



# CS5603 DATA VISUALISATION COURSEWORK



Nassim Benatia  
1238355

## Introduction

The Dataset chosen was an excel dataset detailing orders from a retail company and consists of three tabs in excel: List of Orders, Order Breakdown and Sales Targets. The sheets were joined in R to create the “Edited\_Visualisation\_Dataset”. This merged dataset was also used alongside the original dataset in the visualization itself.

VARIABLE NAME	TYPE	DESCRIPTION & USAGE
ORDER DATE	Date	The date data field will be necessary for the time series element of the visualization
STATE	String	This is State as in locational data. The State field will allow us to drill down to state level in terms of our other variables in the visualization.
SALES	Number	This is the price and/or reward for the product sold. Sales will be summed per month/quarter and be visualized in different contexts depending on the desired insight.
PROFIT	Number	This is sales minus costs. Profits will be visualized in different contexts in the visualization depending on the desired insight.
TARGET	Number	This is referring to the desired sales per period. Target data in the database is given on a monthly basis.
SEGMENT	String	This variable refers to the type of customer of which can take three values: Home office, Consumer or Corporate.
PROFIT MARGIN	Number	This is Sales divided by costs and will be shown at different levels of granularity in the visualization.

### Persona: Senior Management of a Retail Organization

The persona chosen is not an individual but rather a group of individuals. As an example of different persona amongst our group of senior management, senior ranking sales and marketing staff will both take insight from Sales Data. Sales staff will wish to see how they rank against their targets and marketing staff will wish to know of well performing markets in that they can target them for marketing campaigns.

The reason I have chosen a group of individuals is to satisfy my **non-functional objective** of having a visualization usable by other departments. This objective was chosen in the name of efficiency. Companies in light of industry 4.0<sup>1</sup> have been gearing towards central databases to put different departments on the same page in real time. This has come around from the inability of senior managers to filter down information through their organization at the same pace that changes are happening in light of new information.

Naturally following on from the previous non-functional objective is the **non-functional objective** of having Integrated visuals. This has the benefit of seeing data in different contexts which is a necessary tool in the exploration and potential discovery of new information. Both of the above objectives is in harmony with the advice of researchers such as {17 Marr, Bernard 2015;} and {19 Cattell, Rick 2011;} who use the study of data value chains to derive more value from the exponentially increasing availability of data across firms and other organizations. Reference is made to Edward Currys and Millar and Morks Data Value Chains which mention data integration as a central part in the extraction of efficiency out of growing databases.

#### Functional Objectives:

1. To meet sales targets for each region
2. Profit maximization.

#### Planned Questions:

1. Which markets should we target to maximise profits?
2. How have we measured up against our targets

Both of the above planned questions were edited from the proposal in light of feedback

#### User Requirements:

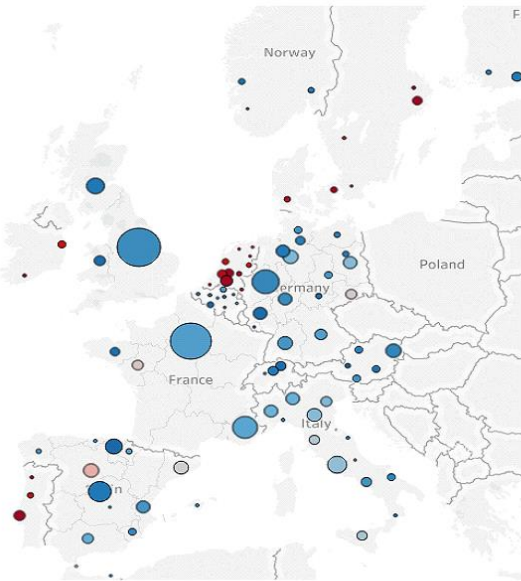
- Be able to compare profit margins across different timeframes and categories of customer.
- Be easily able to compare sales and targets in different periods.
- User friendly design: the interaction element will therefore need play on the natural curiosity of the user. In other words there needs to be minimal effort in realising and making use of interactive features.

## Design

Initial Design (from proposal) shown below on the next page:

---

<sup>1</sup> Industry 4.0 is a name given to the current trend of automation and data exchange in manufacturing technologies. It includes cyber-physical systems, the Internet of things, cloud computing and cognitive computing. Industry 4.0 is commonly referred to as the fourth industrial revolution. (Wikipedia)

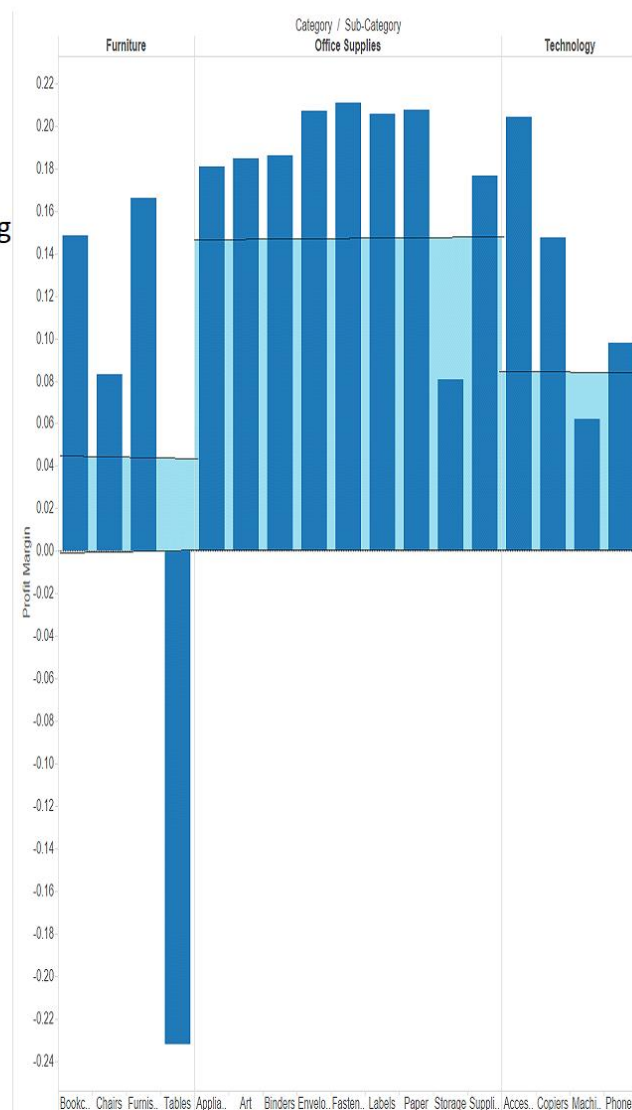


Regional distribution of sales. Circle size indicates sales amount; blue and red colours signify profit and loss margins respectively.

