Business Analytics & Data Visualization

Toolkit Portfolio

Sami Bashir

Dalhousie University

Table of Contents

Chapter 4 SAP Analytics Cloud

Chapter 5 Tableau

Chapter 6 SAP Predictive Analytics

Chapter 7 SAP Analytics Cloud

Chapter 8 Executive Summary

Chapter 1 Microsoft Excel PivotTable & PivotChart

1.1 About Microsoft Excel PivotTable

Sometimes, Data Analysis and Visualization do not always require complex Business Intelligence (BI) software such as Tableau or Power BI. Excel PivotTable and by extension, PivotChart offers a simple, albeit limited, alternative to perform more simple analyses and visualizations. Datasets can be added to Excel in the form of a spreadsheet and a PivotTable can be inserted to filter that data set for more specific analysis. These can then be supplemented with the PivotChart add-on to visualize the filtered data and derive valuable insight from it. This allows analysts to quickly summarize large datasets and uncover patterns within them to help future decisions for decision-makers, proving especially useful for smaller organizations.

1.2 Dataset & Research Questions

Building upon the tutorial provided in the first week of class, Excel PivotTable will be used to analyze the provided data set pertaining to financials for the company known as Global Bikes in the form of an Excel workbook (.xlsx). This dataset contains 18 Columns and 408646 records including key financial variables such as revenues, net sales, cost of goods sold in addition to descriptive variables such as customer descriptions for the time period 2015-2019. The data provided refers to two geographical locations: Germany and the United States. Whilst the dataset was quite comprehensive, I feel as if the analysis could have been even better if the types of Sales Channels were included. Using PivotTable and PivotChart, I will attempt to analyze sales trends for GB and determine their overall profitability and compare profitability across two geographical locations by answering the following questions:

i. Which year had the highest revenue and what was the revenue during that year? How did the revenue compare between the two countries?
ii. How did the revenues for both countries compare for each of the years of data collected? What was the trend in revenues for each individual country?
iii. What were the overall gross margins for each year? Which year had the highest gross margin and how did they compare between Germany and the US?
iv. What were the top 3 customers for GB in the US across the period of data collected? How did their contributions to the overall revenue vary over time?

1.3 Applying Analytical Tools & Results

1.3.1 Which year had the highest revenue and what was the revenue during that year? How did the revenue compare between the two countries?

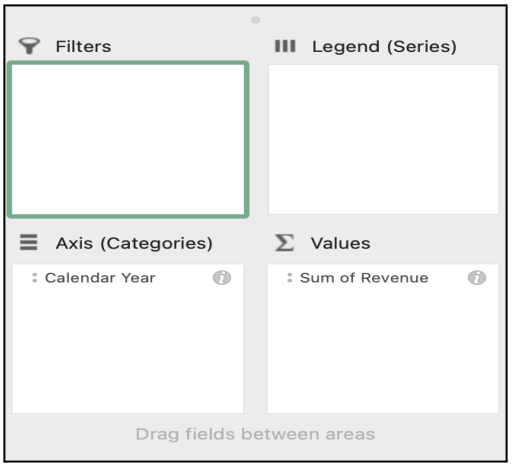


Figure 1: PivotTable Builder

First, a PivotTable was inserted for the data set. In the PivotTable builder, *Revenue* was added to Values and *Calendar Year* to Rows. The number format was changed to *Currency ($)* and the decimal places kept as zero so that the data for revenue could be seen as a financial metric that is more relevant to the overall analysis.

The *'Row Labels'* dropdown arrow was clicked and *Calendar Year* was selected. *'Sort by'* was then set as *'Sum of Revenue'* and the order selected was *Descending* so the revenues for each year are arranged from largest to smallest and as seen in the figures below, the years 2015 and 2016 had the largest revenues.

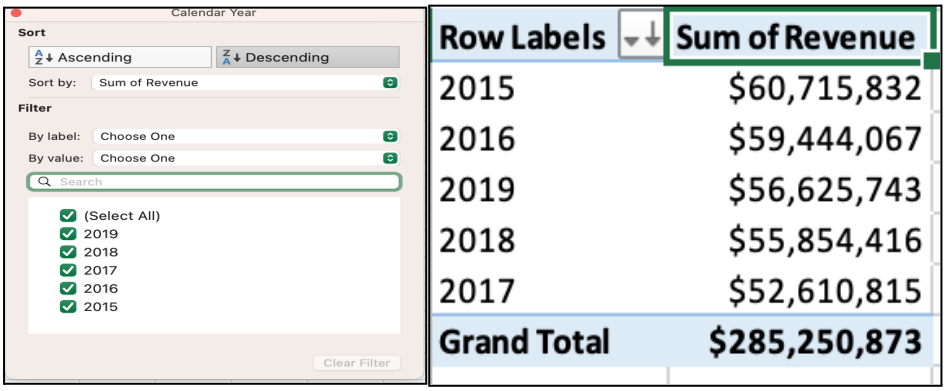


Figure 2: Total Revenues for GB in Descending Order

Lastly, *'Country Desc'* was added to Rows and *Calendar Year* was removed. Similarly, *'Revenue'* was added to Values. Following this, a PivotChart was inserted and the type was set to *Pie Chart*. Both the table and the visualization showed that Germany clearly had the larger share of the revenue during the period of data collected by over $40 million. Perhaps the US division could gather some insight from the German division to reshape their own processes.

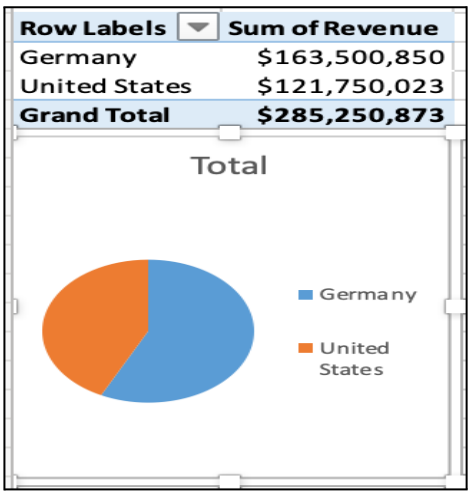


Figure 3: Total Revenues of US and Germany

1.3.2 How did the revenues for both countries compare for each of the years of data collected? What was the trend in revenues for each individual country?

After clearing all the filters in the PivotTable builder, *Calendar Year* was added to Rows, and *Revenue* to Values. *Country Desc* was then added to Columns. Then under *Row Label*, the sorting was done by *Calendar Year* and done in *Ascending* order. This enables a side-by-side comparison of each country's revenues for individual years starting from 2015 to 2019.

Sum of Revenue	Column Labels		
Row Labels	Germany	United States	Grand Total
2015	$30,951,630	$29,764,202	$60,715,832
2016	$31,445,649	$27,998,418	$59,444,067
2017	$31,765,084	$20,845,731	$52,610,815
2018	$34,191,746	$21,662,670	$55,854,416
2019	$35,146,740	$21,479,002	$56,625,743
Grand Total	**$163,500,850**	**$121,750,023**	**$285,250,873**

Figure 4: Revenue comparison between DE and US for years 2015-2019

Following that, it was important to insert a PivotChart and change the chart type to a *Line Graph*. Germany had a steady rise between 2015-2017, and then a larger increase between 2017-2018. Conversely, in the US there was a steady decline between 2015-2016 and then a sharp fall from 2016-2017. There was a slight rise between 2017-2018 and a slight decline between 2018 and 2019. The U.S division has concerning trends in terms of revenue generation.

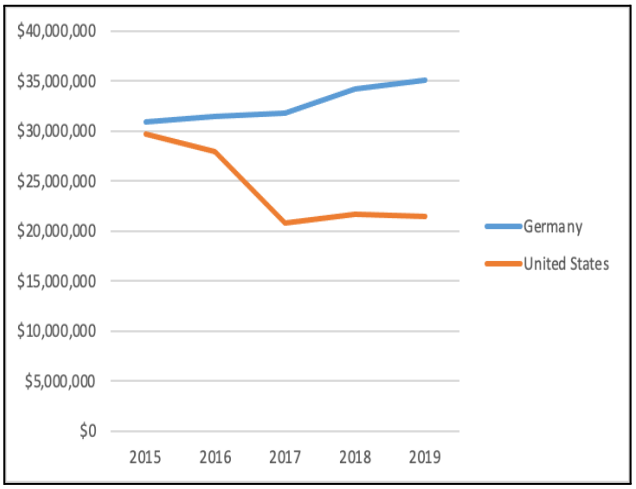
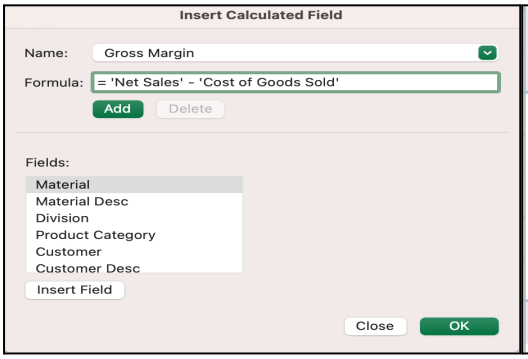


Figure 5: Trends in revenues of GB in US and DE

1.3.3 What were the overall gross margins for each year? Which year had the highest gross margin and how did they compare between Germany and the US?



Sum of Gross Margin	Column Labels		
Row Labels	Germany	United States	Grand Total
2015	$ 15,559,662	$ 13,332,102	$ 28,891,764
2016	$ 16,029,431	$ 12,774,037	$ 28,803,467
2017	$ 16,413,786	$ 9,650,979	$ 26,064,765
2018	$ 17,699,837	$ 10,203,595	$ 27,903,432
2019	$ 18,421,632	$ 10,300,094	$ 28,721,726
Grand Total	**$ 84,124,347**	**$ 56,260,807**	**$ 140,385,155**

Figure 6: Gross Margins of GB in US and DE for 2015-2019

The most important step here was to create a *Calculated Field* and then create a formula for the Value *'Gross Margin'* as seen above. Once the PivotTable Builder filters were cleared, *Gross Margin* was added to Values and *Calendar Year* to Rows. *Country Desc* was once again kept in Columns. As a result, the table in Figure 6 was created and showed the Gross Margin for each country for all of the individual years of data collected.

For visualizing the data, another PivotChart was inserted and it was determined that a *2-D Column* would be best to show a side by side comparison for each year across the two countries. The trend was similar to revenues, with Germany outperforming the U.S once again across all five years. It would be useful for the U.S decision makers to gather some insight from the German division for sales strategies.

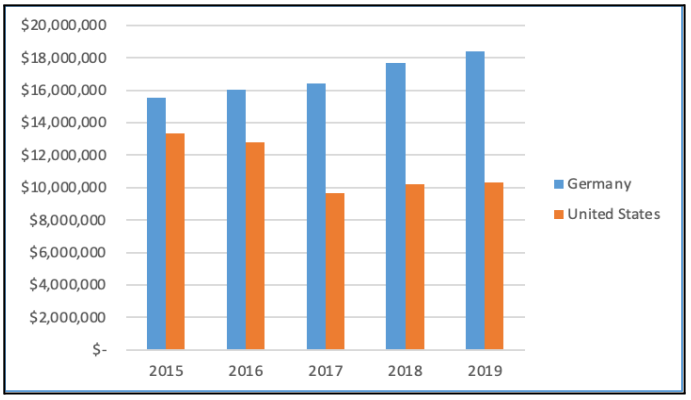


Figure 7: Gross Margin comparison between US and DE for 2015-2019

1.3.4 Who were the top 3 customers for GB in the US across the period of data collected? How did their contributions to the overall revenue vary over time?

After adding *Customer Desc* to Columns and *Calendar Year* to Rows, *Country Desc* was placed in the Filters tab to view each country's data individually. The U.S was then selected in the Excel sheet to display data pertaining to the United States.

Following this, *Column Labels (Customer Desc)* was clicked and then the sorting was done in *Descending* order by *Sum of Revenue*. In *Filter by Value*, Top 10 was chosen and then 3 was entered in the next input box to show the top 3 contributors to revenue in the US.

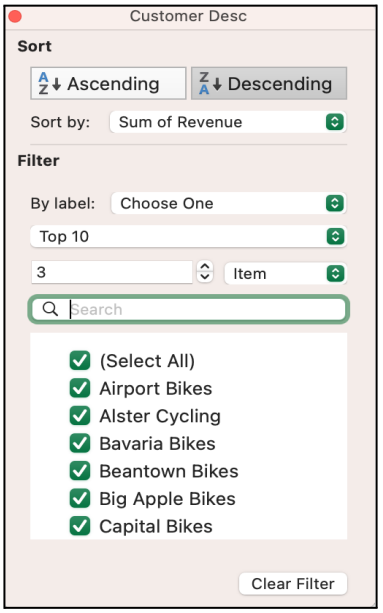


Figure 8: Filtering for Top 3 customers

Country Desc	United States			
Sum of Revenue	**Column Labels**			
Row Labels	**Beantown Bikes**	**Silicon Valley Bikes**	**Big Apple Bikes**	**Grand Total**
2015	$4,265,437	$5,504,421	$2,512,306	$12,282,164
2016	$4,141,214	$5,062,581	$2,515,725	$11,719,521
2017	$2,853,740	$4,216,176	$1,701,089	$8,771,005
2018	$3,486,673	$2,989,108	$1,864,509	$8,340,291
2019	$3,922,744		$2,371,440	$6,294,184
Grand Total	**$18,669,809**	**$17,772,286**	**$10,965,070**	**$47,407,164**

Figure 9: Top 3 GB customers in US (in terms of Revenue)

The table above in *Figure 9* shows the revenues contributed by the top 3 customers in each of the years from 2015-2019.

Lastly, to visualize this data, a PivotChart was inserted and a *2-D Bar* was the chosen method. This was done to show the trend in the revenues for these customers. The biggest takeaway from this was how Silicon Valley Bikes was GB's biggest contributor and its contribution declined from 2015 to them even not contributing to any revenue after 2018. The logical assumption would be that GB lost Silicon Valley Bikes as a customer.

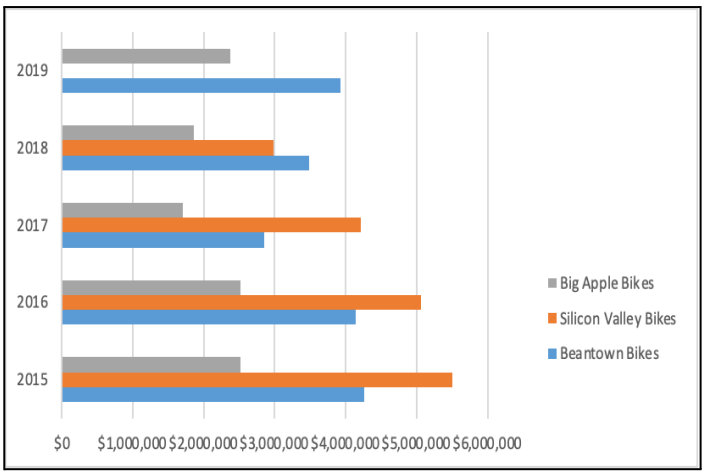


Figure 10: Trends in revenues generated by Top 3 customers for GB in US and DE

1.4 Analysis & Critique of the Tool

I have used Excel in the past for basic spreadsheet creation but this was the first time I used the PivotTable and PivotChart features and I found this was an excellent way to learn the fundamentals of the field. It was very simple to add a dataset to Excel, add specificity to it with the PivotTable and visualize this with the PivotChart feature for further qualitative analysis.

The first thing that stood out was the simplicity of the tool. Having used Excel in the past in academic and professional environments helped but I found the overall interface of the platform quite user-friendly. The tutorial provided by the instructor was very helpful in terms of grasping the basics but the features were quite self-explanatory and easy to become proficient with through some trial and error. For example, it was quite simple to drag and drop variables into the PivotTable builder and then remove them as needed to gain different insight with different combinations of the variables.

Going more in depth into the user-friendliness, the platform was very easy to learn simply because each tab was labeled appropriately and easy to find on my own. There was one instance where the numbers in the dataset were initially displayed as a number with two decimal places. It only took me one or two clicks to change number formatting to *Currency* and decimal places to zero. The essentials for data analysis and visualization were quite easy to locate as the fundamentals were appropriately placed in each tab (i.e. formulas).

While the simplicity of the tool was one of Excel's strengths, it was also one of its

limitations. There were times where I thought the tool was perhaps too simple, especially for visualizing the data. The PivotChart feature was perhaps also not interactive enough for my liking and I believe a platform like Tableau is better in terms of that. It can, however, be used as a supplemental tool to one of Microsoft's more advanced BI offerings like Power BI to add value to organizations.

1.5 Conclusion

Overall, I found using Excel with the PivotTable and PivotChart features a positive learning experience. It was a great way to learn the fundamentals of Data Analysis and Visualization but I look forward to using some more advanced BI software further into this portfolio.

Chapter 2 SAP Lumira Discovery

2.1 About SAP Lumira Discovery

SAP Lumira Discovery is a tool that allows for three key processes: data discovery, analysis and visualization, all within one platform. It is software that must be downloaded onto an individual computer and allows for importing of datasets in various formats (i.e Excel, text files etc). The data can be manipulated in various ways. Columns and Rows can be altered or their values can be grouped into creating entirely new dimensions for analysis. Lastly, Lumira comes with an entire suite of visualization capabilities. Once the dataset is cleaned for specificity, one can switch from the DataView to DesignView and use these variables in the form of Dimensions and Measures. These can then be added to the visualization interface via simple drag and drop and visually represented in the form of columns, charts, graphs, pie charts amongst other techniques. It is one of the BI offerings from SAP's suite and when combined with SAC or Predictive Analytics, can be a very powerful tool to provide insight to organizations in terms of decision-making.

