PORTFOLIO

Monali Mawlikar

B00850426

INFO 6513 – Business Analytics and Data Visualization

Dalhousie University

Executive Summary

Exploring Latest Business Analytics Tools 2021

The success of a business depends on understanding the market, planning strategies, and timely actions, historically these were made on hunch or experience of an individual, but now we understand the importance of substantiating this based on data, this has led to the saying "*Data is the new age Gold/Oil"*.

Firms today want to make data-driven decisions and actions, but this data is stored at various locations, in different forms, used by different tools for various purposes, leading to incorrect analysis and inferences. Although it may be true for one department it may not be the same for other departments viewpoint leading to "*Garbage in garbage out*" making it difficult to find the "*One Single Truth"*. For this, it is imperative that the "right data, for the right user, at the right time" is available, and the business user or the analyst can then slice and dice the data for multi-dimensional analysis and reporting. This can be further enhanced by communicating the information gained visually since a picture can communicate the information of a thousand words. Secondly, it is possible to apply statistical analysis and machine learning algorithms to find hidden insights and correlations and to forecast and create what-if scenarios using predictive analytics. Traditionally, this was done by data warehousing personnel and data scientists by developing scripts and algorithms manually and then applying these to the data. This was often technically difficult, tedious, and error prone.

The business user and their requirements are the most important element of business. Time is another important essence to be considered. Hence it is the need of time to have tools that allow the business user to be able to connect to data assets in real-time, perform the required transformation, gain insights, create visualizations, communicate, and collaborate without relying completely on a technical team. This has led to the development of tools such as IBM Cognos insights, SAP cloud Analytics and Tableau, etc.

This business analytics toolkit contains 7 chapters, showcasing the use of six business analytical tools namely MS Excel- Pivot Table, SAP Lumira Discovery, SAP Predictive Analytics, IBM Cognos Insight, SAP Cloud Analytics, and Tableau Desktop. Each chapter consists of a brief description of the tool, the data used, how the tool was used to answer research questions, and followed by an analysis and self-reflection on the tool ending with a conclusion.

In this portfolio, an attempt has been made to utilize the above-mentioned business analytical tools to analyze data belonging to diverse domains such as business, demography, and healthcare data to derive meaningful insights and explore the use of various features of the tools. The tools were evaluated by the ease of loading data, choice of data source, ability to plot different type of chart to derive meaningful insights, ability to customize these charts to be able to highlight a particular piece of information and to reduce cognitive load on the viewer, ability to collaborate with others, and ability to forecast or perform predictive analytics.

MS Excel offers familiarity, ability to slice and dice data in pivot tables, create and edit the visualization and incorporate them in other Microsoft applications and collaboration using Office 365, whereas SAP Lumira Discovery, IBM Cognos insights, and Tableau offer the business user the ability to connect to existing data assets and merge data for various external and internal

sources and of different formats, to create stunning visualization and derive business insights and present them in a form of “STORY” or interactive “DASHBOARD”. On the other hand, SAP Predictive Analytics aids in knowledge discovery by performing automated analysis of the data to provide valuable insights and puts immense power in the hands of the user by integrated machine learning algorithms using the open-source R programming language.

After exploring these tools, it could be inferred that all business analytics tools have some unique features offering various advantages and have some drawbacks. Some offer the functionalities to cater to the needs of huge enterprises such as SAP while some strive for user-friendliness and simplicity and seek advanced features and connectivity. Hence the analyst or business user must decide to opt for a particular tool depending on their business needs, familiarity with the tool and its ability to integrate with their existing data assets.

Table of Contents

Chapter 4 IBM Cognos

4.1 About IBM Cognos
4.2Dataset & Research Questions
4.3 Applying Analytical tools and Results.
- 4.3.1 How the North American territory of the company performed in the Year 2012? What is the trend of revenue and cost in 2012 in North America? Which month shows the highest profitability?
- 4.3.2 Is there a product category that has generated loss to the company?
- 4.3.3 How are the business channels contributing to this loss concerning the above product (Computer & Tablet)?
- 4.3.4 How are the business channels performing in other territories?
- 4.3.5 What are the top 2 products contributing the highest to the Revenue? Name them and the product contributing the least.

4.4 Analysis & Self-reflection of the tool
4.5 Conclusion

Chapter 5 SAP Cloud Analytics

5.1 About SAP Cloud Analytics
5.2 Dataset & Research Questions
5.3 Applying Analytical tools and Results.
- 5.3.1 What are the top 2 products contributing the highest to the Revenue? Name them and the product contributing the least.
- 5.3.2 Explore revenue contribution of distribution channels. Does the sale of the top 2 revenue generating products (from question number 1) vary as per distribution channel?
- 5.3.3 How each team contributes to revenue for each distribution channel?
- 5.3.4 For Hypermarket, illustrate the top 3 highest sales quantity for RR and QQ team.
- 5.3.5 Should there be more focus on certain areas, we try to find the areas with the highest sales for each of the top 3 selling products of RR, QQ team.

5.4 Analysis & Self-reflection of the tool
5.5 Conclusion

Chapter 6 Tableau

6.1 About Tableau
6.2 Dataset & Research Questions
6.3 Applying Analytical tools and Results.
- 6.3.1 What is the total expenditure on health across Canada and the province? What is the average expenditure per capita across Canada?
- 6.3.2 How are the funds utilized across Canada for Healthcare, and for which purpose?
- 6.3.3 Which sector utilizes the maximum funds in different provinces and where is the maximum expenditure occurring?
- 6.3.4 What is the trend of expenditure across Canada over the years? Provide the projected forecast for 2 years in the same chart.
- 6.3.5 What is the trend of health spending as per age groups and gender?
- 6.3.6 An interactive Dashboard.

Chapter 1 Microsoft Excel Pivot Table

1.1 About Excel Pivot Tables

Pivot tables are excellent tools for summarizing, sorting, reorganizing, and analyzing large amounts of data. They help in arranging and rearranging the data to understand the hidden patterns and derive useful insights.[1]

Microsoft Excel allows importing data from various sources and in various formats. A Pivot table can be created from this data. In the Pivot table, one can drag and drop various variables (dimensions and measures) to create Crosstabs instantaneously and visualize the data in various chart forms (Pivot Charts). These pivot charts can be made interactive by adding certain Filters, helping one to dice and slice through the data at ease, and to visualize the data.[2][3] thereby helping a power user or analyst in creating reports and presenting data in a clear and simple format and saving time, inexpensively.

The data table must be such that the column fields are descriptive attributes and contain values for the descriptive field, in other words, values of the same type or representing the same attribute must be placed in one column.

1.2 Description of Dataset and Research Question

For this chapter, "Global Bike Inc. (GBI)" data is used. The data is in *xlsx* format and was provided in the class is the transactional data of the "Global Bike Inc. (GBI)" company which deals in Bikes and accessories, with customers (B2B) in two countries viz Germany and United States. The data is of 5 years period, i.e., from the year 2007 to 2011. The company has 2 divisions namely the Accessories division and the Bike division, and 5 product categories. The Accessories division has 10 products, and the bike division has 18 products.

The 29188 lines of data include 18 columns of variables such material id, material desc, product category (related to products), Customer Desc (name of customer), Country Desc, (country of operation i.e. Germany and the United States), Division (Accessories and Bike), Sales Organisation, sales organization Desc (Germany South, Germany North, US east and US west), Calendar Year/ Month, etc. which are the Dimensions and Cost of goods M USD, Discounts, Revenue, Sales quantity, Gross margin are the Measures and these values provided are in Dollars. No data cleaning was required.

The purpose of this analysis is to create a yearly report and to identify the possible avenues which may be affecting the growth of the company.

1.3 Applying Analytical tools and Results.

1.3.1 What is the current performance of the company?

The revenue generated by a company is a reliable measure to understand the performance of a company. To understand the company's performance, a pivot table was created from the given excel sheet. A crosstab of "Revenue of the year" 2011(current) was created by dragging "Calendar Year/ Month" to Rows, "Sum of Revenue" to the Values field, and a filter was applied to "Calendar year" by filtering out all years except 2011. Insert a Pivot chart, choose Column chart. (Figure 1.a) In this chart, Manual sort was used to arrange all the months alphabetically. All the labels were removed, Axis was formatted to use units in millions. (Figure 1.a)

To get the contribution of the Countries, a new worksheet was created.
For the crosstabs, "Country Desc" was dragged to Rows and "Sum of Revenue" to Values and a filter was applied to "Calendar year" by filtering out all years except 2011. Insert a Pivot chart, choose a Pie chart, and data labels were changed to show values in percentages. (Figure 1.b)

To get the contribution of the "Product Category" and "Division", For the crosstabs, "Division", and "Product Category" was dragged to Rows and "Sum of Revenue" to Values, and a filter was applied to "Calendar year" by filtering out all years except 2011. Insert a Pivot chart, choose a Pie chart, and data labels were changed to show values in percentages. (Figure 1.c)

1 Divisions
2 Countries
28 Products
24 Customers

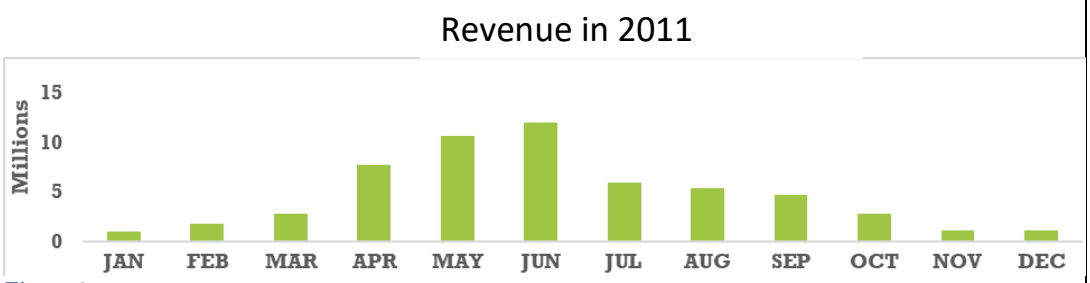


Figure 1.a

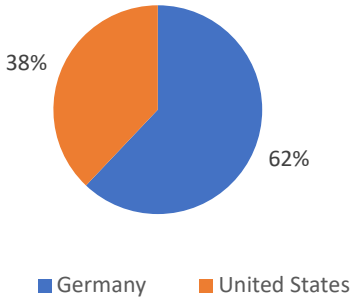


Figure1.b

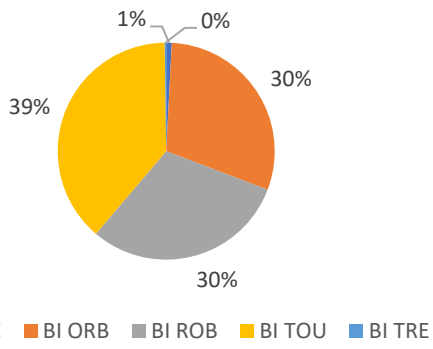


Figure1.c

From the above visualization, it appears that the sales pick up during the spring and the summer season and are less during the winter (Figure 1.a). The German Branches contribute about 62% of the revenue, (Figure 1.b). Among the products, the Bikes division contributes towards most of the revenue, in that the Touring (TOU) product category contributed the highest, whereas the TRE range of bikes and accessories contribute just 1% of the revenue. (Figure 1.c)

1.3.2 *How has the company performed over the years 2007-2011?*

"Calendar year" filter was removed, "Calendar year" was added to Axis (Categories), "Sum of revenue" to the Values field. A line chart was created. (Figure 2. a) showing the trend of revenue over the years 2007-2011. Then "Country Desc" was added to the columns to view how the company has performed according to countries of operation. (Figure 2.b) Then "Country Desc", was replaced with product categories to view the contribution of each product category.

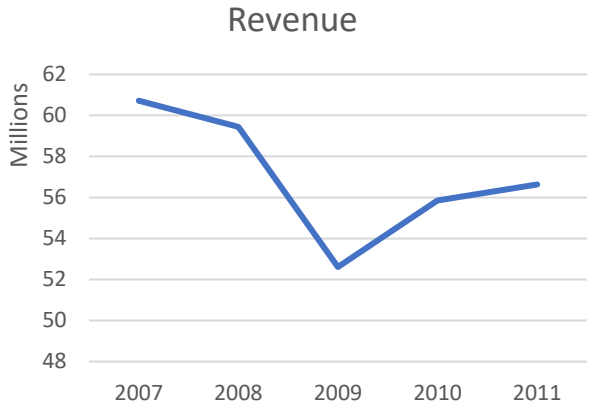


Figure 2. a

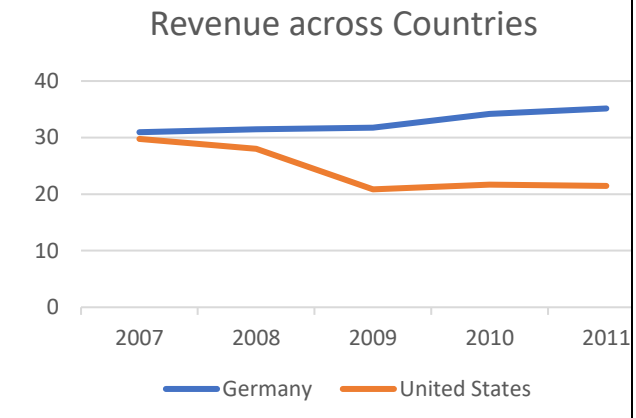


Figure 2.b

The line chart of revenue over the years shows that there has been a sharp decline in revenue from the year 2008 to 2009 and shows recovery afterward. (Figure 2.a) We further diced the data by adding the country desc to visualize the trend over the countries. It is evident that the fall has been majorly due to the decrease in revenue generated by the United States branch, whereas the German branch has been steady and shown growth. (Figure 2.b)

1.3.3 Which sales organization and which customer of that organization has performed poorly over the years?

There are four sales organizations in the company, (2 for each country). To see their performance, the "sales organization" was dragged to the Rows, "Sum of revenue" to values, and "calendar year" into columns, a Pivot chart of clustered column chart was plotted. (Figure 3.a)
For better visualization, the vertical axis, and gridlines were removed, data labels were positioned on top, with no decimals, and represent dollars in millions.

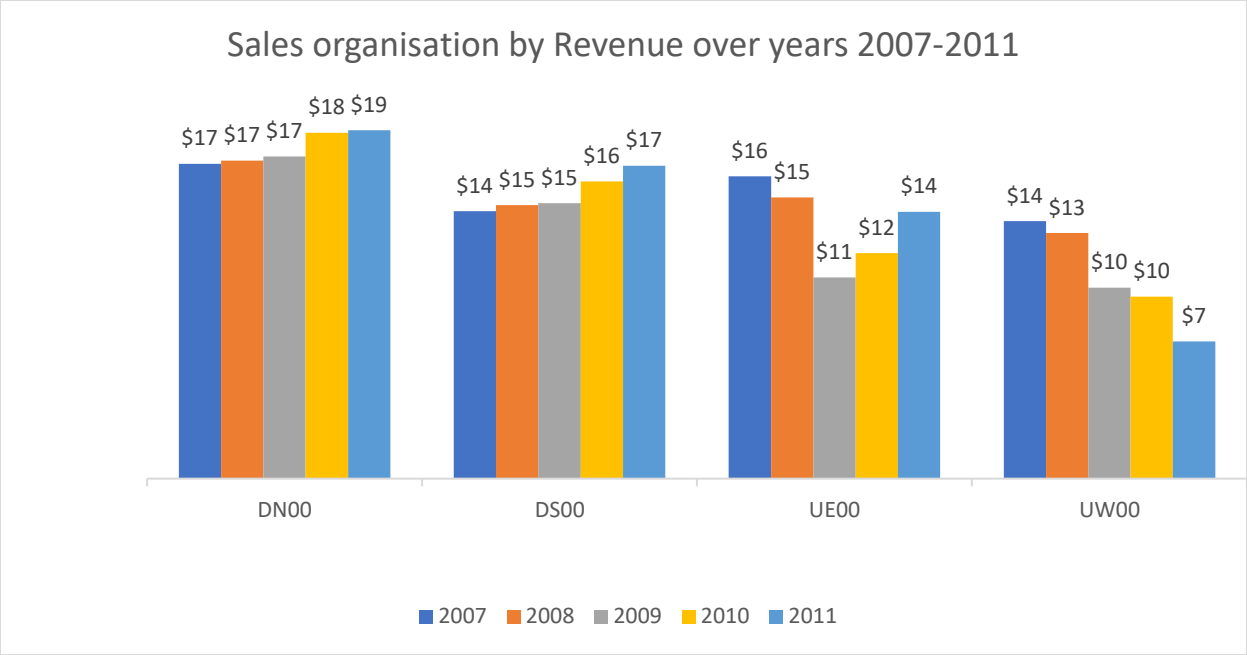
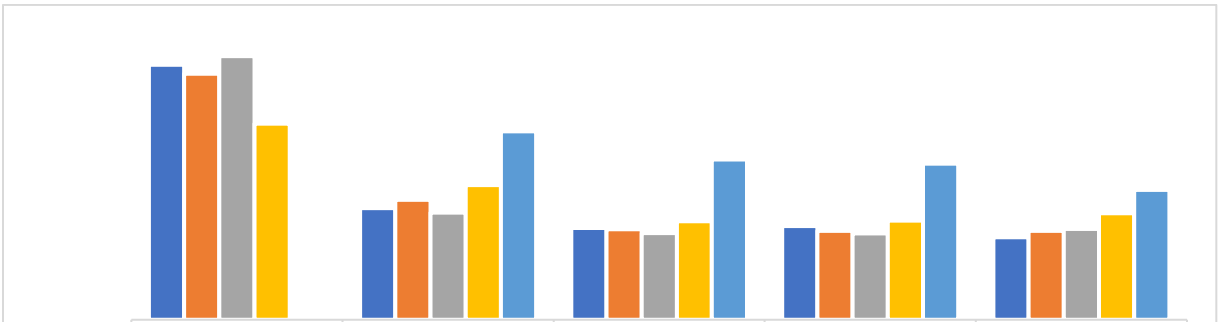


Figure 3.a

The chart (figure 3.a) clearly shows that the UW00 i.e., USA west has been consistently performing poorly, with a decrease in revenue over each year.