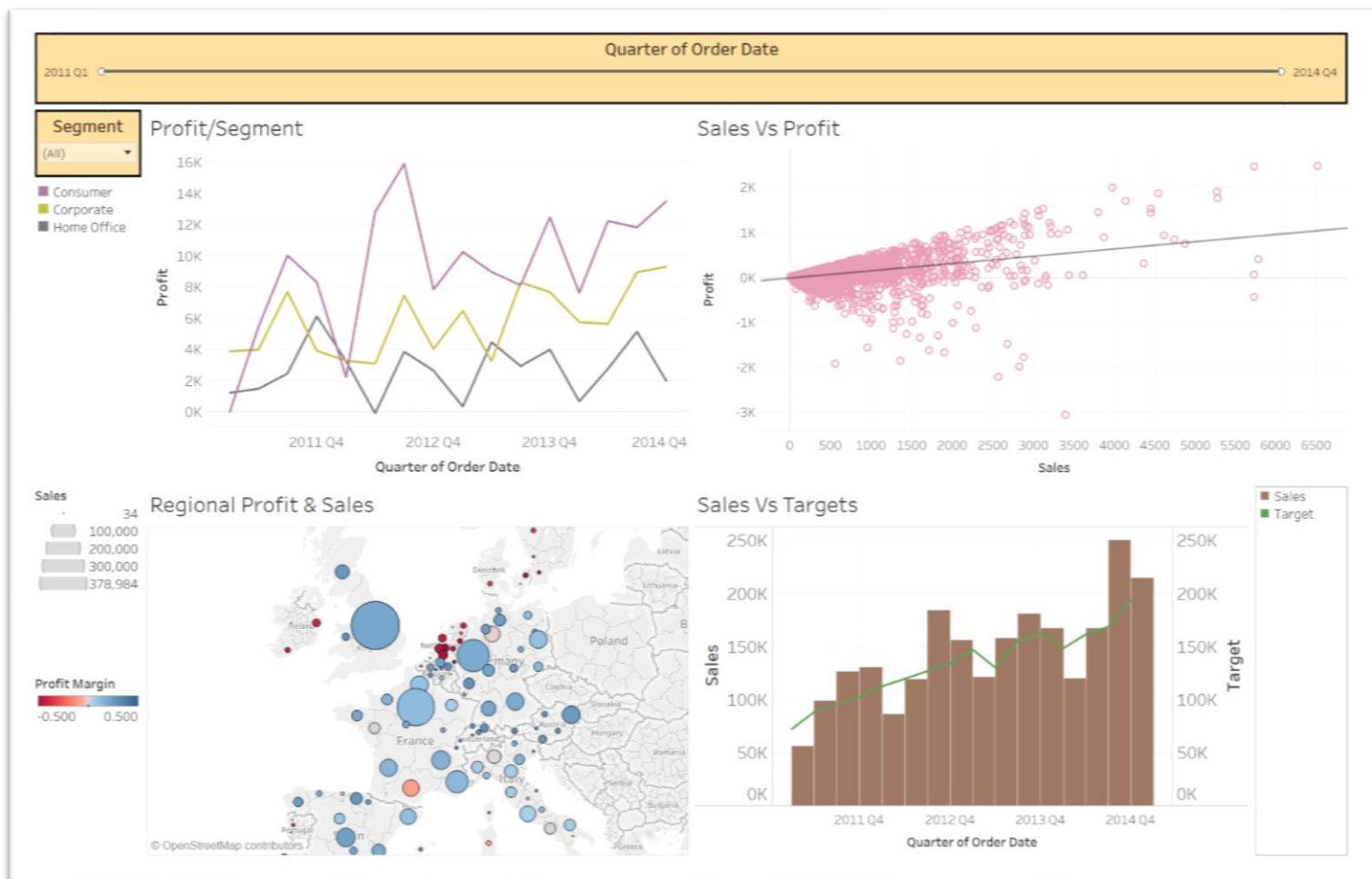


Sales over time compared with targets for the month. Graph will have filter to see required dimensions.

Graph will show profit margins for different categories which can be used to see which categories are performing best and which ones need to be scraped.



## Final Design



From the initial design, the Category/Sub category graph was scrapped. This is because the graph values were aggregated over the three year period and therefore from a business perspective I didn't think it of much use. I also couldn't see a clean way of presenting that information on a timeline and in sticking with three – four chunks of information at a time {Page 2; Lecture Two – Encoding for Perception} I thought it better to remove it.

The original line graph was changed to a dual combination graph with sales represented as bars and targets represented as a line. This was of more use as the eye can more easily distinguish between the bars and the line and the relationship between the two.

The new graphs are two. The first was the Profit/Segment graph, which shows the profit level over time for each customer segment. This aids our profit maximisation objective as the best performing segments can be prioritised.

The last graph is a scatter plot of profit vs sales for each customer. This is useful as we can see what product price level maximises profit margin. We can use this also to see if there is any relationship between product price and profitability. This was done by fitting a regression line to the data. The persona can use this information to make decisions on what products to sell, as will be seen, higher priced products tend to bear more profit.

