**Bike Sales Marketing Project - Where and Whether to Advertise**

-D**escription of dataset**

The dataset outlays sales data of a chain of fictional international bike shops, situated in six countries, with customer demographics, product information and costs/revenue included for each data point (totalling over 113,000).

I chose this dataset as I wanted to investigate whether working with sales data was something that interested me, as this would give me a better understanding of what kind of data analyst job was for me. Further, not only did it include demographic data, it also included product information, so I intuited that it would be a strong dataset for me to investigate, with plenty of potential avenues of exploration.

**-Ask Phase**

The board wants to increase profits so are planning to launch a marketing campaign. They want to maximise the returns of the marketing, therefore they have instructed their data analyst to analyse their sales data over the past 5 years to find out how they should spend their marketing budget, if at all.

For this preliminary ask phase I, brainstormed the following questions to ask the dataset to aid in coming to a conclusion for the shareholders. All the questions are SMART - specific, measurable, action orientated, relevant, time bound).

* Which demographics are leading to the most/least sales/profit? How could this information be used to increase sales/profit?
* Which countries/regions/product categories are leading to the most/least sales/profit? How could this information be used to increase sales/profit?
* When during the year do most of the sales occur? For what reasons may this be?
* Based on this information, WWWWH would you recommend advertising?

**-Prepare Phase**

In this section, I speculated what data will be needed to answer the questions I proposed in the ‘ask’ phase. Further, I considered what metrics could be employed to show the results of my investigation.

The aim of this project is to analyse how the marketing budget should be spent, with the ultimate aim of this being to increase the shop’s profit. Marketing spend is usually allocated to areas already generating the highest profit/revenue, as the theory goes that these areas already performing well have the greatest growth potential, and hence the greatest return from marketing. Therefore analysing where profit is coming from and when should be the first aim of my analysis.

Although a profit variable is already given in the dataset, I believe it is not sufficient to judge which products are performing well/poorly, as it ignores the overall cost of the product. For instance, if the profit of a product costing $100 is $20, and the profit of a product costing $500 is $50, the first product is performing much better, as 20% of the sales revenue is profit, whereas it is only half that for the second (10%). The profit variable included in the data set would not discern this difference, and would actually view the second as being the better performing product. Hence it is vital to use an alternate profit measure to adequately measure product performance.

I believe a productivity variable, calculated in the following way will sufficiently measure the level of profitability of each item, as it takes into account the cost of goods sold. Upon further research, a variable such as this is often called the gross profit ratio (GPR).

Before I proceed, I would like to elucidate fully the possible obstacles faced when using the ‘Cost’ variable included in the dataset. Unfortunately little information is given about the breakdown of how this variable was generated, meaning it is difficult to have full confidence in its efficacy, as it is uncertain which costs are and aren’t included in its total.

In accounting, product cost can be calculated in numerous ways, varying between the highly unsophisticated, where the product’s cost as paid to the supplier is used, to the highly complex, where a detailed breakdown of all the costs of selling each product are calculated and added together. In the latter case, costs such as those associated with lighting the shop, the shop’s rental cost, staffing costs needed to sell the product, and so on, are included in the cost variable of each product. As each item varies in size, product type and shape, the amount these aforementioned costs are apportioned to each sale differs, creating large differences between it and the less technical cost calculated. Depending on how the cost variable is generated, the profitability variable will have more or less weight, with a detailed cost analysis increasing the result’s influence over decision-making.

If the less sophisticated cost calculation was used for the dataset, the following examples of erroneous profit calculation could be found. These would restrict the applicability of the findings:

* Small products, such as bike accessories, are overestimated in cost. This will result in downward bias of such products’ profitability ratio, making these items appear less profitable than they actually are.
* Large products, such as bikes, are underestimated in cost. This will result in upward bias of such products’ profitability ratio, making these items appear more profitable than they actually are.
* Products which take longer to sell, such as high value bikes, will be underestimated in cost. This will result in upward bias of such products’ profitability ratio, making these items appear more profitable than they actually are.
* Products which take shorter to sell, such as low value bike accessories e.g. puncture repair kits, will be overestimated in cost. This will result in downward bias of such products’ profitability ratio, making these items appear less profitable than they actually are.
* Products which require more care to generate sales, such as bikes, as they require salespeople to sell the bikes and maintenance of the in-store examples, will be underestimated in cost. This will result in upward bias of such products’ profitability ratio, making these items appear more profitable than they actually are.
* Products which require little care to generate sales, such as off the shelf maintenance items, will be overestimated in cost. This will result in downward bias of such products’ profitability ratio, making these items appear less profitable than they actually are.

For the sake of my investigation, I will assume that the cost variable is detailed in its make-up and that it has considered all the individual components which make-up costs of sales. This will allow me to formulate recommendations with weight in my conclusion.

**Analyse Plan**

To analyse where profits are being generated, I envisage comparing the profit ratios of different age groups, countries and genders etc. using a range of charts. This should help the shareholders visually see the make-up of profit ratios to help them understand where best to spend their marketing budget. Furthermore, I will employ a range of ratios to compare performances between regions to get a better picture of where it will be best to spend the funds. Data used for this comparison will include GRP (gross regional product) per capita and population. This will allow for truly well performing locations, those that have many sales despite relatively low incomes/populations, to be identified and exploited.

To start with, I will formulate a dataset using python and excel of all the necessary data points needed for the regional analysis. Like before, I will use 2015 data as it is the most recent complete year available. Unfortunately as there were no sales in 2015 in Minnesota, Mississippi, South Carolina and Virginia, these regions cannot be used in the analysis.

To compare the relative performance of each state in the dataset, I will use the following style of variables, where the states variable is compared to the average for that country.

Where the above variable is equal to 1, the state is performing on par with the average performing state. If it is greater than 1,it is performing better than average, and if it is below 1 it is performing worse than average.

Following on from this, I envisage formulating an ‘overall’ performance variable for each statistic (profit ratio, sales and profit) through comparisons of GDP per capita and population between all the states. The thinking behind this is that where these state factors (income and population) are higher than the average of all states, this ought to increase profit etc. Depending on whether the increase in profit is proportional to the increase in income/population dictates the overall profit score. Where the increase is less than expected, the value between the brackets will be less than one, therefore resulting in the performance variable calculating to less than one. Whereas, when the state is performing better than expected, the performance variable will be greater than one.

I was initially going to include a population density statistic, or even an area statistic, as I was thinking that larger states, or those that are more densely populated would have a natural advantage in terms of profit. However, after weighing this up, I decided not to include such a statistic as I thought it would weigh too heavily against the large states. Also, I thought it would give an unrealistic view of these larger states, as although they are large, they are still very densely populated in specific regions, so I decided not to include it in the analysis.

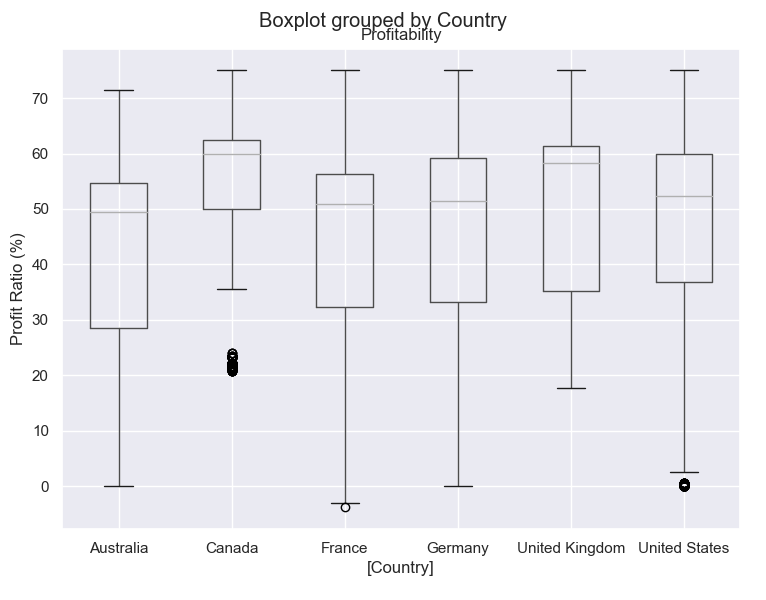
To analyse when sales are happening, the most obvious way of going about it is using a line graph over time, with comparisons in each country and year used to get a full picture of the data. If possible, I would like to analyse which days of the week sales/profits are being made, as this will give me the chance to learn about when to stage the advertising campaign.

To estimate the return, or maximum budget recommended for the advertising campaign will be a challenging task as I don’t have any previous advertising campaigns to work by. To traverse this hurdle, I plan to use secondary data I can find through research, to see how profitable similar firms’ advertising campaigns have been.

**-Process Phase**

In this phase, the data is cleaned. This means removing any duplicate data, biased and/or wrongful data. Fortunately the dataset has already been cleaned, therefore this step of the data analysis process can be skipped.

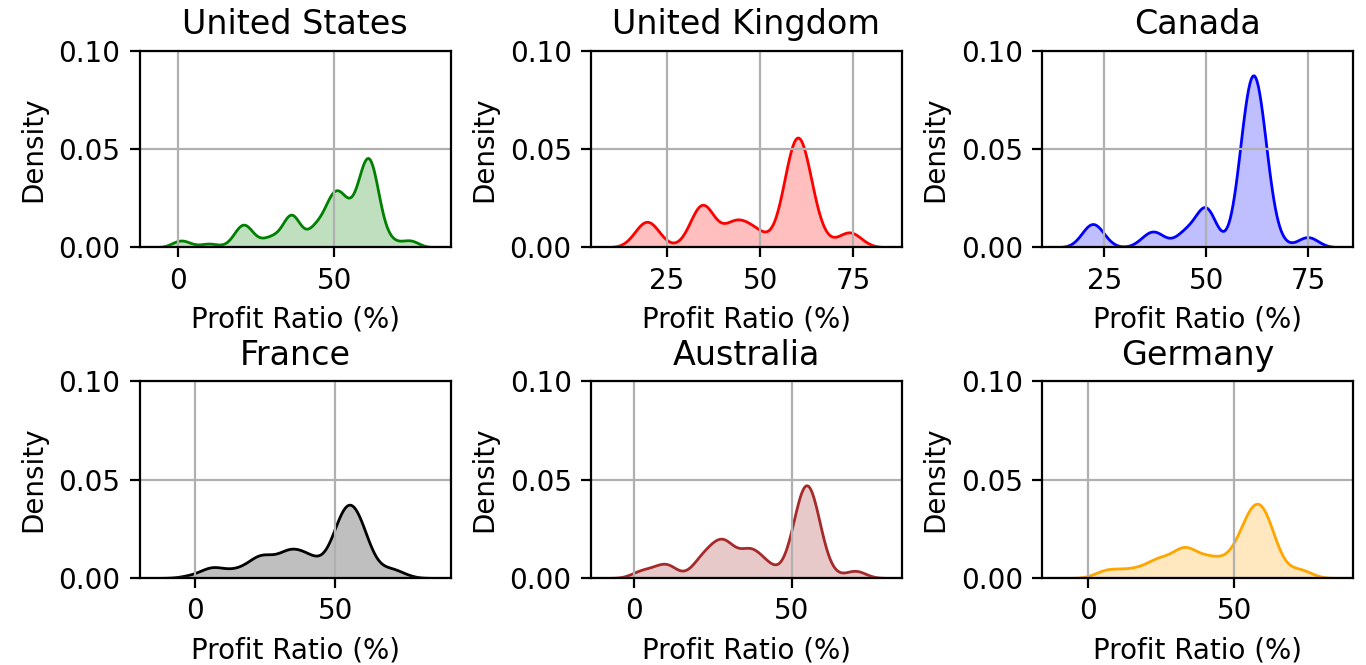
**-Analyse Phase**

This is where the questions proposed in the first section are answered, and the story the data is telling us revealed.