2.2 Dataset & Research Questions

SAP Lumira Discovery will be used to evaluate key statistics from the 2014 FIFA World Cup to derive insights regarding individual and team performance throughout the tournament. The dataset was provided in class as an Excel spreadsheet (.xlsx) and contained 12 Columns and 1404 Records including key variables such as Total Goals Scored, Tackles Made, Solo Runs into Area, Tackles suffered. No cleaning was required in this dataset; however, it was immediately noticeable that the dataset was lacking a very important variable: Assists. This would leave out the best performing playmakers in the tournament and make it harder to determine the best player in terms of overall contributions. I will attempt to answer the following questions:

i. Who was the top scorer in the tournament? Who were the second and third best in terms of goals?
ii. Who was the most efficient scorer in terms of overall minutes played? Who were the best performers on a 90 minute average?
iii. Which player was the most wasteful in keeping possession? Who was the most successful "risk-taker" at the tournament?
iv. Which country's players received the most tackles? Which country was involved in games with the most physicality?

2.3 Applying the Analytical Tool & Results

2.3.1 Who was the top scorer in the tournament? Who were the second and third best in terms of goals?

This was a fairly simple question to answer. After switching from the Dataview to Designview, Total Goals Scored from Measures was added to the Y-Axis and Name was added to the X-Axis. The Chart Type was changed to a simple Column Chart and it showed each player's performance.

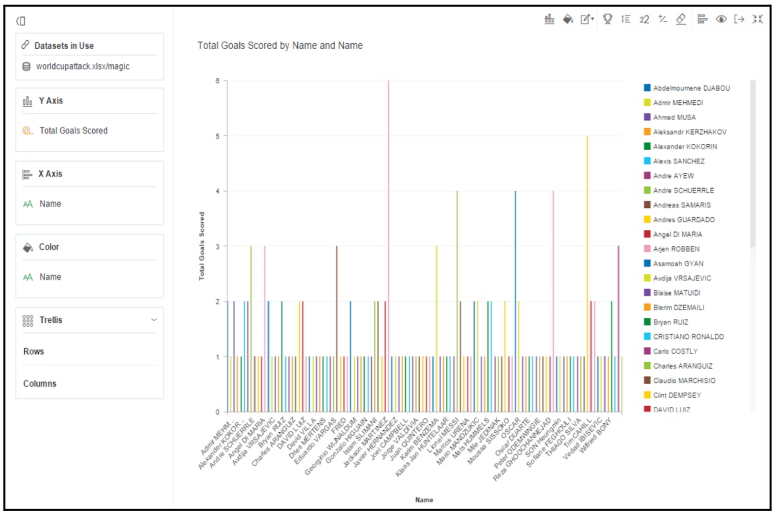


Figure 1: Total Goals by all Players

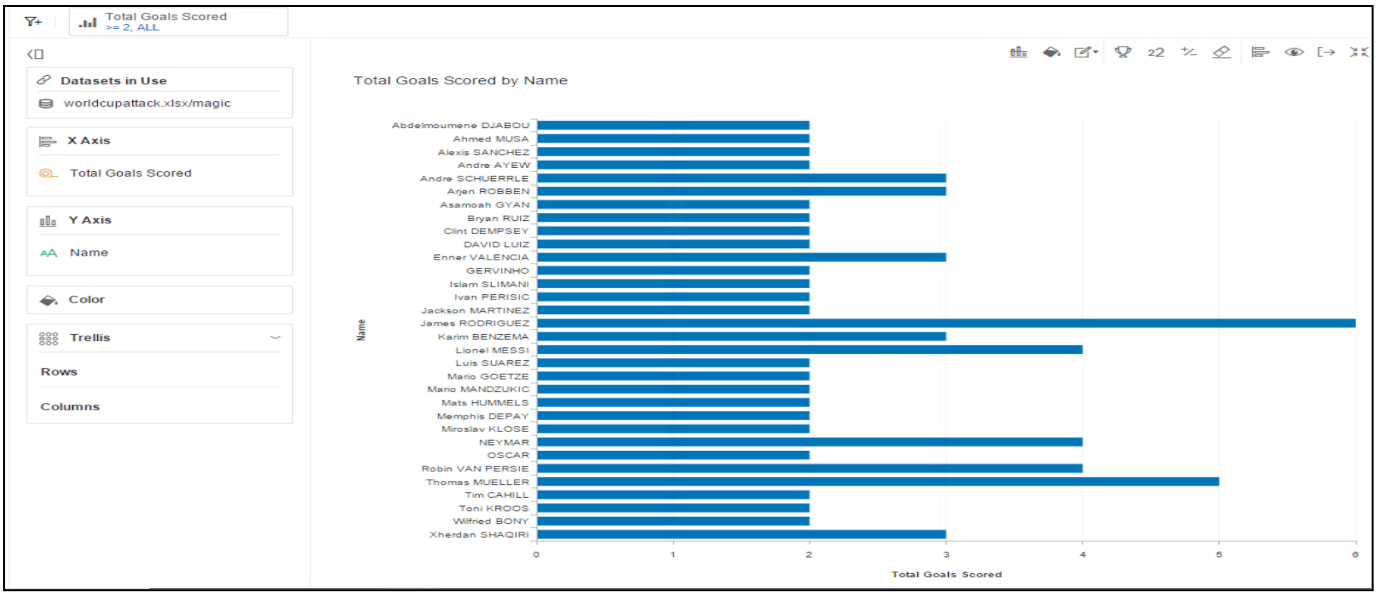


Figure 2: Goals Scored by Players (2 Goals and Above)

However, even though it was possible to tell the top scorers, the visualization looked very cluttered. To improve this, a Filter was added for Goals Scored and the value was set to *Equals 2*. This filtered out any players with goals scored below that value and made for a more appealing graphical visualization. The chart type was set to *Bar Chart*, and it can be seen in *Figure 2* that James Rodriguez was the top scorer with 6 goals followed by Thomas Mueller on 5 goals and multiple players on 4.

2.3.2 Who was the most efficient scorer in terms of overall minutes played? Who were the best performers on a 90 minute average?

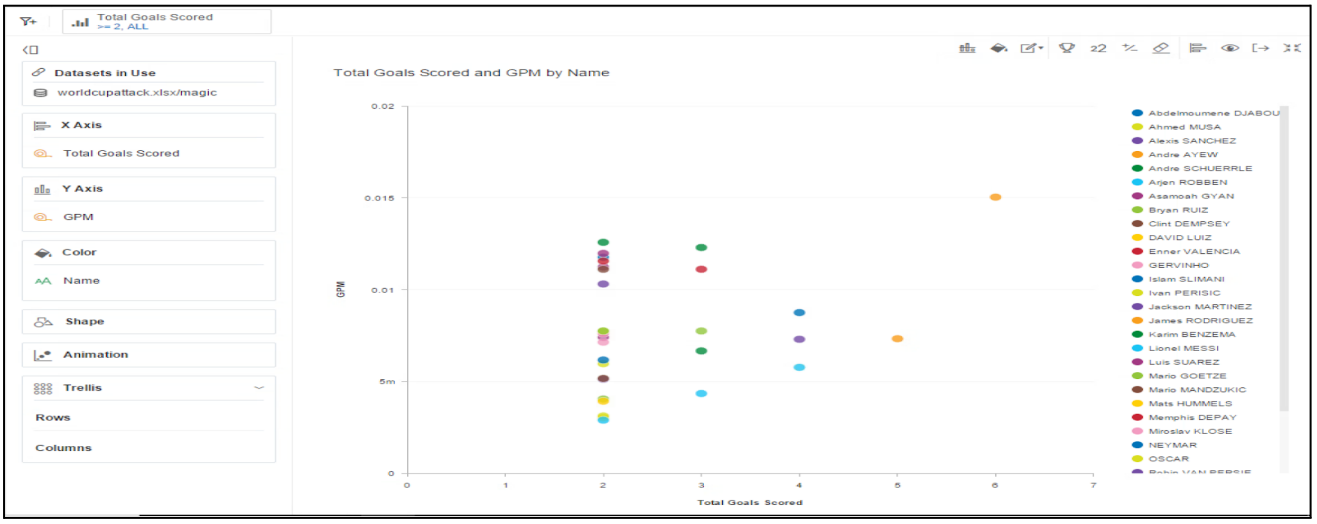


Figure 3: Comparison between GPM and Total Goals Scored

Sometimes goals scored can be misleading in terms of predicting player performance because some players might perform better in more limited minutes. It was important to create a Measure that allows for contributions to be gauged in terms of minutes played. Here, the *Calculated Formula* known as *GPM* was created *(TGS/Minutes Played)*. Following this, the chart type selected was a *Scatter Plot* with *Goals Scored* placed in the X-Axis and *GPM* placed on the Y-Axis. As seen in the figure above, James Rodriguez was clearly the best performer. However, since a game of soccer is ninety minutes long, it might be better to measure contributions per 90 minutes.

Hence, a new calculation was created known as *G/90* which reflected Goals per 90 minutes played. The calculation was done by replicating the formula for GPM as shown above but further dividing the result by 90.

Following the creation of this new measure, *G/90* was added to the Y-Axis and Name was added to the X-Axis. Filter for *TGS* was kept as 2 so that players with one goal and less could be excluded from the visualization. Following that the chart type selected was *Bar Chart* and the resulting visualization (*Figure 5*) was more appealing than the previous one.

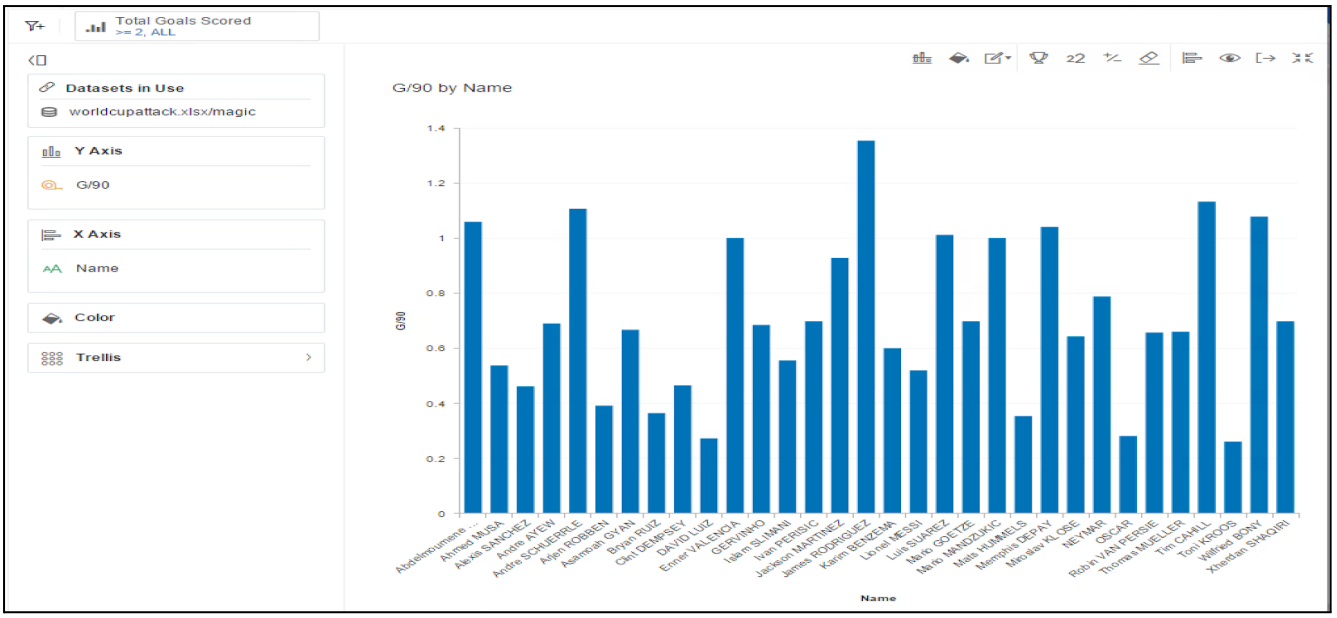


Figure 5: Goals scored per 90 minutes by players with 2 goals and above

While James Rodriguez was still the clear standout, it was interesting to notice players who were second in terms of their contribution per 90. Andre Schuerrle being ahead of fellow countryman Thomas Mueller despite scoring less goals overall suggests his impact was greater.

2.3.3 Which player was the most wasteful in keeping possession? Who was the most successful "risk-taker" at the tournament?

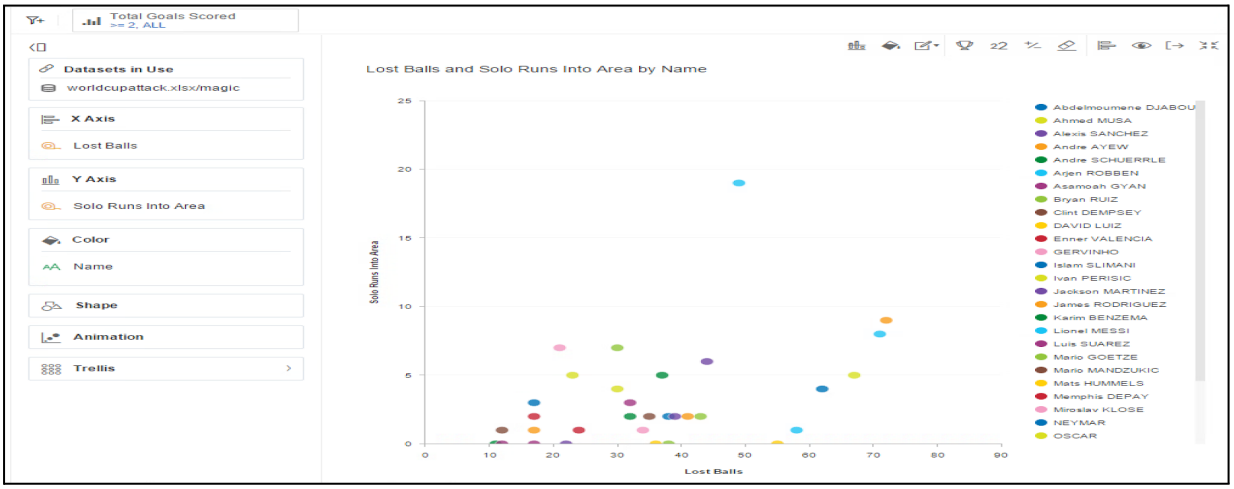


Figure 6: Comparison between Balls Lost & Solo Runs into Area

To assess which player gave away the ball the most, it is important to look at the *Lost Balls* Measure. Hence, it was added to the X-Axis. However, players who are more likely to dribble and take risks are more likely to lose the ball. Hence, it was important

to look at the *Solo Runs into Area* Measure as well and add it to the Y-Axis for a comparative analysis. To represent this data, a *Scatter Plot* was chosen with the Dimension *Name* added to Color so each player has a distinct plot. James Rodriguez lost the ball the most in the tournament followed by Lionel Messi. However, it is interesting to see Arjen Robben in the graph. He lost the ball significantly less than the aforementioned players but also had significantly more runs into the area highlighting how he was more efficient in keeping possession and creating dangerous plays for his team.

2.3.4 Which country's players received the most tackles? Which country was involved in games with the most physicality?

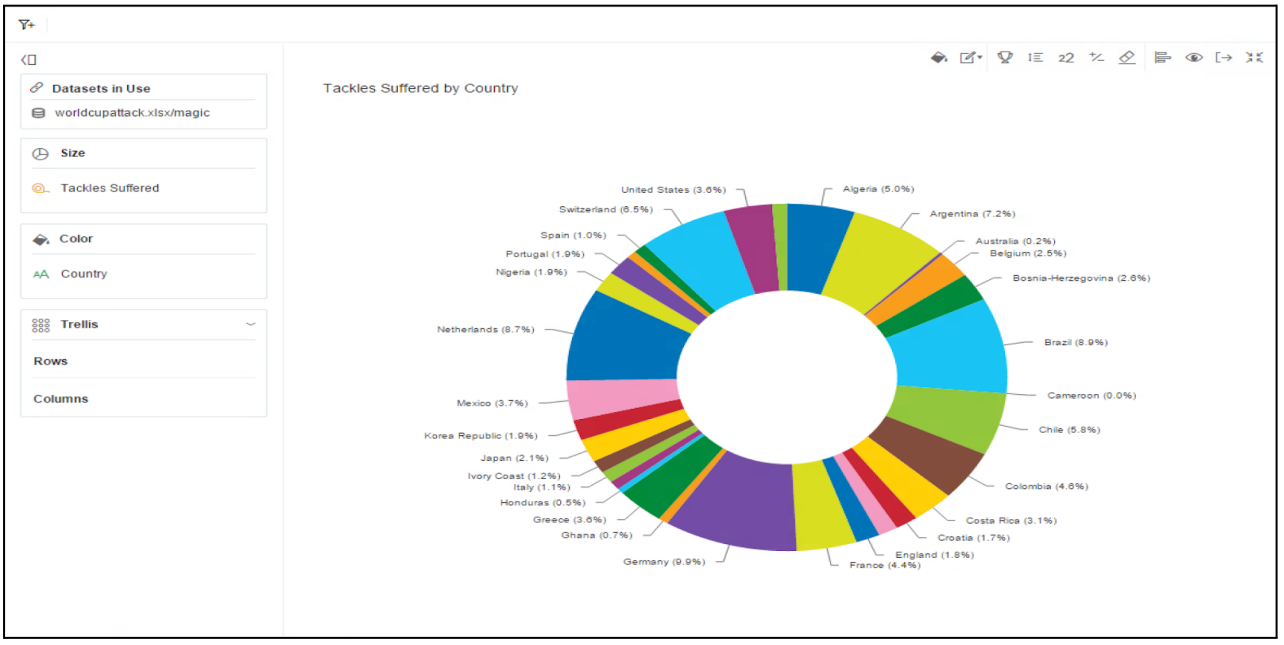


Figure 7: Tackles suffered by Countries (as a Percentage of Total Tackles)

To answer this question, it was important to decide upon a visualization first. A *Donut Chart* was chosen to show the tackles suffered by players from each country as a percentage of the total tackles in the tournament. *Tackles Suffered* was added to Size and the Dimension *Country* was added to Color for this purpose. The resulting visualization as seen above shows German players received the most tackles, followed by Brazil and then the Netherlands. It can be inferred from this that perhaps players from these countries attracted more physicality from other nations in the tournament.