To find which customer is responsible for the downtrend, the "sales organization" was moved to filter (to keep only UW00 customers), and "Customer Desc" was added to the rows field. And the customer Silicon Valley Bikes has shown a significant reduction in revenue over the years and almost negligible sales in 2011. (figure 3.b)



	Silicon Valley Bikes	Rocky Mountain Bikes	Northwest Bikes	Windy City Bikes	SoCal Bikes
2007	40.20%	17.59%	14.45%	14.77%	12.99%
2008	38.79%	18.92%	14.28%	13.99%	14.02%
2009	41.53%	16.86%	13.68%	13.60%	14.33%
2010	30.89%	21.24%	15.50%	15.60%	16.77%
2011	0.00%	29.69%	25.27%	24.59%	20.45%

Figure 3.b

1.3.4 What is the significance of losing the business of Silicon Valley?

To find the contribution of Silicon Valley Bikes, in the overall revenue of GBI, "Country", "sales organisation", and "Customer Desc" were added to the Rows, and "sum of revenues" was added in Values. To focus on the contribution of USA west (UW00), Customers of Germany and USA east (UE00) were rolled up as DE, and UE00 respectively, and a pie chart was plotted.

Row Labels	Sum of Revenue
DE	**57.32%**
US	**42.68%**
UE00	**23.80%**
UW00	**18.89%**
Windy City Bikes	2.99%
SoCal Bikes	2.87%
Silicon Valley Bikes	6.23%
Rocky Mountain Bikes	3.79%
Northwest Bikes	3.01%
Grand Total	**100.00%**

Figure 4. a.

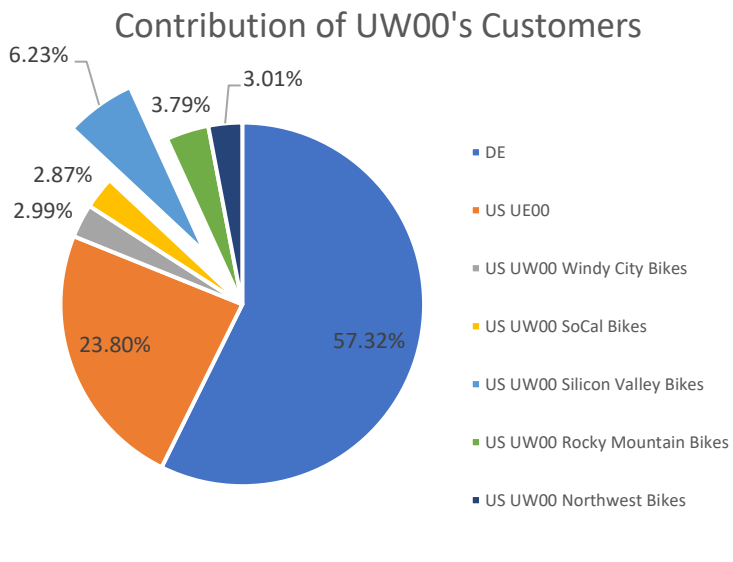


Figure 4. b.

From the crosstab and pie chart, it can be inferred that the customer "Silicon Valley bikes" has contributed over 6% in the total revenue in the last 5 years. Hence their performance has a significant impact on the performance of the organization.

1.3.5 Accessories contribute only 1% of the revenue, then is it worth continuing this product category?

The Revenue although important is not the only measure of performance, a product can be high in revenue but selling in lesser quantity, whereas a product may be of low value but selling in larger volume. To find the sales quantity (volume of sales of accessories) "Product Category" was added to rows, and "sum of sales quantity" was added to the Values field, a Pie chart was plotted, and data values were changed to percentage.

Row Labels	Sum of Sales Quantity
EBI	868
ACC	66941
ORB	38675
ROB	29049
TOU	32887
TRE	434
Grand Total	**168854**

Figure 5. a

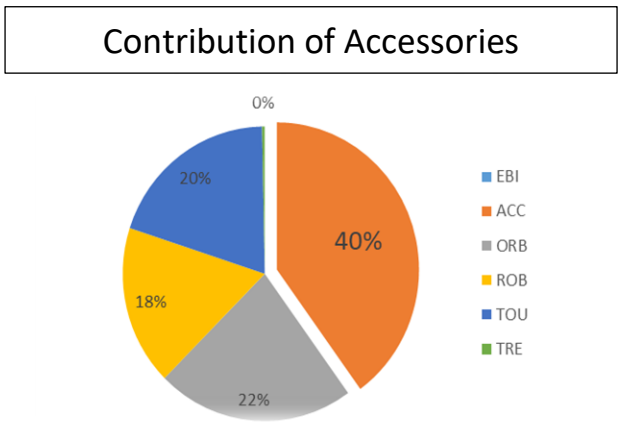


Figure 5. b

Although the Accessories do not have a significant contribution to the revenue, this category has the highest sales in terms of quantity, probably this is getting more footfall to the stores and increase brand awareness hence this product category should be continued.

1.4 Analysis & Self-reflection of the tool

I have been using Excel for many years, for a wide range of purposes from planning a trip to maintaining the budget of my practice. It is an effective tool for small businesses. Although I have used filters, sort, and formulas earlier, Pivot tables were new to me. Pivot tables make it so easy to slice and dice the data, organize in tables of various combinations, and then visualize it.
It allows the user to import data from various sources and even directly from the World wide web. Secondly one can directly copy paste infographics in a word processing tool such as MS word etc. It is a very intuitive tool and allows the user to derive insights by changing various variables, making Excel a swiss knife kind of tool.

Although it is good for some basic visualization, it lacks some features such as, some advanced graph types are not available directly from the pivot table, and one needs to copy the data table in other sheets to do so, this can be time-consuming.
Secondly, since MS Excel is a spreadsheet application it is always possible to inadvertently change the data, which can be catastrophic.

1.5 Conclusion

Overall Pivot table offers the user who is familiar with MS excel, ease of use, intuitive interface, and exceptional ability to work with complex data efficiently.

Chapter 2 SAP Lumira Discovery

2.1 About SAP Lumira

SAP Lumira Discovery is a data visualization tool from SAP, it enables the user to access, transform, and visualize data of large size. It has a simple and user-friendly interface, with drag and drop capabilities. It can be integrated with SAP HANA an in-memory program, making it a very responsive tool.[4] SAP Lumira allows integrating data from various sources such as a Spreadsheet, SQL server, cloud (SAP Lumira Cloud and SAP HANA, etc.) and in various formats with ease. [5] Its strength is the ability to run queries, visualize data in numerous ways with ease, and create a story, dashboard, KPI monitor, and infographics.[4][5] The datasets and reports can also be shared, exported, or printed to collaborate with other team members. Among the tools in the SAP BI portfolio, SAP Lumira Discovery is the only tool that is positioned to handle all 5 aspects of Self-Service BI (Acquisition, Preparation, Visualization, Exploration & Collaboration) making it a useful tool for business intelligence and analysis. [6]

2.2. Description of Data and Research Question.

For this Chapter, two datasets which were provided in the class were used, these data are, a) data of the population of Canada, and data for sales of alcoholic drinks and its provinces, which was originally obtained from the Statistics Canada website. [7]
Both data sets are in the *.csv* format, the population data contains 17 columns and 26585 rows, "Age group", "COORDINATE", "GEO", "REF_DATE", and "VECTOR" appear to be dimensions (values which cannot be aggregated) and "Value" (which is the no of people/ population) appears to be Measures. In the Sales data there are 17 columns and 35281 rows, "Vector", "COORDINATE", "GEO", "Type of beverage", "Origin of product" are the Dimensions and "Value" (which represents the sales and volume) are the measures.

A report published by the CIHI suggests that hospitalization and deaths due to Alcohol abuse are rising. [8]

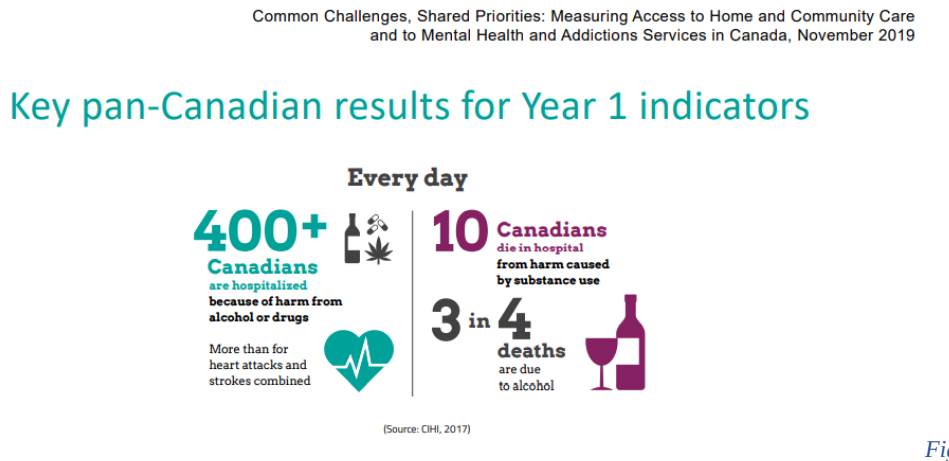


Figure 6

As an official in the health department responsible to run awareness programs on the ill effects of alcohol consumption, this report will try to identify the regions with the highest sales of alcoholic drinks and to identify the type and origin of these drinks to be able to create appropriate public health messages and to priorities the regions to launch appropriate health awareness programs.

2.3 Applying Analytical tools and Results.

2.3.1 What is the trend of population growth in Canada from the years 2005- 2019?

The Population for Alcoholic drinks would be the adults who can legally purchase and consume Alcohol. Hence a calculated measure "Adults" was obtained (the legal age limit of alcohol purchase and consumptions is 19 years and 18 years in Alberta, Manitoba, and Quebec).
To create a chart, "Population of adults" was added on Y-axis, "Date" as Dimension on the x-axis, and a Geo-Filter of Canada. A line chart was plotted.
Figure. 6

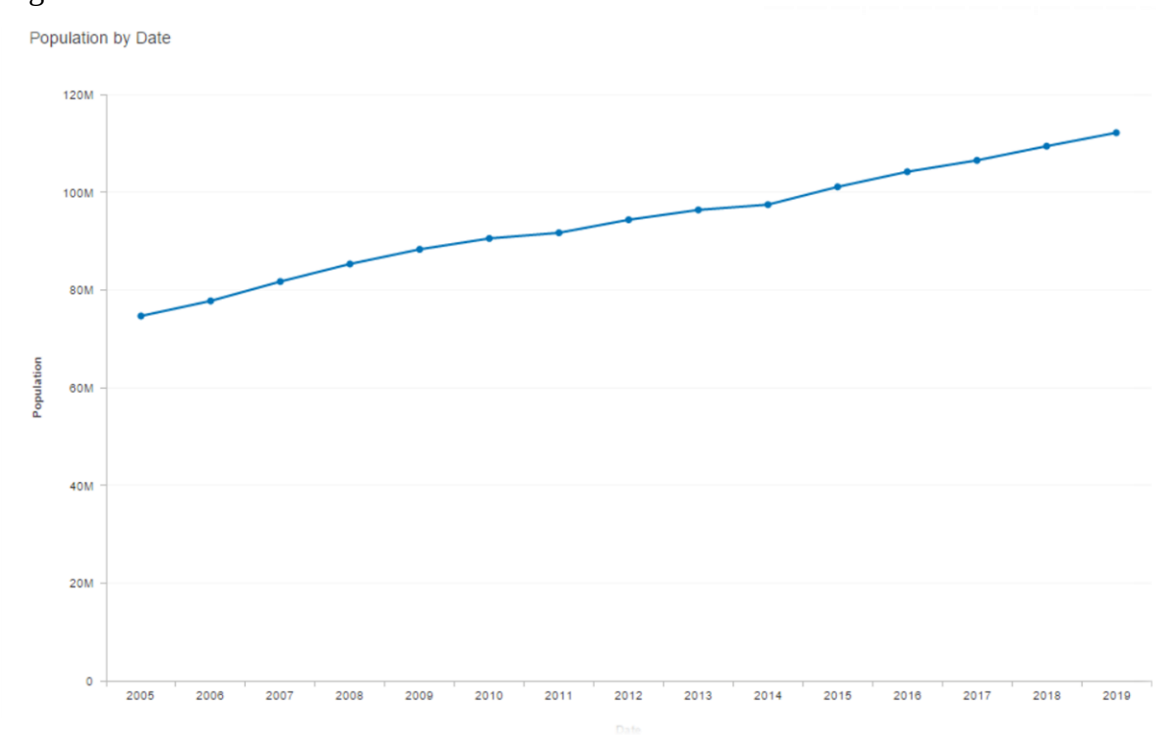


Figure 7

The chart shows that there is a gradual rise in the number of adults throughout 2005 -2019, with a small dip in the years 2011 and 2014.

2.3.2 Is the per capita sales proportional to the growth of the number of adults over years?

To understand how much the sales are per individual of the population, a calculated measure of "Per Capita sales" is obtained by using the formula: Per Capita Sales = {VALUE}/{ADULTS}*1000.

On the y- axis "Adults" was added, "Per Capita Sales" was added as a secondary y-axis. "Date" was added as an x-axis, Geo Filter was used to seeing the population of Canada, and all provinces were filtered out. A column Chart was plotted. *Figure 8*

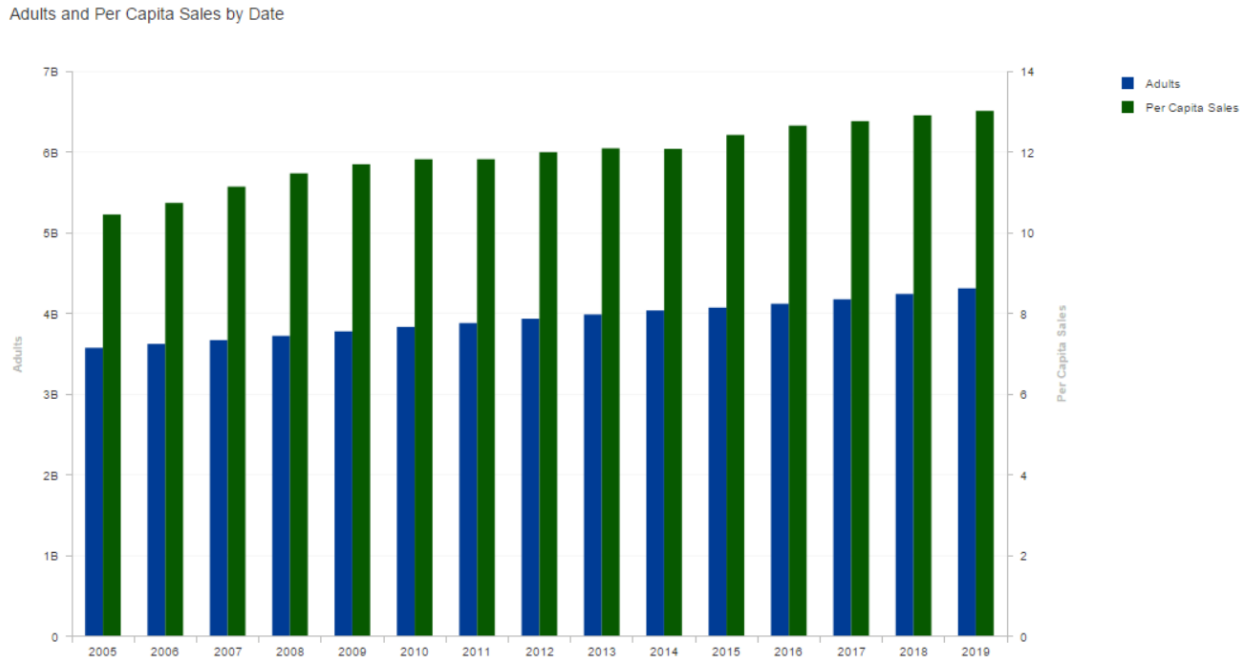


Figure 8

The column chart shows that there seems to be a steady increase in the sales of alcoholic drinks represented here by "per capita sales" and "population" of adults over the years indicating that the consumption amount per individual may have remained more or less the same throughout the years.