**Demographic Analysis**

The boxplot to the side shows the spread of profit ratios in each country. The green line gives the median profitability of each country, and the box gives the upper and lower 25th percentiles. The dots are outlier values i.e. those data points which differ statistically significantly from the rest of the data (outside of the upper and lower outer fences). I used 2015 data only as this is the most recent full year of data available, and therefore most applicable for influencing future decision making.

Box Plot 1 shows:

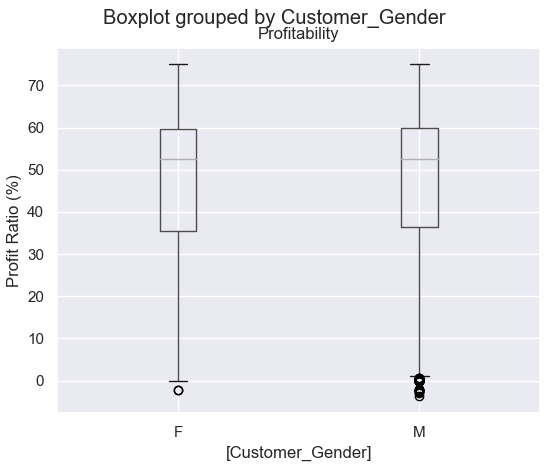
* Canada has the highest median profit ratio of all the countries in the sample, at 60% compared to the UK’s 58%. It also has the lowest spread of profit ratios of all the countries, indicating it consistently outperforms other countries. Even Canada’s outlier profit ratios are still within other countries' interquartile range (IQR), further evidencing this hypothesis.
* Australia is the least profitable of all the countries, with a median profit ratio of 49.5%. It has one of the largest spreads of profit ratio of all the countries, suggesting there are some very well and poorly performing products.
* France only just outperforms Australia, with a profit ratio of 50.5%. Again, France has a large IQR, hinting at its diversity of profit ratios.
* Overall, there are only minor differences between each country’s profitability, however even the small difference of 10% warrants further analysis. It is important to understand what is driving these differences to help direct decision-making in the correct direction. 

The density plot to the side maps each value of profit ratio for each country in order, and where most values occur, a greater ‘hill’ is added to show visually it is more frequently observed.

Density Plot 1 Shows:

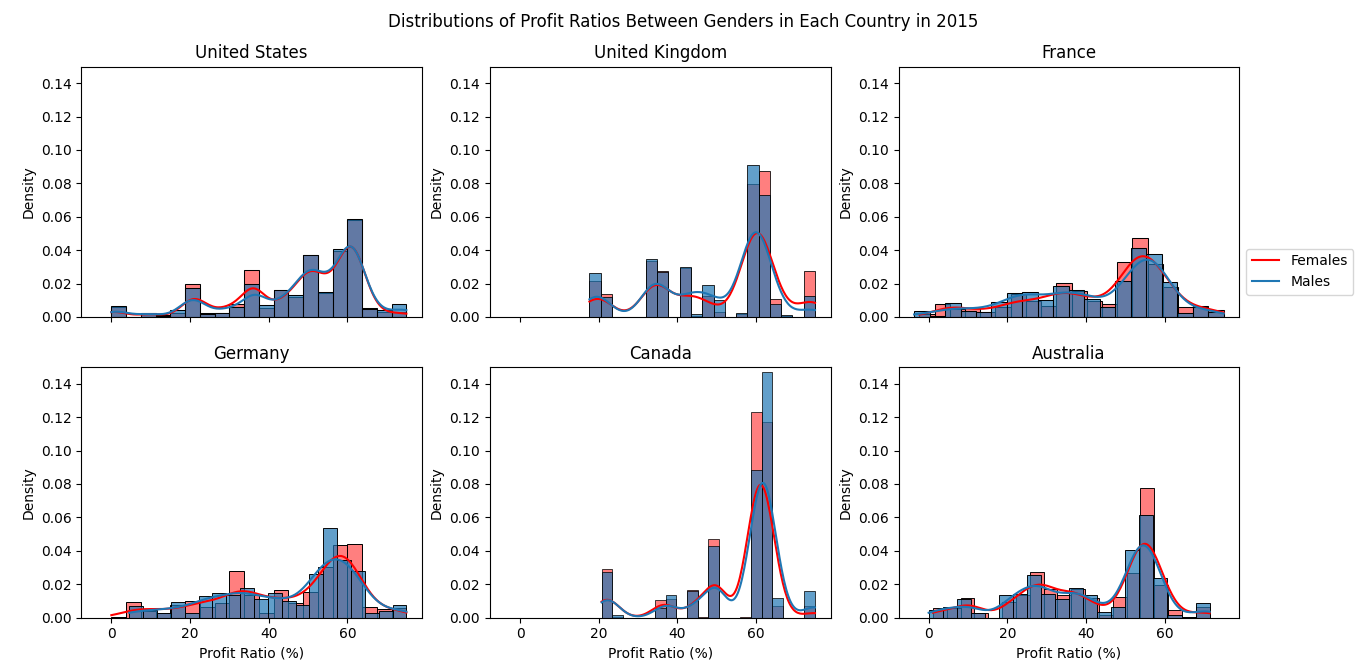
* Canada is the best performing country with regards to profit ratio, with a much higher density of high profit ratios compared to other countries.
* Australia has more low profit ratio sales than the other countries, with a noticeable peak around 25%. This helps to explain why its overall profit ratio is so low.
* Surprisingly, Germany has a fairly high proportion of high profit ratio sales considering it has one of the worst overall profit ratios. This suggests there are many products with a poor profit ratio being sold in Germany, which may warrant a rejigging of products for sale.

I now want to dig deeper into this location data to see which regions in each country are performing the best/worst to get a greater understanding of the potential returns on offer from marketing.

The boxplot below shows the spread of profit ratios in all countries combined in 2015. 

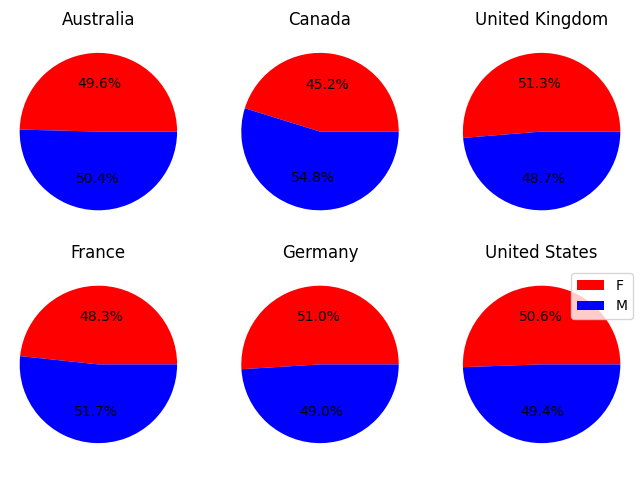
Box Plot 2 Shows:

* There is very little difference in the spread of profit ratios between each gender, suggesting that neither gender is more or less profitable than the other.
* However, the box plot hides potential differences between countries, therefore a little more investigation is necessary.

The histograms below show the distribution of profit ratios of each gender in each country.

Histogram Set 1 Show:

* In most of the countries, the distribution of profit ratios differs very little between genders. This confirms my previous hypothesis. This means that neither gender should be targeted over the other in the marketing campaign as both are as profitable as each other, so there are no extra returns available from targeting either gender more.
* In Germany, there is some slight heterogeneity between each gender, with a slightly higher proportion of high profit ratio products in Women than Men. Given this, it may be worth tilting towards women in Germany to attempt to attract higher profit customers. However, it may be that the proportions of male and female sales are very different in each country, therefore this warrants further analysis.

The pie charts below show the makeup of sales in terms of each gender in 2015 in each country. 

Pie Chart 1 Shows:

* Overall, neither gender dominates sales, however in some of the countries, for example Canada, there is a male majority. Therefore, to maximise sales, men in Canada should be targeted as they form the dominant gender group. However, this should be done with caution, as the proportions are still fairly similar, with a mere 10% gap, meaning that if men are targeted too much there is a risk of a decline in sales, as women could lose interest in a male-centred brand.

I will now begin my analysis into how age affects the profit ratio, to find out whether any specific age groups are more profitable, and therefore more marketing worthy than others.

The scatter plot below shows how customer age is correlated with the profit ratio. A line of best fit is included to better understand the significance of the relationship. The following equation describes the plot line in the following two scatter plots.

Scatterplot 1 Shows:

* As age increases, so does profit ratio, however the relationship is very weak and there is a lot of noise in this relationship. As such, no age group should be targeted just yet.
* Still more research is needed to see whether this trend is observed in all countries.

The next scatter plot is similar to the first, however it splits the data into each country, therefore facilitating the identification of country specific age trends.