2.4 Analysis & Critique of the Tool

Lumira was definitely a step-up from Excel in terms of the interface. The program remains relatively simple to use but what stood out immediately was the ability to

switch between the DataView and DesignView. Lumira also accepts datasets in various formats so it was quite simple to import a spreadsheet in the DataView and then switch to DesignView to visualize it in a full-screen format. The drag and drop feature proved very useful in this regard as it was very simple to take a value from the measures/dimensions tab and drop it into the visualization tab.

It was also noted that making alterations to the dataset was a relatively straightforward process. One could simply switch to the DataView and create new variables or remove existing ones. Creating new fields through calculations was a very useful process because Lumira actually offered suggestions in terms of syntax to ensure that the calculation for the new measure was valid. This made creating Measures such as *GPM* and *G/90* very easy. Furthermore, these Measures and Dimensions could be easily altered in DesignView for ad-hoc analysis.

However, as a Mac user, using Lumira in a Virtual Machine environment was quite a frustrating experience. The fact that it is not compatible with Apple OS was a huge drawback because it meant the application would require additional processing power and run slowly compared to a Windows machine. Furthermore, the work had to be constantly saved and uploaded to a cloud storage platform because the progress would be undone otherwise which made the whole process quite tedious.

2.5 Conclusion

Overall, I still found Lumira to be a positive learning experience. I have to admit that upon loading up the software, the archaic interface looked somewhat off putting and there was also a bit of a learning curve; however, after some time with the software, it proved quite simple to use and was a step up from Excel. Despite this, I would still prefer to use a more advanced tool as my primary BI software.

Chapter 3 IBM Cognos Insight

3.1 About Cognos Insight

IBM Cognos Insight is a desktop-based BI tool that can help organizations analyze and visualize data whilst allowing for interactive reporting. It does this by:

Data Transformation: Cognos Insight is very useful as it allows for various processes like Modeling, Cleaning and Enhancement. Attributes can be mapped, unwanted variables can be removed from the dataset and enhancements are made possible through intuitive calculations and combinations.

Data Analysis and Visualization: Cognos Insight allows the user to conduct these processes simultaneously, with the Split view triggered. Any change made to the dataset is reflected immediately in the visualizations in the same workspace.

Scenario Planning: Data Analysis can easily be harnessed into decision-making within the workspace. The existing workspace, helped by in-application shortcuts, can be altered to create new scenarios (forecasting, scorecards etc).

3.2 Dataset & Research Questions

The two datasets provided by the instructor are text files (.csv) covering Quarterly financial data for a retail organization in the year 2012. The first dataset pertains to financial metrics such as Revenues, Costs for Sales Channels and Product Categories and has 8 Columns and 7840 Records. The second dataset contains the same information but for different geographical locations (Europe, South America etc). This dataset has 9 Columns and 17577 Records. Some data cleansing was needed (i.e the deleting the measure called “Count”) because it does not aid the analysis. Perhaps the datasets could have revealed more insight if product categories were broken up into sub-categories (Computers & Tablets into two separate categories) as this would have allowed the analyst to drill-down further. Using the datasets provided, I will attempt to answer the following questions:

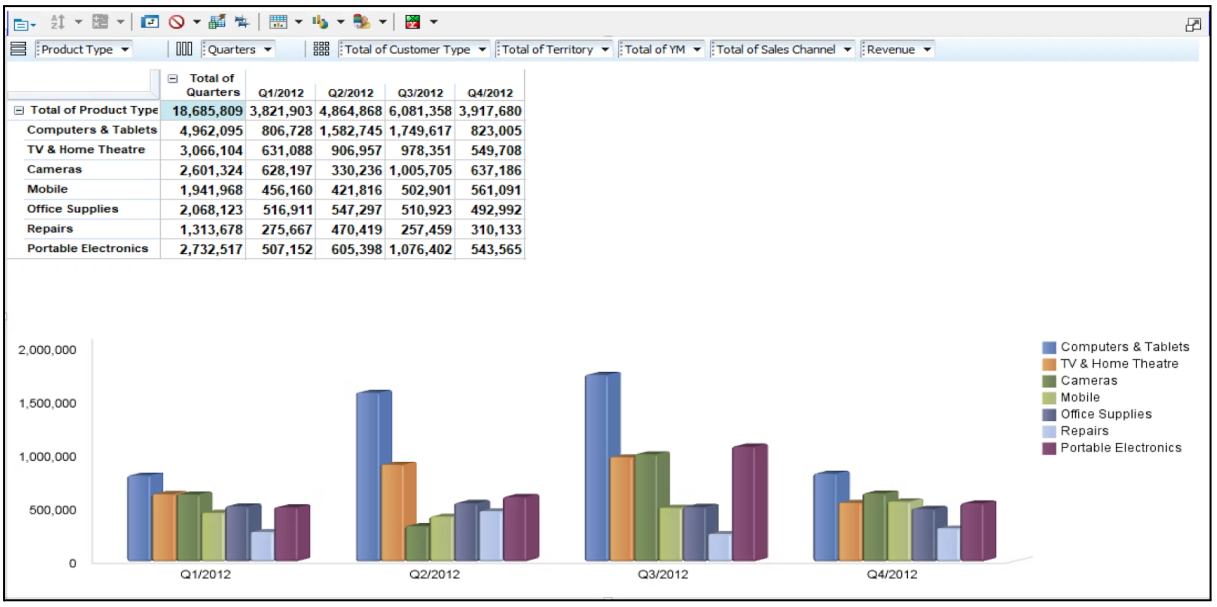
i. What are the trends across all 4 quarters for different kinds of products? What are the trends across Sales Channels?
ii. What are the Profit Margins across different Customer Types?
iii. How are products performing across different Customer Types?
iv. Is the business problem unique to North America?

3.3 Applying the Analytics Tool & Results

3.3.1 What are the trends across all 4 quarters for different kinds of Products? What are the trends across Sales Channels?

Total of Quarters Dimension was dragged to Columns and *Product Type* was added to Rows. Following this the chart type selected was *3-D Column*. We can notice that

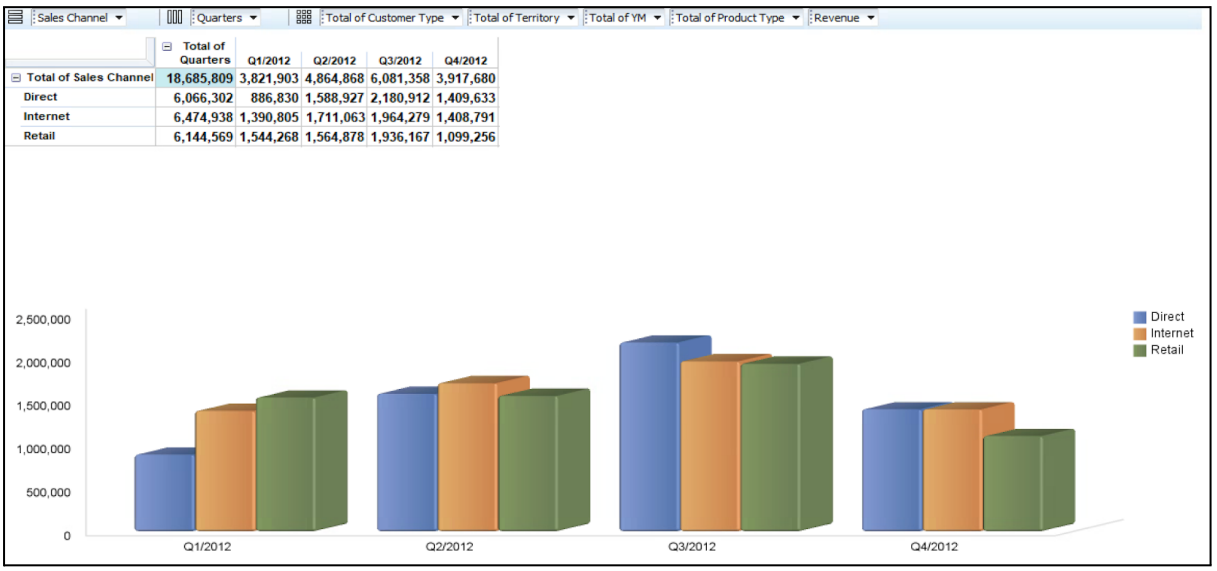
there was an ascending trend in general from Q1 to Q3 (apart from Cameras in Q2). Q3 was the best quarter in terms of revenue followed by an overall sharp decline in Q4 (Mobile and Repairs being the exceptions). Computers & Tablets seem to generate the most revenue overall for the company.



	Total of Quarters	Q1/2012	Q2/2012	Q3/2012	Q4/2012
Total of Product Type	18,685,809	3,821,903	4,864,868	6,081,358	3,917,680
Computers & Tablets	4,962,095	806,728	1,582,745	1,749,617	823,005
TV & Home Theatre	3,066,104	631,088	906,957	978,351	549,708
Cameras	2,601,324	628,197	330,236	1,005,705	637,186
Mobile	1,941,968	456,160	421,816	502,901	561,091
Office Supplies	2,068,123	516,911	547,297	510,923	492,992
Repairs	1,313,678	275,667	470,419	257,459	310,133
Portable Electronics	2,732,517	507,152	605,398	1,076,402	543,565

Figure 1: Sales for Products for 2012 (Q1-Q4)

After replacing *Product Type* on the Rows Dimension with *Total of Sales Channel*, we can see there is a similar trend across Channels in *Figure 2*: An ascending trend with the best performance in Q3 followed by a decline in Q4. The Internet Channel is the best revenue generator for the company.

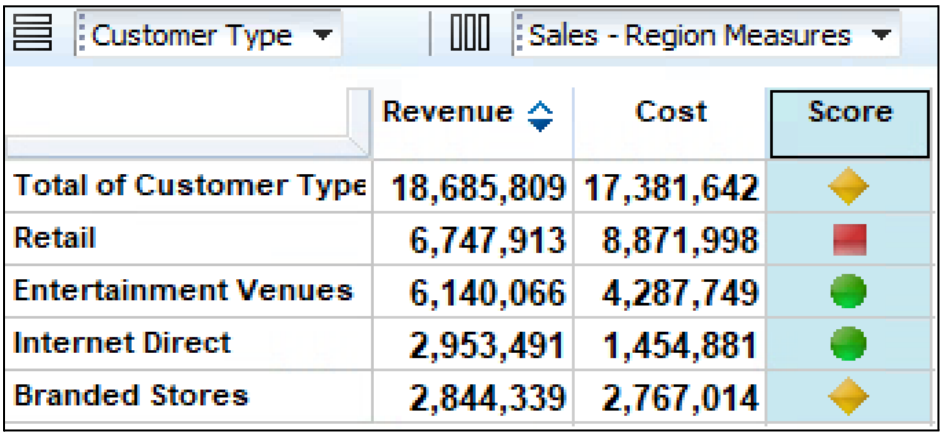


	Total of Quarters	Q1/2012	Q2/2012	Q3/2012	Q4/2012
Total of Sales Channel	18,685,809	3,821,903	4,864,868	6,081,358	3,917,680
Direct	6,066,302	886,830	1,588,927	2,180,912	1,409,633
Internet	6,474,938	1,390,805	1,711,063	1,964,279	1,408,791
Retail	6,144,569	1,544,268	1,564,878	1,936,167	1,099,256

Figure 2: Sales across Channels for 2012 (Q1-Q4)

3.3.2 What are the Profit Margins across different Customer Types?

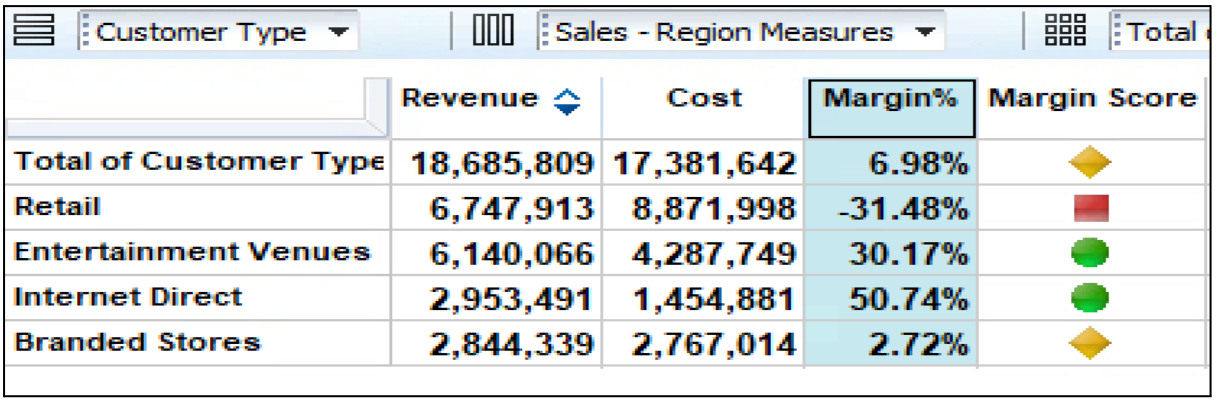
After dragging *Revenue* to Columns and keeping *Customer Type* in Rows. both *Revenue* and *Cost* were selected, and from the dropdown menu, *Calculate -> Compare Revenue vs Cost* was chosen. After naming the new Column *Margin Score*, *Revenue* was clicked and sorted in Descending order. This simple scorecard shows poor performance in the 'Retail' *Customer Type* and unimpressive performance across 'Branded Stores'.



	Revenue	Cost	Score
Total of Customer Type	18,685,809	17,381,642	
Retail	6,747,913	8,871,998	
Entertainment Venues	6,140,066	4,287,749	
Internet Direct	2,953,491	1,454,881	
Branded Stores	2,844,339	2,767,014	

Figure 3: Margin Scorecard for different Channels

Both the *Revenue* and *Cost* columns were selected again, and the Calculation menu was opened to do *Cost/Revenue*. This is then renamed to *Margin %*. This calculation needs to be modified so after clicking Edit This Calculation, the expression is modified by placing a "1-" in front of the equation. To format this properly, *Format Measure* was clicked and then *Percentage* was chosen with 2 Decimal Places.



	Revenue	Cost	Margin%	Margin Score
Total of Customer Type	18,685,809	17,381,642	6.98%	
Retail	6,747,913	8,871,998	-31.48%	
Entertainment Venues	6,140,066	4,287,749	30.17%	
Internet Direct	2,953,491	1,454,881	50.74%	
Branded Stores	2,844,339	2,767,014	2.72%	

Figure 4: Margin Percentages across all Channels

What can be seen here is whilst 'Retail' customer type might be the highest in terms of generating revenue, the associated costs are far too high and this results in a negative *Margin*, leading to overall losses in this type. 'Branded Stores', too, are unimpressive. Management should investigate their Retail sector urgently followed by Branded Stores.

3.3.3 How are products performing across different Channels?

The display style is changed to *Crosstab* and *Total of Product Type* is moved into Rows. Following this, *Total of Customer Type* dimension is moved to Columns as first value and *Sales - Region Measures* on Columns as the second value. After switching back to Split view, hide the *Margin%* columns and rename the *Margin Score* to *Score* for a cleaner table and more appealing visualizations

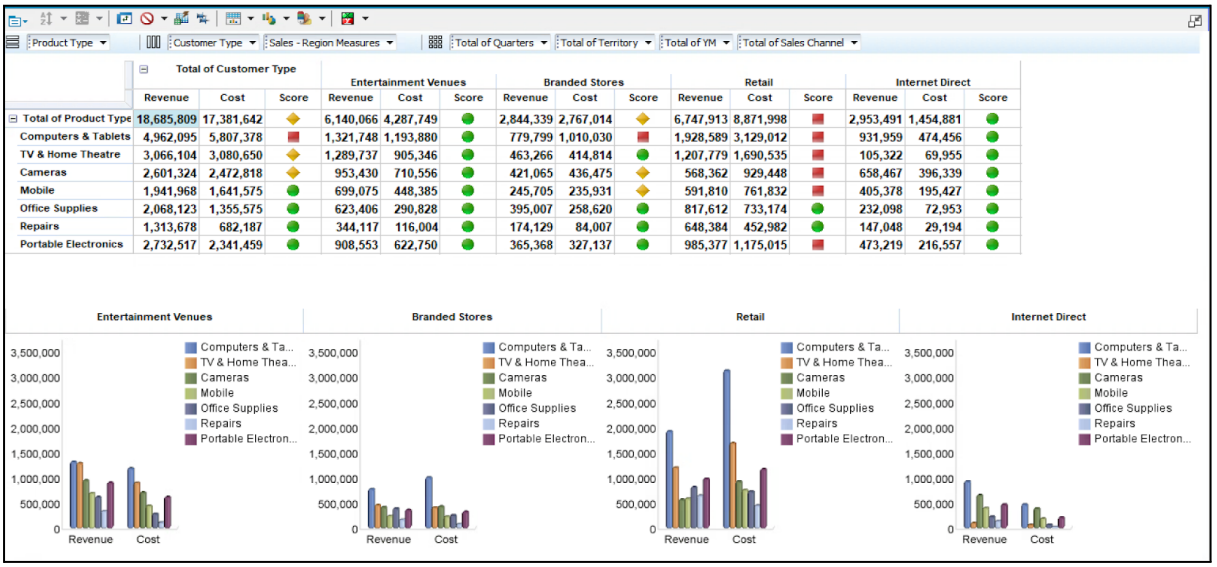


Figure 5: Margin analysis for Product Categories across different Channels

What is immediately noticeable is that there is a critical problem within the Computer and Tablets line of products. Upon further inspection, we can see that these products are especially performing poorly in Branded Stores and Retail Customer types. Also it is important to note that the retail channel as a whole is performing poorly other than Office Supplies and Repairs. Perhaps this is where management could step in and reallocate resources to cut costs and allocate these products to more profitable channels (Internet Direct or Entertainment Venues).