2.3.3 How do alcohol sales vary across Canada? Has any province shown a rise in sales of Alcoholics drinks?

"Date" was dragged to the x-axis, "GEO" on the y-axis, and "per capita sales" (pcs) was placed in color, from the Filter, Canada was filtered out and all provinces were selected, a Heat map was plotted (*Figure 9*)

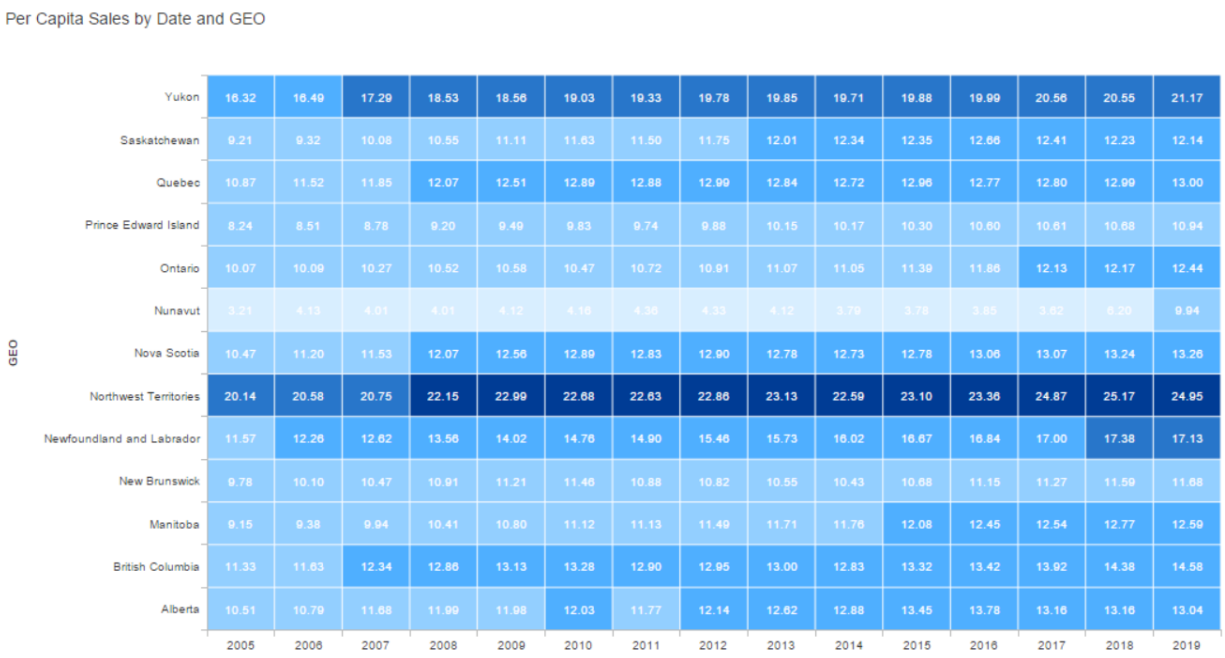


Figure 9

From the above chart, it can be inferred that the highest consumption of alcoholic beverages is in Northwest territories followed by Yukon, chart shows that there was a slight increase in the sales of Alcoholic beverages in the province of Nunavut in the year 2005-2006, and from 2006 it was steady but then it spiked up from the year 2017 (3.6 pcs) to 2019 (9.94 pcs) this spike is due to lifting of alcohol restrictions in certain regions Dec 2018.[28]

2.3.4 Which type of beverage has contributed the most to this uptrend in Nunavut?

To visualize which products are contributing to this sharp rise, a stacked bar chart was plotted with the "Date" on the y-axis, "Per capita sales" on the x-axis, and "Type of beverage" was added to the color. To be able to visualize the changes more clearly, we reduced the date set by adding a Date Filter by selecting years 2015-2019 and filtering out the rest. To focus on the Nunavut province, a filter of "GEO" was used and only Nunavut was selected, and the rest provinces were filtered out. (*Figure 10*)

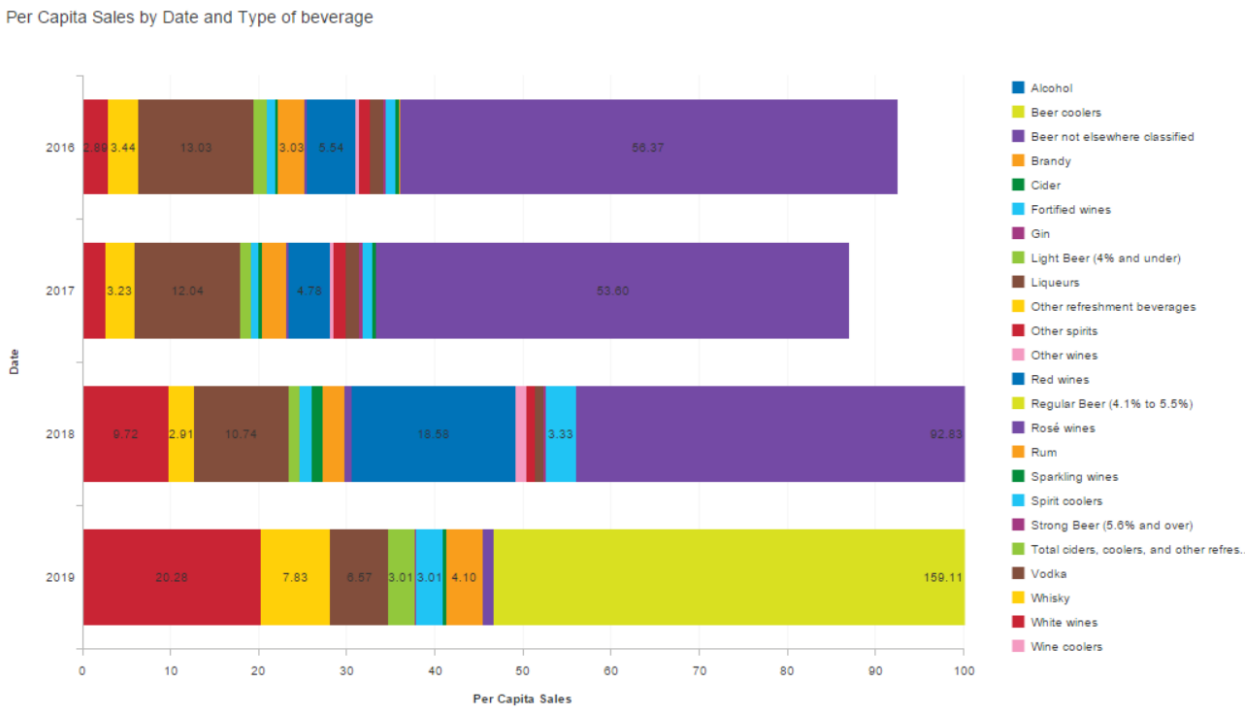


Figure 10

The bar chart illustrates that Beer, Red wine, White wine, and whiskey are contributing to this increase in "Per Capita Sales" from the year 2017 onwards.

2.3.5 Among these contributing beverages, except for beer is there a preference for imported or domestic products?

To the above data set, two more filters were applied "Type of Beverage" to select only Red wine, White wine, and Whiskey and "Origin of product" where "total products" were filtered out. "Type of Product" was also added to the colour. In Trellis, "Origin of product" was added in columns. (Figure11)

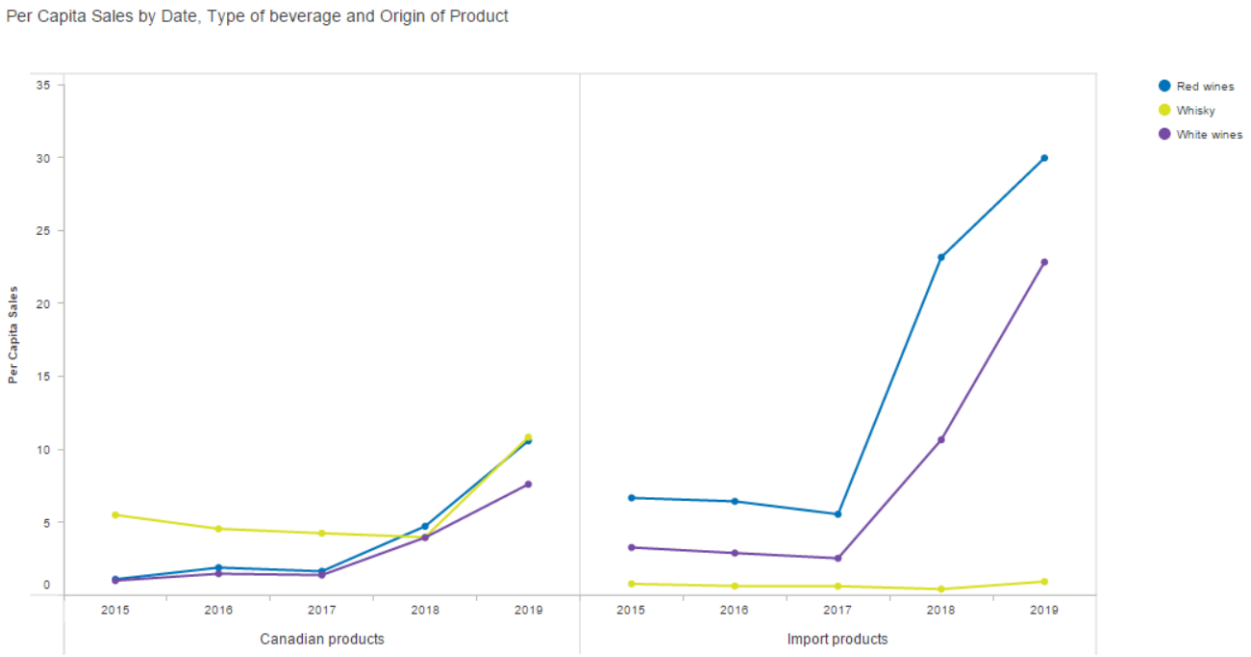


Figure 11

The trellis line chart illustrates that the imported wines (Red and white) are preferred over domestic wines while domestic whiskey is preferred over imported ones.

From the analysis it can be deducted that Northwest territories, Yukon and Newfoundland, and Labrador have the highest per capita sales of alcohol and that there has been a sharp increase in the sales of alcoholic beverages in the province of Nunavut, this probably is due to the lifting of restrictions on the sale of alcoholic drinks and there is an increase in demand for imported wines. Hence the awareness campaigns with a focus on the ill effects of consumption of wines as they contain about 12% alcohol should be run first in these locations.

2.4 Analysis & Self-reflection of the tool

SAP Lumira was a bit different from MS Excel. The installation was a little hassle, but once it was properly installed, the tool was easy to navigate. The data can be imported from various sources. At the same time, one can choose which columns to include, thereby reducing the size of the working file, memory usage and increase the focus on the variables needed for analysis.
The tool offers exceptional control while merging the data from various sources, one can choose how to join the different data without writing a single line of SQL code. Equally modifying the data such as introducing a calculated field is a breeze, helping one to extract, transform and load huge data. Secondly, The Facet view allows one to look at the whole data set in a neat and summarized form. The design view offered numerous possibilities for creating a visualization. I particularly like the ability to create secondary axis and charts which are based on geographic location without entering the latitude and longitudes.

The only drawback, I encountered was that if I used the program for a longer time, it sometimes crashed or hanged up, requiring me to restart the application; but this may be happening because of my inexperience with this application.

2.5 Conclusion

I found SAP Lumira discovery a very powerful tool for extracting, transforming, and loading data, One can slice, dice, and create hierarchy conveniently. It is a user-friendly tool and offers numerous options to visualize the data and create reports.

Chapter 3 SAP Predictive Analysis

3.1 About SAP Predictive Analytics

SAP Predictive analytics is a data exploration and manipulation tool, to slice and dice data to discover the hidden insights and relationships in the data and build prediction models. Such models aid in making an accurate prediction of future events. The data acquisition and manipulation features are inherited from SAP Lumira. The data used on SAP PA can be current or historical, in various forms such as a database or flat files. It can process large datasets such as big data or the Internet of Things (IOT) [10][12]. The integration with R enables to customize the functionality whereby users can add their own R scripts (Expert Analytics). Some of the key features are automated data preparation, modeling, discovering, and visualizing the results of hidden insights. [9] [11] It is an advanced analytical BI tool used by Data scientists for businesses.[10]

3.2 Dataset & Research Questions

For this chapter, "Global Bike Inc. (GBI)" data is used. The data is in *xlsx* format and was provided in the class is the transactional data of the GBI company which deals in Bikes and accessories, with customers (B2B) in two countries viz Germany and United States. The data is of 13 years period, i.e., from the year 2007 to 2019. The company has 2 divisions namely the Accessories division and the Bike division, and 5 product categories. The Accessories division has 10 products, and the bike division has 18 products.

The 171,010 lines of data include 23 columns of variables such as City, Country, Division, Sales Organisation, Product category, Customer details, Year, Month, Day, etc. which are the Dimensions and Cost of goods, Discounts, Revenue, Sales quantity, Gross margin are the Measures. No data cleaning was required. The revenue was converted to USD to be used for analysis.

Applying the Pareto principle to business means that 80 percent of a company's revenue is derived from 20 percent of its products or services or dealers or employees. A company can emphasize the value of these core products/ dealers/sales team to achieve better results the lessons learned from these achievers can be shared with other employees/dealers to improve their performance. [13] For this purpose, the company wants to explore its wholesale division performance for the last 13 years with a special focus on Germany, on extracting and illustrating the best performing product and customer of the company for the last year i.e., 2019, and to build a strategy for the future.

3.3 Applying Analytical tools and Results.

3.3.1 What is the total revenue across both countries and Sales organization for 2007 - 2019? Provide a detailed table for the last 3 years.

To get the revenue trend for both the countries, a line chart was plotted with "Revenue USD" in the Y-axis, "Year" in the X-axis, and "Country" in Legend color.

On another sheet, the Donut chart was selected as a choice of chart type, "Revenue USD" was added in Pie Sectors, and "SalesOrg" in Legend color.

On the third sheet, a Crosstab was created by adding “RevenueUSD” in measures, “Year” in Columns, and “Country”, “Salesorg” in Rows. Filter on the “Year” 2017, 2018, and 2019.
On SAP PA, I then clicked on compose tab and selected a blank sheet to create a single sheet representing an overview of the company’s overall revenue and last 3 years' performance. I moved all 3 sheets on the blank sheet. Compose Tab gives the freedom to format the images (color, font, text, etc.) to our personal requirement. The year “2019” is demarcated with a red box in the crosstab to highlight.

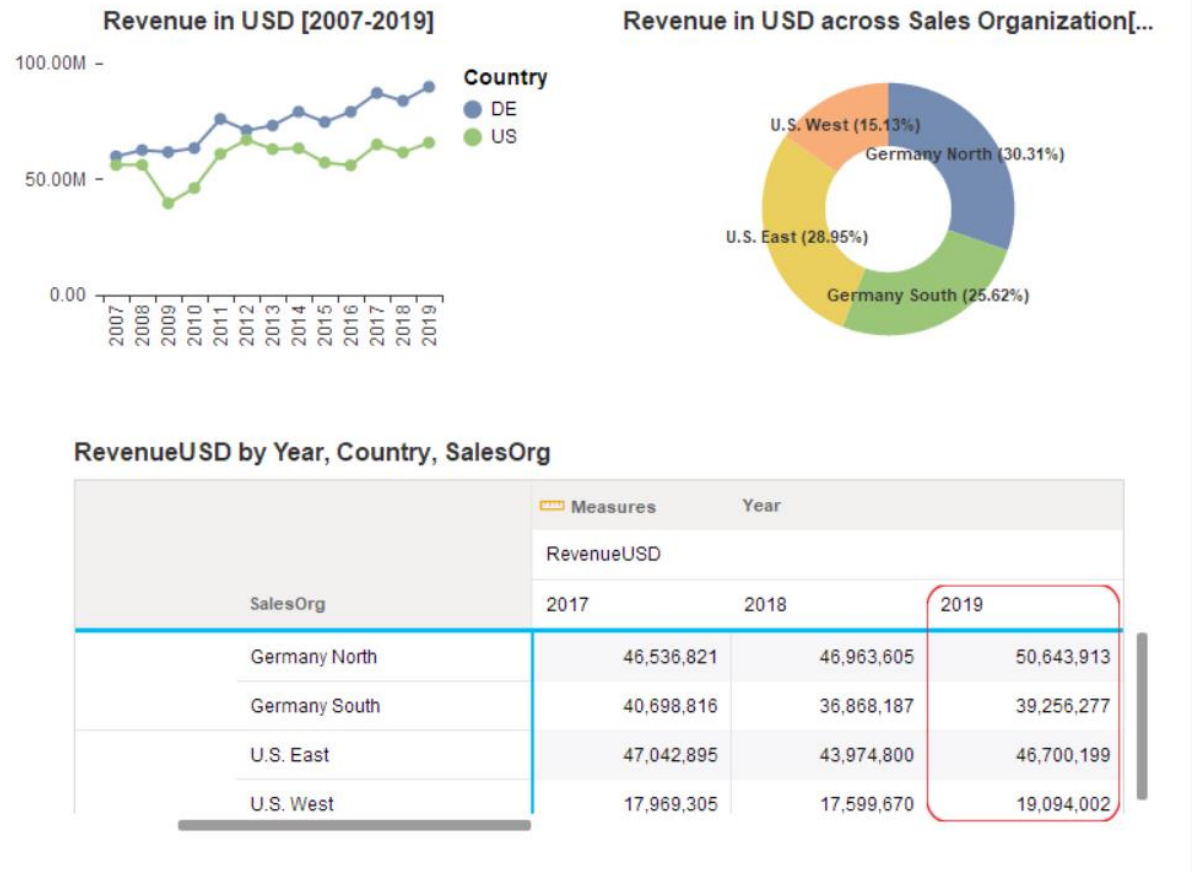
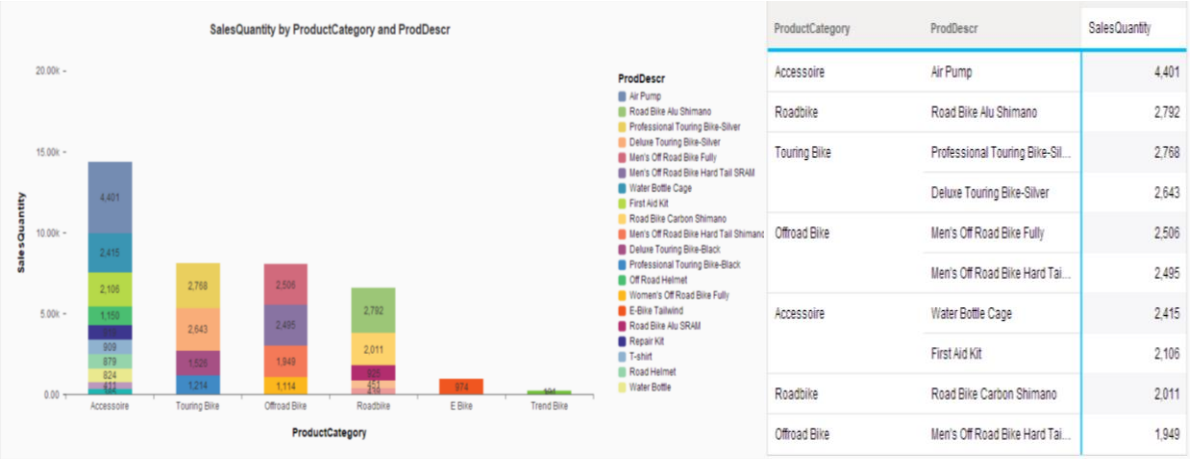


Figure 12

From the above visualization, the Line chart shows that Germany has shown a steady increase in revenue generation from 2007 -2019. While the U.S shows a slight increase in revenue when compared to revenue in 2019 to 2007. It also shows how the U.S experienced lots of downfall and uphill events in revenue through the years. The pie chart shows that out of the 55 % revenue contribution by Germany, Germany North has a maximum contribution of 30% and the remaining 25% from Germany South. The U.S has a lesser revenue generation of only 44% with 29% from East Sales organization AND 16 % from West Sales organization. The highlighted area on the crosstab gives in-detail revenue for all sales organizations.