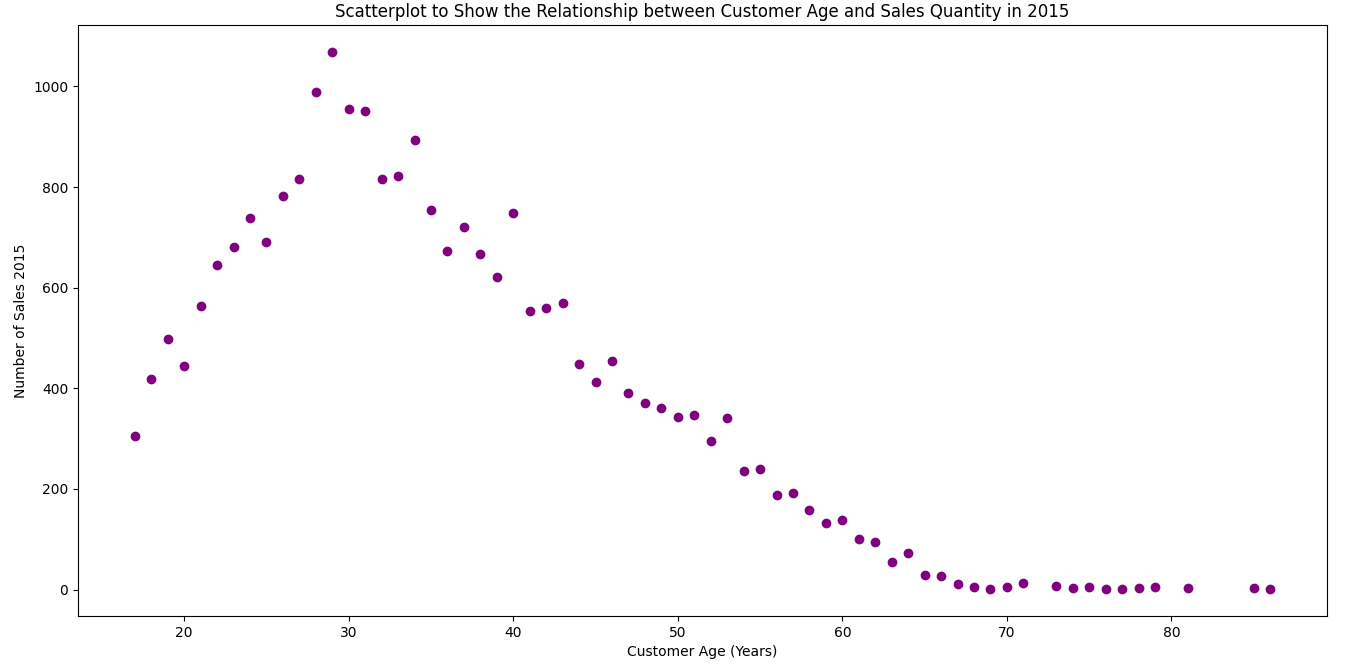


| Country | Constant () (1 d.f) | Age Effect () (2 d.f) |
| --- | --- | --- |
| Canada | 54.1 | 0.04 |
| Australia | 38.0 | 0.11 |
| United States | 47.6 | 0.02 |
| Germany | 44.3 | 0.05 |
| France | 42.8 | 0.02 |
| United Kingdom | 50.1 | -0.02 |

Scatter Plots/Table Shows:

* As in the previous scatter plot, for the most part, age has a very weak positive effect on the profit ratio. This reinforces the idea that certain age groups should not be targeted in the marketing campaign.
* Australia is the outlier of the dataset, with a much larger age effect variable observed compared to the other countries. This means that for a customer to be ten years older, a sale in Australia can expect to generate 1.1% more profit. This is still small, however still large enough to be taken into account.
* Overall, as the values are all so small, age should be ignored as there isn’t strong enough evidence that it affects profitability in a significant way.

Although profitability is not correlated with the age of customers, sales might be. Hence, I have produced the following graph to study the overall relationship in 2015.

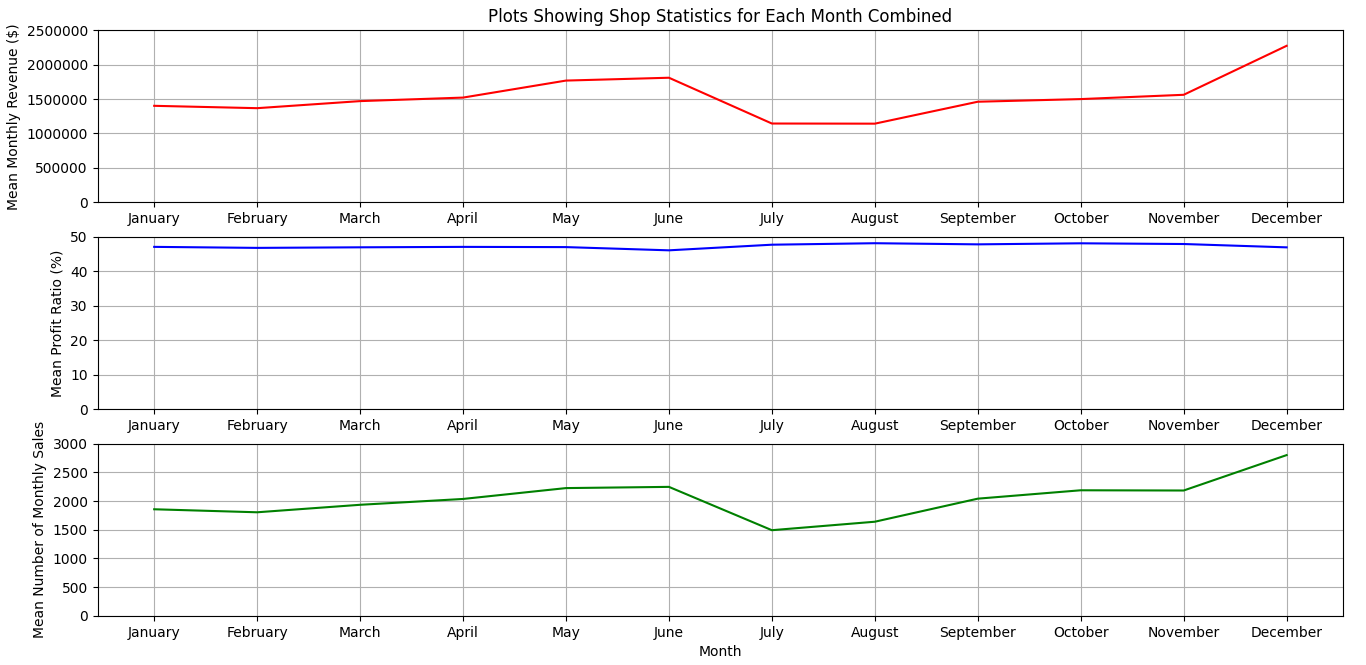


Scatter Plot 3 Shows:

* There is strong evidence that the age of customers has a discernible relationship on the number of sales, with the most sales coming from customers aged 27-35. Customers between these ages should therefore be targeted. After this age, increasing age reduces sales.

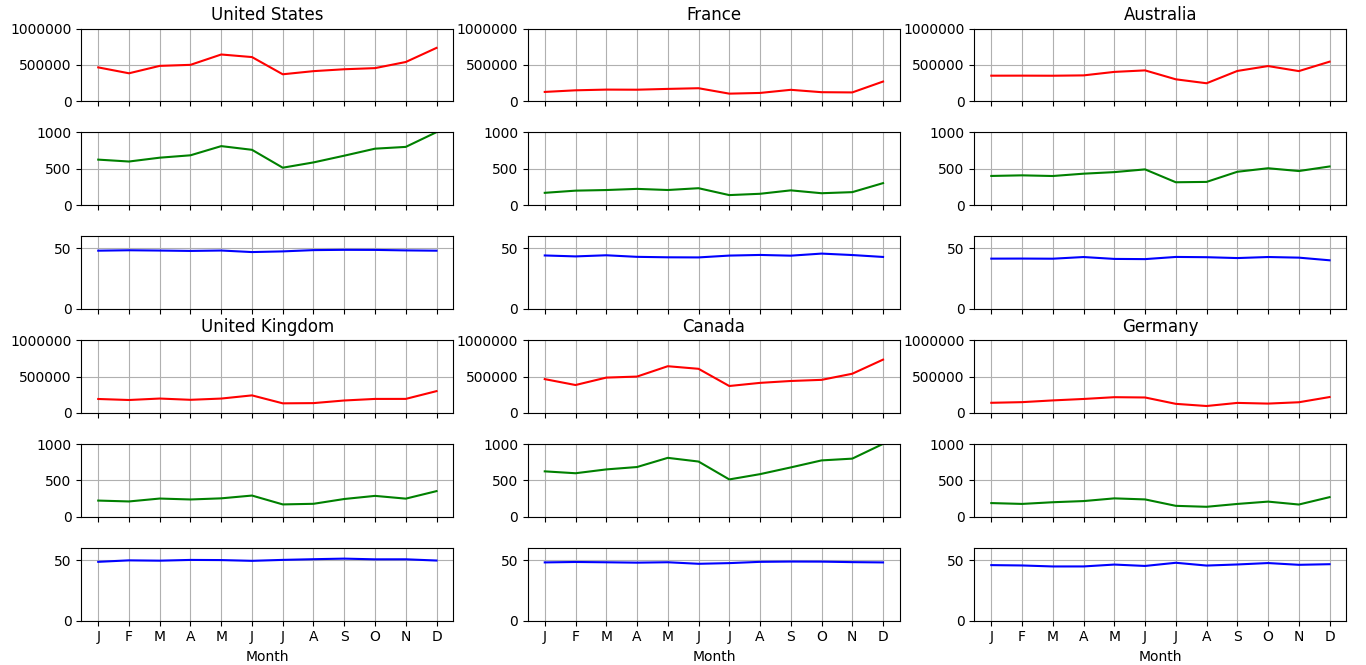
**When are Sales Occurring**

This is an important question to ask the dataset, as it will help to guide when the marketing campaign should take place by discovering when it will have the largest effect. To investigate this, I will start with a review of the monthly shop statistics (revenue, profit ratio and number of sales) in all countries combined to see whether there are any months where the shop performs better than in other months.

Line Graph 1 Shows:

* Average monthly sales and revenue are proportional to one another, meaning that where there are more sales there is also higher profit. This is expected given the results of the profit ratio analysis, which shows that the profit ratio is almost completely constant throughout the year.
* There are peaks in revenue in May, June and December. This may be due to sportswear being purchased in the summer months in preparation for good weather which permits more use of sports equipment. The increase in sales in December is likely due to Christmas, with the event being widely celebrated in all the countries in the dataset. Given this information, it would be wise for the shop to plan their advertising for around these times to try and maximise the returns they can receive.
* July and August are particularly weak times of year for the store, with the lowest revenue, and therefore profit experienced. This suggests these months should be avoided in the store’s marketing strategy as not doing so may result in low profit.