## Implementation

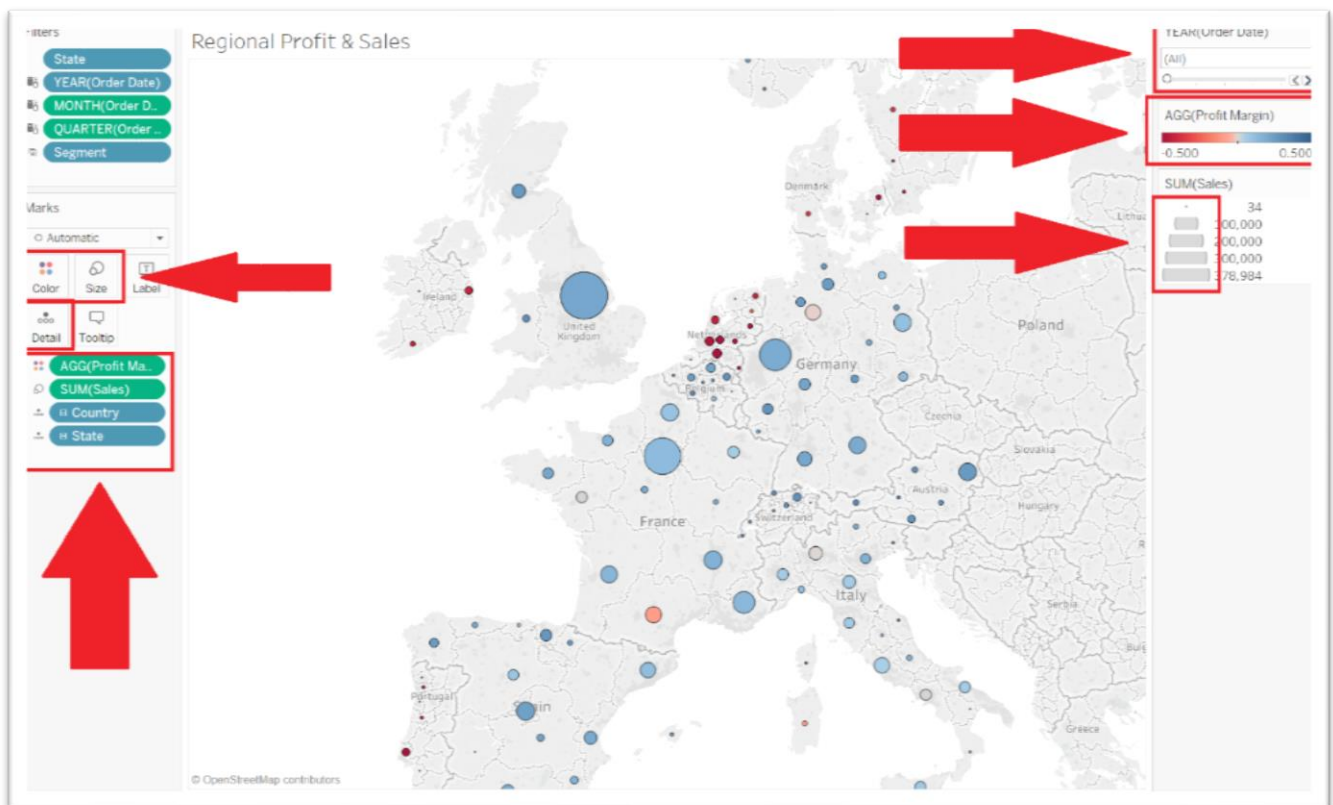
My design was implemented using Tableau 2018.3.

## Loading the data

The sheets from the original excel dataset had been merged to include a time element for each product sold. I had an issue however with the targets variable as the monthly targets were duplicated across the different items in the dataset. This resulted in incorrect sums in the visualisations. I corrected this problem by uploading the original dataset (the excel file with three different sheets) and used the data blending feature in tableau which allows tableau link datasets without the need of unnecessary duplication of values.

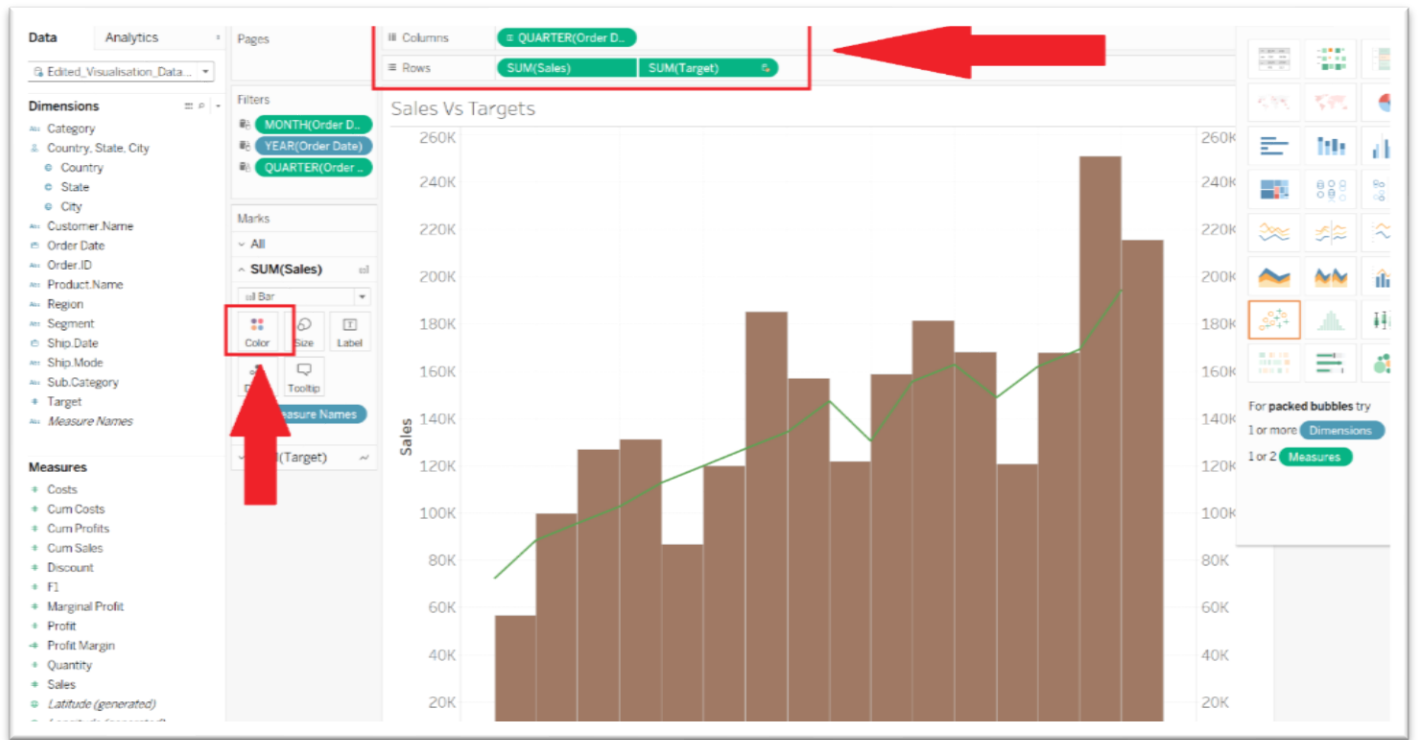
## Regional Profit and Sales

I dragged the state dimension into the sheet which mapped out the different regions onto the world map. I then dragged the sales variable onto “size” as seen below so that the size of the resulting circles were related to the amount of sales in the region. Next I wanted to see the profitability of each state. At this point I didn’t have a profit margin variable and so I created it by creating a calculated field which can be accessed through the analysis tab. I then proceeded to show profitability by dragging the profit margin variable onto colour, in the marks tab, to show through changes in colour the profitability of each region. I then proceeded to edit the colours, making anything under 0 profitability show as red and anything above 0 show as blue. The lower the profit margin is below zero the darker the red, and the higher it is above 0 the deeper the darker the blue. I lastly added the sales and profit margin legends through the analysis and legends tab. I think this graph best takes advantage of pre-attentive perceptions {Page 6; Lecture 2; Encoding for Perception} as size and colour are “processed pre-attentively into iconic or sensory memory”.