3.3.4 Is the business problem unique to North America?

After clicking on the *Sales-region cube* and choosing “Import into cube Sales - Region”, we will select the second dataset (WW Monthly Sales). To properly map the data, we will click the Advanced button and drag and drop *Quarter* from the new dataset into the cube. This ensures *Quarter* and *Quarters* are mapped. After finishing the import process, we proceed to the new data.

	Total of Territory	North America	Europe	Asia Pacific	South America
Total of Product Type	14,427,525.33	1,304,167	8,203,921.89	3,459,325.94	1,460,110.5
Computers & Tablets	3,547,701.58	-845,283	2,871,870.54	1,090,951.97	430,162.07
TV & Home Theatre	2,751,170.26	-14,546	1,800,982.32	631,227.85	333,506.09
Cameras	750,540.54	128,506	491,329.83	94,091.59	36,613.12
Mobile	1,495,738.57	300,393	607,583.95	375,384.61	212,377.01
Office Supplies	2,514,184.83	712,548	1,146,756.33	499,194.83	155,685.67
Repairs	1,634,193.85	631,491	521,181.07	315,423.22	166,098.56
Portable Electronics	1,733,995.7	391,058	764,217.85	453,051.87	125,667.98

Figure 6: Margin data for Product Type across different Territories

We will create a new tab to separate this from the data analysis for NA. The earlier process to measure profitability will be repeated in order to create a calculation for *Margin* from the *Revenue* and *Cost* Columns (*Calculate -> Revenue - Cost*). After choosing *Product Type* as Rows, *Total of Territory* as Columns and *Margin* in the Context dropdown, we can see how products compare across all of the territories.

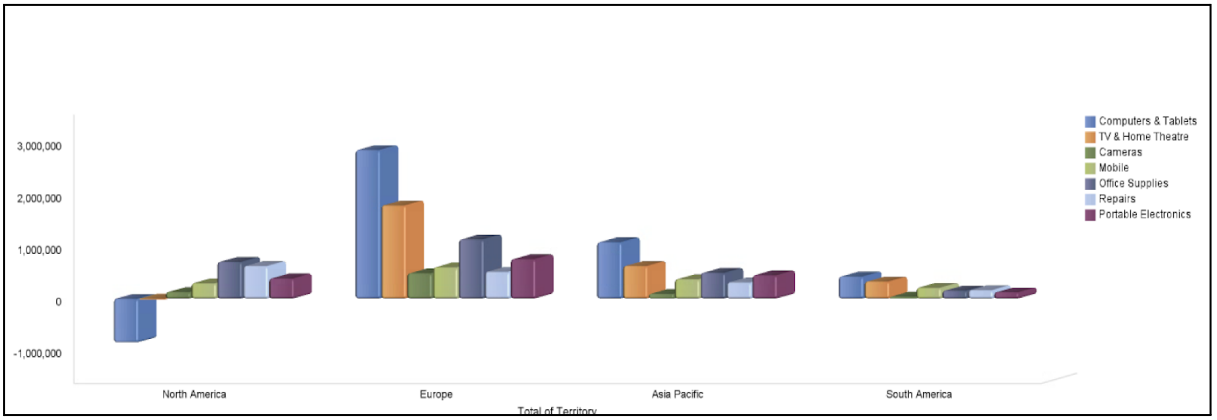


Figure 7: Margin visualizations for Product Categories across Territories

Here we can see the trend is only common to NA as the Computers and Tablets category generates a positive margin in other territories. Europe, in particular, is performing remarkably better than NA.

Total of Sales Channel | Total of Territory | Margin | Total of Quarters

	Total of Territory	North America	Europe	Asia Pacific	South America
Total of Sales Channel	14,427,525.33	1,304,167	8,203,921.89	3,459,325.94	1,460,110.5
Direct	6,058,604.97	487,008	3,305,758.33	1,705,038.66	560,799.98
Internet	6,368,145.21	1,856,799	3,017,450.68	1,036,406.8	457,488.73
Retail	2,000,775.15	-1,039,640	1,880,712.88	717,880.48	441,821.79

Figure 8: Margin data for Sales Channels across different Territories

Following this we will see how the trends are across Channels. In Rows, Total of *Product Type* will be replaced with *Total of Sales Channel*.

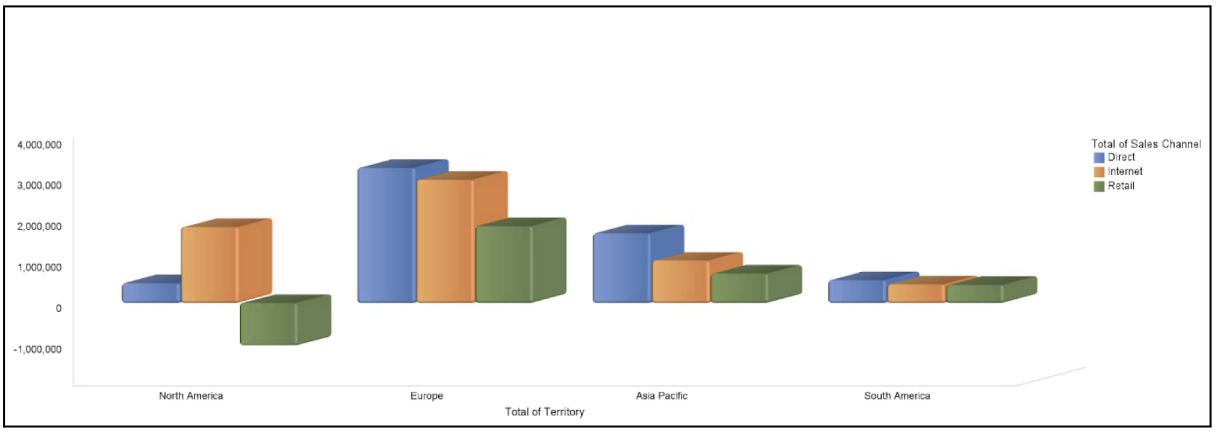


Figure 9: Margin visualization for Sales Channels across Territories

Once again, all the other territories have positive margins across all three channels. The 'Retail' channel in NA is the only one with a negative margin. Europe is once

again, by far, the best performer. This only reinforces the need for management in NA to investigate the ‘Retail’ channel and Computers and Tablets product category further. Perhaps they could seek some consulting from the European division and compare their business practices to see if there is room for improvement.

3.4 Analysis & Critique of the Tool

One of the positives of using Cognos Insight was how it allows for ad-hoc analysis and generates updated visualizations based on changes. The Split view was incredibly useful as it allows for manipulation of the dataset as needed and this is reflected in the visualizations in the same workspace. This makes the analysis process quite interactive and intuitive.

A second advantage relates to the overall simplicity of the tool. It was very simple to import a dataset into the application and make changes. The Columns, Rows features are very simple to use with a drag and drop method, but what especially stood out was the Context feature. This added a whole new layer to the analysis process as simple changes to attributes within it revealed a whole new level of insight (i.e Total of Sales Channel could be seen as either Revenue or Margins).

One noticeable downside of the tool was that the platform looked outdated and the interface seemed very archaic compared to some more modern, elegant BI tools i.e Tableau and Qlik. The platform also seemed cumbersome to use at times (even more so in a Virtual Machine environment) with slight delays in processing changes. Visualization capabilities also seemed very limited as the platform seems to be built to suit Bar/Column graphs.

3.5 Conclusion

I found IBM Cognos to be a challenging, yet rewarding experience. It is very useful in terms of simulating an environment where one feels like they are working within the financial analysis department of an organization and derive insights that can help decision makers with the bottom line; however, the somewhat outdated interface means it would not be my go-to BI tool.

Chapter 4 SAP Analytics Cloud

4.1 About SAC

SAP Analytics Cloud is a versatile cloud-based BI platform that creates a secure public cloud experience to simplify data-driven decision making by allowing a variety of processes within one platform such as Data Modeling, Data Visualization and Predictive Analytics. Paired with other SAP offerings like PA or Lumira, it can provide immense value to organizations by combining trend and predictive analyses.

Data Modeling: Once Data is imported into SAC, it can be altered through a variety of functions. Additional measures can be easily created through the insertion of calculations and existing ones can be altered by using the built-in transformation functions.

Data Visualization: SAC allows users to create graphs and other visualizations like other popular BI tools. One can simply add a page to the workspace and create and style visualizations using the Chart Builder feature. SAC opts for a selection interface as opposed to a drag and drop one.

Predictive Analytics: By incorporating both manual processes and utilizing machine learning to identify trends not easily visible to the human eye, users can look for patterns in data to help predict future outcomes. The impact of Key Influencers on certain attributes is one such example.

4.2 Dataset & Research Questions

The dataset provided as a resource for this exercise is ERPSim game data and is provided in the form of an Excel spreadsheet (.xlsx). It contains 10 Columns and 65590 Records along with information such as the Teams involved, the type of Product (different variations of Muesli) and Regions and Distribution Channels. The dataset required some enhancement for the Area and Distribution dimensions to make the analysis more useful. It was noticeable that Cost of Goods Sold was missing from this dataset. I believe if it was available, Margins could have been created to better reflect profitability or determine pricing strategies. Based on the dataset, I will attempt to answer the following questions:

i. Which combination of entities contributes to the highest revenue from sales of Muesli?
ii. Which Distribution Channel has the highest Average Revenue? How did this vary for different price bins in this channel? Why use average Revenue instead of total Revenue?
iii. What team had the highest Revenue? Which product had the highest Revenue?
iv. What is the Market share of each team by product?
v. Are there any products that do not sell in specific Distribution Channels?

4.3 Applying the Analytics Tool & Results

4.3.1 Which combination of entities contributes to the highest revenue from sales of Muesli?

After Running a *Smart Discovery*, *Revenue* was selected as the *Target* and all the Dimensions were Selected except *Day* and *Round* as Entity. After filtering for the top 10 across all Dimensions, we find the top 10 revenue generators for a combination of different entities.

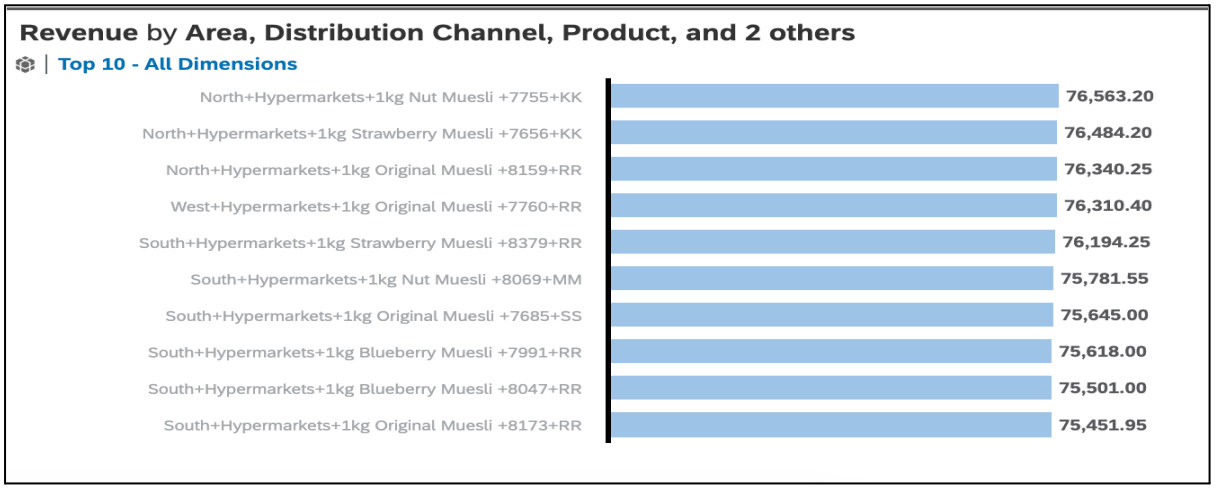


Figure 1: Revenues for Combination of Entities (Top 10)

It is important to note here that the top 3 do have the North region and Hypermarket as the channel of distribution in common along with 1 kg versions of Muesli. However, since there is not a huge revenue difference between the 1st and the 10th combination, it is fascinating that the two most recurring patterns in this data are the distribution channels (Hypermarkets) and the weight of the product (All are 1kg versions of Muesli).

4.3.2 Which Distribution Channel has the highest Average Revenue? How did this vary for different price bins in this channel? Why use average revenue instead of total revenue?

After clicking on the Key Influencers tab and selecting *Distribution Channel* from List A of Key Influencers, we can see that the highest average revenue generated is through Hypermarkets. It generates, on average, around $26,000 more than the second best channel: Grocery chains.

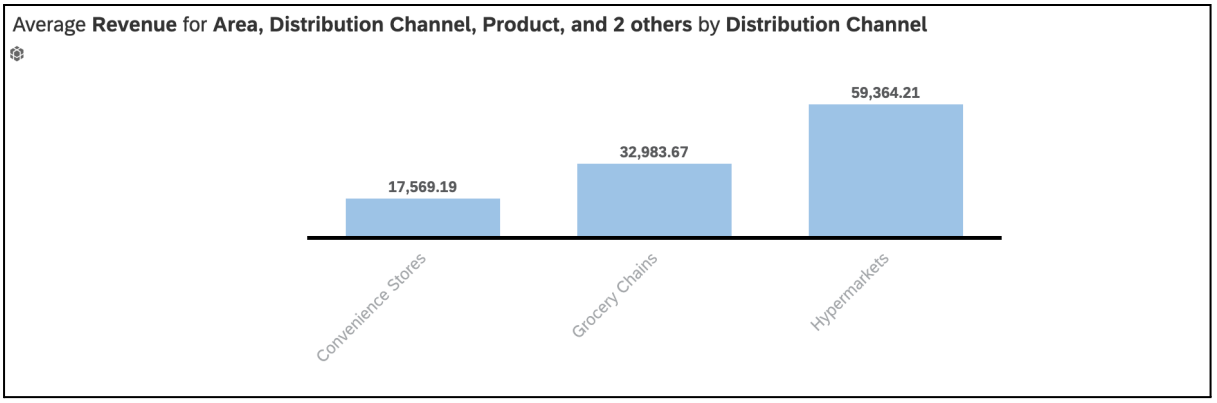


Figure 2: Average Revenue across Distribution Channels

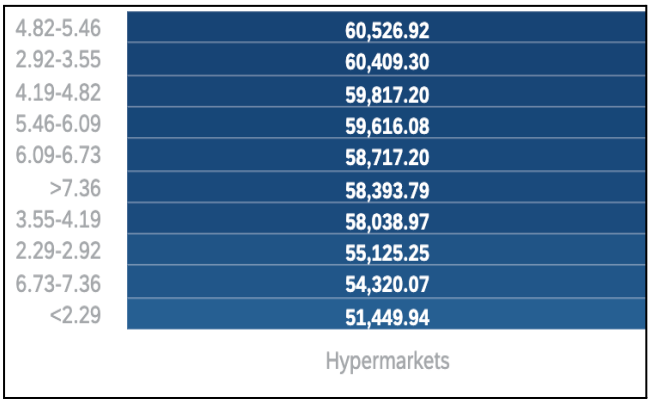


Figure 3: Avg. Revenues for Price Bins

I scrolled down and selected Price Bin from List B of Key Influencers and sorted by *Revenue* (Highest to Lowest), it can be seen that the Prince Bin Ranges 4.19- 4.82, 4.82-5.46 and 5.46 - 6.09 are in the top 4 for *Average Revenue* for the Hypermarkets *Distribution Channel*. Perhaps management could look at this data in developing pricing strategies and realize that the price ranges between 4.19 - 6.09 might be optimal for maximizing revenue generated.

Using total revenue in this case might have been misleading if we're trying to accurately gauge the effectiveness of a particular distribution channel as a whole. For example, if there were a lot more grocery stores than there were hypermarkets then they may present a higher total revenue; however using average revenues creates a clearer picture with regards to resource management. The managers can gauge the effectiveness of each channel and reallocate resources/budget as needed to maximize profitability. In this case, it might make more sense to prioritize resources toward hypermarkets as opposed to convenience stores.

4.3.3 What Team had the highest Revenue? Which Product had the highest Revenue?

After creating a new Page, click *Insert Chart*. Here, a *Column Chart* was decided as the best option. In Measures, *Revenue* was added and in Dimensions, *Team* was selected. *Chart Orientation* was then set to vertical and *Revenue* was sorted from Highest to Lowest. In terms of revenue, RR was the best performing team followed by NN and then KK.