The next set of line graphs show the same data as the previous line graph, but this time the data is separated into each country.



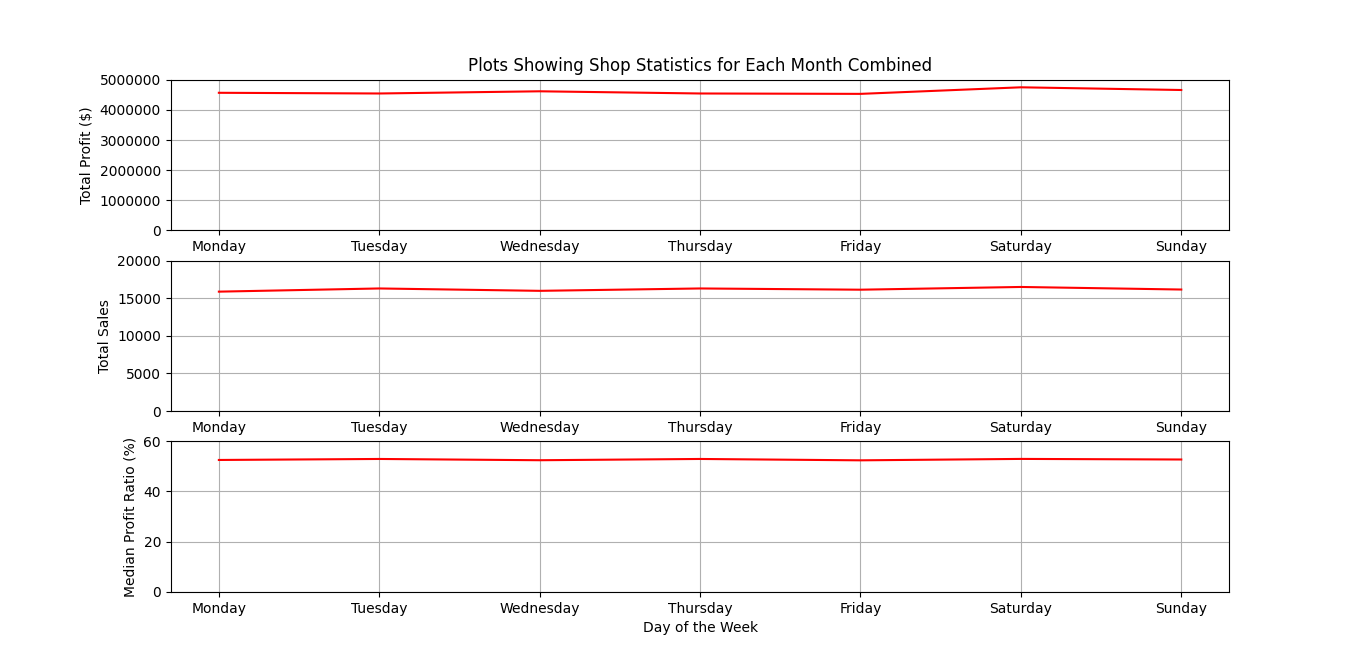
Line Graphs 2 Shows:

* There is some slight heterogeneity between countries’ performance, with some months being stronger in some countries than others. For instance, in Australia, December and October are the two strongest months and August is the weakest, whereas in the USA and Canada, May and December are the strongest months, and July is the weakest.
* Once again, there is very little fluctuation in the profit ratio over the year, suggesting that any profit fluctuations are coming exclusively from more revenue from more sales.

| Country | Best Performing Month(s) | Worst Performing Month(s) |
| --- | --- | --- |
| USA | * May * December | * July |
| UK | * June * December | * July |
| France | * June * December | * July |
| Canada | * May * December | * July |
| Australia | * October * December | * August |
| Germany | * May * June * December | * August |

* The general findings of the line graph are that December is the strongest month for all countries, however after this, there is quite a lot of variability in month performance. Depending on the country, the month of advertising should be selected based on where sales/revenue is already highest, as this will maximise returns.
* July and August should be avoided in the marketing campaign, as these are both poorly performing months across the data set.

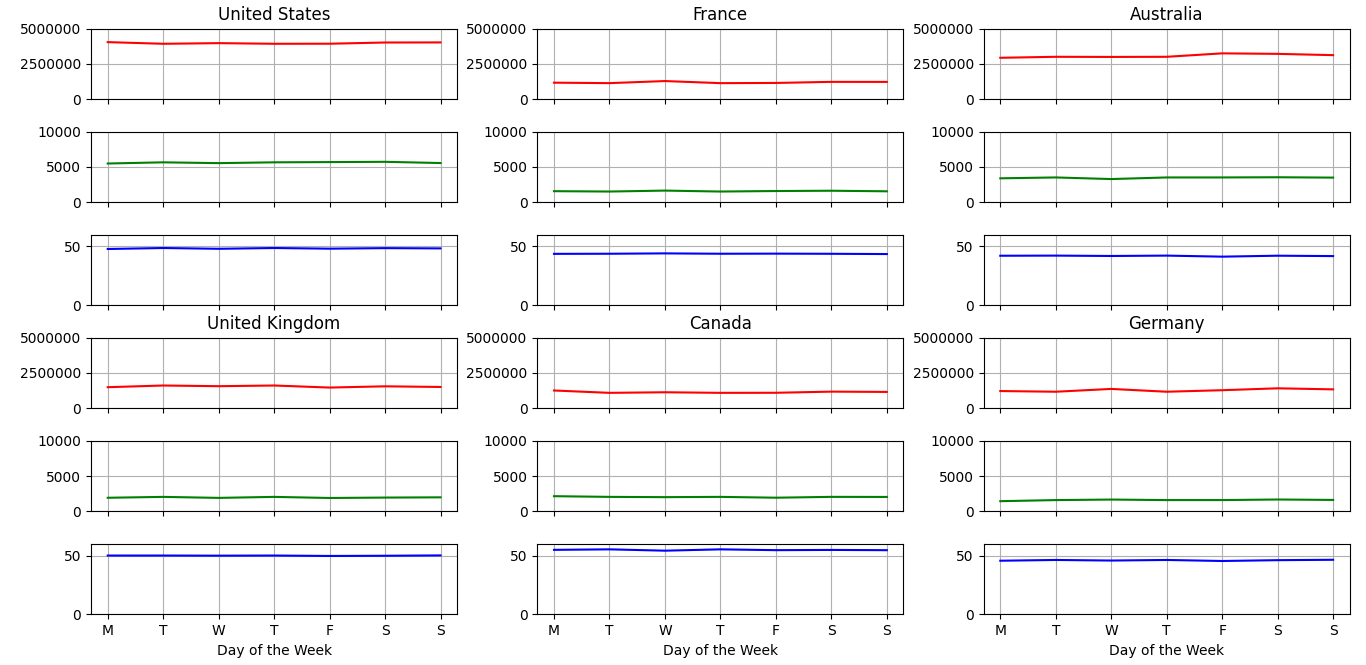
I now want to go even more in depth with the data and find out whether any day of the week is more or less profitable than another. If there is a day of a week much more profitable than the rest, this would incline the bike store to market for this selected day to attempt to maximise sales.



Line Graph 3 shows:

* There is no specific day that is performing better or worse on average than the rest. Over the whole dataset there is no discernable difference in the level of profit over each day of the week.

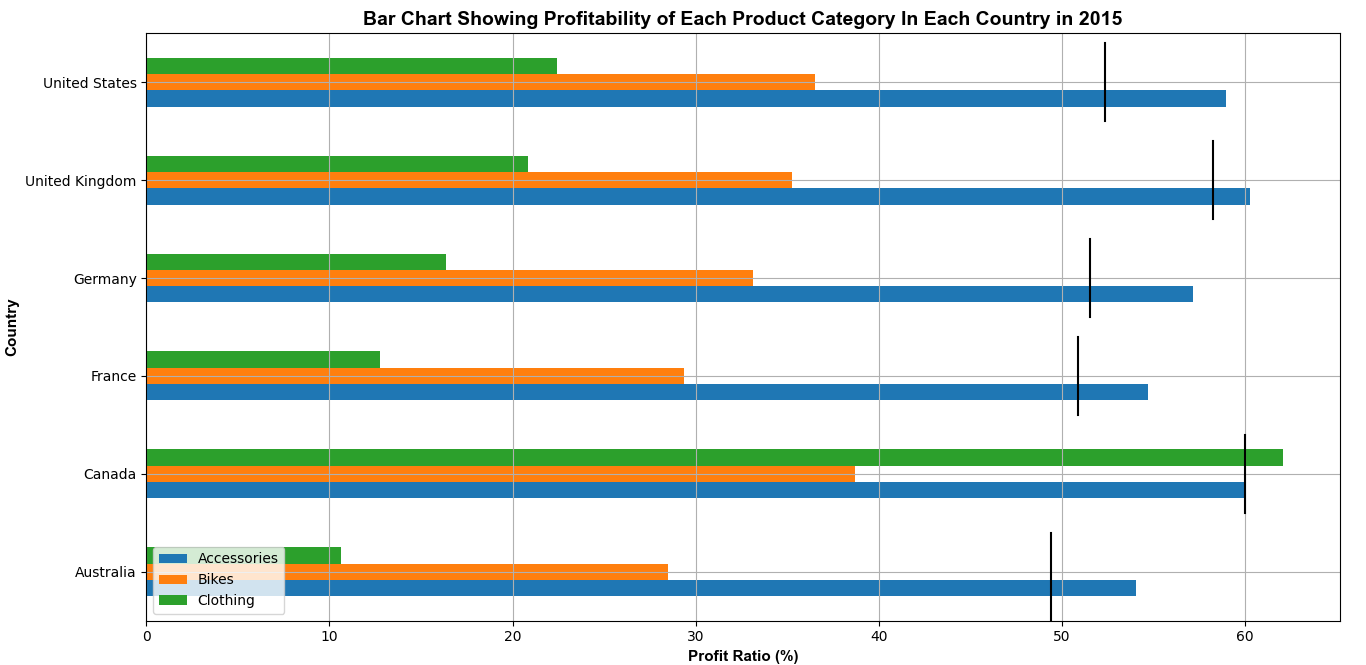
The below line plot gives the same information, however this time separated by each country.