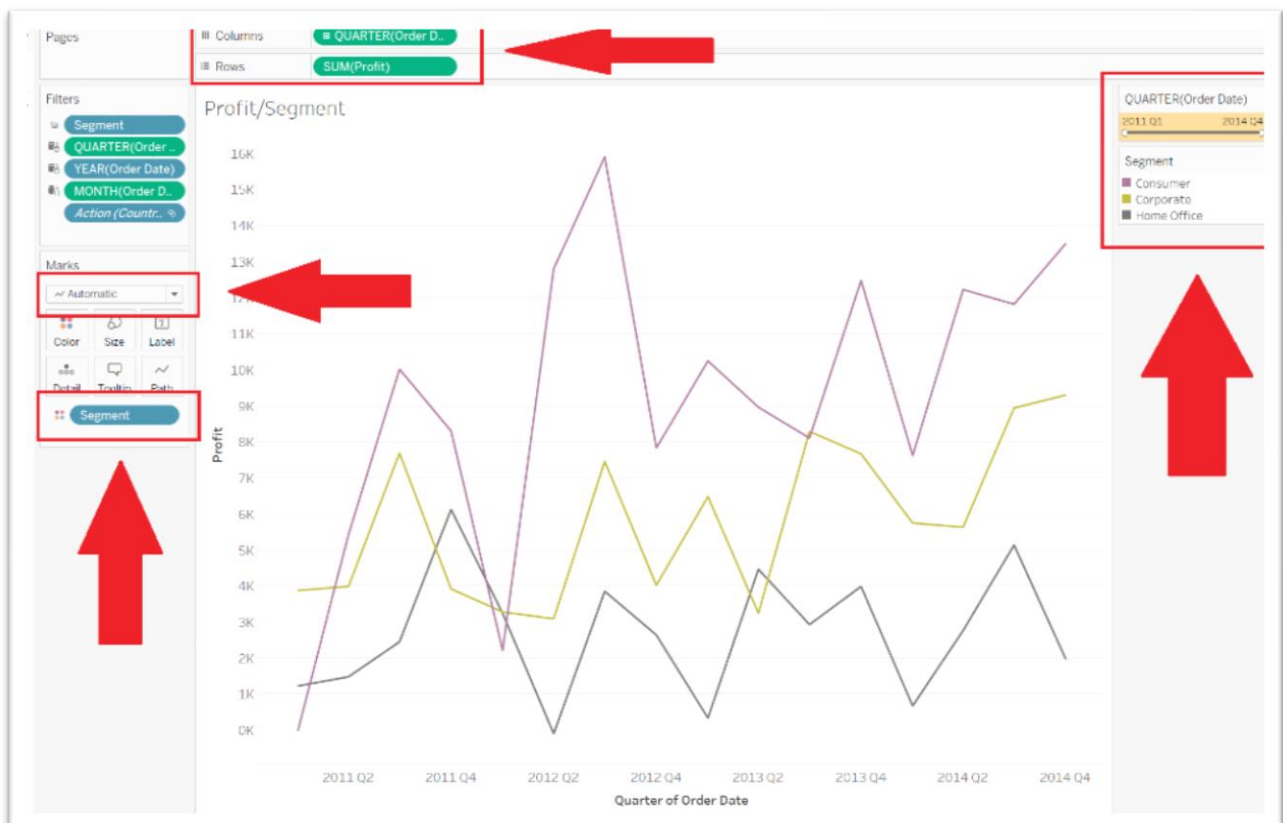
## Sales Vs Targets

To get this visual, the first step was to put the order date in the columns tab and then arrange by quarters. The sales and targets variables were dragged into the rows tab and summed. The show me tab was clicked and the dual combination chart was selected. I then edited the axis to put them on the same scale.



## Profit Per Segment

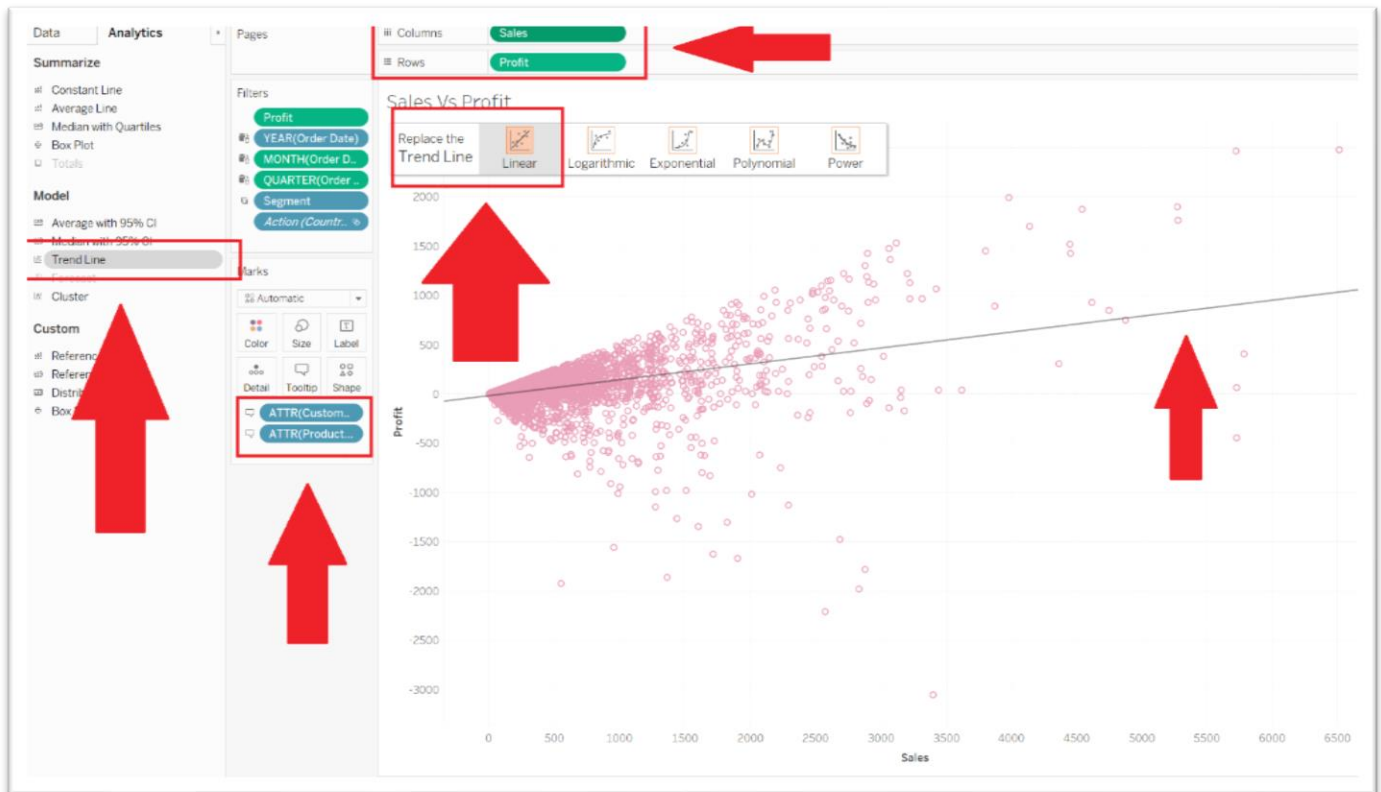
For this chart I dragged order date into columns to give the time-series element of the graph. For uniformity the data is drilled down to individual quarters from the yearly calendar. I then put profit into the rows to get a profit. Tableau automatically gave me a line chart which is what I was looking for. I then drilled down to individual segments by dragging the segment variable into the colours tab and this gave me three separate trend lines for each segment. a filter was added for individual quarters, and I changed the mechanics of the filter so that a range of dates can be selected.





