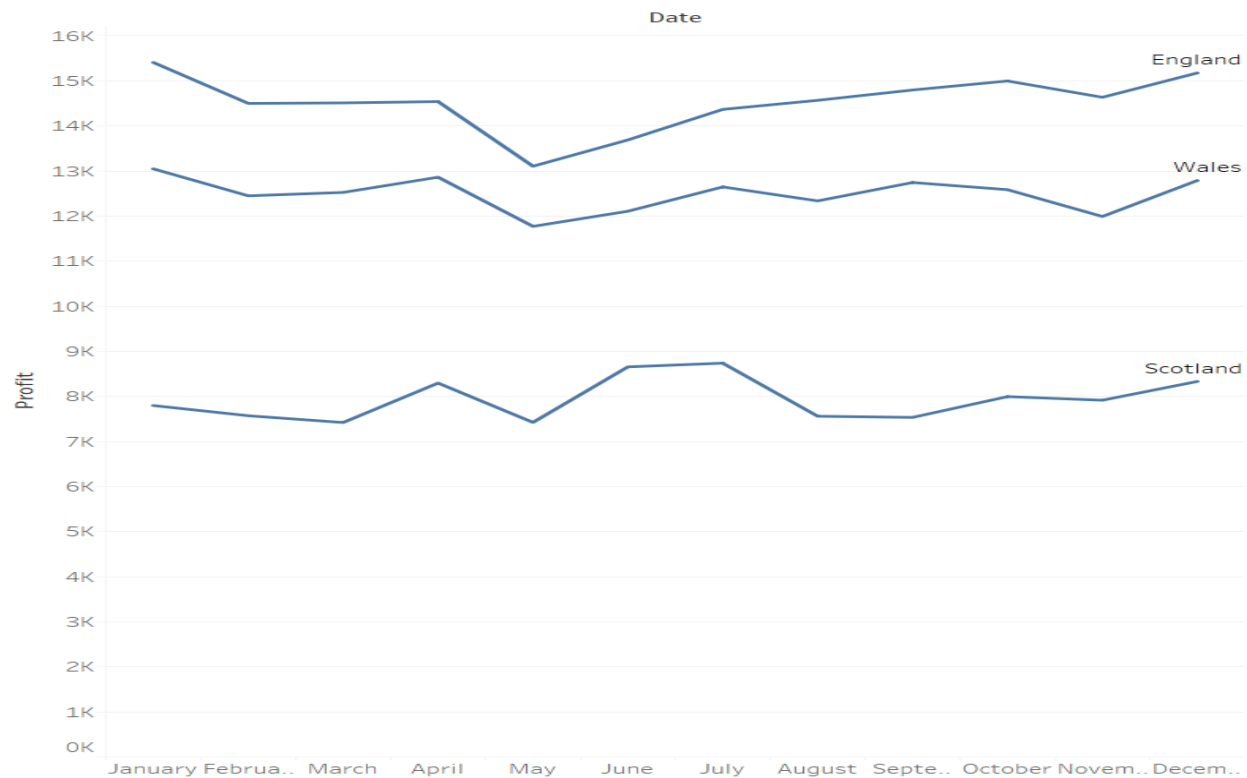


Figure 17 Profit changes over the years

As it can be seen in figure 17, the profits of the company across all three states have stayed the same throughout the years. This means that the company has stayed stagnated, the past three years without any visible growth. Staying stagnated over the years is actually an issue because of inflation and price gouging [7]. Given that the inflation in 2020 was on a record-high 7%, the company has to reconsider some of its strategies and plan ahead on how to tackle the issue [8].

3.4.4 Section 4 – What if and suggestions

- The company has to investigate why they have the budgeting issues after summer. Fixing this issue will perfect their budgeting abilities.
- Resolving the stagnated sales problem as a long term strategic goal. It can be deduced from our findings that increased marketing spending leads consistently to increased sales, and increased sales is the key to get out of the stagnated profit situation. This can be considered as a priority as failing to address this issue early on, may lead reduced market share and purchasing power.

4. Conclusion

4.1 Summing up the report

To conclude, Lupita's café could greatly benefit from a business intelligence solution. By developing a prototype the team was able to detect various problematic areas, their short and long term effects on the business, as well as to propose pragmatic and measurable solutions for a number of them.

A BI solution will furnish a real-time evaluation platform to the business to find and test the ideal processes to follow, in order to steer itself into their rough but ever changing free market.

4.2 Lessons learned

Data visualization is a necessary process to every business and organization to deduce knowledge and insight by using pure data. Data visualization revolutionizes problem detection and prevention, as we can detect crucial problems and tackle them earlier, when their cost and effects can be easily managed.

BI can follow very standard processes but we should always remember that every business and domain is different.

Finally we learned how we can transform our results and findings into insight, ready to be consumed by non-technical for the presentation and technical for the report people.

References

- [1]"Coffee, tea and cocoa purchase trend in the UK 2005-2020 | Statista", *Statista*, 2022. [Online]. Available: <https://www.statista.com/statistics/300812/annual-expenditure-on-coffee-tea-and-cocoa-cream-in-the-united-kingdom-uk/>. [Accessed: 10- Mar- 2022]
- [2]"Advertising spending in the UK 2020 | Statista", *Statista*, 2022. [Online]. Available: <https://www.statista.com/statistics/262754/advertising-revenue-in-the-uk/>. [Accessed: 10- Mar- 2022]
- [3]"Change in Inventories: Retail: £M: CP: SA - Office for National Statistics", *Ons.gov.uk*, 2022. [Online]. Available: <https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/faha/pn2>. [Accessed: 10- Mar- 2022]
- [4]"What Does It Mean To Have Negative Inventory?", *IRVINE BOOKKEEPING*, 2022. [Online]. Available: <https://www.irvinebookkeeping.com/post/2017/12/19/what-does-it-mean-to-have-negative-inventory>. [Accessed: 10- Mar- 2022]
- [5]"Leveraging Color to Improve Your Data Visualization", *Tableau Public*, 2022. [Online]. Available: <https://public.tableau.com/en-us/s/blog/2013/10/leveraging-color-improve-your-data-visualization>. [Accessed: 11- Mar- 2022]
- [6]"ROI Formula (Return on Investment)", *Corporate Finance Institute*, 2022. [Online]. Available: <https://corporatefinanceinstitute.com/resources/knowledge/finance/return-on-investment-roi-formula/>. [Accessed: 11- Mar- 2022]
- [7]J. Kozlowski, L. Veldkamp and V. Venkateswaran, "The Tail that Wags the Economy: Belief-Driven Business Cycles and Persistent Stagnation", *SSRN Electronic Journal*, 2015.
- [8]"As inflation soars, a look at what's inside the Consumer Price Index", *Pew Research Center*, 2022. [Online]. Available: <https://www.pewresearch.org/fact-tank/2022/01/24/as-inflation-soars-a-look-at-whats-inside-the-consumer-price-index/>. [Accessed: 12- Mar- 2022]