Line Graph 4 shows:

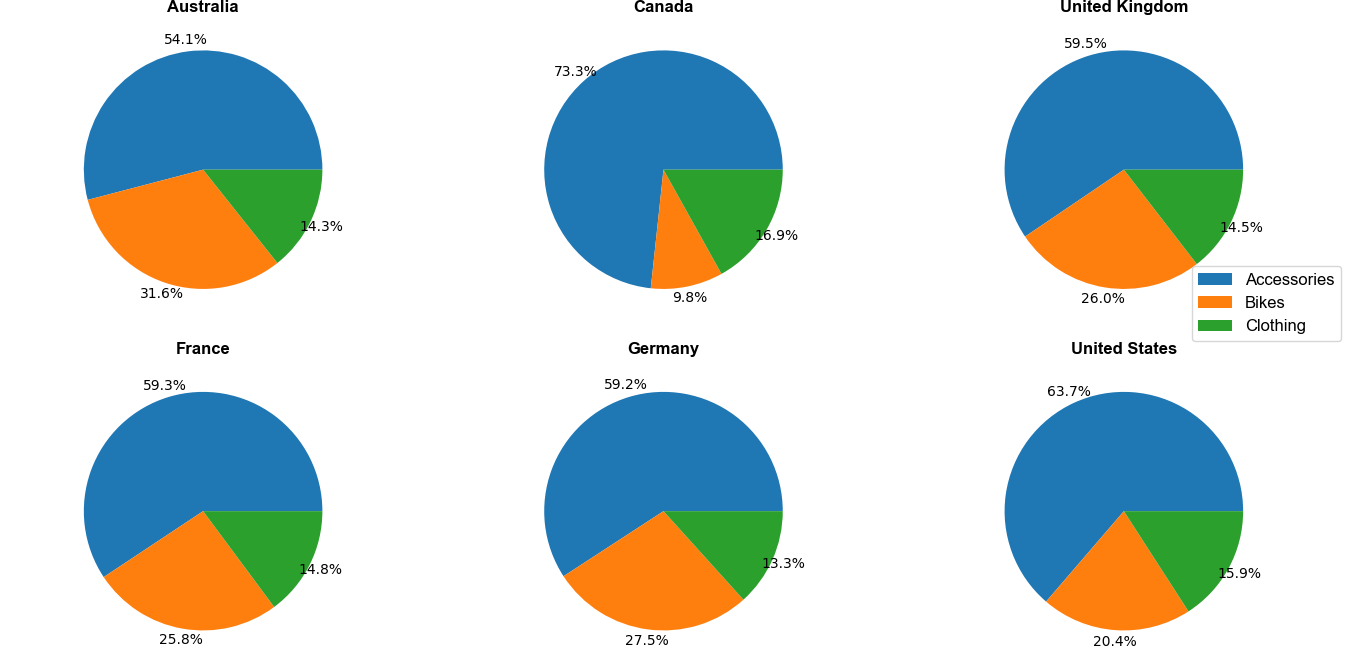
* In all of the countries, there is no place that has specific days where there are more sales or profits made. This means that none of the days should be selected over each other for targeting in the marketing campaign as none of them are likely to garner more or less sales.

**Where are Profits/Sales Coming From**

Now I know when sales/profits are being made, I want to find out where in terms of product category they are coming from. The below graph compares the profit ratio of each category in each country to see where there are differences and similarities. Where profitability of categories is higher, these products should be targeted more in the ad campaign.



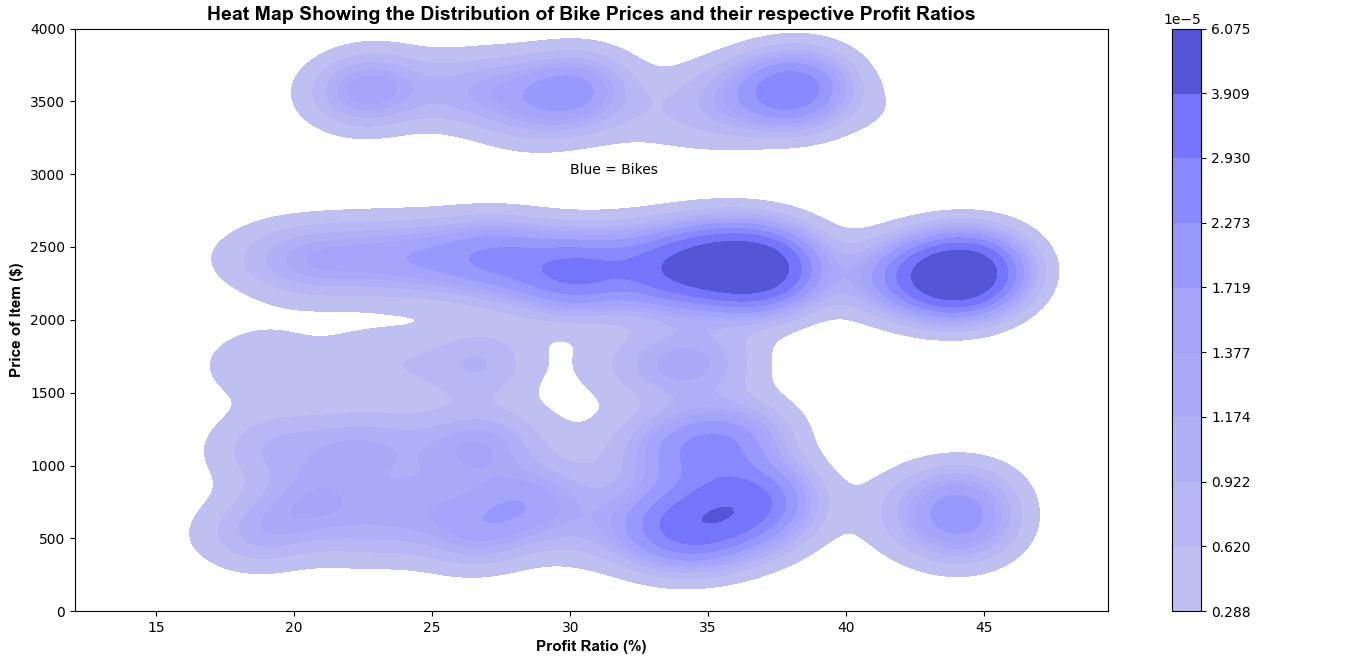
Bar Chart 1 Shows:

* Profitability of each profit category differs significantly between countries.
* Australia has the lowest median profit ratios of each product category. This at least partially explains why its overall median is also the lowest in the dataset.
* Conversely, Canada has the highest profit ratios of each category, which helps to explain why its overall profit ratio is the highest.
* Accessory sales are consistently the most profitable of all categories in each country. These should therefore be targeted above other categories in marketing.
* Clothing is the least profitable product category in each country apart from in Canada, where it is surprisingly the most profitable. This suggests it should be avoided in the marketing campaign in all countries apart from Canada.

The below pie charts give a breakdown of where the sales, in terms of product category, are coming from in each country.

Pie Charts 2 Show:

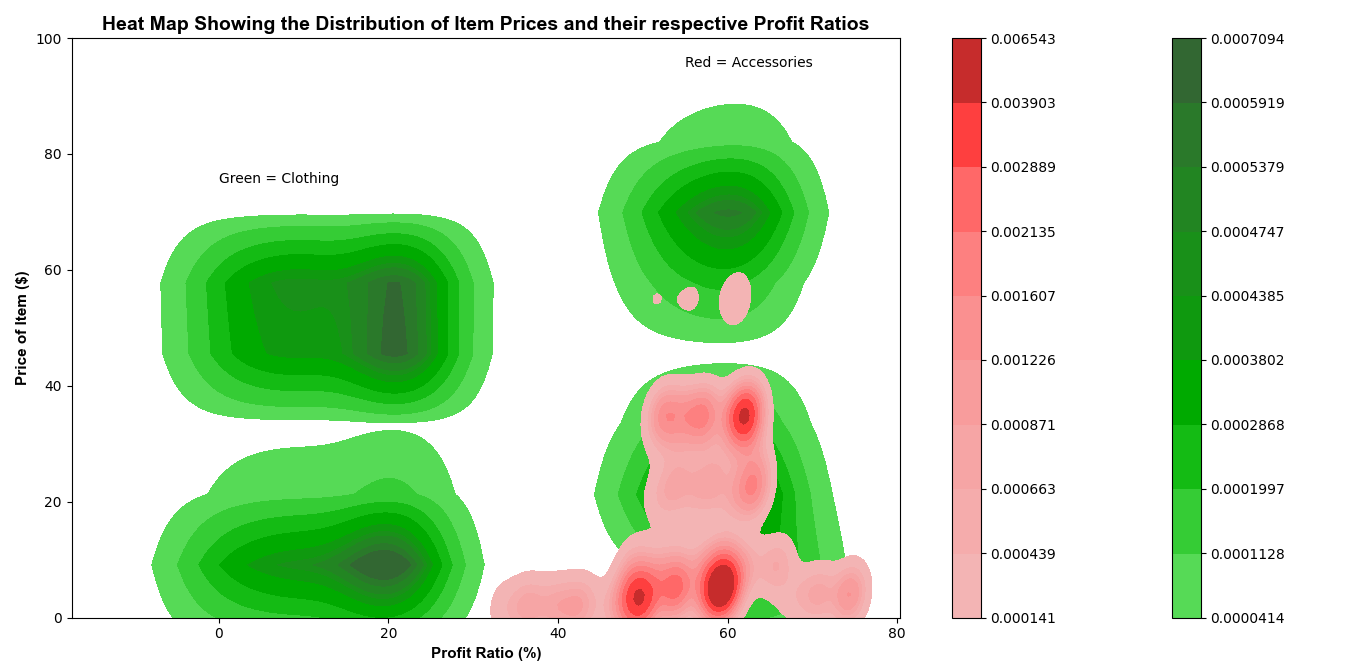
* There is quite a significant amount of variation in the make-up of sales between countries, however accessories are consistently making up the largest share of sales in all the countries. This suggests accessory sales should be targeted most in the advertising campaign.
* After this, bikes make up the second largest share of sales in all of the countries apart from Canada.

Now I know where the profits are being made, I want to investigate which products in each product category are the most profitable in terms of price. To do this I have plotted the following heat map to visualise which are the best products to market.

Heat Map 1 Shows:

* Bikes that are the most profitable are those costing between $2000 and $2500, and those from $300 to $1000. These are those in the low and mid ranges. Those of the highest range tend to be slightly less profitable.
* There is a large distribution of profit ratios within each price range. This may suggest it is worth dropping some of the current products sold in favour of the ones that are producing the higher profits.