3.3.2 *Which product category and which product from that category has maximum quantity sales in 2019 in Germany?*

To see the maximum quantity of product sold, "Sales Quantity" was added in the Y-axis, "Product category" was added in X-axis, "ProdDescr" in Legend color. Two filters were added, for "Year" – 2019 and "Country" – DE(Germany). The Chart type chosen is a Stacked column chart. To see the quantity of each product, data labels were selected by right-clicking on the chart.



ProductCategory	ProdDescr	SalesQuantity
Accessoire	Air Pump	4,401
Roadbike	Road Bike Alu Shimano	2,792
Touring Bike	Professional Touring Bike-Sil...	2,768
	Deluxe Touring Bike-Silver	2,643
Offroad Bike	Men's Off Road Bike Fully	2,506
	Men's Off Road Bike Hard Tai...	2,495
Accessoire	Water Bottle Cage	2,415
	First Aid Kit	2,106
Roadbike	Road Bike Carbon Shimano	2,011
Offroad Bike	Men's Off Road Bike Hard Tai...	1,949

Figure 13

The stacked column chart represents "Accessories" as the category with maximum sales and in Accessories, 4401 "Air Pump" were sold.

3.3.3 *In the bikes (BI) division, which bike has the highest sales Quantity and Gross margin and revenue in 2019 in Germany?*

Products with higher margin and high sales help bring in more profit to the company, primary exploration of the data suggested that the margins of the Bike Division were significantly more than the Accessories division, I used a Bubble chart to visualize the maximum profit-generating products, both quantity, and gross margin by adding "Revenue USD" in the Y-axis, "Sales Quantity" in X-axis, "Gross Margin" in Bubble Width, and "ProdDescr" in Legend color. By selecting 2019 in "Year", DE in "Country", and BI in "Division", filters were set.

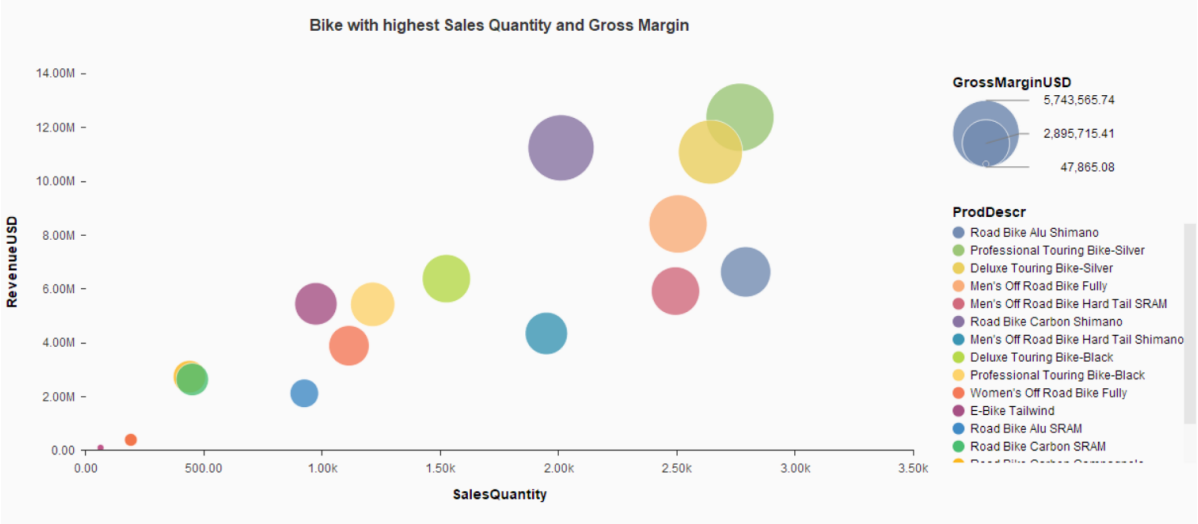


Figure 14

Road bike Alu Shimano seems to have the highest sales quantity, but Professional Touring Bike - Silver has almost the same sales quantity and offers higher revenue and gross margin.

3.3.4 Is there a correlation b/w Sales Quantity and Discount?

There is usually common thinking, the reason for increased sales of a product is due to the higher discount applied on it. To be able to see if there is any correlation between them, I used a Scatter plot and added "Discount USD" in X-axis, "SalesQuantity" in Y-axis, and "ProdDescr" in the legend color, and products were filtered to see only bike division.

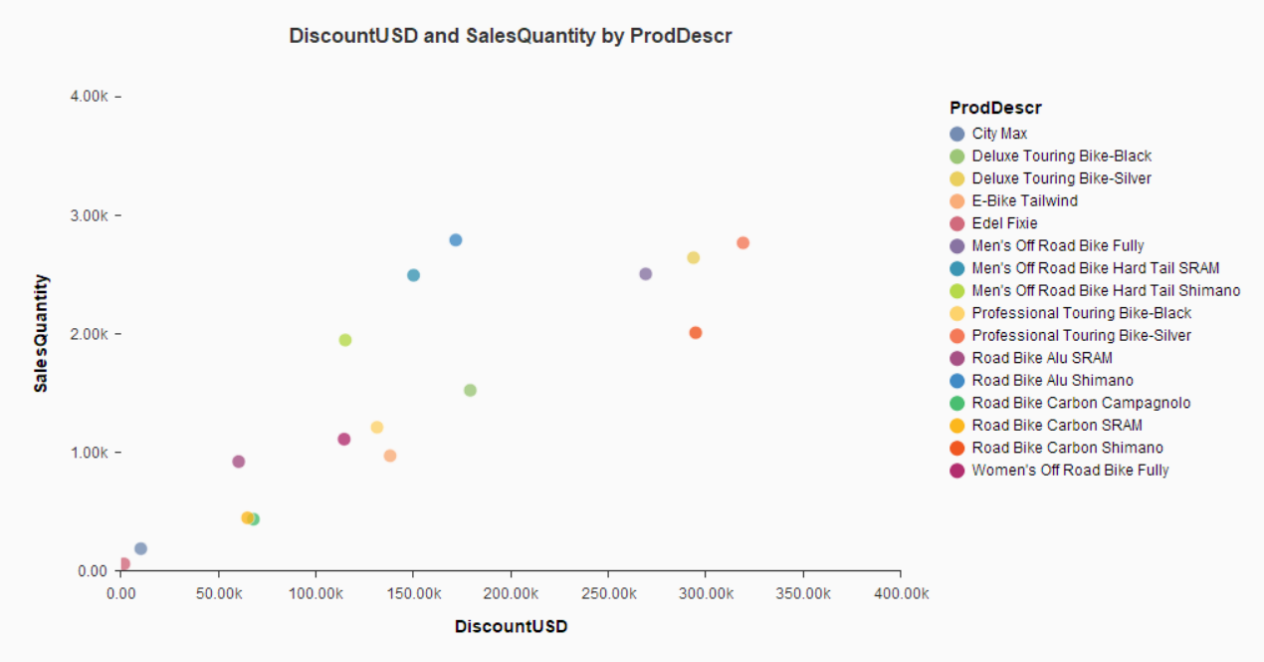


Figure 15

Figure 16a

Figure 16b

ProdDescr	DiscountUSD	SalesQuantity
Professional Touring Bike-Sil...	319,179.73	2,768
Road Bike Carbon Shimano	294,828.11	2,011
Deluxe Touring Bike-Silver	293,717.39	2,643
Men's Off Road Bike Fully	269,207.10	2,506
Deluxe Touring Bike-Black	179,203.20	1,526

Figure 17

From the scatter plot it is evident that Road Bike Alu Shimano has higher sales than the top 5 discounted bikes viz the Professional Touring bike, Road Bike Carbon Shimano, Deluxe Touring Bike Silver, Men's Off-road bike, and Deluxe Touring Bike, Thus, there seems to be more perceived value for money or demand for Road bike Alu Shimano, this may be not linked to preference to Road bike, as one sees that the sales of Road Bike Carbon Shimano, has lower sales even though it is discounted more than the Road Bike Alu Shimano, maybe the Carbon version is considered or promoted as a premium version and hence has fewer buyers.

Secondly by comparing figure 16a and figure 16b, one can see that the Professional Touring bike, Road Bike Alu Shimano gets more revenue and is also promoted the most (by giving heavy discounts), which may drop if the discount on them is discontinued.

3.3.5 Provide an overview of Germany's revenue contribution for the company and their best performers for the year 2019.

For the 2019 summary of Germany, six different visualizations were created. Two Numeric Point charts were created, one was the total revenue of GBI in the year 2019, by dragging "RevenueUSD" in value and filter for "Year" – 2019. Another Numeric Point was the same but with an additional filter of Germany for "Country". Two-column charts were prepared to illustrate the best-selling product and best revenue-generating product. For the 'best-selling product', "RevenueUSD" was moved to Y-axis, "ProdDescr" was moved to X-axis, and two filters "Country" – Germany and "Year" – 2019 was set. For the 'best revenue-generating product', "SalesQuantity" was moved to Y-axis, "ProdDescr" was moved to X-axis, and two filters

"Country" – Germany and "Year" – 2019 was set. Both the best performer column charts were sorted from largest to smallest by the X-axis value and ranked to visualize the Top 5 performer. The next two charts were to represent the 'Best Customer' for Germany in 2019. The first was crosstab to illustrate the top 5 revenue-generating customers, for which "CustDescr" was added in rows and "RevenueUSD" in Measures and then were sorted from largest to smallest and ranked to visualize the Top 5 performer. The next visualization was 'Tag cloud' where "RevenueUSD" was added in word weight and word color and "CustDescr" was added in word. For both the best customer visualization, the filters set for "Country" and "Year" remained the same.

All the six visualizations were then added on the blank sheet in Compose tab and an annual summary report of Germany for 2019 was created. The title and GBI company logo were added.

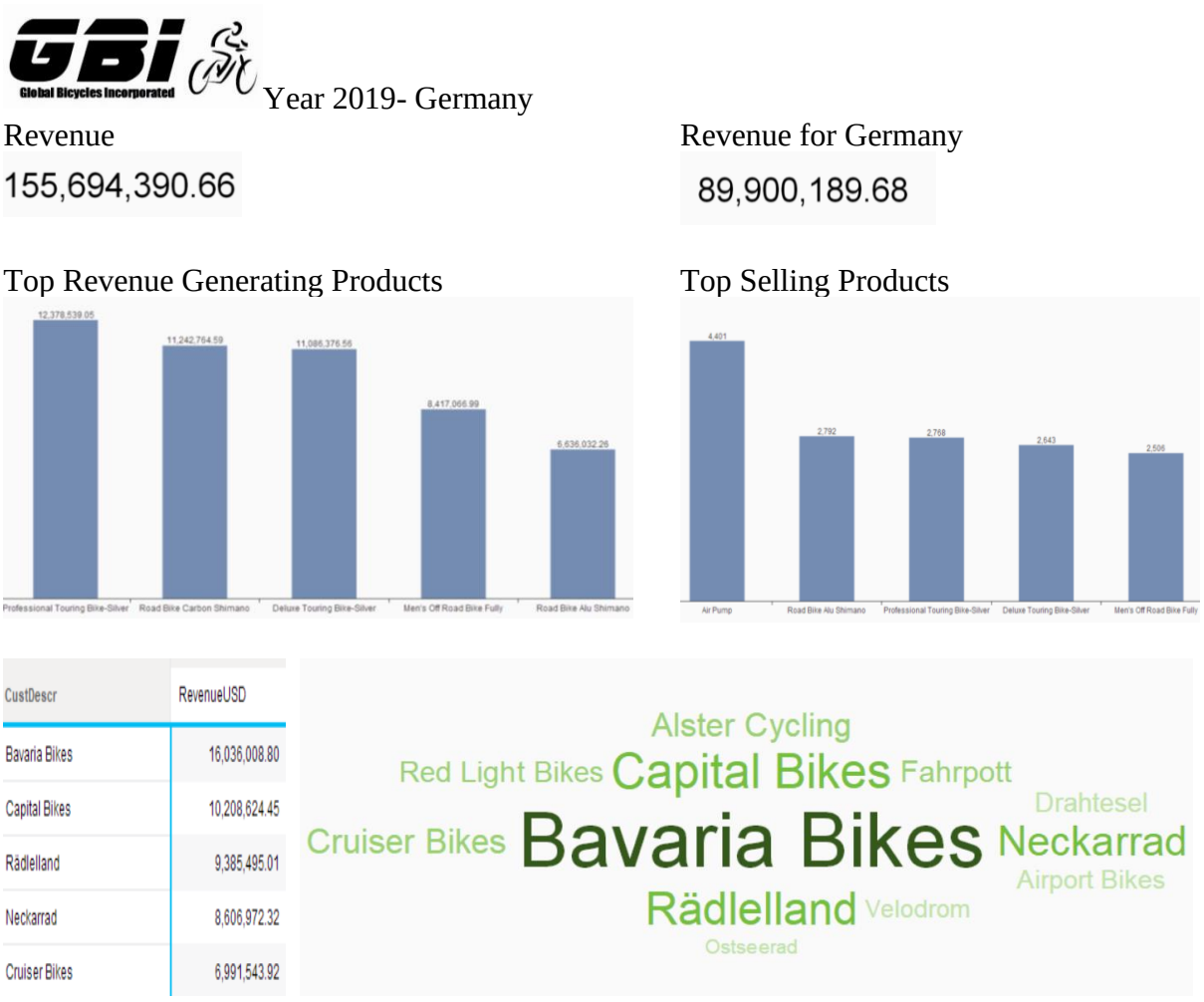


Figure 18

The overview gives a summary for 2019, the total revenue value for the company, and Germany. It represents that for Germany in 2019, the Professional Touring bike – Silver is the top revenue-generating product and Air Pump is the maximum selling product and Bavaria bikes are the highest

revenue-generating customer, it would help the company incentivize them to encourage them to continue their success and understand how Bavaria Bikes are conducting their business and try to replicate their methods to a wider customer base.

3.4 Analysis & Self-reflection of the tool

Working on SAP PA was a wonderful experience, it is possible to work with different types and sizes for files from a flat file to a complex database, it is easy to merge data from various sources and create, discover, and visualize results. Although the visuals /charts created by this tool were beautiful, I found it very difficult to format the visualization, this may be due to the limited experience I have with this tool. The formatting of visualization can only be done when the images are moved to create a story (on Compose Tab), It takes a little longer to load data as the software tries to analyze the data by itself, the other issue I experienced was that this tool seems to crash a lot of time, then compared to previously used BI tools such as Microsoft excel and Sap Lumira, requiring reloading the data and redoing all the unsaved analysis and visualization.

3.5 Conclusion

My analysis for this chapter was majorly around slicing and dicing the data and analyzing to discover the insights to visualize the results. SAP PA tool was a convenient tool to answer any queries and create a report for the organization. My future interest will be to explore the possibility of connecting to an operational database, use of automated analysis, and using the prediction functionality of SAP PA which seems to be its most important and interesting feature, helping to make an accurate prediction of future events for an organization.

Chapter 4 IBM Cognos

4.1 About IBM Cognos

IBM Cognos is a personal analytical tool empowering data analysts and business managers to conveniently load, explore, analyze, visualize, and share results without dependency on IT managers.[14]. It helps build custom dashboards and applications and provides good support to team collaboration to solve individual and enterprise-level challenges. [15]

IBM Cognos allows users to seamlessly connect, import and analyze data from a variety of sources such as CSV files, Microsoft Excel spreadsheets, ODBC data sources, IBM Cognos BI Reports. [16]

The user can connect to pre-existing data assets such as IBM Cognos Statistics, IBM SPSS Modeler, and IBM CognosTM1, get real-time information, develop meaningful and factual context, additionally it allows the user to be able to collaborate with co-workers and create discussion threads add comments, and questions. [16]

4.2 Dataset & Research Questions

I have used two datasets for this chapter, the Sales- Region & WW Monthly sales. Both the datasets were provided for the class lab.