## Sales Vs Profit

I dragged the sales variable into columns and the profit variable into rows. I created the scatter plot by changing the sales variable into a dimension. The same was done for profit to arrive at the scatter plot. I added the regression line by clicking analytics tab then by dragging "Trend Line" into the visual and clicking the "Linear". The regression equation, R Squared value and P value can all be seen in a tooltip by hovering the mouse over the line. I changed the colour to a light pink as I thought it more neutral than the blue which came before and which has already been used in the Regional Profit & Sales visual. Customer/Product Name variables were dragged into tooltip to give more information when the mouse is hovered over an individual point.



## Walkthroughs

My visuals were created with my persona, questions and functional non-functional objectives in mind. With profit maximisation in mind, I needed create a visual where I could easily see which market was performing best. I therefore put order date (time series element) on the horizontal axis and profit on the vertical. For each segment of customer I used an individual trend line and the result allowed me to view the performance of each segment over time. This visual also allowed me to add a time filter and select any desired timeframe to view the visual over, which is desirable where one need see both short and long term performance and to gauge trends over specific periods. This is of especially good use to treasurers/financial accounting senior staff whose investments and risk appetite will depend on the past performance of the firm over days, weeks, months and years.

Having now gauged the most profitable segments, I thought it necessary to gauge the most profitable regions. The dimensions of most importance was Profitability and State. Tableau made easy work of this as by merely dragging state onto the screen I was presented with a map of Europe (where my company operates) and small circles which I knew could represent profitability through their colours. The aim of seeing most profitable regions was therefore realised. Tableau let me use the size of the circles to see sales amounts for the different regions also. This lead to an unintended discovery, where-by a trend can be seen between regional profit and regional sales. We can see that the bigger circles (insinuating higher sales) tend to be the bluest (signifying higher profitability) therefore the regions

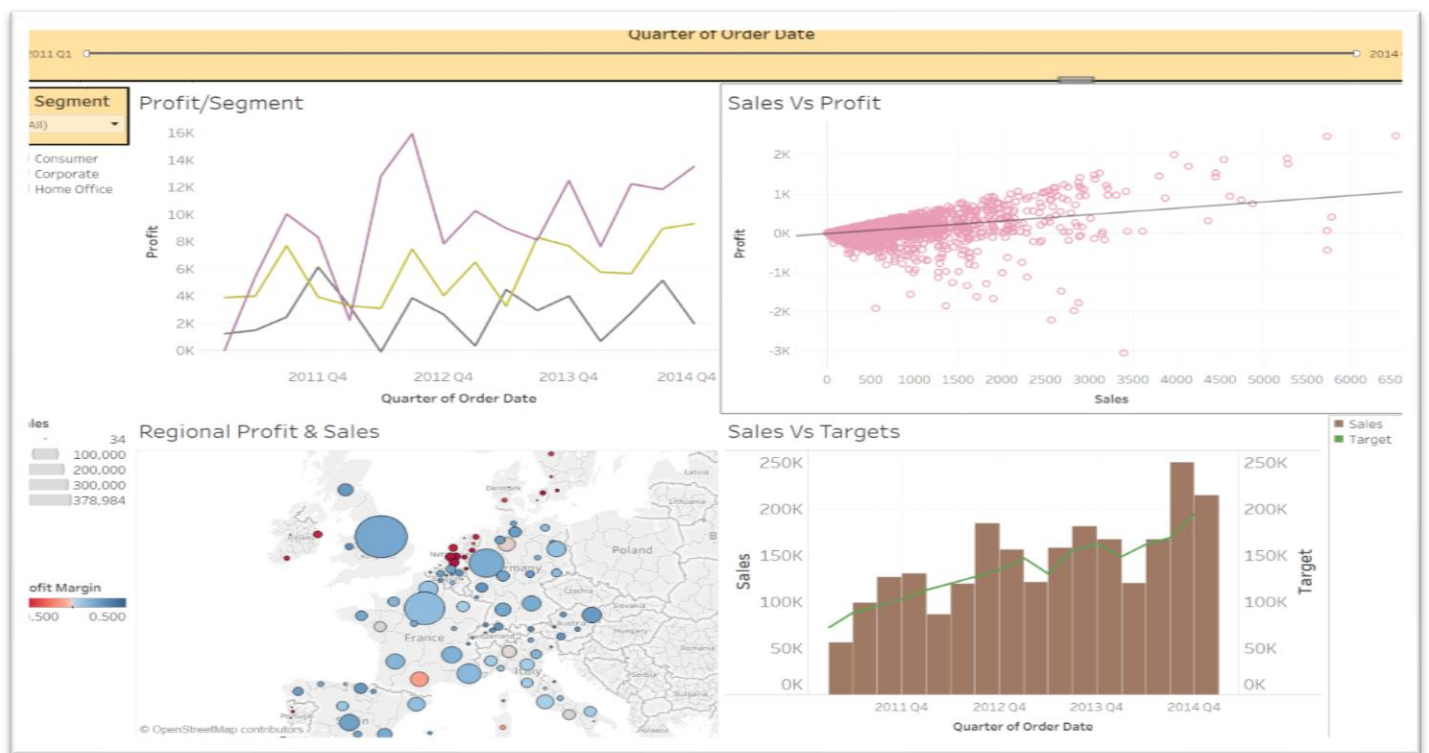


that shop more often tend to buy the more profitable products for the firm. An insight that can be taken from the senior staff personas therefore is that they should target areas with higher sales if they want to increase their profitability. This fits in with the functional objective of profit maximisation.