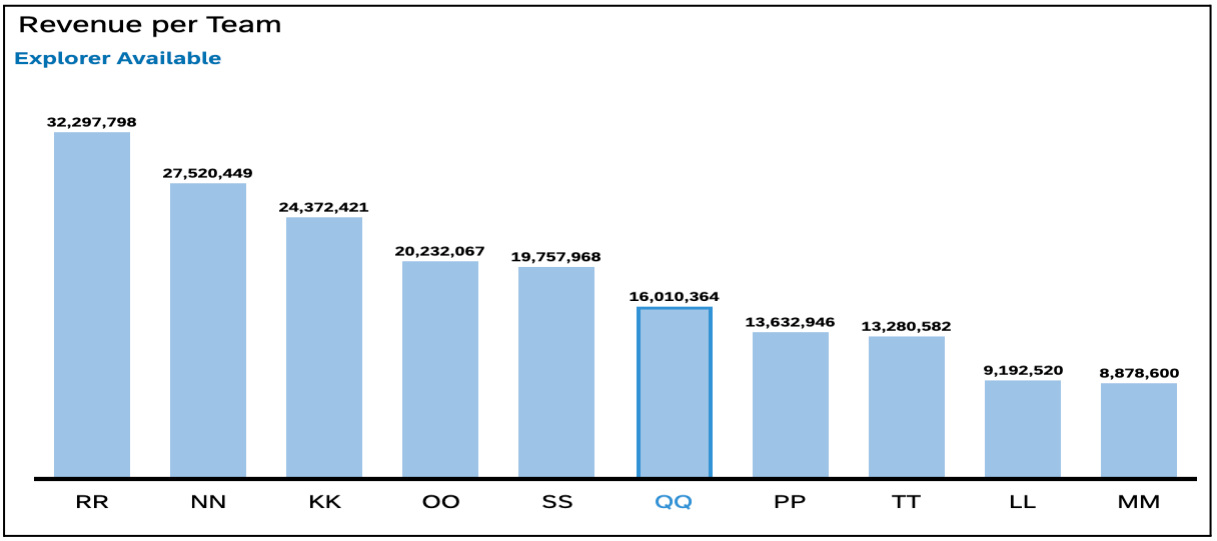


Figure 4: Revenue per Team

The process from above was then repeated; however, in the Chart Builder, *Product* was selected in Dimensions instead of *Teams*. As seen in the visualization below, the 500g Nut Muesli proved to be the most popular product as it generated the most revenue. The 1kg version of Nut Muesli was third so the Nut Muesli might be the most popular product offering regardless of size.

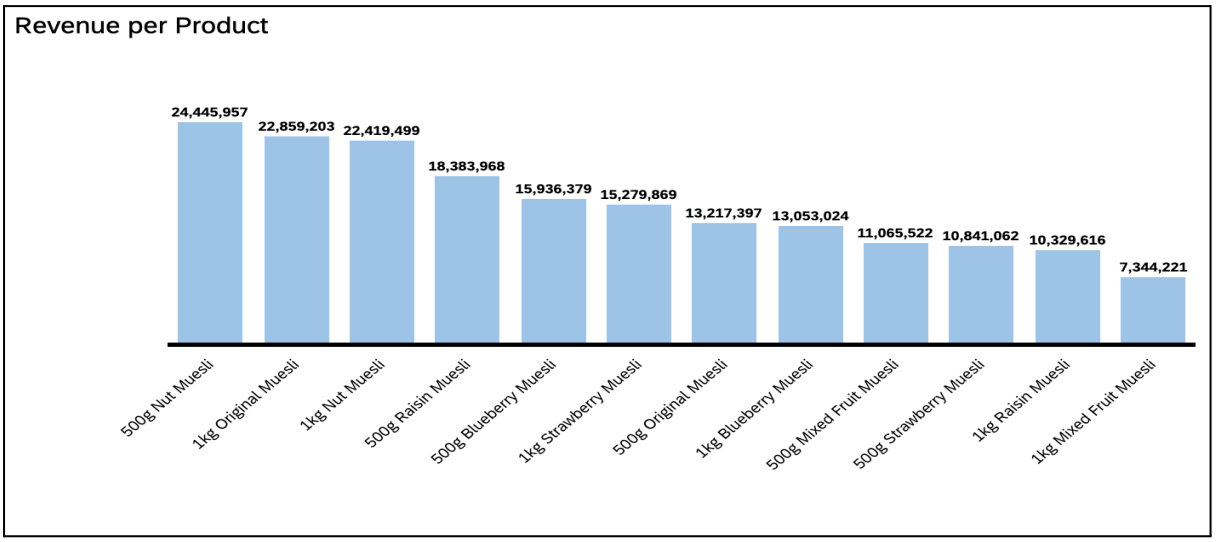


Figure 5: Revenue per Product

4.3.4 What is the Market share of each Team by Product?

After creating a new page, and inserting a new chart, a *Stacked Column Chart 100%* was chosen as the best option. In Measures, *Revenue* was selected and in Dimensions: *Product*. Lastly, in Color, *Teams* were added.

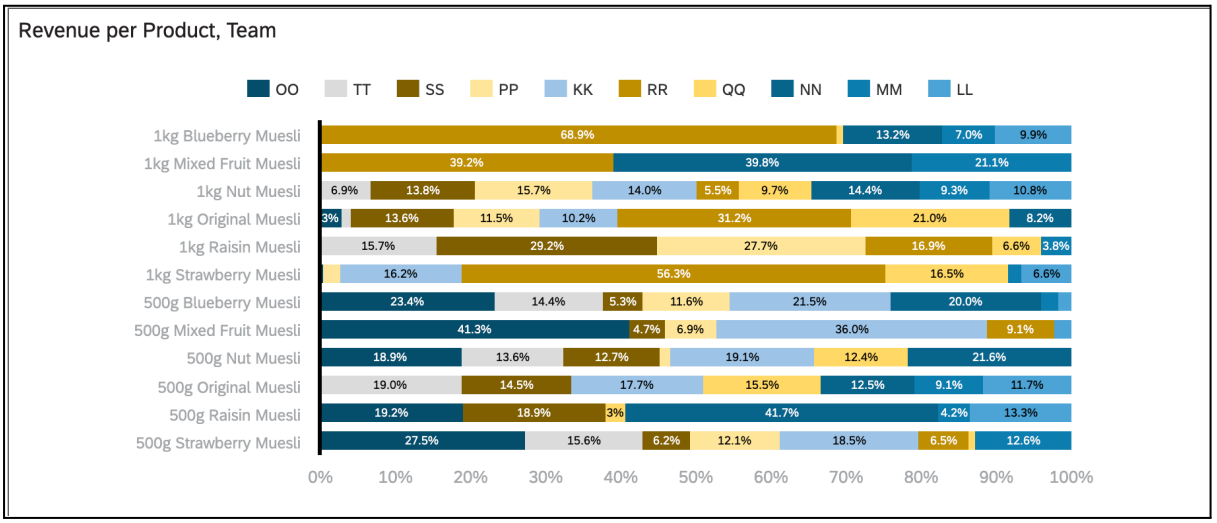


Figure 6: Market Share of Teams by Product Type

Figure 6 shows that RR was the highest performing team in terms of revenue and has significant market shares in multiple 1kg Versions. NN was the second best performing team and has a more varied market share across both 500g versions and 1kg versions. This highlights a difference in strategies. One opted for a strategy of concentration within a certain product type (1kg) while the other one used a strategy of diversity (noticeable shares in all types of products).

4.3.5 Are there any Products that do not sell in specific Distribution Channels?

After creating a new page and inserting a new chart, the Chart Type selected was *Heat Map.* In Dimensions, *Distribution Channels* were added to X-Axis and *Product* was added to Y-Axis. Lastly, *Quantity* was added to Color and the results can be seen below.

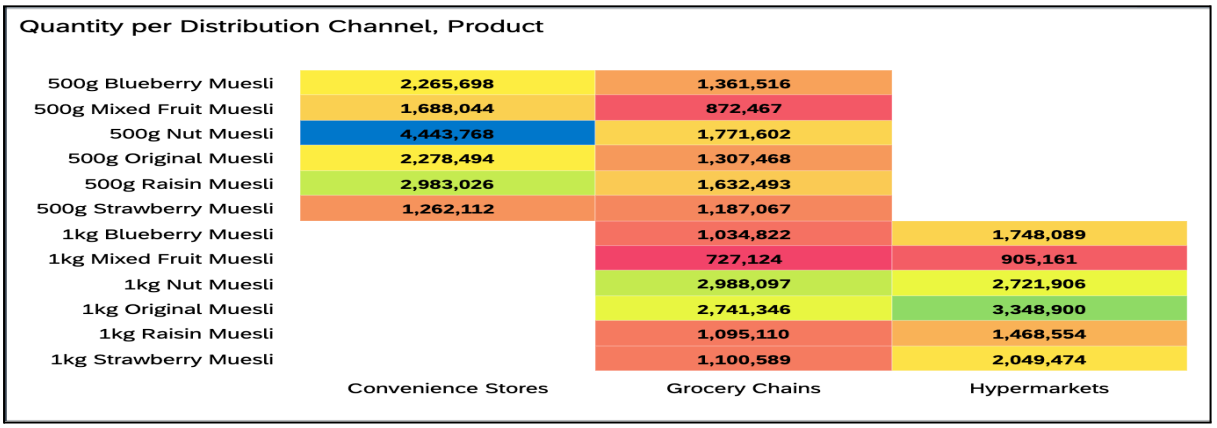


Figure 7: Quantities of Products Sold across Distribution Channels

The 500g variations of Muesli are unavailable in Hypermarkets and the 1kg variations are unavailable at Grocery. This might be a strategic decision as Convenience Stores have higher prices and usually carry smaller variations of products due to them buying smaller quantities at higher per-unit prices. Hypermarkets, on the other hand, offer better value for money for larger variations and perhaps decided that they should only keep the 1kg Muesli in stock. An additional insight generated from this are the performances of 500g Nut Muesli in Convenience Stores and 1kg Original in Hypermarkets. Grocery Chains tend to be doing poorly overall other than two variations of the product. Perhaps the decision makers could reallocate resources to the other two more profitable channels accordingly.

4.4 Analysis & Critique of the Tool

What immediately stood out is how SAC automatically generated insights by simply running the Smart Discovery feature. It generated four pages of insight (Overview, Key Influencers, Unexpected Values and Simulations) before adding new pages into the workspace and manually manipulating data and generating visualizations to

derive insight. This can not only add to the analytics process but also guide it by exploring those areas further.

I was very impressed by the Key Influencers page that was automatically generated. Not only does it show what could influence an important metric such as revenue but the fact that it was possible to run comparisons between List A and List B in Key Influencers section was so straightforward and can allow the user to derive a lot of useful information especially pertaining to KPIs.

While the interface was designed to look simple, it can be quite intimidating with the sheer number of options on display. SAC was not very user-friendly in that sense as the information overload might be too much and the user needs to learn how to use the tool properly before being able to harness its potential. This might be off-putting for people who have not been formally trained to use the platform. One feature that was especially frustrating was finding the Chart Builder after creating a chart. There seemed to be no straightforward way to access it if it was hidden other than using the Edit Styling option and then switching over to it.

4.5 Conclusion

I found SAC to be a challenging experience yet a rewarding one, nonetheless. Learning the platform proved difficult at times but the insight it could generate if one properly learns the capabilities of the tool are outstanding. It is certainly a tool I would like to explore in further detail to gain mastery of it.

Chapter 5 Tableau

5.1 About Tableau

Tableau is a powerful business intelligence platform that simplifies data-driven decision making. The state-of-the-art visualization capabilities allow analysts and organizations to uncover patterns in data and harness them in decision-making.

Data Import and Blending: Tableau accepts Data sources in various formats such as Excel, Text files and even PDF. However, what elevates Tableau is the additional ability to connect to servers and databases to retrieve data. Data from multiple sources can be blended together through methods such as forming relationships.

Real Time Analysis/Visualization: Arguably, the most popular feature of Tableau is the ability to simplify what can be a complex process: Analysis and Visualization. Once the dataset is ready, Tableau uses a drag and drop interface to create powerful visualizations from a variety of dimensions and measures to uncover insights.

Creating Stories: Tableau does not only allow the user to create visualizations but also lets them arrange these in a sequence to create an interactive journey through multiple visualizations. This ability to organize individual visualizations into a 'bigger picture' streamlines reporting and decision-making processes.

5.2 Dataset & Research Questions

The dataset pertains to Sales Reports for a large retail company across all states in the United States for the years 2018 to 2021. The dataset was provided in the tutorial in the form of an Excel spreadsheet (.xls) and is quite large (21 Columns and 209895 Records). It contains metrics such as Sales, Orders, Profit along with descriptive data such as Product Categories, Sub-Categories and geographical data such as Regions, States etc. This dataset was quite comprehensive in terms of the sheer detail it provided, but if I had to push, I would have liked the Cost of Goods Sold within the dataset to uncover more insight by comparing Costs and Sales. Through a combination of financial and geographical analyses, I will attempt to determine the profitability of the retail organization by answering the following questions:

i. How does Profit Ratio compare across all States?
ii. What are the Sales Trends across Different Categories?
iii. What are the trends for Profit across poorly performing Products?
iv. What are trends for the Sub-Category 'Machines' across Regions?
v. How does profitability vary across States in the South?

5.3 Applying the Analytics Tool & Results

5.3.1 How does Profit Ratio compare across all States?

A new sheet called Profit Ratio by Geography is created. In the Data Pane, select the dropdown menu and create a Calculated Field known as *Profit Ratio* with the formula: *SUM([Profit]) / SUM([Sales])*. After selecting *State* and *Profit Ratio* simultaneously, click on Show Me and select *Map* to generate a visualization that

shows profit ratio in each of the States. In the Marks Card, the context menu for Profit Ratio - Format was selected and switched from Axis to Pane. The number format was changed to *Percentage* Decimal Place set to 1. The profit ratio now appears as a percentage. The final step involves clicking on Color and then Edit Colors. The color palette is set to *Orange-Blue Diverging* and under Advanced, the Start and End points are set as -0.5 and 0.5 respectively. The data can now be visualized as seen below.

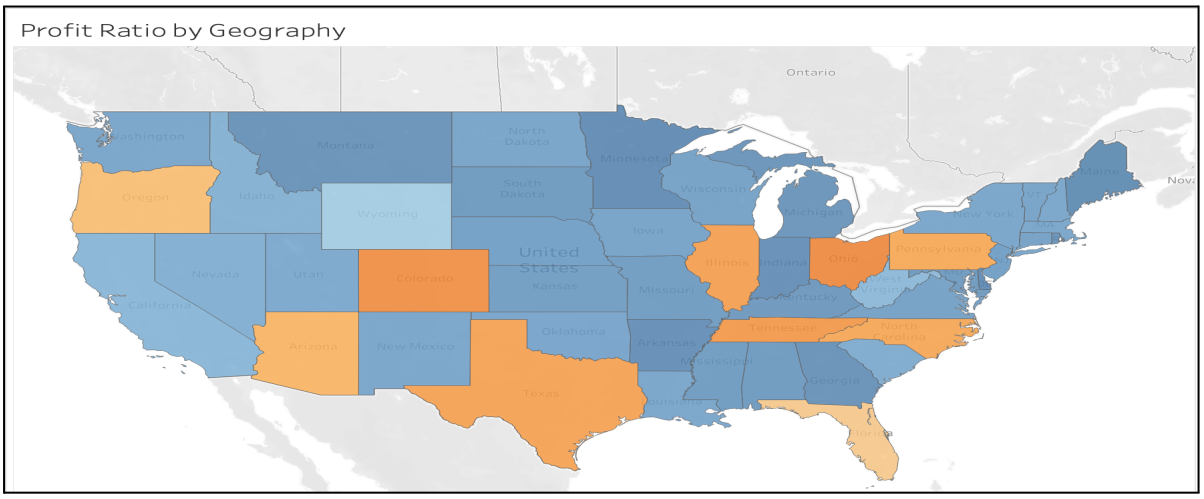


Figure 1: Profit Ratio Map of all States

Colorado, Ohio, Texas and Tennessee immediately stand out as some of the worst performers in this sense and may warrant deeper analysis into their poor performance. To filter the best performing states, *AGG(Profit Ratio)* was selected from the Marks card and after clicking Edit Filter, Navigate to At Least and under that select Start as 0.3. This will show the states with a profit ratio of 30% or greater.

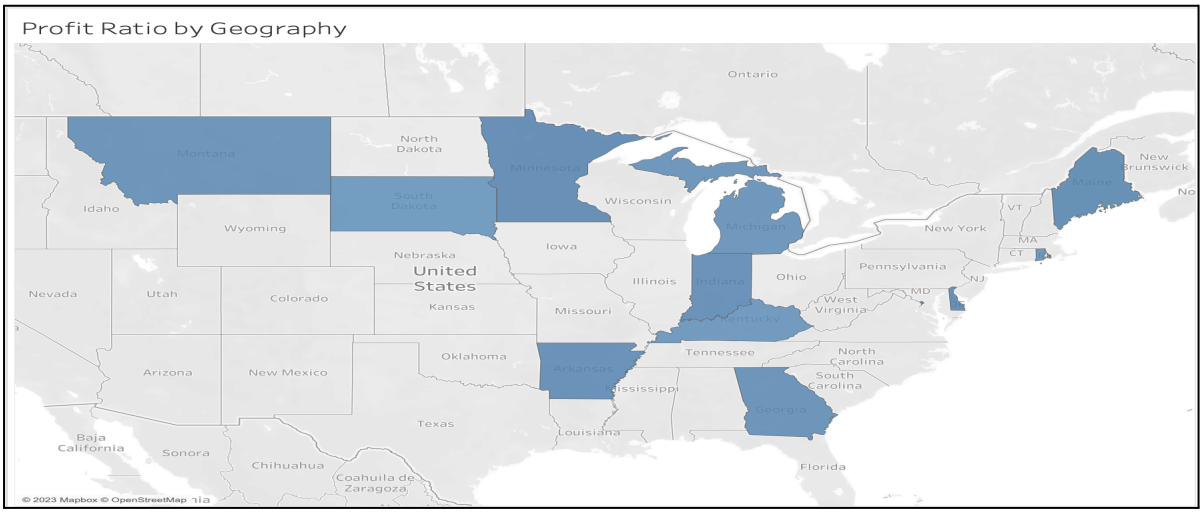


Figure 2: Profit Ratio Map with States showing Profit above 30%

We can see that 12 states were able to generate a Profit Ratio of greater than 30% and these are the best performers across the United States.