The Sales – Region data is in *xlsx* format having 979 rows and 8 columns namely Product type, Sales Channel, Territory, Quarters, YM, Customer, Revenue, and Cost. The products are Computers & Tablets, TV & Home theatre, Camera, Mobile, Office supplies, Repairs, and Portable Electronics. These products are sold through 3 channels Direct, Internet, and Retail by Customers of types such as Entertainment Venues, Branded Stores, Retail, and Internet Direct. The transactions are of the year 2012 of North American territory.

The WW Monthly sales data is of the same *xlsx* format having 1952 rows and 8 columns as same as Sales – Region and an extra column of 'Week', which are transactions details for each week of the 2012 year. Other than this difference, the data also has records of territories such as Europe, Asia Pacific, and South America other than North America.

As a business analyst, the company wants me to explore and visualize the overall performance of the company and find the causes of the downtrend in sales of any products while comparing across the territories.

4.3 Applying Analytical tools and Results.

4.3.1 How the North American territory of the company performed in the Year 2012? What is the trend of revenue and cost in 2012 in North America? Which month shows the highest profitability?

Primarily, I used 'Sales-Region' data, to evaluate the overall performance of the company. With the first graph, I looked at total sales made by the company in the year 2012 for NA. For this Bar chart was selected as a graph for visualization, and "Sales-Region Measure" was dragged into rows and "YM" into the columns. On the cross tab, "YM" was drilled up to have a view of total revenue and cost.
To visualize the trend of sales over the months of the year 2012, changed the chart type to Line chart and drilled down the "YM".

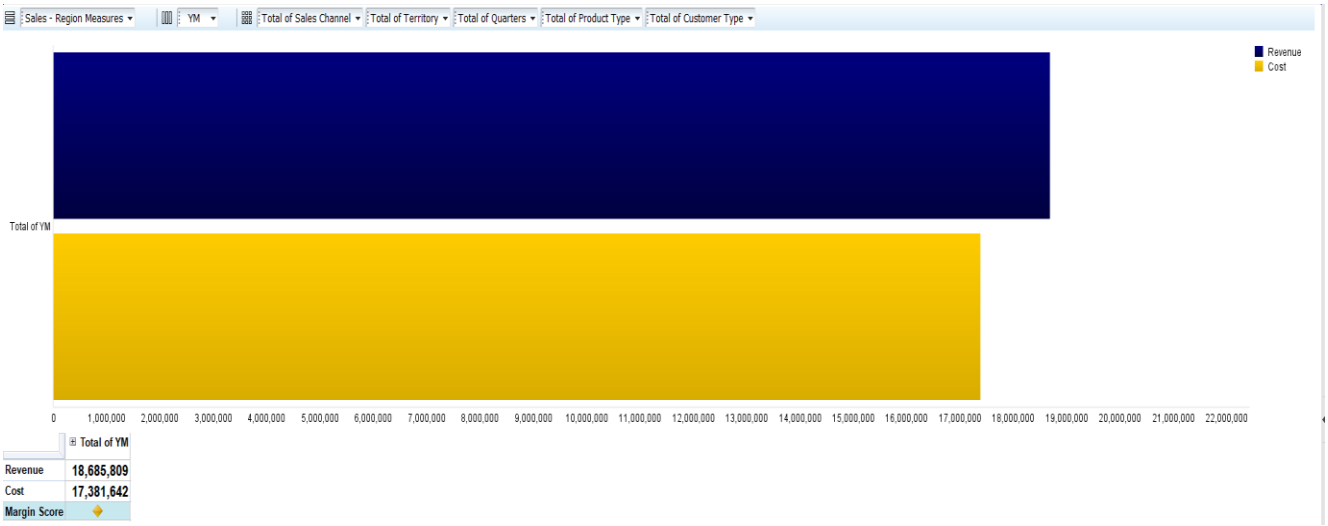


Figure 19

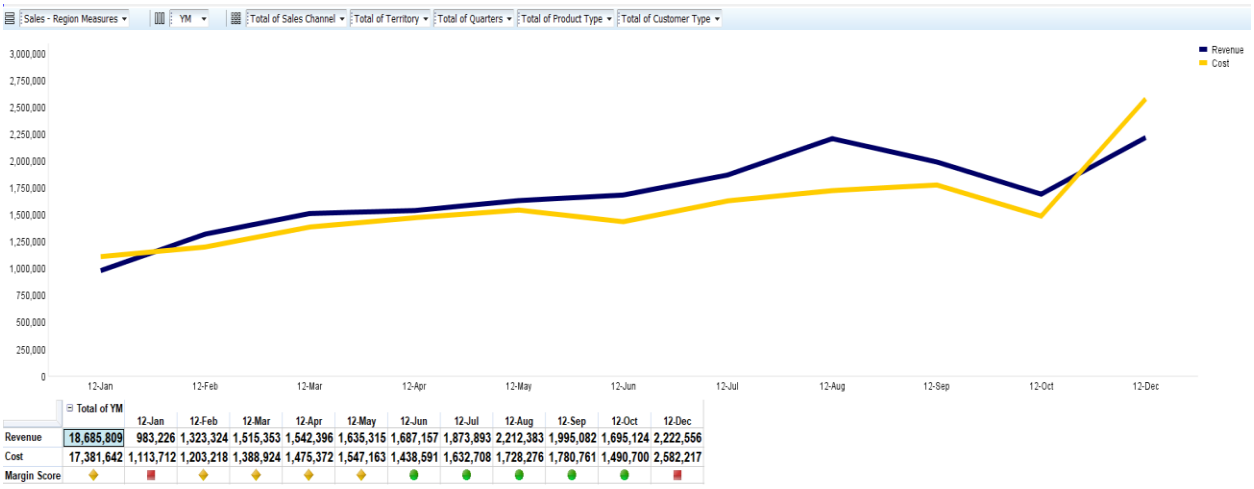
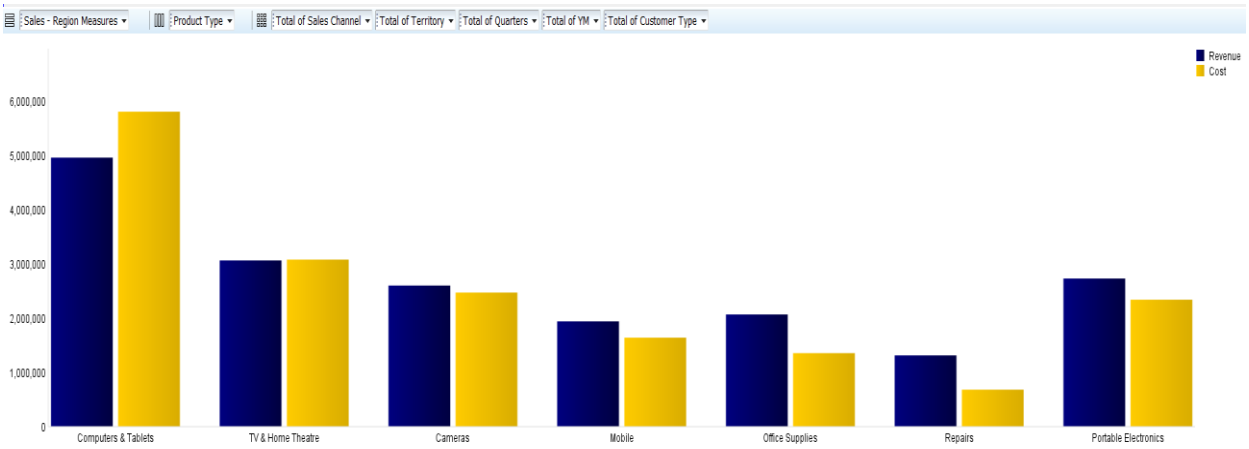


Figure 20

From the first visualization, the company's total revenue generation for 2012 in NA is more than the Cost i.e., the company was in profit.
To identify months with profitability and to visualize this a calculated field the "margin score" was created, Margin Score acts as a scorecard, comparing the revenue to cost. Were the months ranked green, yellow, and Red, to depict the relation of revenue and cost.
The line graph shows that the year begins with a little loss in January (indicated by Red square), then a marginal increase in revenue from February till May (Months with Yellow Diamonds), suggesting that the business was at a steady pace, beyond which there seems to evident increase in revenue (represented by Green circles), with August month shows the highest profitability. And ends with losses in the month of December.
An interesting observation is that the month of November was not seen in the Chart.

4.3.2 Is there a product category that has generated loss to the company?

The column chart was selected as a choice of visualization. The "Sales- Region Measures" was dragged into Rows and "Product Type" into Columns.



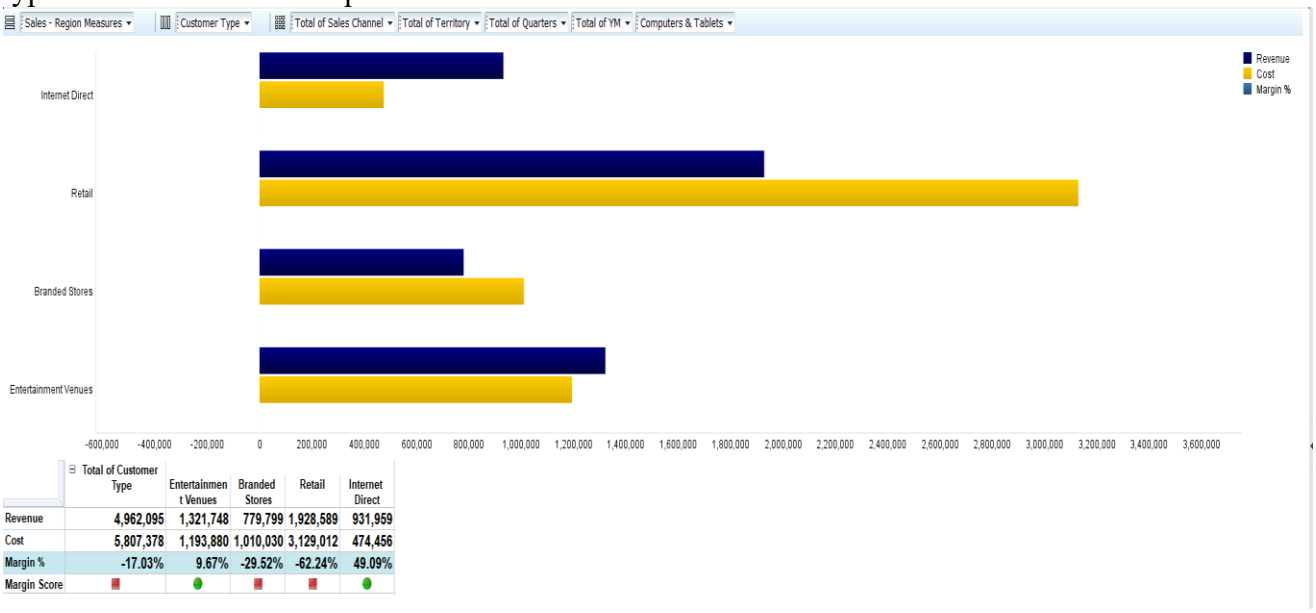
Total of Product Type	Computers & Tablets	TV & Home Theatre	Cameras	Mobile	Office Supplies	Repairs	Portable Electronics
Revenue	18,685,809	4,962,095	3,066,104	2,601,324	1,941,968	2,068,123	1,313,678	2,732,517
Cost	17,381,642	5,807,378	3,080,650	2,472,818	1,641,575	1,355,575	682,187	2,341,459
Margin Score	♦	■	♦	♦	●	●	●	●

Figure 21

From the crosstab within the image, the red square in the Margin Score for the product 'Computer & Tablets' represents the negative margin score. The same is visible in the column chart where this product's revenue is far less than the cost of the product. Thus, for North America, out of all the products such as 'TV & Home Theatres', 'Cameras', 'Mobile', 'Other supplies', 'Repairs' and 'Portable Electronics', product 'Computer & Tablets' has generated loss to the company in the year 2012.

4.3.3 How are the business channels contributing to this loss concerning the above product (Computer & Tablet)?

To illustrate the performance of various business channels in the product category 'Computer & Tablets', and to be able to find the impact of these business channels, I created a new calculated field "Margin %" which is a ratio of cost to revenue / **Cost/Revenue.**
Then "Sales- Region Measures" was added in rows, "Customer type" in columns, and "Product type" was filtered to 'Computer & Tablets' on Bar chart.



Total of Customer Type	Entertainment Venues	Branded Stores	Retail	Internet Direct
Revenue	4,962,095	1,321,748	779,799	1,928,589	931,959
Cost	5,807,378	1,193,880	1,010,030	3,129,012	474,456
Margin %	-17.03%	9.67%	-29.52%	-62.24%	49.09%
Margin Score					

Figure 22

From the crosstab and the bar chart, it is clearly visible that the loss was due to the loss generated in the 'Retail' and 'Branded Store' channels. Whereas the Internet-Based business generated good profit followed by Entertainment Venues.

4.3.4 How are the business channels performing in other territories?

To compare with other regions, additional data was added where there are records of other regions. The 'WW Monthly sales data' was merged into the original Sales – Region data. The "Sales-Region Measures" were added to rows, "Total of territory" and "Customer type" into columns, and the filter was added by selecting "Computer & Tablets". I used a combination chart of columns

and lines. To represent the information more clearly, charts and cross tab were separately created.

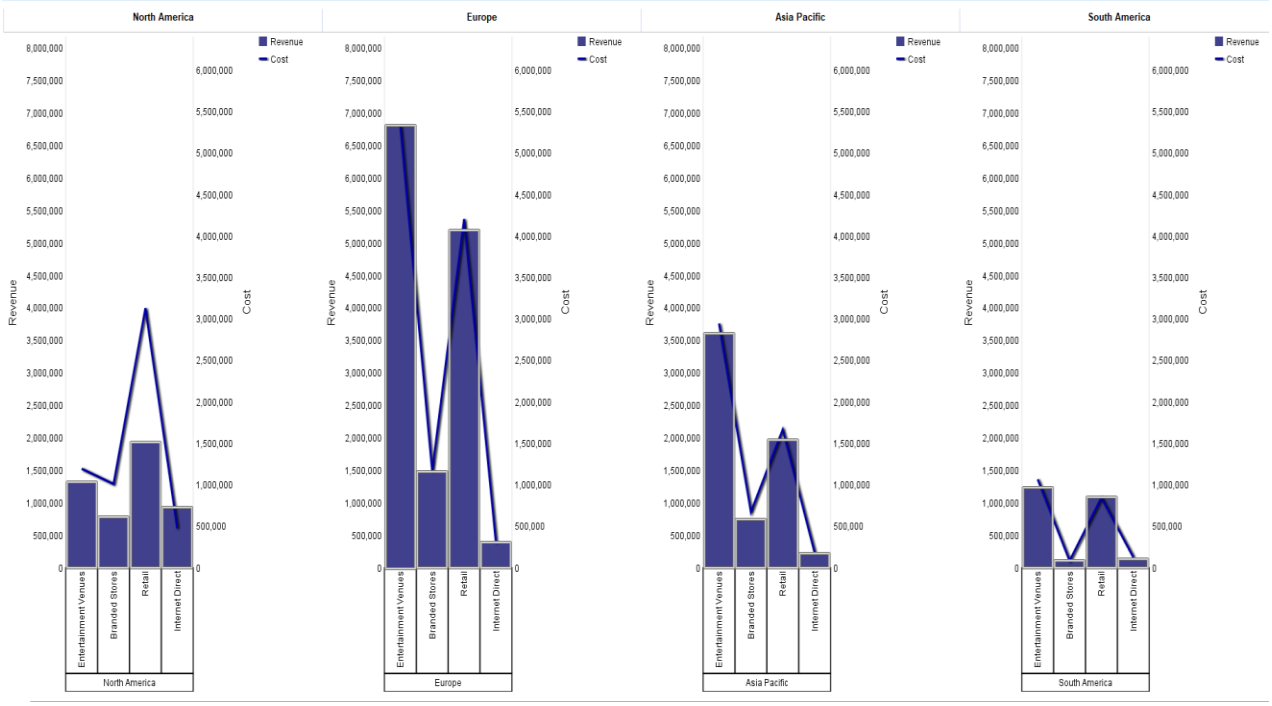


Figure23

	Total of Territory	North America					Europe				
	Total of Customer Type	Total of Customer Type	Entertainment Venues	Branded Stores	Retail	Internet Direct	Total of Customer Type	Entertainment Venues	Branded Stores	Retail	Internet Direct
Revenue	27,957,213.27	4,962,095	1,321,748	779,799	1,928,589	931,959	13,879,487.44	6,817,811.36	1,480,090.52	5,193,260.97	388,324.59
Cost	24,409,511.69	5,807,378	1,193,880	1,010,030	3,129,012	474,456	11,007,616.9	5,337,467.63	1,165,418.81	4,203,186.46	301,544
Margin %	12.69%	-17.03%	9.67%	-29.52%	-62.24%	49.09%	20.69%	21.71%	21.26%	19.06%	22.35%
Margin Score											

Figure24

Asia Pacific					South America				
Total of Customer Type	Entertainment Venues	Branded Stores	Retail	Internet Direct	Total of Customer Type	Entertainment	Branded	Retail	Internet Direct
6,560,019.24	3,610,942.46	751,436.37	1,974,221	223,419.41	2,555,611.59	1,231,069.61	107,116.05	1,082,938.5	134,487.43
5,469,067.27	2,945,672.61	671,837.17	1,667,717.45	183,840.04	2,125,449.52	1,067,221.24	84,541.34	845,864.58	127,822.36
16.63%	18.42%	10.59%	15.53%	17.72%	16.83%	13.31%	21.08%	21.89%	4.96%

Figure25

The chart and the table, both illustrate that the negative margin of revenue to cost observation is only evident in North America. Where there is a total of -17 % negative margin.
Secondly, it can be seen that “Internet Direct” business which is performing well in most of the territories/region except for South America, where it is the lowest profit-making channel.