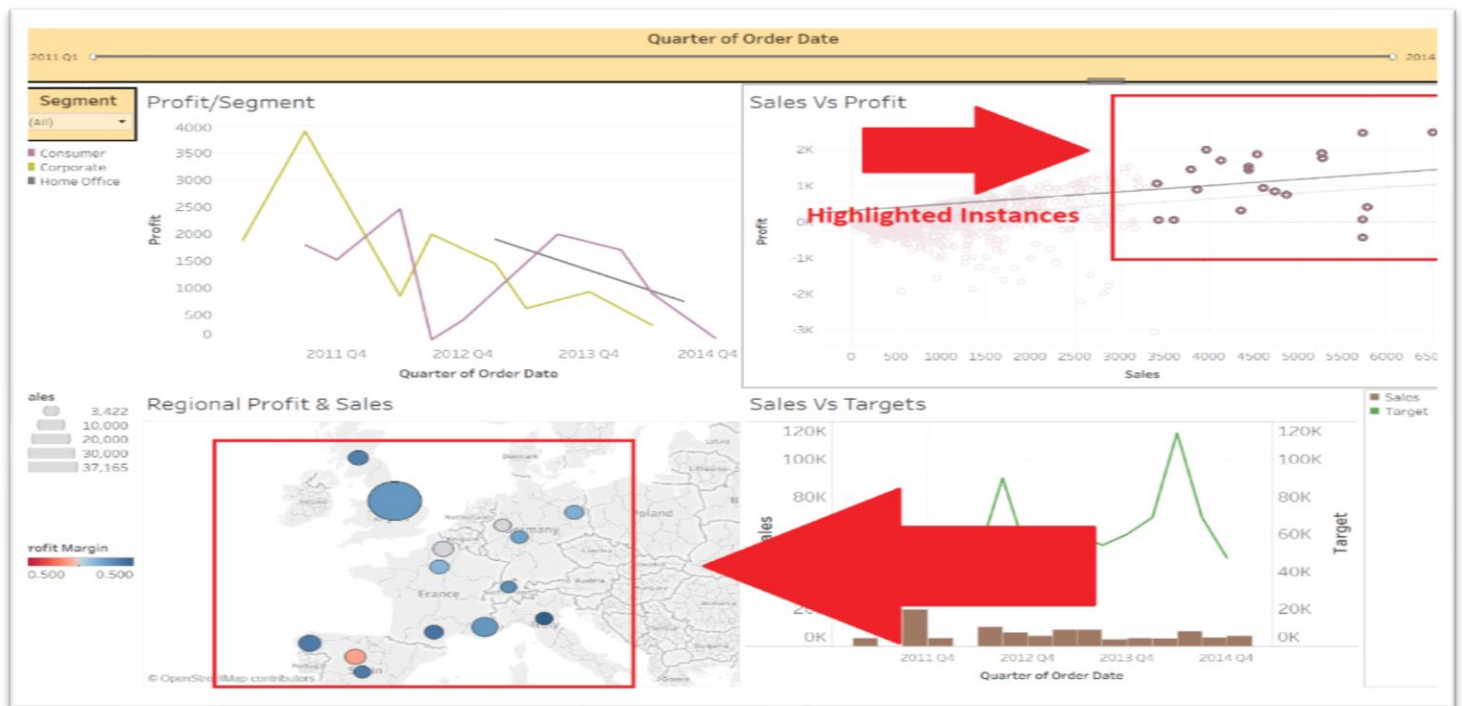
With this new insight in mind, I began to wonder how the profits differ for different priced products. This would be of use to the senior staff personas also as they can see which products are more profitable, then by highlighting that area in the dashboard they can see how it changes the other graphs also. I therefore took it upon myself to create a scatter plot of the different prices of sold products vs their profitability. This is where the Sales Vs Profit Graph was created. Upon plotting the data I couldn't see a general upward or downward trend, but could see that the cheaper the product the more the company sells of it, as there is a high concentration of values on the left side of the graph. I could also see that the more expensive the product, the higher the variability of the profit. At this point I thought it wise to plot a linear regression line to the graph to tease out any potential relationship. The result was an increasing line with a significant P value, suggesting that generally profitability per product increases the more expensive the product is.

With the above idea in mind that more expensive products are more profitable, I proceeded to use this graph as a filter for the dashboard, and highlighted the most expensive instances, to see how the Regional Profit and Sales Graph changes. The output can be seen below:

First showing with no highlights



Now with highlighted instances



We can see in the first screenshot above that the highlighted instances in the Sales Vs Profit (Scatter) Graph have outputted the regions of highest profitability and sales. We already know that these regions have the highest sales and profits. In light of the above information however we now know that the reason for this is that those regions tend to buy more expensive products and we know from the scatter plot that the more expensive products are also more profitable, therefore those regions are more profitable. Now the Senior Staff know, not only to target those regions if they want increased profitability, but also to target those regions with more expensive products also.

Having answered my first question it was time to answer the second "How have we measured up against our targets". This led to the creation of the "Sales Vs Targets" graph. Having had sales targets for each category of product, I initially wanted to find a way to compare the combined sales of each category to their respective targets. I could have done this by utilising a bar graph with different coloured bars for sales and category, however this means the data would have been summed over the whole three year period. This is not of much use to our personas who will need to be able to see sales vs profits on a short term basis to see if their performing. Therefore a time series element was necessary. I intended to solve this problem by using a line graph with different trend lines for each category, however this would result in 6 trend lines, a sales and a target line for each category. This would have broken the 3-4 chunks of information rule. In order to stick to good practise I used combined sales and targets to have only two variables to deal with. Instead of using two trend lines I used a bar graph for sales and a trend line for targets. This works well as the variables are distinguished from one another.