5.3.2 What are the Sales Trends across Different Categories?

From the Data Pane, *Order Date* was moved to the Columns shelf and *Sales* from the same pane was put into the Rows shelf. After this, *Category* was moved to the right of *YEAR(Order Date)* in Columns.

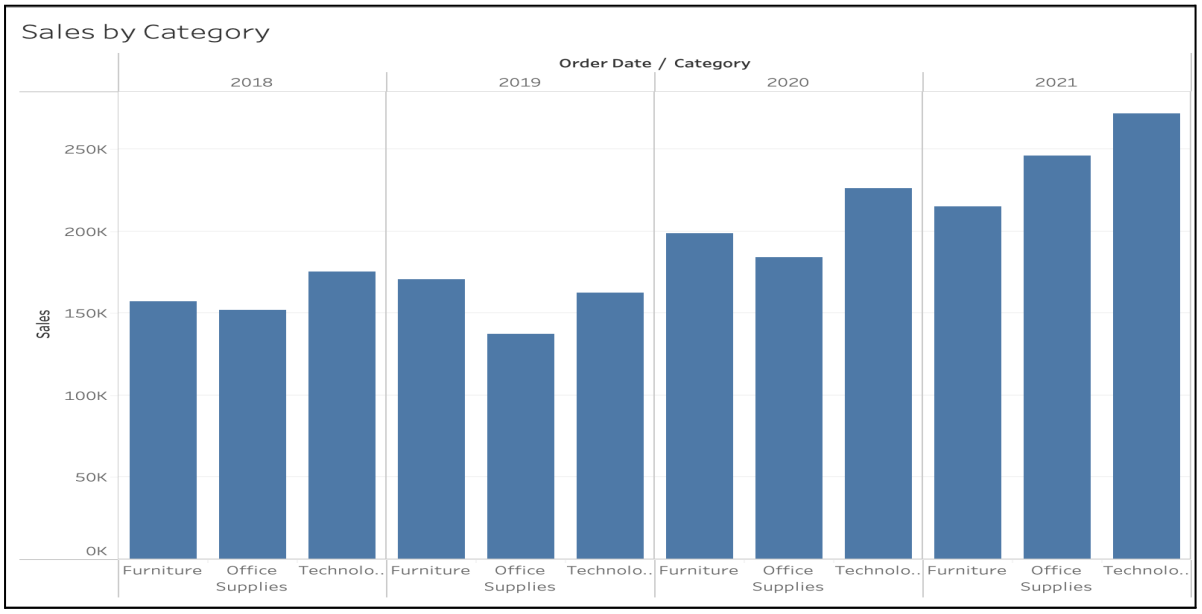


Figure 3: Sales across Product Categories for 2018-2021

Office Supplies and Technology declined from 2018 to 2019 but grew each year in 2020 and 2021. Technology was the best performer in 2021 for Sales followed by Office Supplies and then Furniture. However, Furniture showed growth year on year.

For focused analysis, Sub-Category was taken from the Data pane and moved into Color in the Marks card. This generates the visualization in *Figure 4.*

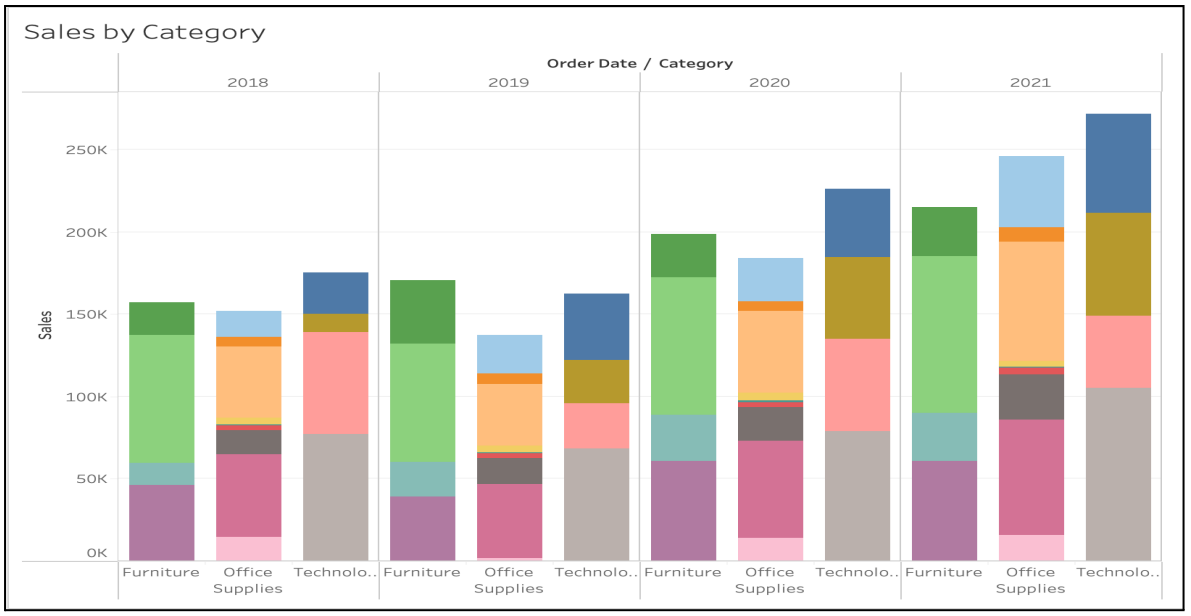


Figure 4: Breakdown of Sales by Sub-Categories across Product Type for 2018-2021

Within Furniture, Chairs seem to generate the most sales and this trend continues over the years. In Office Supplies, Binders and Storage seem to form the bulk of the sales and this trend is repeated from 2018-2021. Lastly, in Technology, Machines have shown inconsistent sales across the years as they formed a large share of sales in 2018, followed by a sharp decline in 2019, followed by growth in 2020 and then declined again in 2021. It was interesting to see the growth in Accessories and Copiers over the years, forming relatively significant shares of the overall sales by 2021. Perhaps these are two sub-categories management can focus on even further to generate more sales.

5.3.3 What are the trends for Profit across poorly performing Products?

Sales were added to Rows. The following three were then added to Columns in this order: *Order Date, Category, Sub-Category*. From the Data pane, *Profit* was moved to Color on the Marks card. It was found that there is negative profitability across Tables, Bookcases and even Machines. To focus on products with negative profitability, all the check boxes in the Sub-Category filter card except Bookcases, Machines and Tables are cleared. The results can be seen in *Figure 5.*

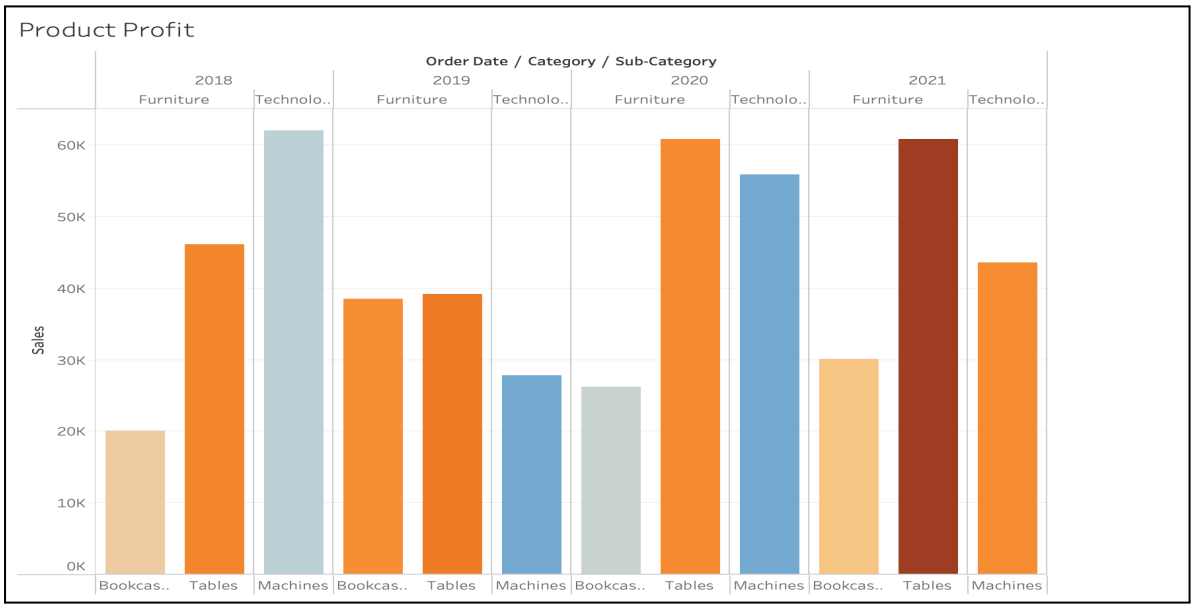


Figure 5: Sales & Profit comparison of Sub-Categories (Machines, Bookcases, Tables) for 2018-2021

Bookcases have been poor performers throughout the years with either low profits or negative profits. Machines are very inconsistent in their profitability as they generated a negative profit in 2018, followed by two years of positive profits and once again negative profits. Tables; however, have generated negative profits each year and perhaps, the management team could look into eliminating this line of products. Whereas, further investigation into Machines and Bookcases might be warranted to see if they can generate positive margins.

5.3.4 What are trends for the Sub-Category 'Machines' across Regions?

Figure 6: Profit from Machines in all Regions

As learned in the previous question, Machines have shown inconsistent results across the years and it was decided to drill down into this data. All of the Columns and Rows are kept identical as the last question. In the *Sub-Category* filter card, only Machines are selected. From the Data Pane, Region is added to the left of *Sales* in Rows.

Figure 6 reveals that Machines in the South are resulting in a higher overall negative profit than other regions for the period of 2018-2021. Management could drill down into this further and look for patterns within *Cities* and individual *Product Types*.

5.3.5 How does profitability vary across States in the South?

In a new worksheet, *State* from the Data pane was added to *Detail* in the Marks card which generates a map visualization. To enhance this, *Region* was moved to Filters shelf and then filtered down to the *South* only. The map will then focus on the South Region. Add the profit measure to Color on the Marks card. After clicking SUM(Profit) on the Marks Card and selecting Format, the format was changed to *Currency (Custom)* and decimal places set to 0.

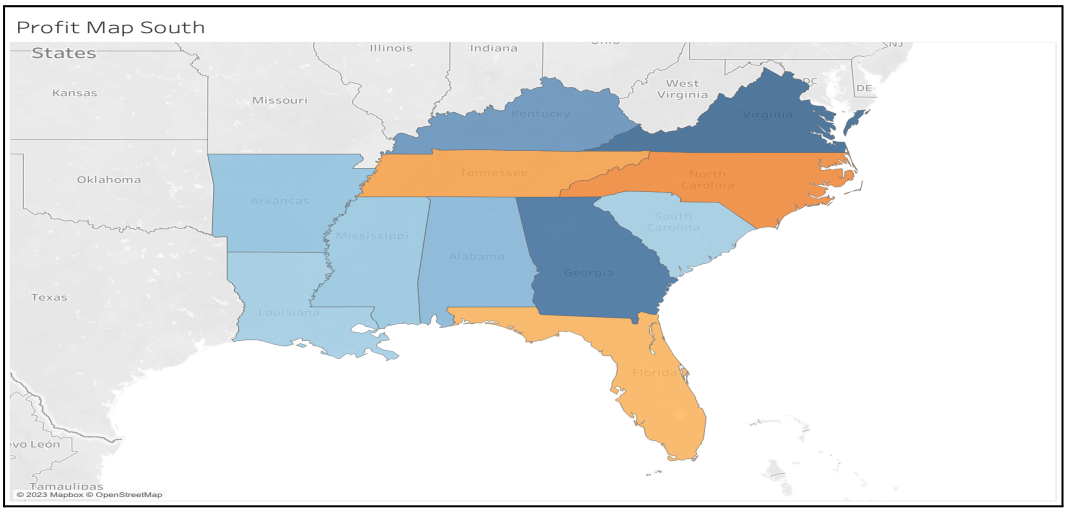


Figure 7: Profit Map for the South Region

Tennessee, North Carolina and Florida seem to be the poorest performers as they are all generating negative profits for the company. Managers or Analysts could explore this problem further by drilling down into data for these regions (Cities, Product Categories) to uncover more insights regarding poor performance and take corrective measures.

5.4 Analysis & Critique of the Tool

Tableau was one of the tools I was looking forward to learning and it did not disappoint. What immediately stands out is the interface. It is both simple yet looks elegant and the Data Pane was placed perfectly next to the visualization area. The drag and drop feature may be common amongst BI tools yet Tableau's ability to visualize that data sets it apart from its contemporaries. The platform automatically generates a visualization based on the type of data added to the workspace. This alone can generate meaningful insights, especially for people without knowledge of analytics, but skilled analysts can make changes to perform ad-hoc analysis. It is, in sum, very user-friendly.

Secondly, the ability to generate multiple visualizations and separate them into different worksheets is a feature that stood out. This allowed me to not only keep my work organized, but allows for a step-by-step analysis and linking it all together. It also helps that worksheets can be duplicated into a new one and then altered to generate a new visualization without starting over and preserving the previous one. The ability to organize all the worksheets and then create an interactive story (A presentation of sorts) just sets it apart from others as a BI tool.

It is hard to find much fault with Tableau but if I had to really push it would be the limited data preprocessing. Tableau does allow you to do basic procedures such as joining and blending data but data cleansing is an essential step before loading it. Tableau is strictly a visualization tool but to perform that, one would need to use Tableau Prep or some other tool. Also, not being able to save my work locally was frustrating.

5.5 Conclusion

Throughout this portfolio, Tableau has been my favorite BI tool to use so far. It helped that I had used it in the past but what makes it a favorite of mine and many other analysts is the very simple learning curve. Although, it definitely does help but a user does not need to have a formal analytics education to be able to generate valuable insights from Tableau. The balance between user-friendliness and BI capabilities is outstanding.

Chapter 6 SAP Predictive Analytics

6.1 About Predictive Analytics

SAP Predictive Analytics is a multi-faceted analytical tool that allows for multiple data-related processes such as data analysis and data mining. It contains multiple interfaces: Automated Analytics and Expert Analytics. For this Chapter, Expert Analytics was used to uncover trends within sales data for GB.

Data Preparation: Expert Analytics enables the user to prepare data before loading in a variety of ways. Through the Prepare tab, the dataset can be transformed through creation of new dimensions and measures, formation of hierarchies (i.e time, geographical). Existing datasets can also be merged with other datasets to create new ones.

Data Analysis and Visualization: Once the dataset is prepared, there is a Visualize tab that contains a variety of techniques such as Heat Maps, Scatter Graphs, Line Charts and Tree Maps amongst others. All the essential data visualization techniques are present in the Expert Analytics interface and visualizations can be added as new pages.

Predictive Analysis: Through a combination of built-in predictive algorithms or allowing for the import of statistical frameworks such as R programming language, models can be created and trained to offer predictive insight (i.e as forecasting).

6.2 Dataset & Research Questions

There are two datasets for this chapter: Customers and GB_Data_GM. They are both in the form of .xlsx files. The Customers dataset is quite simple and simply tells us where the names of the Customers of GB and where they are located (US/DE and cities). This contains 150 records. The GB_Data_GM dataset is more comprehensive and contains transaction data such as Net Sales, COGS, Discounts and Sales Quantities etc. This dataset contains 36,968 records for the period between 2015-2018. The data had to be transformed (i.e creation of new measures such as Gross Margin) and the two datasets needed to be merged before analysis could occur. I will attempt to uncover trends within the financial data for GB which could help decision-makers understand how their organization is performing in terms of profitability.

i. Which product has the highest sales quantities? Does it appear that the ratio of sales by product changes over time?
ii. What are some of the customers with the highest gross margin ratios on average?
iii. Which product has the highest sales quantities? Does it appear that the ratio of sales by product changes over time?
iv. Which product contributes the most to total Gross Margin?

6.3 Applying the Analytics Tool & Results

6.3.1 What does a heat map tell us about seasonality at GB? What other type of chart might be more appropriate for identifying seasonality?

After selecting Heat Map from Map Charts, Revenue in USD was added to the Area Color. Month was added to the Area Name and Years was added to Area Name 2.