4.3.5 Are they any causes for this weak sale or revenue loss in NA?

In Bar chart, “Sales- Region Measures” was dragged into rows, “Total of territory” and “YM” into columns, and a filter was added by selecting “Computer & Tablets”. Here again chart and crosstab were done separately.

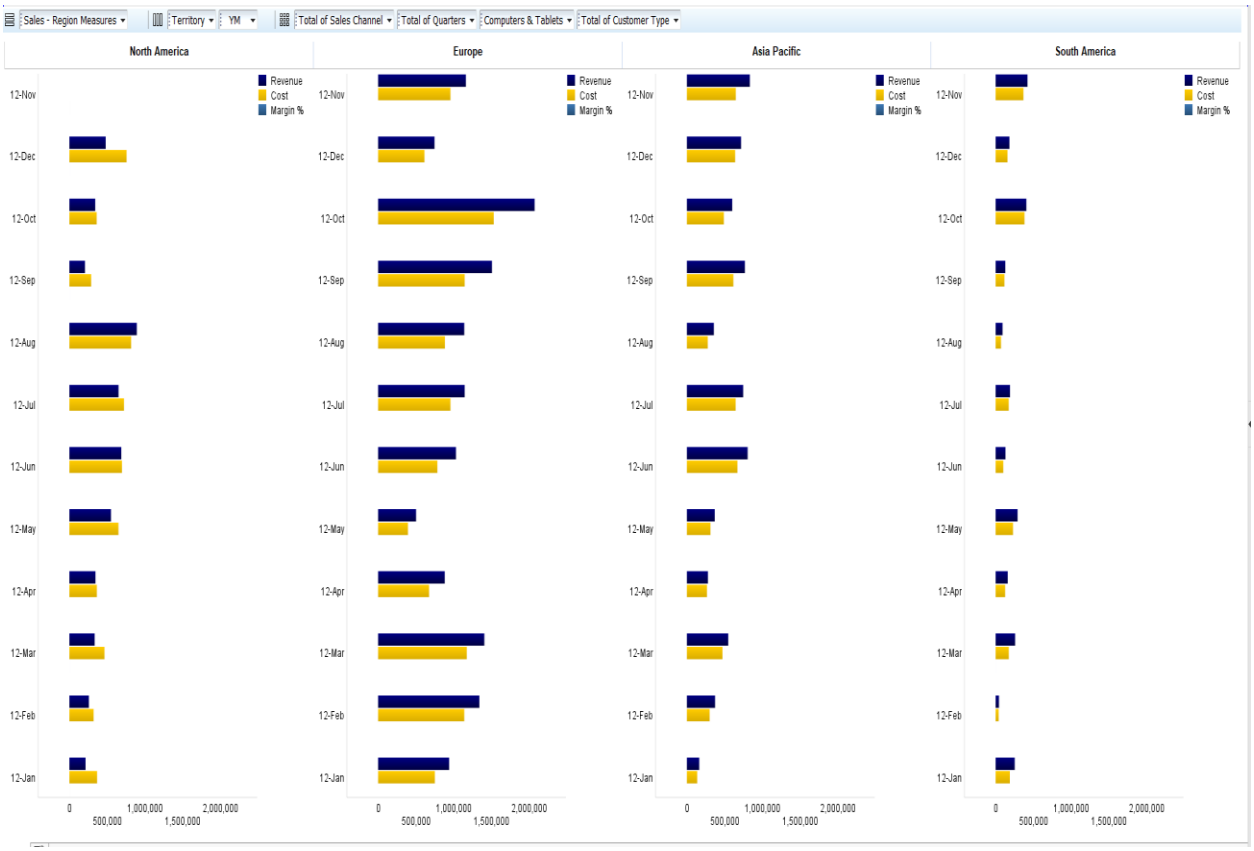


Figure26

The above bar chart represents compare to other territories such as ‘Europe’, ‘Asia Pacific’ and ‘South America’, there seem to be no sales in the month of November and the months of Jan, Feb, March, May, July, December generated loss and rest of the months showed slim profits in North America.

4.4 Analysis & Self-reflection of the tool

IBM Cognos Insight is very easy to download, use, learn and handle for a new user. The interface is also user-friendly and can be mastered within few minutes of handling it. It has the beautiful feature of representing the cross tab and chart at the same time on a single screen, making the results easy to comprehend. For deeper analysis, it is very easy to merge datasets. It also has the 'What-if Scenario' analysis feature which helps present ways of how to impact the business outcomes. Though the tool is easily available and easy to learn and understand, it also has many limitations like the quality of data visualization seems to be sub-standard when compared to other BI tools such as Microsoft Excel, SAP Lumira, or SAP PA. As well the choice of chart types and control on data formatting is limited.
Looking forward I would like to work with live data sources, access live operational databases such as IBM Cognos Statistics, IBM SPSS Modeler, and IBM CognosTM1, and try out the online collaboration capability.

4.5 Conclusion

I initially enjoyed using this tool as it was easy, from part of data loading to visualizing the results. For a new user like me, there is very less cognitive load while working with this tool. I really like the feature of being able to see the crosstab(data) and visualize the same in the chart at the same time. Additionally, I also like that I could control this feature by choosing to only visualizing charts or crosstab. However, I came across a lot of shortcomings such as when trying to visualize advanced charts such as tree maps or heat maps. It also falls short in analyzing and visualizing when handling multiple variables. Overall, it is a perfect tool for newcomers or business managers who would like to present and share the company's performance instantly.

Chapter 5 SAP CLOUD ANALYTICS

5.1 About SAP Cloud Analytics

SAP Cloud Analytics (SAP CA) is a cloud-based data visualization tool for Business Intelligence. [17] Apart from visualization, it also has planning and predictive analytics functionalities. It can access data from both, by importing data as a file or from the database and from direct access i.e., real-time data. It has added the feature of adding indicators, inserting formulas while data modeling. SAP CA has a Smart Discovery feature where it can predict with integrated Artificial Intelligence (AI). The recently added Planning feature makes it possible to create strategies and plans based on real-time financial data and operational data. [18] It is cloud base enabling a user to access and share information from any location and to collaborate with others.

5.2 Dataset & Research Questions

The data used for this chapter is ERP Simulation (ERPSIM) data, obtained from the ERP Simulation game. The ERPSim game simulates a commodity market wherein each team must run a company to gain the highest company valuation compared to other teams. It is played by teams over several rounds (up to 12 rounds of 20 virtual days each). The products sold by teams are all muesli cereal in various flavors and quantities. The teams must forecast demand, decide on the MRP, procure, produce, price, and market their products for sale. The data used is static and historical data in *xlsx* format.

The data has 6558 rows and 10 columns, the quantitative variables such as Price, Quantity, and Revenue are the measures whereas descriptive variables such as Round, Day, Distribution Channel, SalesOrder, Team, Area, and Product as dimensions. There are 10 teams namely KK, LL, MM, NN, OO, PP, QQ, RR, SS, and TT from three different areas NO, SO, and WE. For better understanding, after data loading, areas are renamed for better readability, NO- north, SO – south, and WE- west, and Distribution channels are renamed 10: Hypermarkets, 12: Grocery Chains, 14: Convenience.

The main objective of this analysis of the ERPSIM data is to explore the major contributors in revenue generation and suggest strategies to improve the revenue generation of the weak contributors.

And review and study the teams of the distribution channel with lesser revenue to be able to suggest strategies in the future.

5.3 Applying Analytical tools and Results.

5.3.1 What are the top 2 products contributing highest to the Revenue? Name them and the product contributing the least.

To get the top two and least revenue contributor products, “Revenue” was selected in Measures, “Products” in Dimension, and Bar/Column was selected as the chart type. The revenue was sorted from largest to smallest. By using the design function, the top two revenue generators and the least were formatted for better visualization.

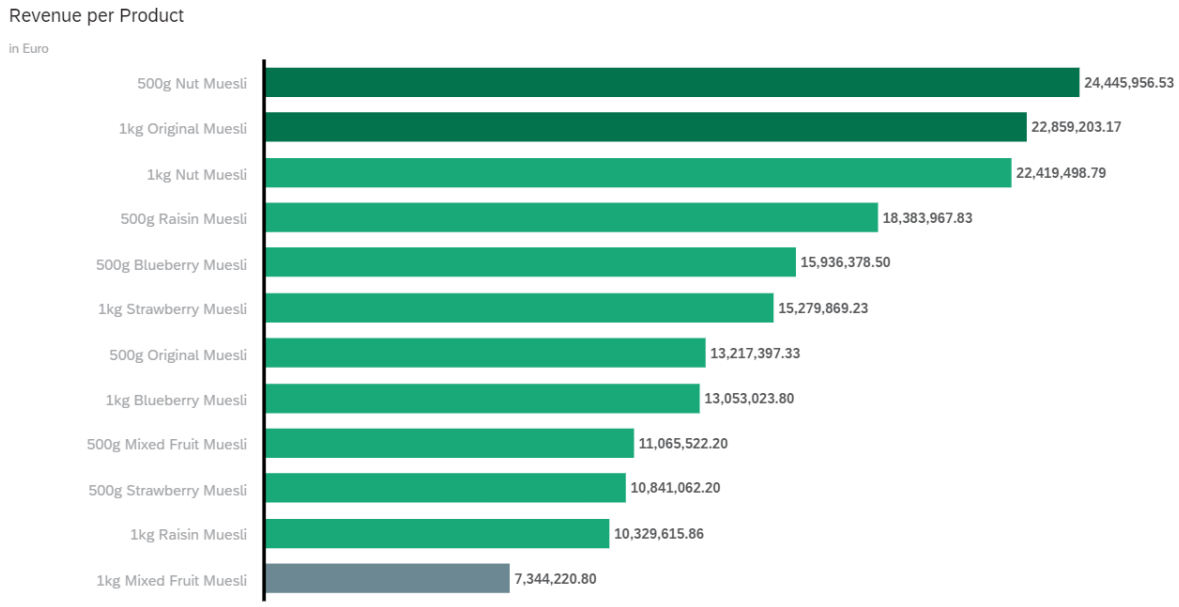


Figure27

The 500g Nut Muesli and 1kg Original Muesli are the first and second revenue contributor, respectively. While 1kg Mixed Fruit Muesli is the lowest revenue generator.

It would be prudent to have more stock of 500 g nut muesli and 1kg original Muesli.

5.3.2 Explore revenue contribution of distribution channels. Does the sale of the top 2 revenue generate products (from question number 1) vary as per distribution channel?

To visualize the total amount of revenue for each distribution channel and the Revenue of each product across all Distribution Channels, “Revenue” was selected in Measures, “Distribution Channel” and “Products” in Dimension. Heat Map was selected as the choice of chart.

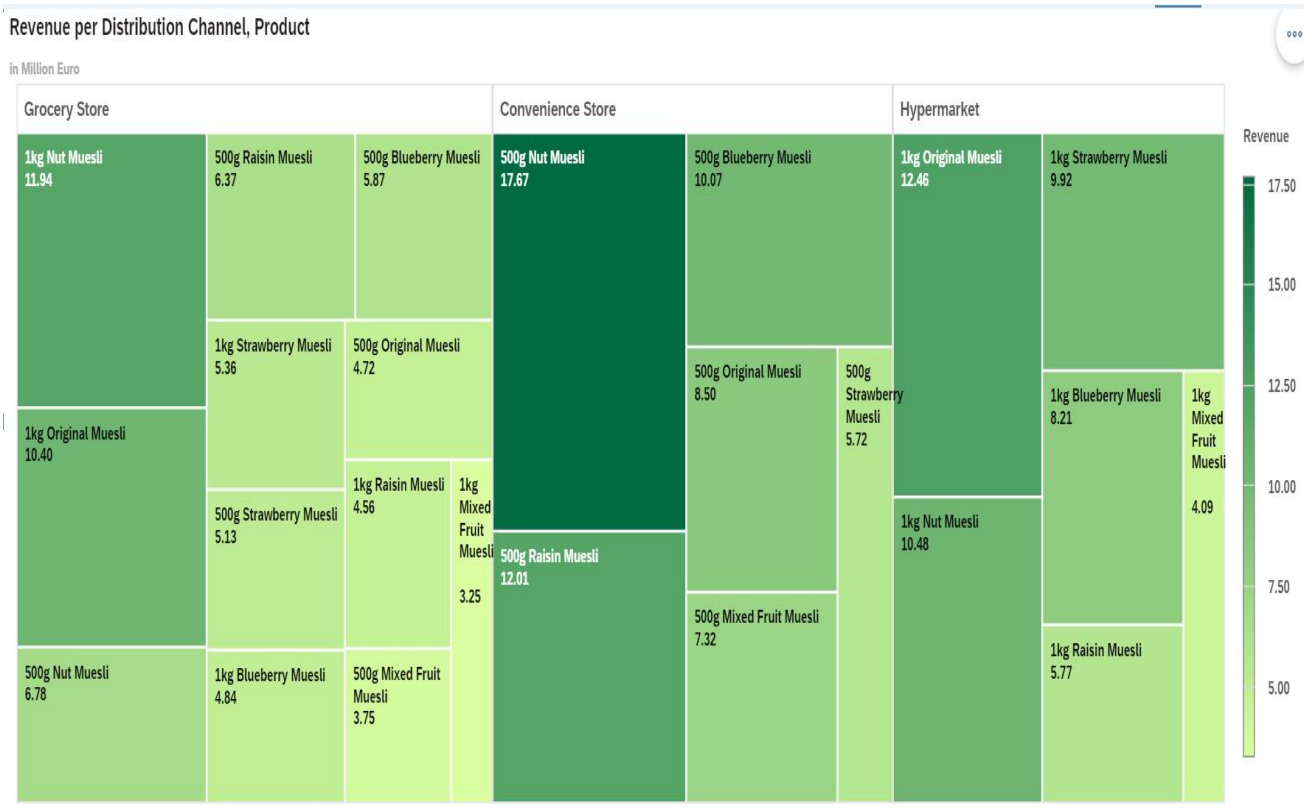


Figure28

From the above Heat map, one can infer that 'Grocery stores' has the highest and 'Hypermarket' has the lowest total revenue. we can see that the top 2 rankers '500g Nut Muesli' and '1kg Original Muesli' from question 1 are not the top product for all distribution channels. The top product for each distribution channel is as follows; Grocery stores – 1kg Nut Muesli, Convenience store - 500g Nut Muesli and 'Hypermarket'- 1kg Original Muesli.

From this chart, it becomes clear that the teams would have decided the product mix depending on the distribution channel they wish to target.

5.3.3 How each team contributes to revenue for each distribution channel?

"Donut chart was selected as the choice of chart. Revenue" was selected in Measures and "Team" in color. To be able to represent them for all Distribution Channel, Trellis was added as a new field in the Designer by clicking on the chart and then on the three dots on the top right side of the chart and selecting 'add field'. In Trellis "Distribution Channel" was added.

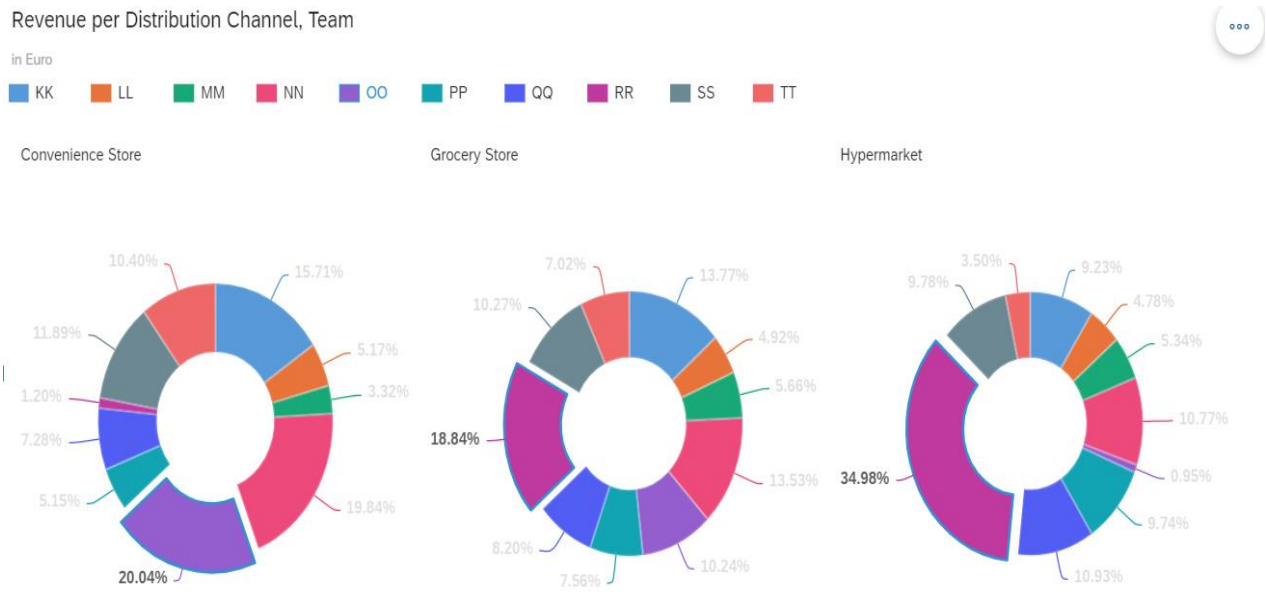


Figure 29

The pie chart reveals that the RR team generates the highest revenue in both Grocery stores (18.84%) and Hypermarket (34.98%). In the Convenience store, Team OO is the highest revenue generator.