Heat Map 2 Shows:



* The higher priced accessories have higher profit ratios, suggesting that if accessories are targeted in the ad campaign, the higher priced items should be selected.
* Generally, the higher the price of clothing, the higher the profitability of each sale.

**Where are Sales/Profit Coming From**

Now I want to dig deeper into where the sales/profits are occurring in terms of regions. To do this, I will employ the aforementioned performance variables to see which locations are performing the best/worst.

| Country | State | Profit 2015 YT: Average Profit 2015 YT | Number of Sales 2015: Average Number of Sales 2015 | Region Mean Profit Ratio 2015: Average Region Mean Profit Ratio 2015 | Profit overall | PR overall | Sales overall |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Australia | Victoria | 2.47 | 2.66 | 0.82 | 1.78 | -0.43 | 1.93 |
| Australia | Tasmania | 0.36 | 0.42 | 0.84 | 0.64 | 2.35 | 0.94 |
| Australia | South Australia | 0.81 | 0.77 | 0.85 | 1.06 | 1.15 | 0.94 |
| Australia | Queensland | 2.88 | 2.37 | 0.93 | 2.37 | 0.11 | 1.98 |
| Australia | New South Wales | 5.21 | 4.49 | 0.94 | 3.02 | -0.41 | 2.72 |
| Canada | Alberta | 0.02 | 0.03 | 1.01 | -7.49 | 0.21 | -7.09 |
| Canada | Ontario | 0.00 | 0.01 | 1.16 | -12.17 | 0.77 | -9.76 |
| Canada | British Columbia | 5.37 | 6.15 | 1.20 | 3.71 | 0.71 | 3.99 |
| France | Yveline | 0.21 | 0.37 | 0.78 | -1.54 | 1.09 | -0.39 |
| France | Seine et Marne | 0.13 | 0.15 | 0.86 | -2.16 | 1.61 | -1.87 |
| France | Seine Saint Denis | 0.50 | 0.67 | 0.86 | 0.13 | 1.21 | 0.71 |
| France | Seine (Paris) | 0.78 | 1.02 | 0.90 | 0.53 | 0.81 | 1.07 |
| France | Nord | 0.70 | 0.72 | 0.94 | 0.57 | 1.17 | 0.63 |
| France | Val de Marne | 0.03 | 0.07 | 0.95 | -5.41 | 1.47 | -3.68 |
| France | Garonne (Haute) | 0.09 | 0.07 | 0.96 | -3.26 | 1.59 | -3.76 |
| France | Val d'Oise | 0.08 | 0.12 | 0.97 | -3.03 | 2.00 | -2.24 |
| France | Loiret | 0.21 | 0.17 | 0.97 | -0.50 | 2.54 | -0.92 |
| France | Pas de Calais | 0.01 | 0.02 | 1.00 | -7.11 | 2.14 | -5.48 |
| France | Somme | 0.03 | 0.06 | 1.01 | -3.72 | 3.51 | -2.21 |
| France | Moselle | 0.14 | 0.13 | 1.03 | -1.53 | 2.39 | -1.70 |
| France | Essonne | 0.57 | 0.42 | 1.06 | 0.63 | 1.87 | 0.04 |
| France | Loir et Cher | 0.05 | 0.05 | 1.06 | -2.50 | 3.58 | -2.59 |
| France | Hauts de Seine | 0.60 | 0.47 | 1.07 | 0.65 | 1.79 | 0.14 |
| France | Charente-Maritime | 0.10 | 0.08 | 1.11 | -1.81 | 3.06 | -2.14 |
| Germany | Saarland | 0.88 | 1.14 | 0.86 | 1.91 | 1.87 | 2.43 |
| Germany | Nordrhein-Westfalen | 0.87 | 0.99 | 0.96 | -1.14 | -0.94 | -0.89 |
| Germany | Brandenburg | 0.13 | 0.09 | 1.01 | -2.67 | 1.41 | -3.43 |
| Germany | Hessen | 1.04 | 0.92 | 1.02 | 0.12 | 0.09 | -0.12 |
| Germany | Hamburg | 0.95 | 0.81 | 1.07 | 0.81 | 1.04 | 0.50 |
| Germany | Bayern | 0.79 | 0.57 | 1.12 | -1.22 | -0.50 | -1.86 |
| UK | England | 6.62 | 5.95 | 1.08 | 1.73 | -1.90 | 1.52 |
| USA | Texas | 0.00 | 0.01 | 0.80 | -13.61 | -2.00 | -11.77 |
| USA | Massachusetts | 0.01 | 0.00 | 0.82 | -10.88 | -0.72 | -12.75 |
| USA | Florida | 0.02 | 0.00 | 0.83 | -9.07 | -1.30 | -11.98 |
| USA | Utah | 0.01 | 0.01 | 0.87 | -9.61 | 0.55 | -8.38 |
| USA | Washington | 2.90 | 4.87 | 0.87 | 1.86 | -0.53 | 2.90 |
| USA | Wyoming | 0.01 | 0.00 | 0.89 | -7.98 | 2.05 | -8.75 |
| USA | New York | 0.00 | 0.00 | 0.91 | -14.36 | -1.55 | -12.40 |
| USA | Illinois | 0.03 | 0.01 | 0.92 | -8.09 | -0.89 | -9.26 |
| USA | Oregon | 1.79 | 2.27 | 1.00 | 1.61 | 0.45 | 2.09 |
| USA | Alabama | 0.00 | 0.00 | 1.08 | -16.45 | 0.73 | -11.85 |
| USA | California | 11.57 | 9.83 | 1.13 | 2.96 | -1.69 | 2.64 |
| USA | North Carolina | 0.00 | 0.00 | 1.17 | -14.85 | -0.01 | -11.37 |
| USA | Arizona | 0.02 | 0.00 | 1.18 | -7.56 | 0.52 | -10.85 |
| USA | Ohio | 0.01 | 0.02 | 1.18 | -10.15 | -0.19 | -8.55 |
| USA | Montana | 0.00 | 0.01 | 1.22 | -9.14 | 2.46 | -8.17 |
| USA | Missouri | 0.00 | 0.01 | 1.25 | -11.26 | 0.68 | -10.00 |
| USA | Kentucky | 0.00 | 0.00 | 1.26 | -15.00 | 1.10 | -11.78 |
| USA | Georgia | 0.00 | 0.00 | 1.27 | -12.89 | 0.11 | -11.40 |

The regions with the most sales/profits above average were California, England and British Columbia. These would therefore be great candidates for marketing, as these places would likely see the greatest returns from the advertising. The poorest performing locations in this regard were Alabama, Kentucky and North Carolina.

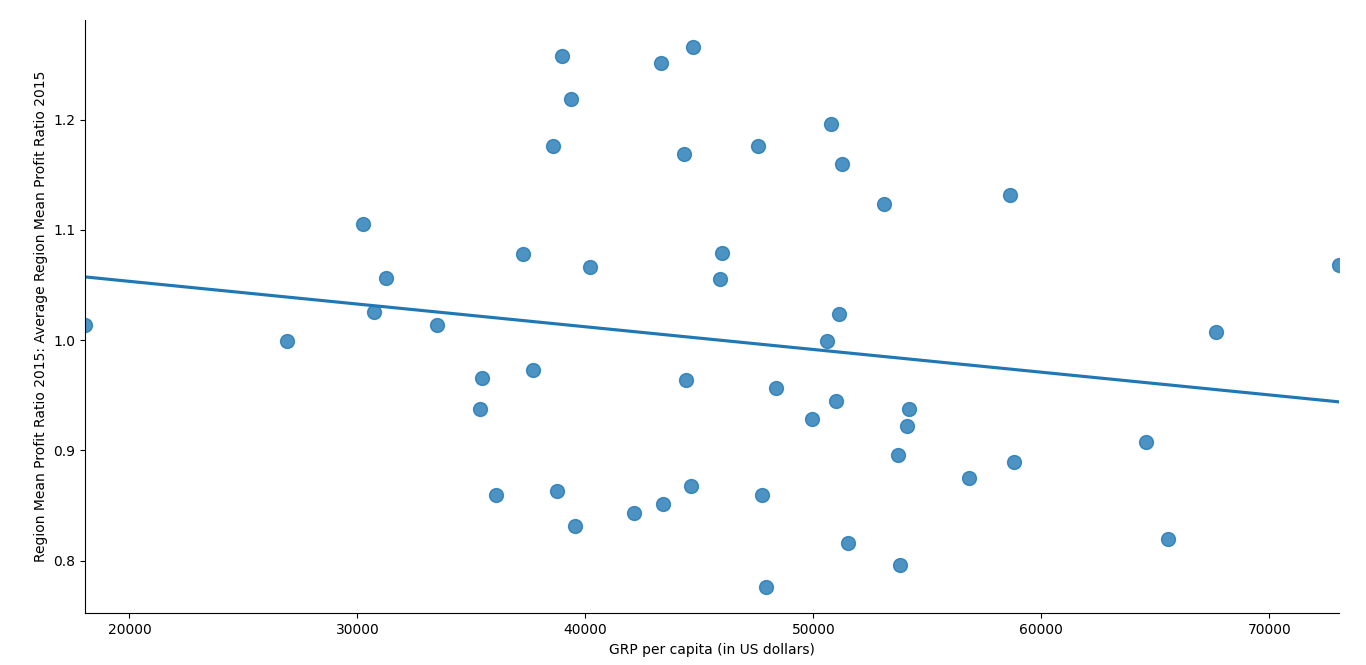
The profit ratio comparison gave a very different view of performances in each region, with the better performing places in terms of profit ratio often doing the worst in terms of profit/sales and vice versa. Georgia, which has one of the worst profit and sales performances, was the most profitable in terms of its profitability variable of all the regions, and Kuntucky with the second worst before had the second highest profit ratio. Depending on the shareholders' aversion to risk, places like Georgia could be marketed in, however I believe it's best to focus on the places already providing high sales and profit.

The profit performance parameter was highest in British Columbia, and was shortly followed by New South Wales and California. This suggests that even when population and GRP p/c is taken into account, these regions are still performing very well, so should be invested in according to the data. On the other hand, Alabama, Kentucky and North Carolina are all performing the worst, so these places should be avoided by management.