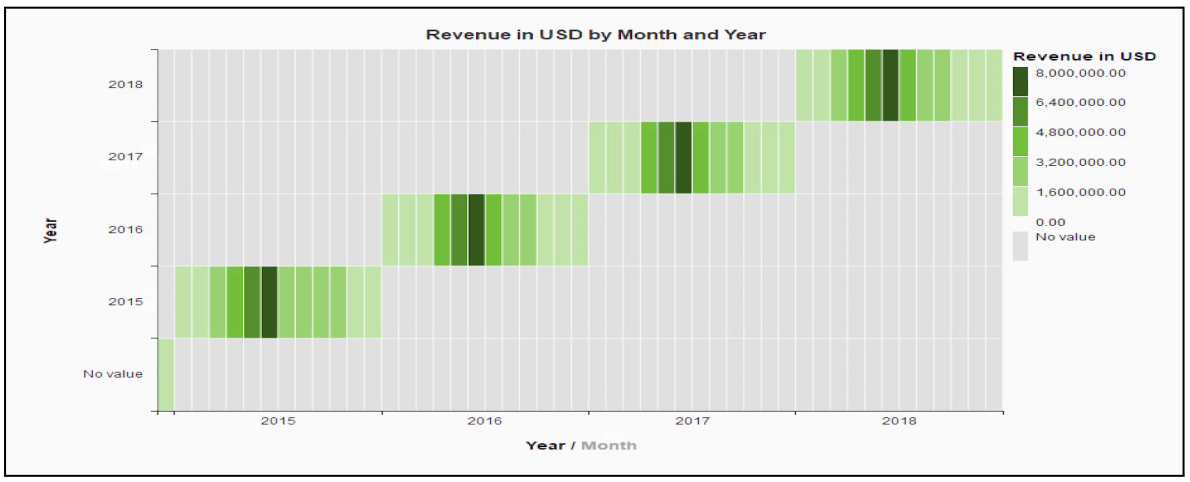


Figure 1: Heat Map of Revenue across Years (& Months)

The difference in color schemes between the bars in each rectangle highlight that there is seasonality in Revenue. There is an upward trend from January to a peak in June. For 2015, Revenue for April is around $4.35 million followed by approximately $5.4 million in May and $7.0 million in June. After that it falls toward the end of the Year. The trend is almost exactly the same in every following year.

After creating a duplicate chart, A Line Chart was selected and it transferred over dimensions from the Heat Map without any additional effort.

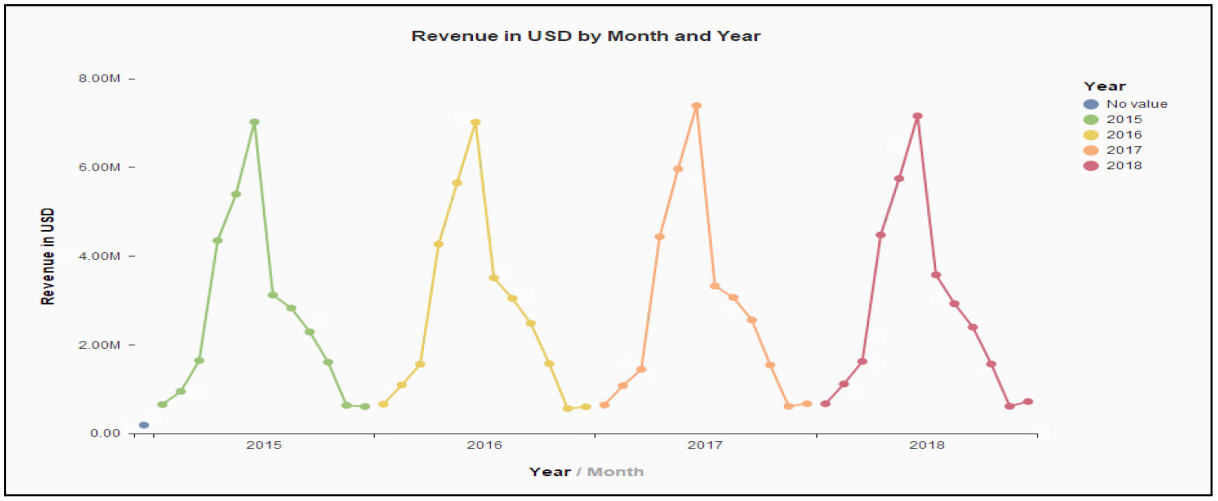


Figure 2: Line Chart depicting Seasonality

The line chart is much easier to comprehend for decision makers and users alike as it clearly shows peaks and troughs that reflect seasonality unlike the Heat Map during different seasons. The trends are the same as the Heat Map.

6.3.2 What are some of the customers with the highest gross margin ratios on average?

Under Scatter Charts, Bubble Chart was chosen. Sales Quantity was added to the X-Axis and Revenue in USD on Y-axis. Gross Margin Ratio was selected as the Bubble Width and the Aggregation Method selected was AVG. Customer was then added to Legend Color and a Filter for Gross Margin Ratio was added with a Between setting and with values 0.3 as Start Point and 0.4 as End Point.

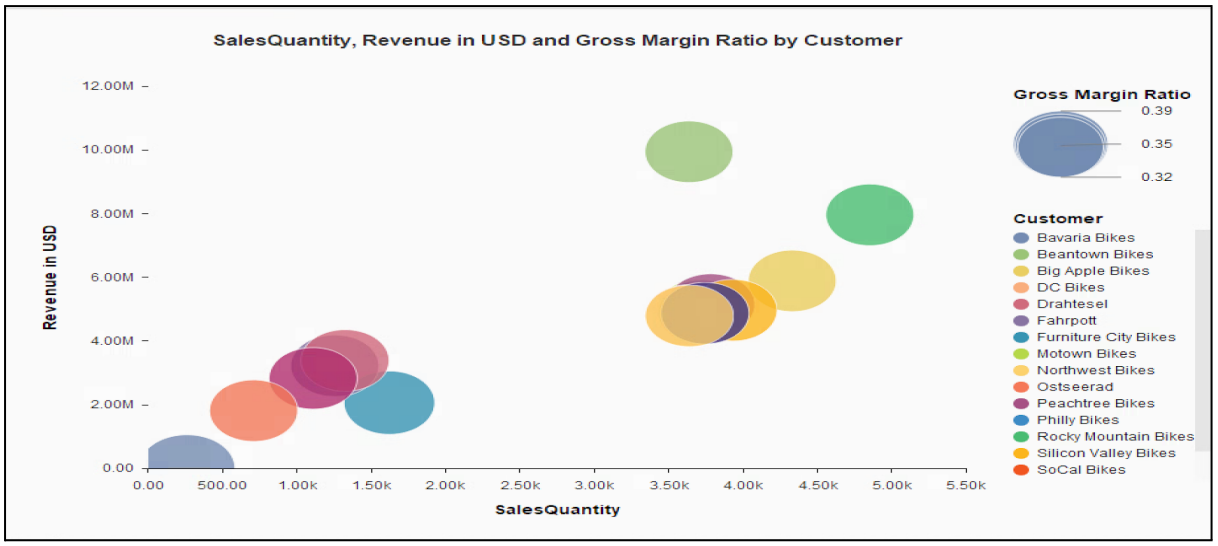


Figure 3: Bubble Chart with Revenue, Sales Quantities and GMR

There is no clear relationship between Revenue/Sales Quantity and Gross Margin Ratio (GMR) from the Bubble Chart. There are Customers that generate lower Revenue overall but still provide the same GMR as some of the Customers on the higher end of the Revenue/Sales Quantity axes. One explanation might be that those on the higher end of the axes get larger discounts; therefore, accumulate more Sales and Revenue, but the GMR remains similar.

6.3.3 Which product has the highest sales quantities? Does it appear that the ratio of sales by product changes over time?

A Pie Chart was selected. Sales Quantity was added to Pie Sectors and Product to the Legend Color. Sorting was then done in Ascending Order.

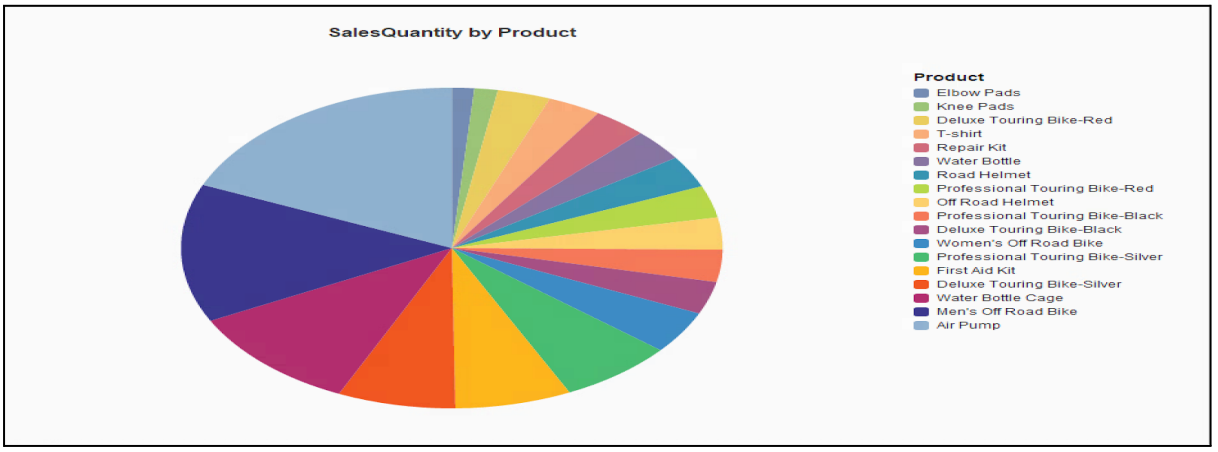


Figure 4: Pie Chart showing Sales by Products

Both the Product Legend and actual Chart indicated the same: Air Pump clearly had the highest ratio of sales followed by Men's Off Road Bike. Hence they seem to be the most popular Product offerings. After unsorting the Pie Chart, the Trellis was expanded and Years dimension was added in Columns.

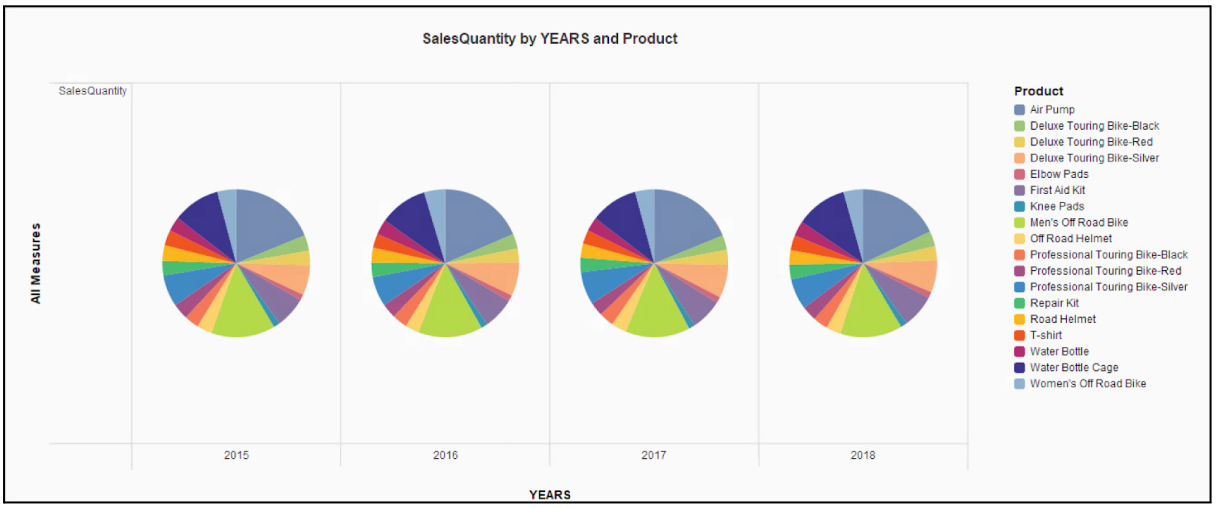


Figure 5: Pie Chart showing Sales by Products from 2015-2018

It appears over time that the Ratio of Sales by Product remained relatively similar over the years with Air Pump maintaining the largest share of Sales Quantity and Men's Off Road Bike staying in second. This highlights a consistency in Sales patterns for the company and is a positive sign.

6.3.4 Which product contributes the most to total Gross Margin?

From Other Charts, a Waterfall Chart was chosen. Gross Margin in USD Measure was added to the Y-axis, and Product was added to the X-axis. The sorting was done in Descending Order.

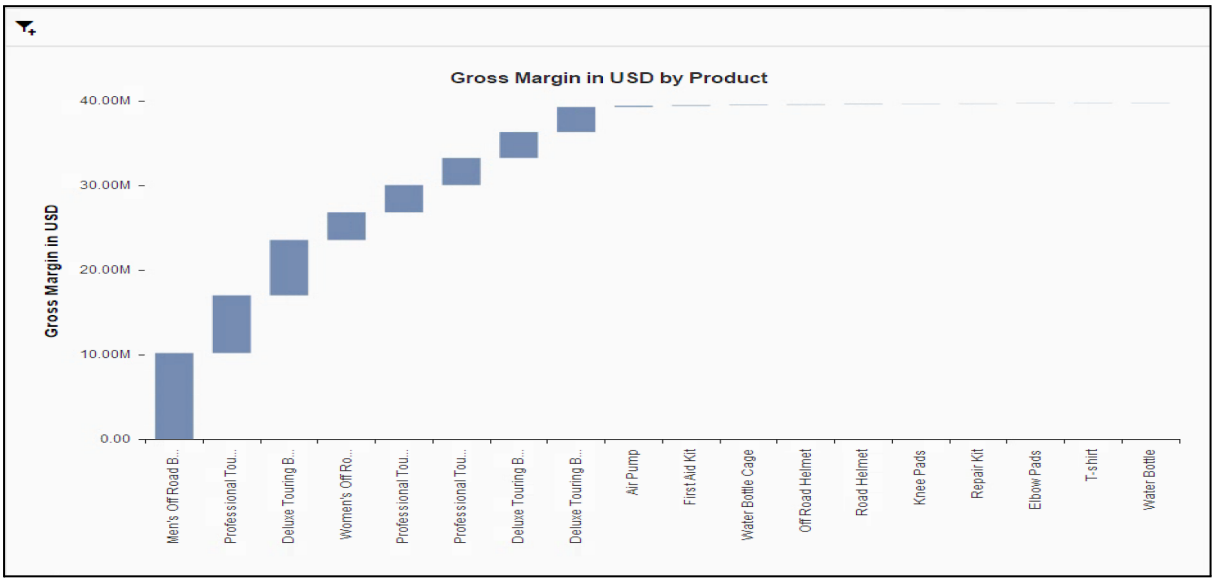


Figure 6: Gross Margin by Product Type

The Waterfall Chart shows that Men's Off Road Bikes were the largest contributors to the overall Gross Margin of GB followed by a wide variety of other Bikes. It was interesting to note that noticeable Margins came from Bikes yet accessories like Air Pumps and Water Bottles had little to negligible Margins. Perhaps, decision-makers could look into this further to understand why that is so and reallocate resources as needed.

6.4 Analysis & Critique of the Tool

I found SAP Predictive Analytics to be an intuitive application in certain areas like the interface to the actual analytics and visualization. The user-interface was very simple and easy to become familiar with. It was very useful that the tool acts as a 2 in 1 and allows you to not only merge/transform datasets but also visualize them. The visualization capabilities were also more impressive than expected with a variety of Chart and Graph types that can be separated into different pages.

Secondly, when I mention intuitiveness, it applies to multiple facets. Firstly it was noticeable that when I switched over from a Heat Map to a Line Graph, PA automatically transferred over my settings from the previous chart and allowed me to create a more suitable visualization without any additional effort. Secondly the Compose tab is very useful if this analysis needed to be presented in a condensed report form. It allows users to create infographics from existing visualizations using a drag and drop feature and makes the reporting process very simple.

There are not too many flaws with PA as it covers all the essentials for data transformation and analysis quite well. The biggest drawback was likely having to use it in a Virtual Machine environment since it does not seem compatible with Mac OS. This made the application a bit sluggish and made saving my work somewhat cumbersome. This might be problematic for even larger datasets.

6.5 Conclusion

Overall, SAP PA was one of the better tools I have used so far in this portfolio. After an initial learning curve, it becomes simpler to use as one gets more experience with it. Although, I would have liked to use the predictive capabilities to form a full opinion on it. Perhaps, that is something I can look at doing in the future.

Chapter 7 SAP Analytics Cloud

7.1 About SAC

Although SAC has been used previously in this portfolio, it was used for data analysis and visualizations. In this chapter, SAC will be used in a different context: statistical analysis (i.e regression) and for its predictive capabilities. It can provide a lot of value to organizations to use past trends to predict future ones.

Creating Predictive Model: SAC allows users to create predictive models from existing datasets. A dataset (i.e Excel) can be converted into SAC format and used to train a predictive model (regression) that can be applied to generate predictive insight (i.e forecasting).

Applying Predictive Model: Once the model has been created and the Predictive Confidence and Root Mean Square Error (RMSE) indicators are deemed sufficient in terms of accuracy, they can be applied to generate predicted values for given scenarios. This can be useful in terms of forecasting.

7.2 Dataset & Research Questions:

The first dataset (GB_AnalyticsData) is in the form of an Excel file (.xlsx) and is a very large dataset. It contains 25 Columns and 2736136 records along with financial measures such as Revenue, Discounts, Costs and Sales Quantities. It also has Dimensions such as Customer/Product Descriptions, Categories and Geographical Data (Country). This dataset covers the period of sales from 2007-2019 and will be used to train the regression model so it can be applied to the second dataset which is QuantityPrediction.xlsx. This just contains one record with certain attributes (Transaction for the Year 2021). Perhaps, the first dataset could have been even better if it listed the type of Sales Channels for Sales as that would also impact the Sales Quantities. By creating and applying the regression model, the following questions can be answered:

i. Which variable contributes the most to the regression model?
ii. What is the predictive confidence of the model? What is the Root Mean Square error? Based on these values, is this a reliable model?
iii. Which customer (CustDescr) has the most Influence on the model?
iv. After creating a 2nd regression model, which variable contributes the most to the regression model?
v. Looking at the Performance indicators, which model is better?
vi. For Sales Org Germany North, (DN00), what are the expected sales of Professional Touring Bike for a transaction with the given attributes?