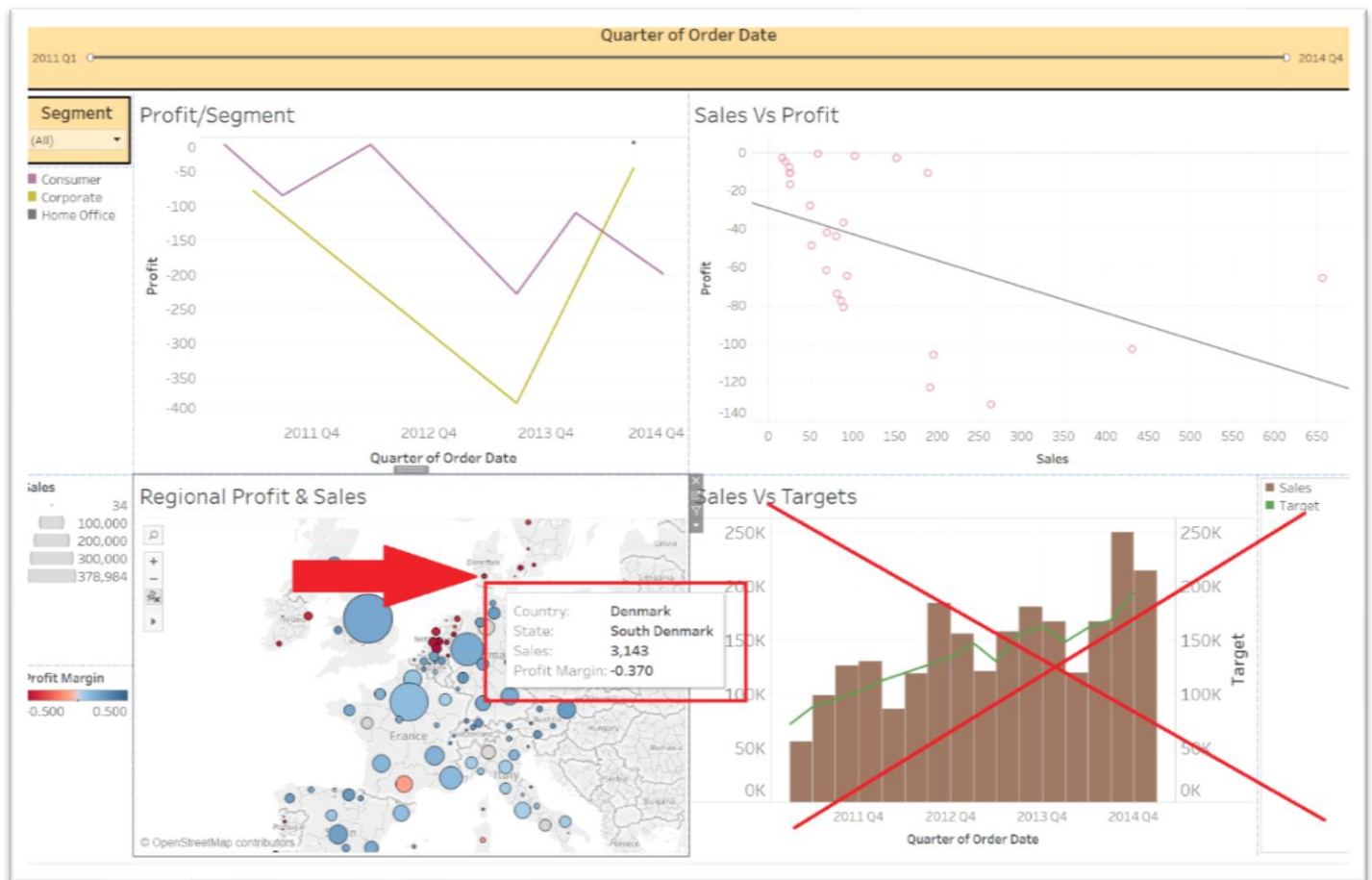
The last thing to do was to deal with our non-functional objective of integrated visuals.

In order to integrate my visuals I had to think about which graphs were appropriate as sources and which filters within the graphs were appropriate also. Before this however, it was clear that date was to become a filter, as the different personas amongst the group would have different interests in terms of the date. I therefore added a date filter and customized it to have the option to view a range. This filter can be seen across the top of the page, and applies to every graph of the page.

The next filter I added was one for Segment type, which can be seen at the top left of the screen. I applied this filter to all graphs except the Sales Vs Targets graph as that graph doesn't have targets per segment and therefore doesn't apply.

I applied another interactive element where a user hovers his mouse over a particular state in the Regional Profit and Sales Graph. The remaining graphs (except Sales vs Targets) are filtered to state level. For the Sales Vs Targets

graph there is no state level target available from the data therefore I excluded it from the target sheets. The filter was applied using the actions option which can be located within the dashboard tab. An example of usage is present below:



## Evaluations/Discussions

Initially I was tasked with finding a dataset with enough relative features to use in the visualisation. I had chosen a different dataset with my group beforehand but came to realise that its not just about having relative features but also having well structured data that can translate to fruitful comparisons in visualisations. An example of this is where I made my own dataset “Edited\_Visualisation\_Dataset” in which the targets variable ended up having had duplicated instances which wasn’t useful. This lead me to discover Tablaeu’s blending function which meant I could use multiple datasets in creating visuals. This was of much use as It meant my datasets needn’t be merged and had alleviated the problem of duplicated instances. This is definitely a highlight I will take away with me, as database operations can become tedious and unwieldy.

Using the regression as an example, I came to realise after that some visualisations rely on prerequisite mathematical assumptions which I didn’t test beforehand e.g. correlations, normality of the variables etc. In future I will perhaps explore how to use tableau for data exploration and see if there’s any live tool that can alert the user to violations of these assumptions, at which point the visualisations may not be useful. I will also explore if tableau has tools for data exploration e.g. principal component analysis. With principal component analysis in mind, I may explore Tablaues tools for dimensionality reduction also, as this is something I haven’t explored in my work.

### Picking my Persona, Questions, Functional Objectives & User Requirements

Upon seeing that the data was retail business oriented, I know immediately that the visuals will have profit maximisation as an objective. I had yet to write them down before exploring visuals with my variables, and this lead to me writing questions based on my visuals and not the other way around. In reality I was trying to show my knowledge of tablaeu’s tools instead of answering a question which is not what will happen in a business context. The feedback showed me that my questions were unclear as it removed that bias from the work.

I also realised that instead of writing a question for analysis, I was answering my profit maximisation question in the question itself. Therefore I simplified the questions to include only one focus per question e.g. profit maximisation and achievement of targets.