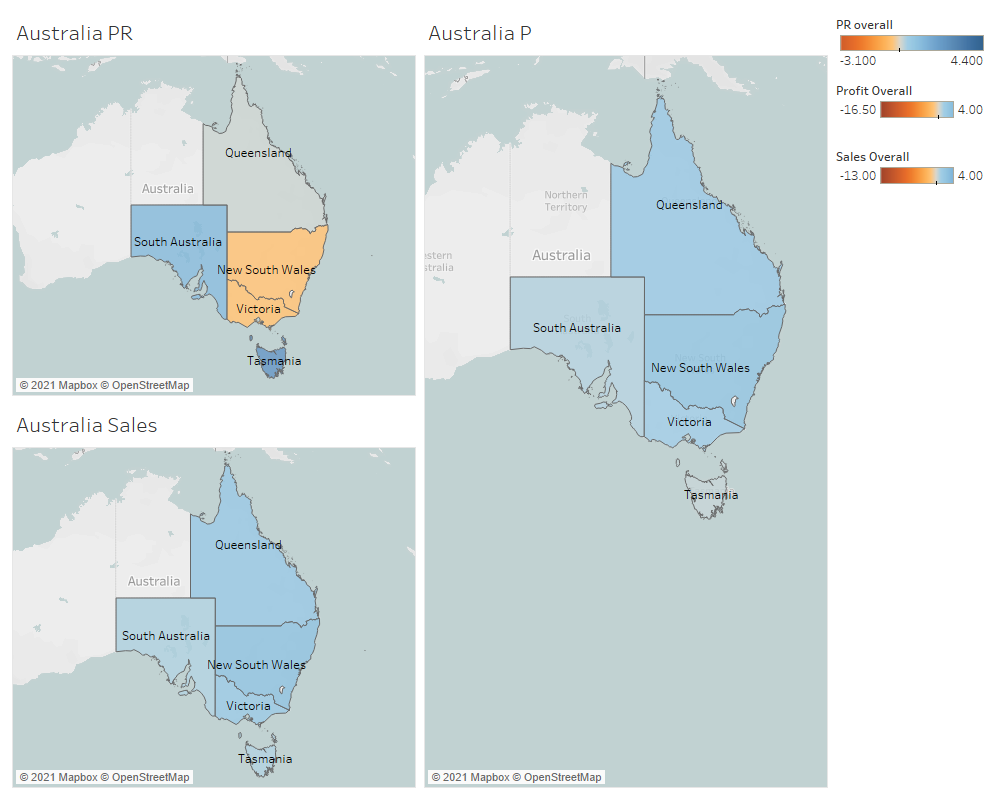
The sales performance parameter was highest in British Columbia, and was shortly followed by Washington and New South Wales. This suggests that even when population and GRP p/c is taken into account, these regions are still performing very well, so should be invested in according to the data. On the other hand, Massachusetts, New York and Florida are all performing the worst, so these places should be avoided by management.

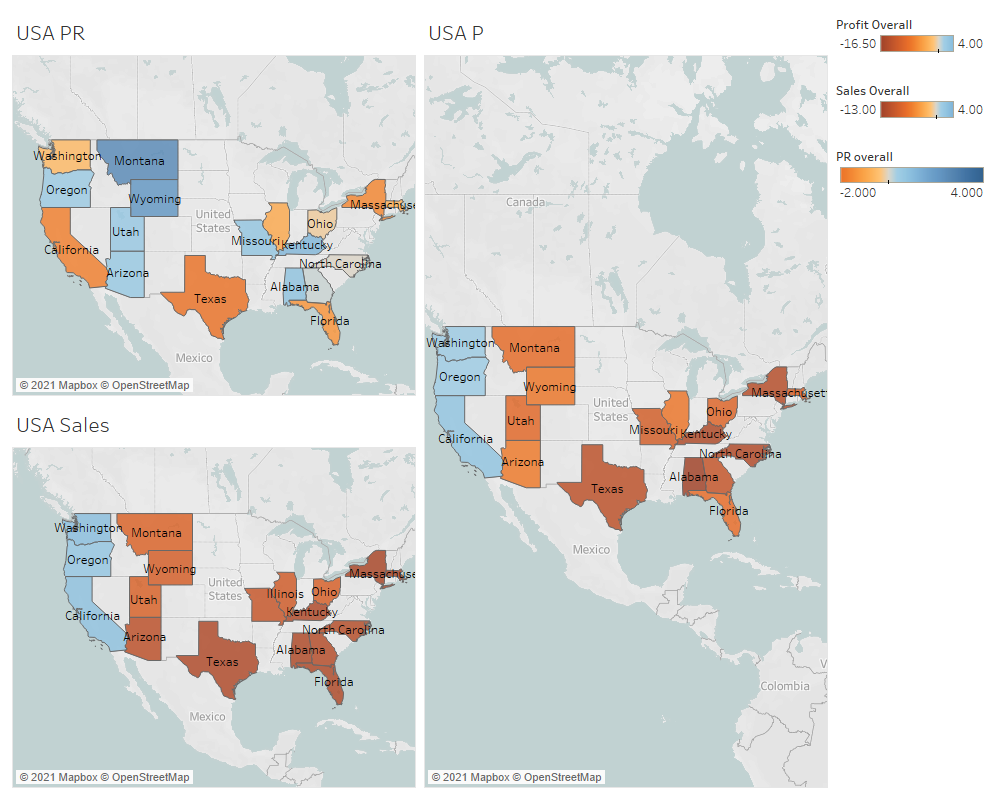
The profit ratio performance parameter was highest in Loir et Cher, and was shortly followed by Somme and Charente-Maritime. On the other hand, Texas, England and California are underperforming in this regard. This is unexpected given England and California are doing so well in profits/sales, therefore this warrants further investigation. To look into this I will check to see whether profit ratio is positively correlated with population and GRP p/c as expected. If it doesn’t, this suggests that there is a more complex relationship going on, which would invalidate the profit ratio statistic.

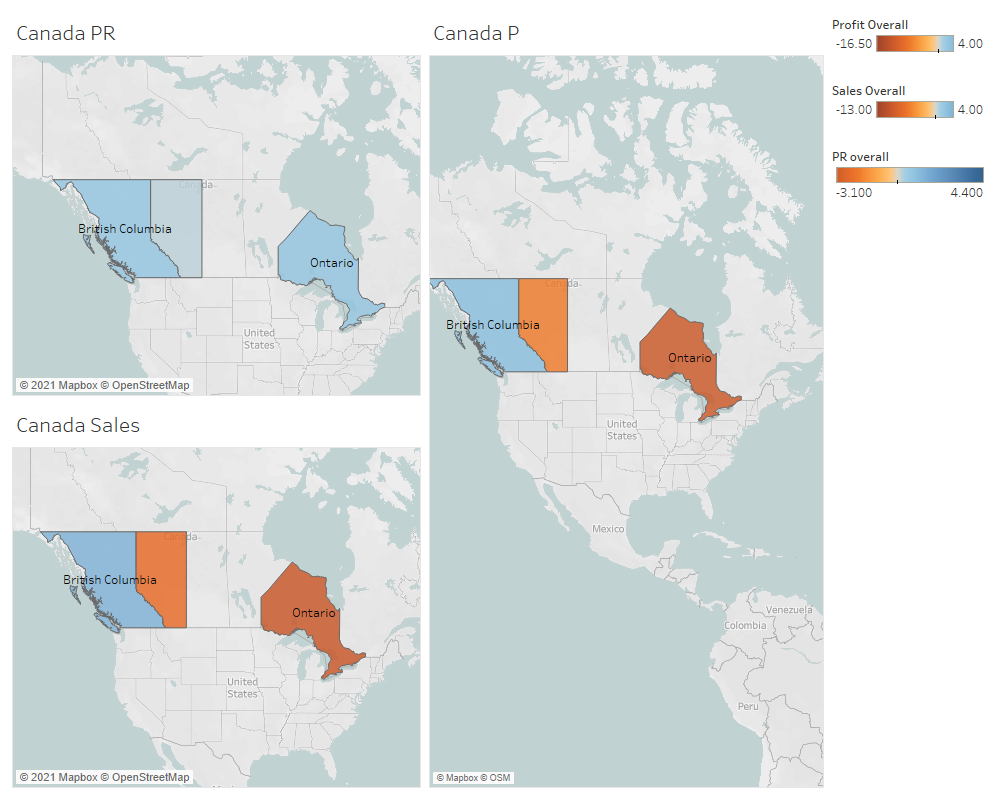
The plot below confirms my worries about the profit ratio statistic. Clearly my thinking, where I thought higher GRP p/c would increase profit ratio is flawed, as there must be something else going on. I progress with the rest of my region analysis however will ignore the profit ratio statistic.