7.3 Applying the Analytical Tool & Results

7.3.1 Which variable contributes the most to the regression model?

To create the first regression model, Predictive Scenario was selected from the menu. The Target under Predictive Goal was SalesQuantity. Since this is Regression Analysis of Sales Data with Selected Influencers, all influencers were excluded

except for Month, CustDescr, City, SalesOrg, ProdDescr, CatDescr, SalesQuantity. The model was then trained.

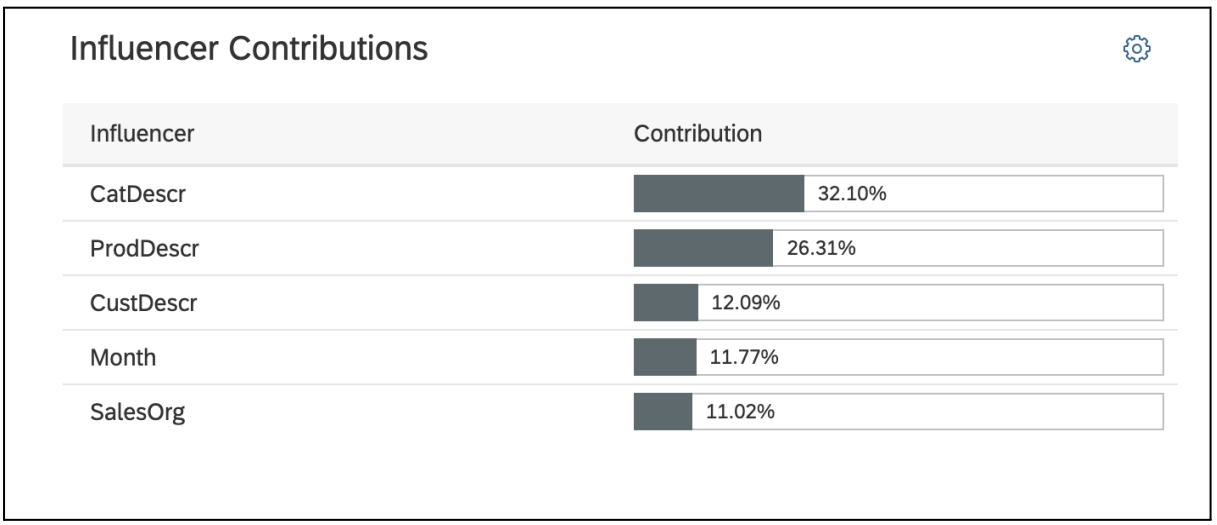


Figure 1: Influencer Contributions for first Regression Model

After creating the first regression model, we can see that the Influencer with the largest contribution was *CatDescr*.

7.3.2 What is the predictive confidence of the model? What is the Root Mean Square error? Based on these values, is this is a reliable model?

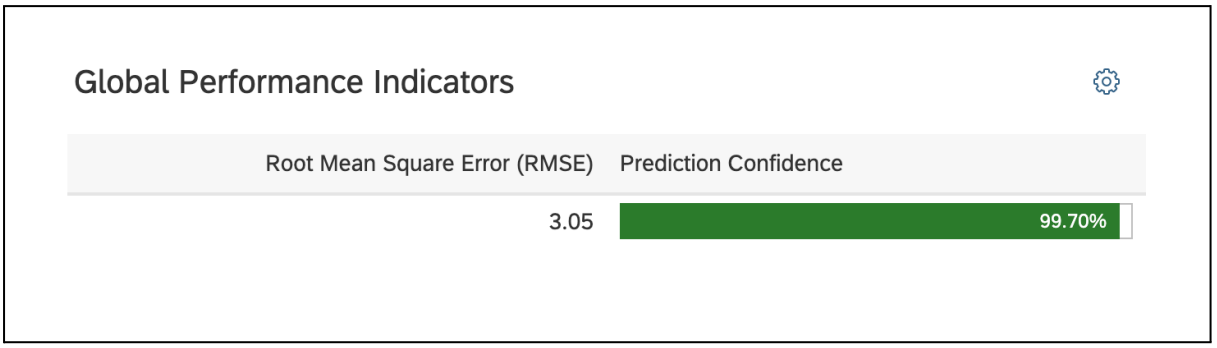


Figure 2: Performance Indicators for first Regression Model

In the Overview tab, we can see that the *Prediction Confidence* of the model is 99.70% which is very high and should speak to the ability of the model as an accurate one; however the *RMSE* is 3.05 which is significantly above 1. Ideally the model could be trained to be better for an *RMSE* in the range of 0-0.5 to be more reliable.

7.3.3 Which customer (CustDescr) has the most Influence on the model?

In the Influencer Contribution Tab, we set *CustDescr* in the Grouped Category Influence.

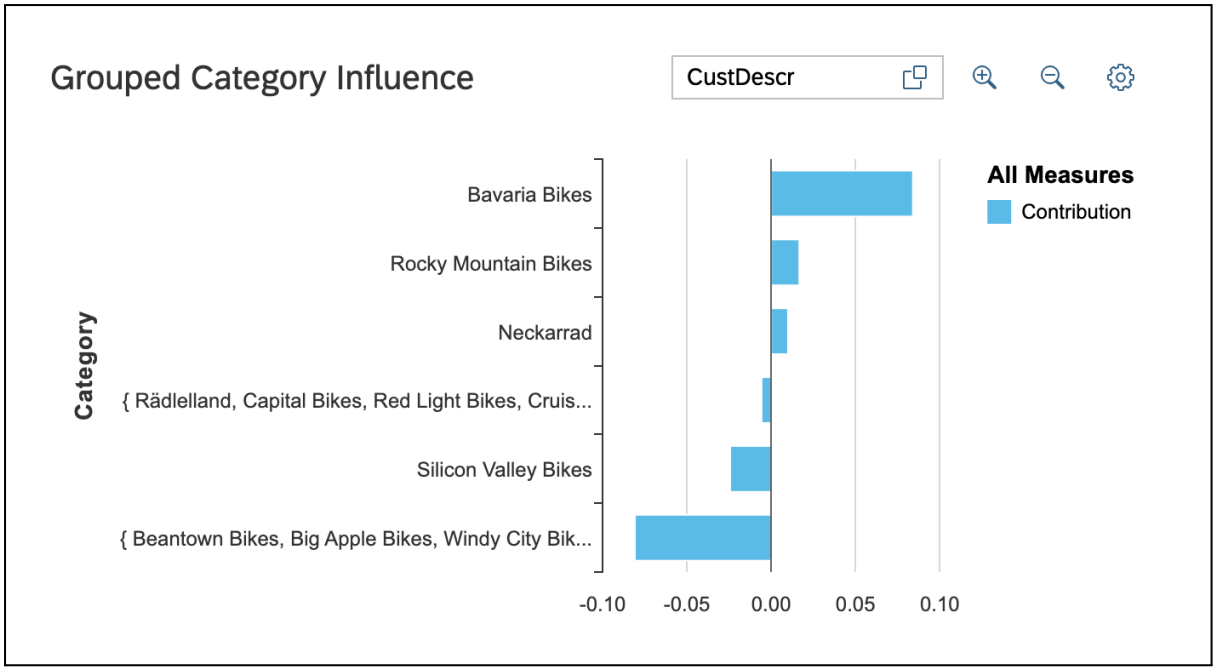


Figure 3: Influence of Customers on Regression Model

We can see that Bavaria Bikes has the largest influence on the model.

7.3.4 After creating a 2nd regression model, which variable contributes the most to the regression model?

To create a second regression model, the steps required to create the first model were replicated; however, since this is a regression model without choosing influencers, no influencers were excluded and the Limit Number of Influencers was Toggled On. The default setting of 10 was kept.

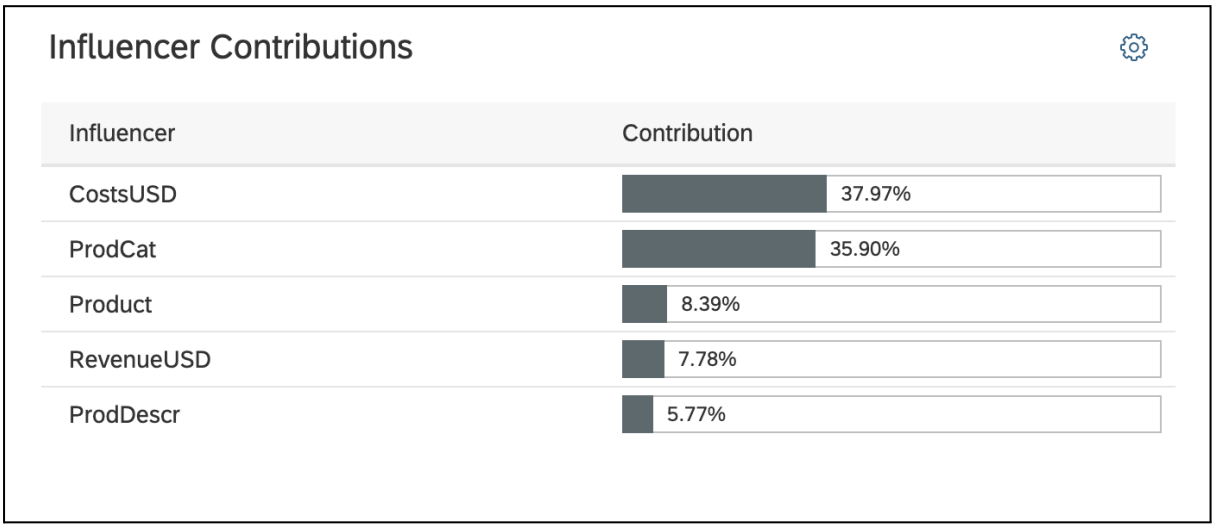


Figure 4: Influencer Contributions for second Regression Model

Here we can see that Costs in USD had the highest contribution with 37.97% followed closely by Product Category with 35.90%. This does make sense as often the Costs associated with a product can reflect the quality of the product and therefore, sales. Product Category having a large influence also makes sense as Products in certain categories tend to sell more than others.

7.3.5 Looking at the Performance indicators, which model is better and why?

As seen below, the *Prediction Confidence* of the second model is 99.99% which could almost be considered perfect. What makes this model better than the first is an *RMSE* of 0.27. This falls between the 0-0.5 range and is significantly lower than 3.05 in the first one. This speaks to greater accuracy of predictive capabilities of this model.

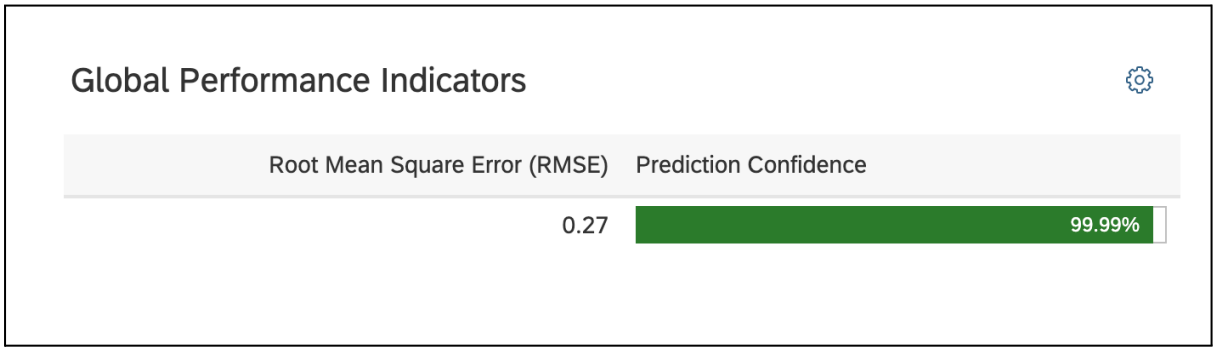


Figure 5: Performance Indicators for second Regression Model

7.3.6 For Sales Org Germany North, (DN00), what are the expected sales of Professional Touring Bike for a transaction with the given attributes?

Once the second dataset is created in SAC from the Excel File known as QuantityPrediction.xlsx, and the Measures and Dimensions are correctly determined to match those in the transformed *GB Sales* dataset, the dataset is saved. After selecting the Regression scenario, *Apply Predictive Model* is clicked and the first model is used for this.

The Data Source is the *QuantityPrediction* dataset and Replicated Columns are *Year, Month, SalesOrg, Product, ProdDescr, ProdCat, CatDescr and Division*. In Statistics & Predictions, *Predicted Values* is chosen and the Output As is “GB Regression Results”.

Month	SalesQua...	SalesOrg	Product	ProdDescr	ProdCat	CatDescr	Division	Predicted ...
1	*null*	DN00	PRTR1000	Professional Touri	TOU	Touring Bike	BI	7

Figure 6: Predicted Value for transaction after applying Regression Model

After successfully applying the Regression Model to the dataset, we can see from the screenshot above, that the predicted value of Sales Quantity in that given transaction is 7.

7.4 Analysis & Critique of the Tool

One of the things I really appreciate about SAC in this context was how simple it was to create a predictive model. Having an existing dataset for time-related data made the entire process very intuitive. The major inputs I had to provide on my part was making sure the dataset was in SAC format and that the Measures and Dimensions were correctly identified. From that point, SAC generated a model with a few clicks with impressive reliability metrics. I also appreciated how it showed metrics from the training data to gain more insight into the effectiveness of the model and what contributes the most to it through Classification/Regression visualizations.

Applying the Regression Model was also a very simple process. One simply has to make sure the correct model is selected from the list and click the Apply Predictive Model button. From there, it is only a matter of providing three inputs to generate predictive insight. It also gives users the select Influencers in that model for the user automatically, and this is very useful for people who do not have a formal education in Data Science or Machine Learning and can allow them to create simple predictive models without a very steep learning curve.

There were not too many flaws with SAC given its simplicity but, perhaps, it might have been useful to include the significance of performance indicators. This would be very beneficial for people without formal analytics training to understand what each performance indicator actually means (i.e RMSE and Prediction Confidence).

7.5 Conclusion:

Overall, I find SAC to be a very enjoyable analytics tool. Whilst in the past, it was used for data visualizations, it made statistical analysis a very simple and non-tedious process. After using SAC in this context, I would like to apply regression to another dataset and create more models. Not only that, I would like to explore more tools that allow predictive analytics in the future and compare them to SAC.

Chapter 8 Executive Summary

8.1 Overview of the Portfolio

This portfolio gave me the opportunity to learn and showcase a variety of skills related to Business Analytics and Data Visualization over the course of four months. I performed several processes such as Data Preprocessing, Data Modelling, Data Transformation to ensure the data was ready for analysis. This portfolio shows the results of these analyses, using a variety of analytics tools and visualizations.

8.2 Excel

Excel is synonymous with data processing and analysis. This was an opportunity to learn some of the more advanced features of Excel (PivotChart and PivotTable). This was a very important step as it allowed me to learn the fundamentals of data analysis and generate simple visualizations from sales data. This knowledge was very useful as it was transferable to the more advanced tools in the portfolio and made the learning curve for those more manageable.

8.3 SAP Lumira Discovery

Lumira was one of the first SAP tools that I had the opportunity to use. It was the first data visualization tool used in the portfolio and was used to derive insights on player performance at the FIFA World Cup 2014. Here, it was especially important to become familiar with processes such as the creation of calculated fields and how they are able to provide more insight by adding additional layers to the analysis process.

8.4 IBM Cognos Insight

IBM Cognos was very useful as it allowed me to learn the fundamentals of financial data analysis and report creation. This was very suitable for ad-hoc analysis as it allowed me to make alterations to the data table and these changes were represented instantly in the visualization pane (which could be seen in the same window). I learned how to drill-down into financial data and uncover negative trends across channels and products here that could potentially help decision-makers reshape their strategies.

8.5 SAP Analytics Cloud

This was the first of two times SAC was used in this portfolio. Using ERP simulation data was nice as it allowed me to analyze the performance of teams involved and the effectiveness of sales channels and how they contributed to overall revenues for a given product. This chapter was quite important as it taught me the fundamentals of data analysis and how it applies to resource allocation and business strategy.

8.6 Tableau

Tableau is one of the most popular BI platforms in the world and I was especially looking forward to using it for this portfolio as it is something I had used in the past.

The goal with this chapter was to become adept at uncovering patterns within large datasets through a variety of visualization techniques, and then bring them together in the form of a cohesive report that could provide insight to decision-makers and inform future decisions and investigate poorly performing divisions/locations. This was also a chapter where I learned how to do geographical analysis.

8.7 SAP Predictive Analytics

Another one from SAP's suite of offerings, SAP PA (Expert Analytics) was used for its data analysis and visualization techniques in this chapter. It was used immediately after Tableau so that I could compare the two in terms of their efficiency when it came to deriving insight from data. Although Tableau was clearly superior when it came to visualizing data, Expert Analytics did a solid job in analyzing the dataset that was used. Still, I would like to explore its predictive capabilities in the future.

8.8 SAP Analytics Cloud

This was the second time I used SAC in this portfolio; however, here it was used in a forecasting context as opposed to analysis/visualization. I really appreciated how simple the tool made it for me to train a Predictive Model (Regression) from a given dataset and then apply it to predict sales for a future transaction. In fact, it is something I am going to use for another analytics project. After using it, I want to learn further about statistical techniques and predictive analysis.

8.9 Conclusion

Overall, I am very happy with the progress I have made as I progressed through this portfolio. Starting with Excel, my knowledge of Data Analytics and Visualization was quite rudimentary but after using several analytics tools in various different contexts, I believe I have become quite adept at understanding the insight data analysis can provide to organizations in terms of decision-making, especially when paired with report creation. Perhaps, the next logical step for me is to build upon this and take a Data Analytics certification in the future.