To be able to provide strategies and to understand the distribution channel 'Hypermarket', we will now focus on analyzing the 'RR' and 'QQ' team, the largest and second-largest revenue generator in 'Hypermarket' as seen in the above diagram.

5.3.4 For Hypermarket, illustrate the top 3 highest sales quantity for RR and QQ team.

"Quantity" was selected in Measures, "Product" in Dimension, and "Team" in color. In Trellis, the "Team" was added. Two filters were added, for "Distribution Channel" – 'Hypermarket' and for "Team" - RR & QQ. Using the rank function, Top 3 are selected.

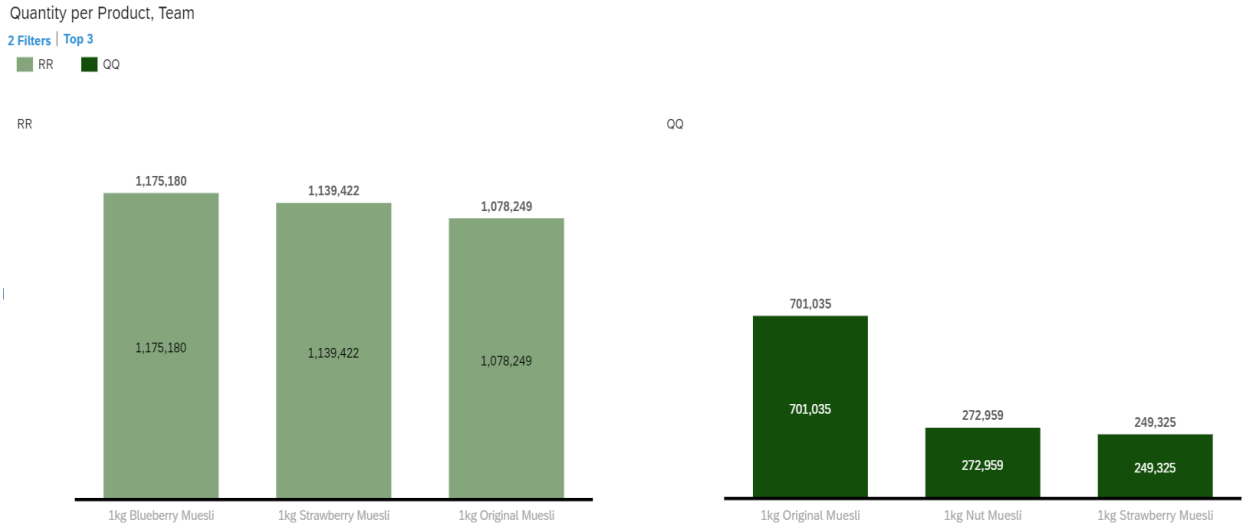


Figure 30

From the above column chart, we can appreciate the three products of maximum quantity sold by team 'QQ' and 'RR' at Hypermarket. For RR in decreasing order, they are namely 1kg Blueberry Muesli, 1kg Strawberry Muesli and 1kg Original Muesli. While for the 'QQ' team they are 1kg Original Muesli, 1kg Nut Muesli and 1kg Strawberry Muesli. It seems that 1kg packing is the preferred packing in the Hypermarkets and Original Muesli and Strawberry Muesli flavors of choice.

5.3.5 Should there be more focus on certain areas, we try to find the areas with the highest sales for each of the top 3 selling products of RR, QQ team.

The stacked column chart was chosen to represent the information. "Quantity" was added in Measure, "Product" in Dimension, "Area" in color, and "Team" in Trellis. In filter for "Distribution Channel"- 'Hypermarket' is chosen and for "Team" 'QQ' and 'RR' are selected.

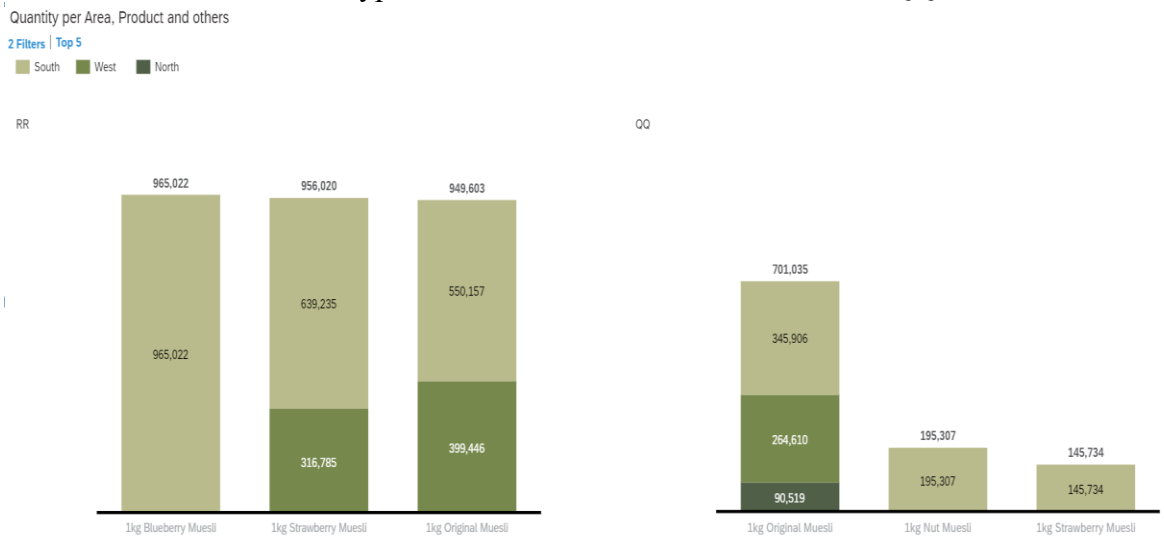


Figure 31

The first view of the stacked column chart shows 'South' seems to be the major area with sales of all the top 3 products for the team 'RR' & 'QQ' in 'Hypermarket'. We can also infer that the top product for 'RR' is only sold in the 'South' area while for the QQ team almost 50% is sold in the 'South' area and 30 % in 'West' and remaining in 'North'. Or one can say that Team RR focused most of their efforts in the South and West region and did not focus on the North region.

5.4 Analysis & Self-reflection of the tool

The SAP Cloud Analytics tool seemed easy and intuitive to use. The interface is very user-friendly. The best feature is that it is on a cloud, allowing access to multiple systems which generate data, the user can access the information when needed from various locations, and the ability to collaborate with team members. I also explored the Smart Discovery feature for personal learning, it helped me retrieve a lot of insights within few seconds. (Images in Appendix).
The most troublesome part of working on such a cloud-based tool is that the speed of performance is affected by a slow internet connection. Another issue is that there is no undo/redo option resulting need to repeat the steps if any mistakes occur.

5.5 Conclusion

Compared to the earlier two tools SAP PA and IBM Cognos Insight, SAP CA seemed to have better visualizations, even to represent multiple variables. I found it has better user control, especially when formatting the visualizations. For now, I have only explored the visualization functionalities. I wish to learn and try out connecting the software to a live database, collaborating with colleagues and use its predictive and planning functions.
I personally think SAP CA is a perfect "all in one" tool for business intelligence and analytics offering with visualization, planning, and predictive capabilities, its ability to integrate with live data and data assets from remote location allowing accurate and on time analysis of the current situation and its ability to collaborate, comment and discuss among colleagues makes it a tool which is apt for the current times and future,

Chapter 6 Tableau

6.1About Tableau

Tableau is a data visualization and data analytics tool. It was created with a business user in mind and is a very intuitive tool, making it convenient even for a non-technical user to create interactive data visualizations and dashboards. Additionally, the analytical platform, allows users to manage, explore, discover insights from the data from various sources. Tableau's ease in use and potential of creating powerful visualization to understand data has made it a leading choice for BI Industry. It has five different products **Tableau Desktop** – for professional and personal use, providing great visualization, data stories, and data analysis, **Tableau Public** – for cost-effective users where everything is shared with the public (no privacy) and nothing can be stored on a local computer, **Tableau Online** – where data is stored in the cloud and managed securely by Tableau administrative servers, **Tableau Server** – data visualization created on desktop can be shared within an organization when uploaded on the server and accessible only be a licensed user and **Tableau Reader** – free to use, having capability to view and modify the visualization created on Tableau Desktop. For data analysis, data must be clean, transformed and various tables need to be joined for this purpose Tableau offers **Tableau Prep** which allows users to perform ETL visually. [19] [20]
There are two data sources to which Tableau can connect: To a file and To a Server, in the former option, it can connect to local text files, MS Excel, PDFs, JSON, Spatial and Statistical files while in latter it can connect to servers like Tableau Server, MySQL Server, Microsoft SQL Server, etc. Tableau repository stores the saved workbooks for future access. [21]

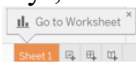
6.2 Dataset & Research Questions

The 'National Health Expenditure' dataset was taken from the Canadian Institute for Health Information (CIHI) website for this chapter. [22] The data is the annual 2020 release/publication of health expenditure in Canada, provided by Canada's provinces and territories. The data is in *xlsx* format and contains two tables, consisting of actual detailed information from 1996 to 2018 and predicted values of 2019, as the 2020 forecasted values were not available from all provinces and territories due to the pandemic, 2020 health spending projections are not included in the dataset.
The dataset has two data tables,
Table 1 - Health expenditure by use of funds and source of finance, by province/territory and Canada. The table has 18900 records and fields such as 'Year', 'Forecast Category', 'Province', 'Sector', 'Use of Funds', Current dollars', 'Current dollars per capita', 'Constant 1997 dollars' and 'Constant 1997 dollars per capita'.
Table 2 - Provincial/territorial government health expenditures by age and sex and use of funds, by province/territory and Canada. The table has 18900 records and fields such as 'Year', 'Jurisdiction, 'Sector', 'Use of Funds', 'Age group', 'Gender' Current dollars', and 'Current dollars per capita'.
In the 'Province', all the provinces, as well as Canada, were included. The 'Sector' included four sectors namely Private, Provincial Government, Public, and Territorial Government. The data

includes expenditure on following categories, Hospital, Other Institutions (residential facilities), Physicians, Other Professionals (dental, vision, and non-paramedic), Drugs (prescribed and non-prescribed), Capital, Public health, Administration, and other Health Spending (health research, home and community care, ambulance, and hearing aids).

Data Loading and Data Preparation

The data was loaded from the local computer on Tableau by selecting Microsoft Excel on the main screen under To a File. The Data interpreter is used to cleanse and visualize the interpreted results. For initial exploration (question 1 – 3) table 1 is dragged to the 'Drag sheet here' section. The Columns 'Forecast Category', 'Constant 1997 dollars' and 'Constant 1997 dollars per capita' were hidden.

By clicking on 

, data visualization and analysis were initiated. For Question 4-5, the table1 was merged with table 2 by clicking on the Data Source -> dragging and dropping table 2 adjacent to table 1 -> 'Year' is selected to create a relationship between the tables. Figure

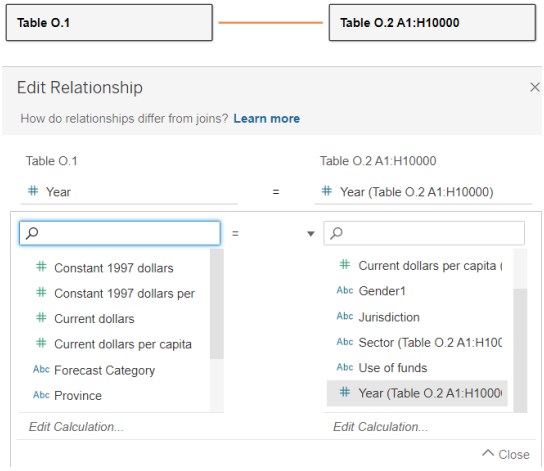


Figure 32

In data preparation, from table 1 fields like 'Forecast Category', 'Constant 1997 dollars' and 'Constant 1997 dollars per capita' were excluded as they didn't contribute to our analysis.
As a data analyst for the Canadian Health organization, with required to create an annual report, the primary aim was to identify trends in health expenditure and fund utilization in various sectors across provinces and to create a dashboard.

6.3 Applying Analytical tools and Results.

6.3.1 What is the total expenditure on health across Canada and the province? What is the average expenditure per capita across Canada?

The dimension "Province" is added in columns and "Current dollars" was added in rows to visualize the dollars spending on healthcare across Canada. The "Province" is sorted in ascending order by "Current dollars".

From figure 33 a, it seems that there is an uneven distribution of funds. So to find the expenditure for the population, per capita expenditure "Current dollars per capita" was dragged in rows, the "Current dollar per capita" was changed from the sum to average. A reference line was plotted by clicking on Analytics ->Custom -> drag Reference Line -> . A pop-up window to customize the position of the reference line opens. In this window, under the line, value- R.L, and in tooltip -Custom is selected and 381 value is added to keep the national expenditure by selecting "Canada" as a reference point.

For both the charts, a column chart was plotted.

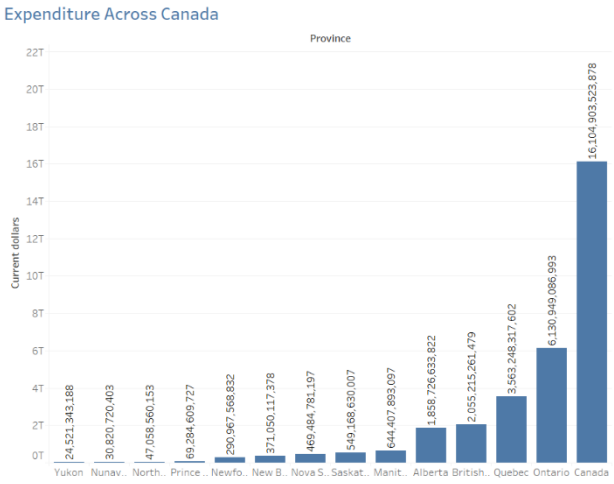


Figure 33.a

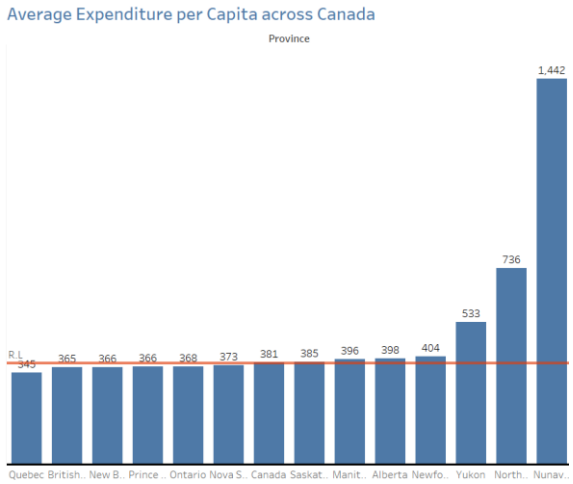


Figure 33.b

The resulting chart (figure 33.b) illustrates that most of the provinces have an annual expenditure like the national average and provinces such as Yukon, Northwest Territories and Nunavut (the highest) are above the national average.
It can be inferred from this that the Government is trying to boost the health care system in these regions.

6.3.2 How are the funds utilized across Canada for Healthcare, and for which purpose?

A Tree map is selected as a choice of visualization, "Use of Funds" is added to Label and "Current dollars per capita" is moved to Size and Label and "Province" is moved to Color and Label. A filter is added on "Use of Funds" to exclude the 'Total' field.