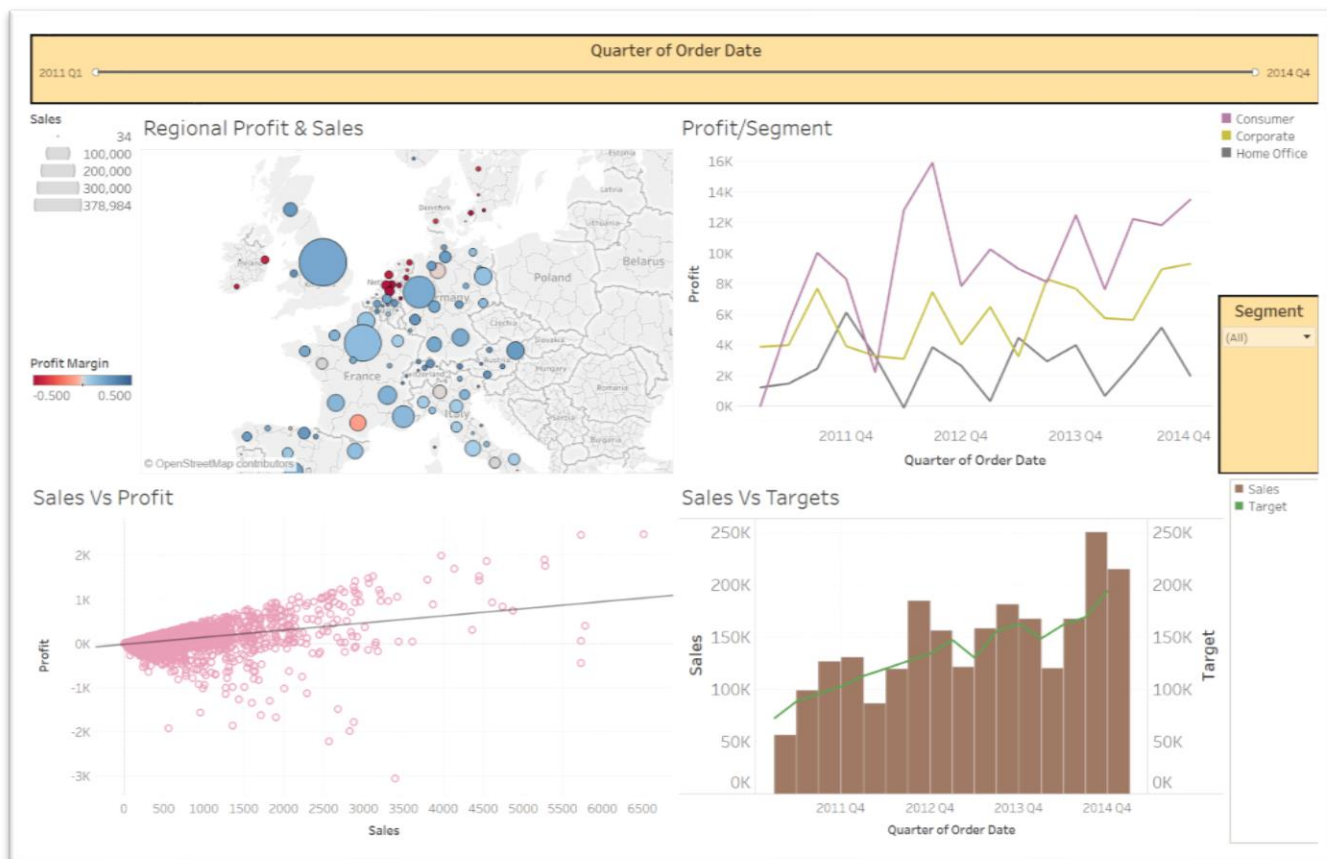
### Evaluation of Implementation

Although tableau is a great and clever resource, I came to realise that its not completely fool proof for a few reasons which are highlighted below:

If you use a view as a filter on the dashboard (for example clicking a country on the Regional Profit & Sales view), there is a problem in that if you have other filters in the sheet of the view and not on the dashboard itself, it will still link. This may be confusing for the end user as their data may be filtered without them knowing, as there's no viewable filter in the dashboard to tell you the data has been filtered. To solve this I made sure any active filters on each sheet were also available on the dashboard itself.

Another issue is that I had a global action filter on the dashboard with the "Retail Profit and Sales" graph being the source. When I hovered over say, England, the "Sales vs Targets" graph would adjust for sales, but not targets, as the data doesn't provide a means of knowing target per state. This could have lead to Change blindness (Rensink, 2002) taking place where the user is so focused on the change in sales that they don't realize that targets haven't adjusted. I solved this by taking the latter graph off of the target worksheets so that it doesn't adjust according to the filter. This can be confusing for the user though as sales doesn't change, therefore it would be good to know of a means of deleting desired worksheets in response to a filter. This is something I would like to find out in future.

Upon looking at the mark scheme I was reminded of appropriate view coordination. I therefore chose to edit the dashboard putting regional sales and profit top left, as this is the way we read in the target language (English) and I wanted this to be the first view that meets the eye. When the user then explores by naturally hovering their mouse over an instance of the view, there is an immediate interaction filter which results in a change in the two other related views, which prompts the users natural curiosity (one of my user requirements). Also, with {Woods, 1984} visual momentum concept in mind, I wanted Profit/Segment to come after Regional Profit & Sales so that the user can keep the profit dimension in mind as they naturally follow the visual across. The resulting view is shown below:



I also highlighted the global filters in a light neutral color to prompt the user to interact with the filters. As an evaluator point however this could conflict with the intention of first seeing the regional profit and sales graph and could lead the user to assume that all filters are global and need be accessed via some toggle or drop down as in the “quarter of order date” and “segment” filters.

A last evaluator point is that I missed out a graph detailing profit margin for different categories. This is no doubt an important feature in answering my first question and I could have included a graph on this however I believed anything more than 4 views would have become clutter. I therefore had to choose between this graph and my Profit/Segment graph. In the end I chose the Profit/Segment graph as I thought it of more use to study the trends of specific customers rather than specific products. In a real application of this problem I would have likely consulted the end users/Personas on which of the two would have been of more interest. in future I will likely explore how to include more views without overwhelming the target audience.

I think overall I did well to maximize data ink {Page 6; Lecture 4: Presentation} and enhance data pixels {Page 7; Lecture 4: Presentation} . I think I can definitely improve on using display space effectively {Page 10; Lecture 4: Presentation}. One way to do this in future might be through Tableaus’ story telling feature, where I can help with cognition by telling a story through the data, this way having more visuals to play with.

### **Bibliography**

- Rensink, R.A. (2002) Change detection. Annual Review of Psychology, 245-277.
- Woods, D. (1984) Visual momentum: a concept to improve the cognitive coupling of person and computer. International Journal of Man-Machine Systems, 21(3), 229-244.
- Marr, B. 2015, Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance, John Wiley & Sons.
- Miller, H.G. & Mork, P. 2013, "From data to decisions: a value chain for big data", It Professional, vol. 15, no. 1, pp. 57-59.
- Cattell, R. 2011, "Scalable SQL and NoSQL data stores", Acm Sigmod Record, vol. 39, no. 4, pp. 12-27.
- Curry, E. 2016, "The Big Data Value Chain: Definitions", Concepts, and Theoretical,