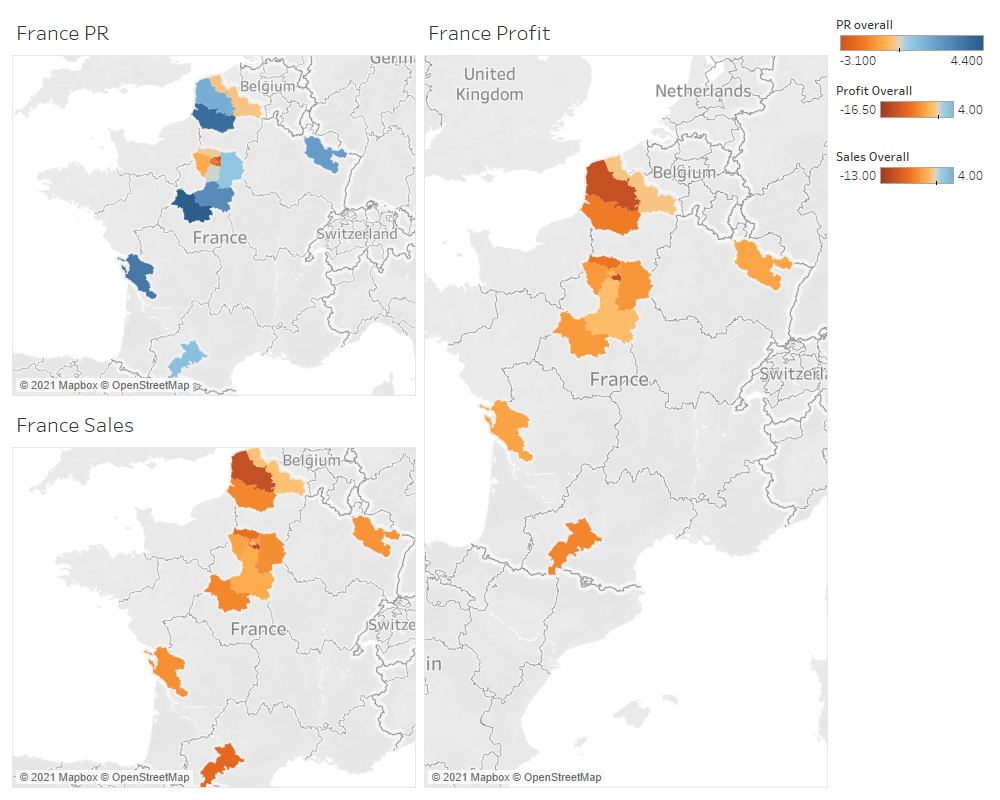


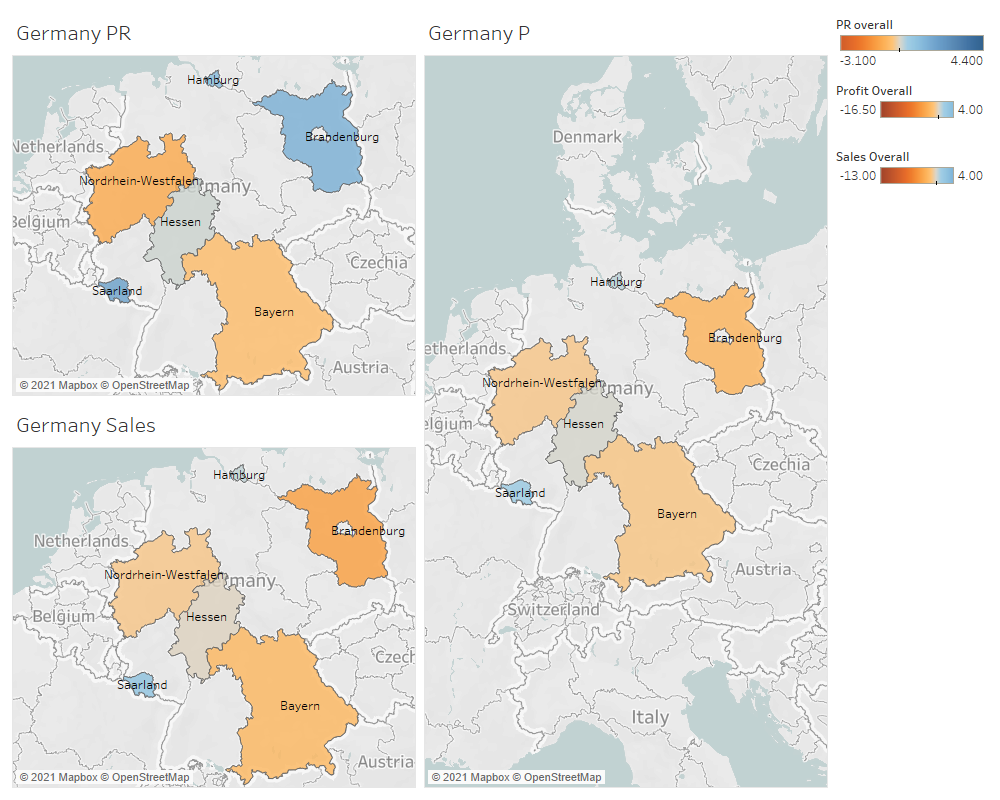
Below are the tableau visualisations I made with the data to show visually the information above. The regions in blue are the ones performing the best, and those in shades of orange are the ones performing worst.

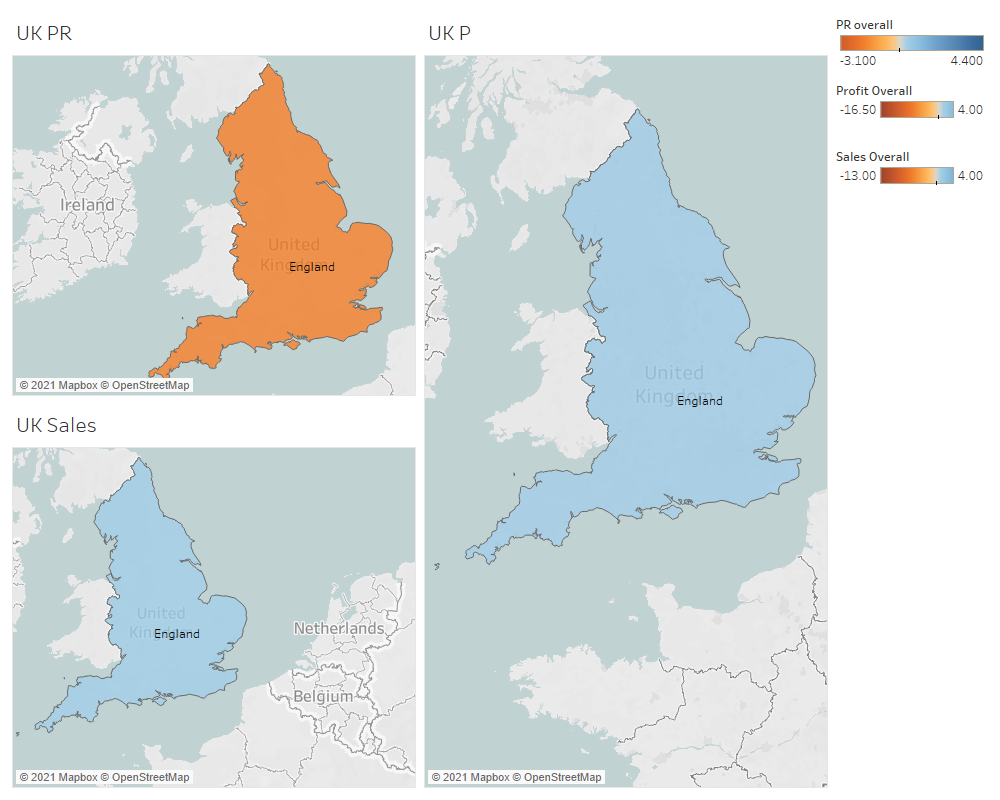


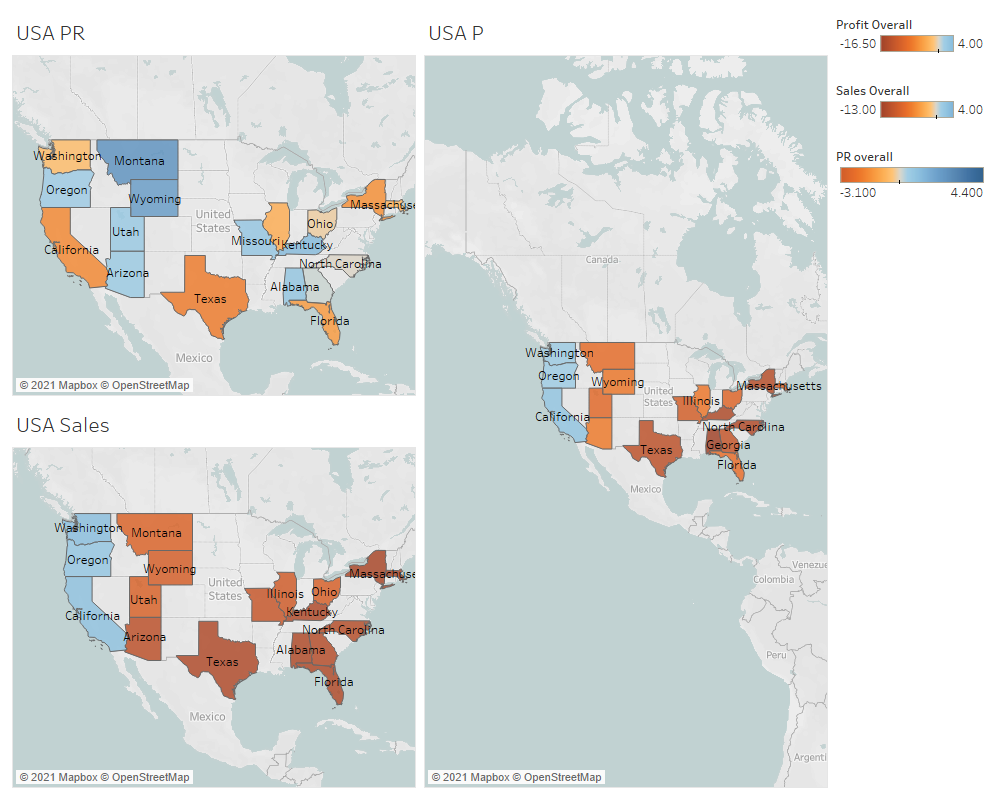












**Analysis Conclusion**

* The best performing regions in terms of profit were Califtornia, England and British Columbia. Likewise, these places also performed very well in terms of sales. However, British Columbia is the stand out region here, with an excellent profit ratio performance compared to the others in the pure and log statistics, suggesting BC would make an excellent choice to market in.
* There is no significant difference between the profit ratios of either gender in any of the countries. This suggests that neither of the genders should be favoured over the other in any of the countries based on this statistic. There was a slight difference in sales make-up between genders in Canada, with men forming a greater proportion of the sales compared to women. Hence men should be moderately targeted in Canada.
* Age has none to little effect on the profit ratio in all the countries apart from Australia, where there is a slight positive effect. This means that no age groups, apart from the older demographic in Australia should be targeted over other age groups.
* Sales are however correlated with age, with 27-34 being the age group with the greatest share of sales of all the customers. This demographic should therefore be targeted in the advertising campaign.
* May, June and December were all good performing months in the dataset, suggesting these should be the times of year that are selected for marketing purposes. The worst performing months were August and July, hence these months ought to be avoided.
* No day of the week is more profitable than any other in any of the countries, therefore,if there is an option for advertising on specific days, it should not be taken, with advertising in specific months more favourable.
* Accessory sales consistently have the highest profit ratios in all the countries, and they also were the greatest component of sale. They should therefore be advertised over other products to maximise returns. The best performing accessories are those with the highest prices, hence these should be selected in the advertising.

**-Share Phase**

In this section, the data I have analysed will be shared with the stakeholders. This will be done using a powerpoint presentation (shown in a video).

**-Act Phase**

This is where, given my main findings, I recommend to the stakeholders my recommended course of action.

Recommendations

* Advertise in British Columbia with a slight focus on men. In terms of age, 27-34 year olds should be the primary target. There should be focus on highly priced accessories to maximise the returns of the marketing. This should be done in either May or December to maximize the returns from the investment of marketing.
* Other suitable regions include England, California and possibly Hamburg.