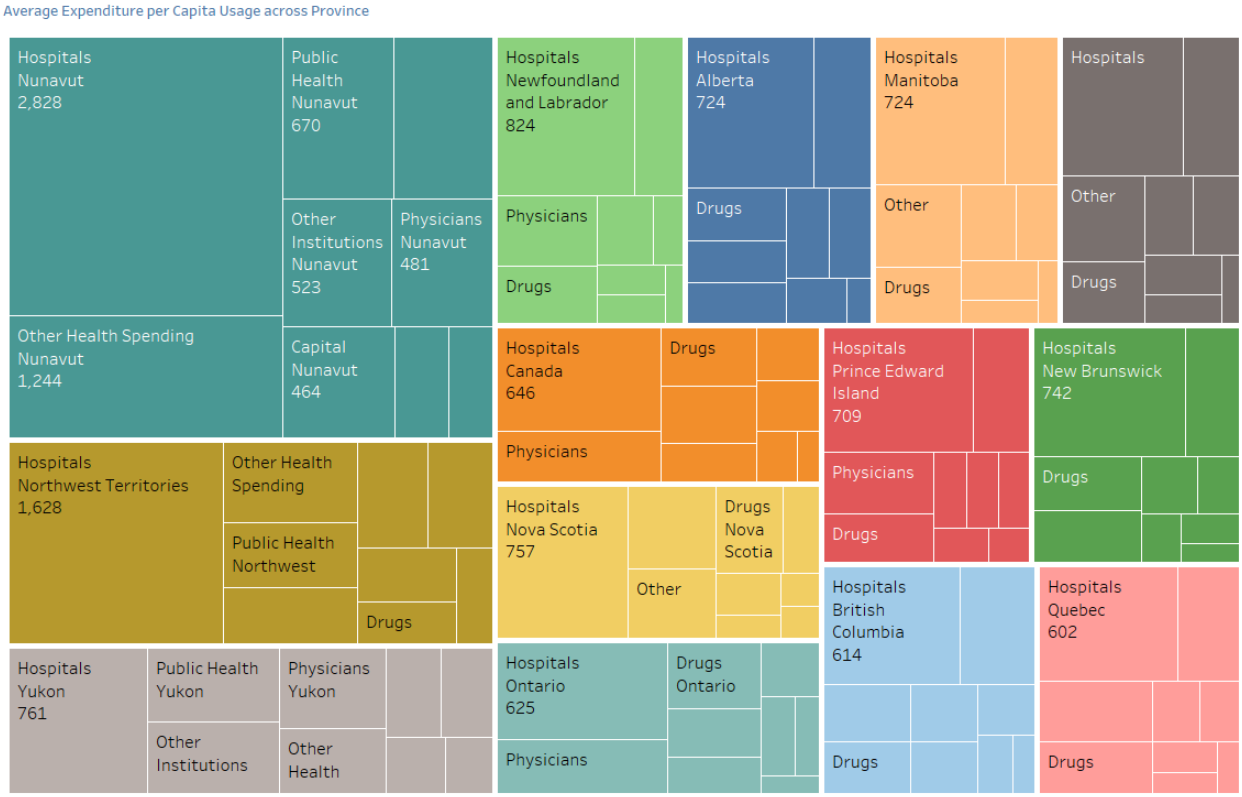


Figure 34

The chart shows that the highest per capita expenditure on health is in the province of Nunavut, followed by northwest territories and Yukon and lowest in British Columbia and Quebec, and the maximum, expenditure is in hospitals across all provinces and territories.

6.3.3 Which sector utilizes the maximum funds in different provinces and where is the maximum expenditure occurring?

A packed bubble chart was plotted with the various sectors "Sector" as "color", "current dollar per capita" expenditure to "size" of the bubbles, "Sector" "province" and "current dollar per capita" was added to the label.

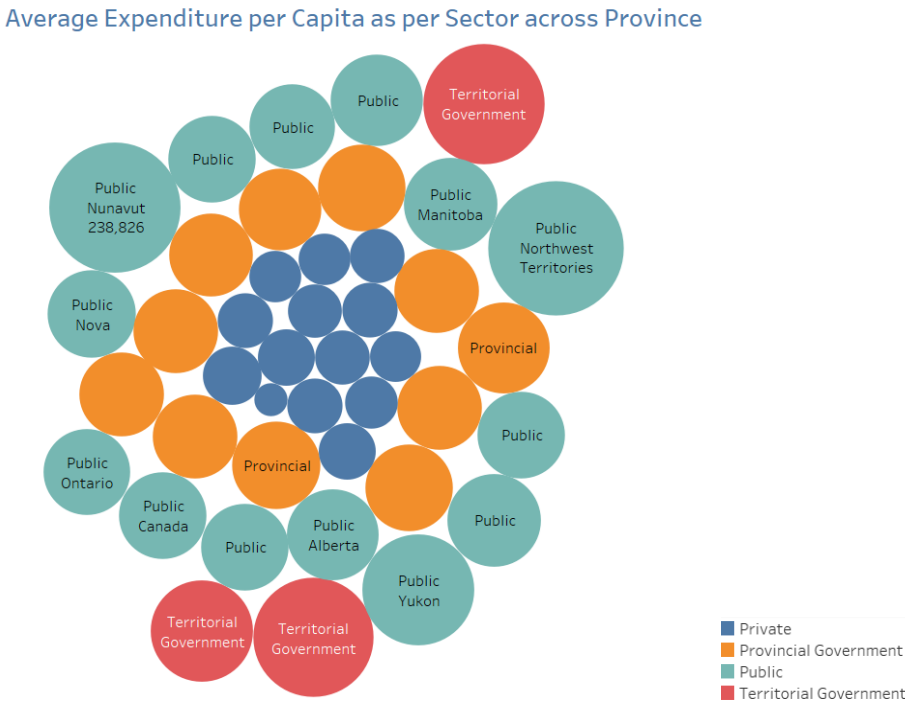


Figure 35

From the above chart, it is evident that the maximum expenditure is happening in the public sector across the country followed by the Territorial government, and the lowest expenditure is happening in the private sector. The maximum expenditure in the public sector is happening in the province of Nunavut and Northwest Territories.

6.3.4 What is the trend of expenditure across Canada over the years? Provide the projected forecast for 2 years in the same chart.

For the next two questions, merged data was used. The dimension "Year" is dragged to columns and "Current dollars per capita" to rows and "Province" is moved to color. From the dimension "Province" only Canada is selected others are filtered out. To visualize, right-click on the trend line and on Forecast -> Show Forecast ->forecast options ->until 3 periods.

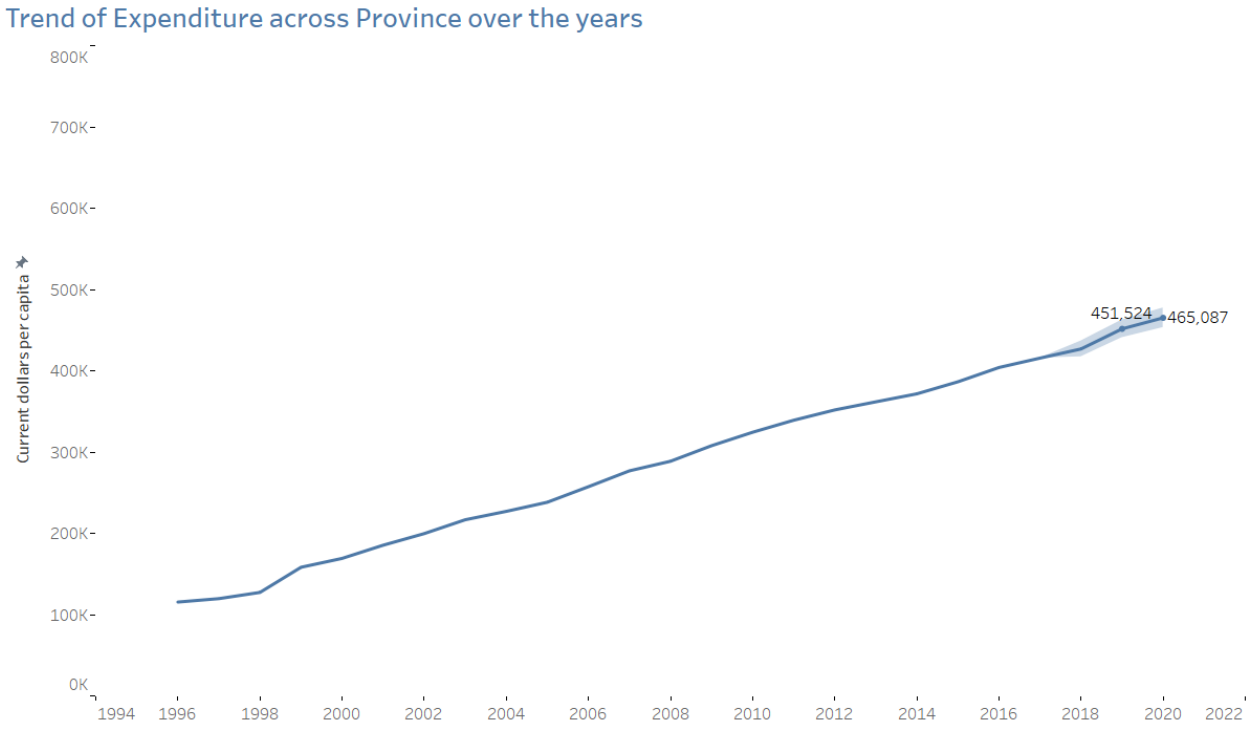


Figure 36

The predicted per capita expenditure is $451, 524 for the year 2019 and $465,087 for the year 2020.

6.3.5. What is the trend of health spending as per age groups and gender?

To see the trend across age groups, a line chart was chosen, with the "Age group" in the columns and "Current expenditure per capita" in the rows to differentiate between the expenditure according to gender, "Gender" was dragged to Color.

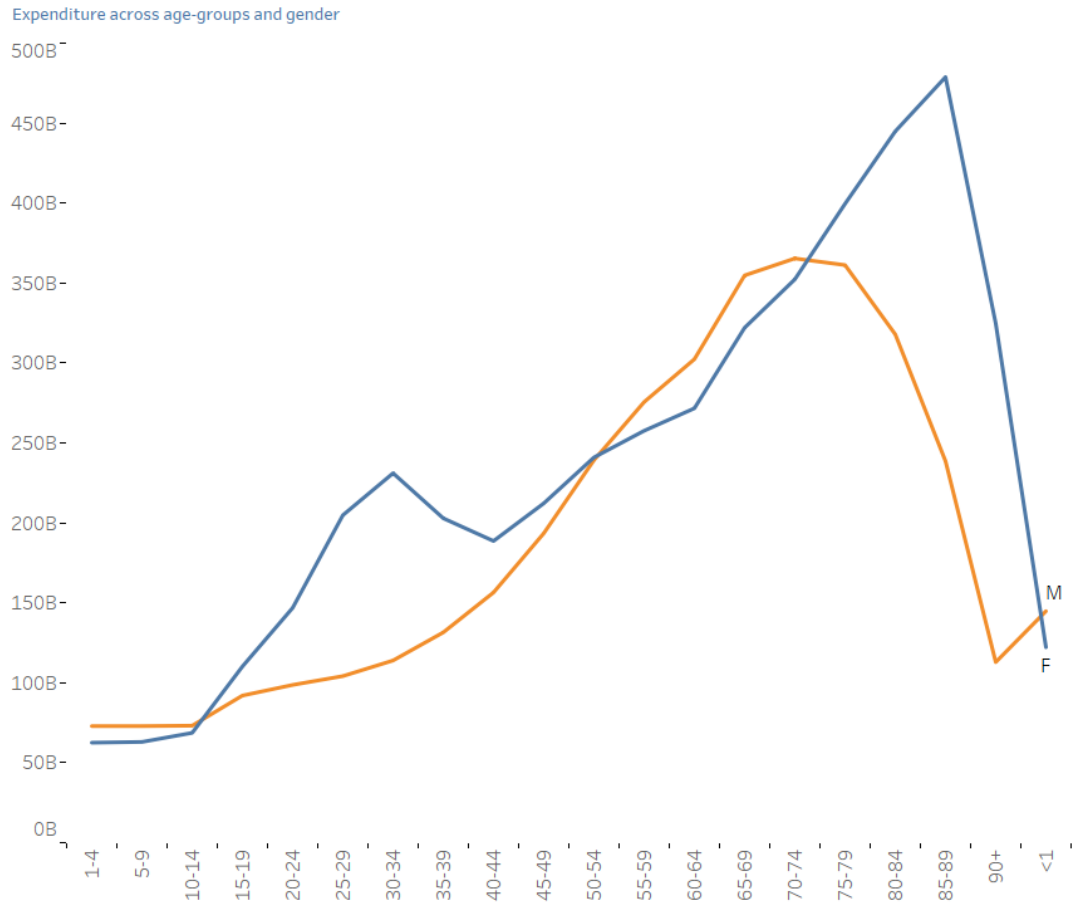


Figure 37

The resulting chart shows that the expenditure on males is slightly more from childhood to teenage, after that there is more health expenditure in case of females till the age of 50 years peaking in the age group 25 to 35. After the age of 50 more expenditure is in males, which drops significantly after the age of 70-75 and highest expenditure happens for females in the age group 85 -90.

6.3.6 An interactive Dashboard representing key indicators.

To Create a Dashboard, the map cap showing the expenditure by provinces and the line chart showing expenditure by age group and gender were dragged on the blank dashboard while holding down the shift key to create floating charts. Two additional charts were created, A bar chart to show expense across various sectors (public, private, provincial, and territorial) and a heat map to show where the funds were used viz hospitals, physicians, administration, drugs, etc. (“use of funds” “current per capita” Expenditure in rows in the column, a filter of use of funds to remove total) and these were also added to the dashboard.

An image for the logo of Health Canada was download and added to a text box. To filter data according to the years a filter for years was added, and a slider option was chosen. To be able to choose the province to view information pertaining to a particular province/ territory only. The map chart was selected and then clicking on the *filter icon* *use as filter* option was selected, same was done for the Bar Chart (expenditure across the sector) and Heat map (Where Fund was used). The resulting dashboard is shown here as *figure 38*, this offers the user the option to visually see how the expense was according to province, in which sector and where it was used. On selecting only one sector from the bar graph all other graphs change showing how the expenditure occurred related to that sector and if one wishes to see how much was spent on hospitals/ drugs etc. one can do so by clicking appropriate zone on the heat map. (more images in appendix 2)

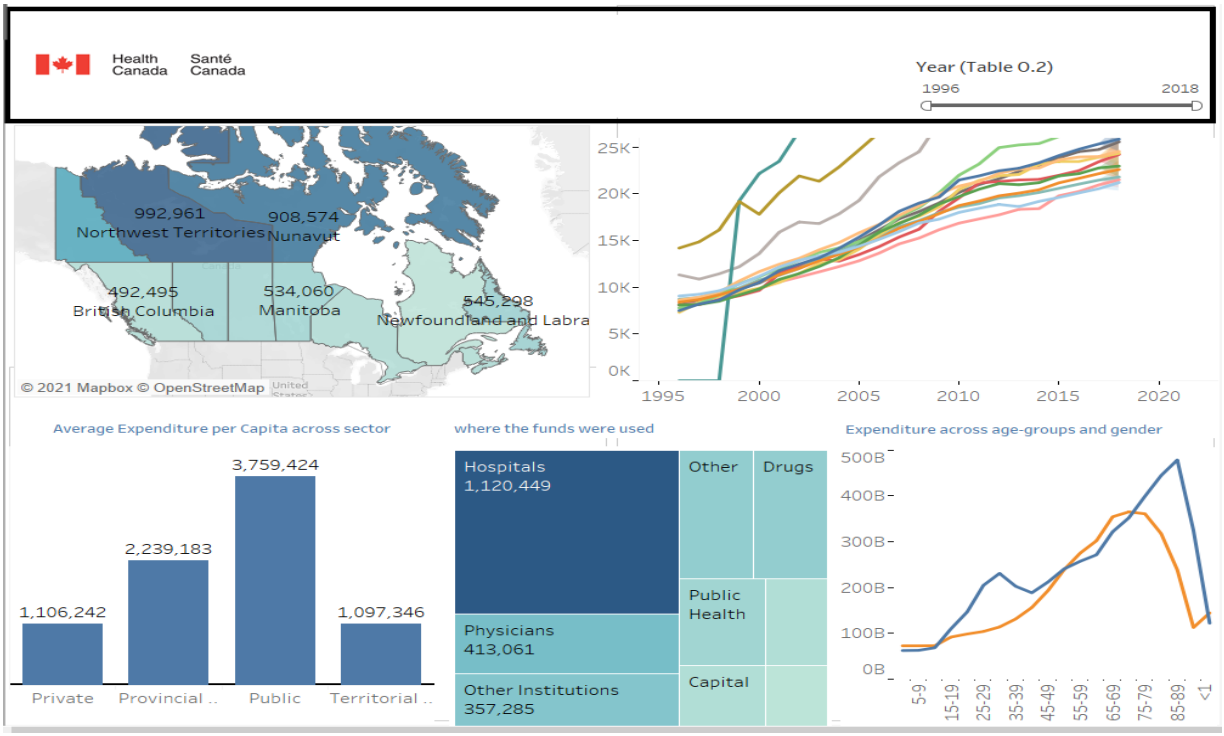


Figure 38

6.3 Analysis & Self-reflection of the tool

I found Tableau Desktop to be a very intuitive tool. It is possible to load multiple data sets, both flat files and databases from various sources. It can plug in live data, operational database, and Web API. I was able to Load the dataset fast. The software layout is clean and easy to understand. Creating a visualization is very simple, by using the drag and drop feature. The Show me tab/marks card suggests the type of graph most suited for the data provided. The best feature was the ability to optimize the chart or format it, such that I was able to remove grid lines, adjust axis, ticks, add a reference line, choose colors to highlight individual datapoints with ease, which I found difficult

with other software except for MS Excel. The dashboard provides an at-a-glance view to get insights from the data. Tableau desktop also has an inbuilt feature which allows to preview and optimize the dashboard for various platforms. The analysis tab has the forecasting option which by default is “8” but can be adjusted to the period of choice, making it extremely easy to make a forecast depending on previous data. Other than forecasting it can also perform clustering for descriptive analytics.

However, Tableau when compared to SAP PA lacks integrating programming languages such as R or python for analysis. Thus, it lacks in building prediction models, such as creating a decision tree, performing association analysis, building a classification model. Tableau desktop does not allow collaboration as SAP cloud analytics does.

6.5 Conclusion

Tableau desktop by itself is a very intuitive and user-friendly tool for business analytics. Hence a popular tool among business users. Although it lacks some features in this version, the software is available in different forms such as Tableau Online, Tableau Server, Tableau Public, and Tableau reader to suit the needs and demands of various kinds of users. With the addition of Tableau Prep, it is also able to perform ETL easily and visually. I would like to explore these options in the future.

Chapter 7 SAP Predictive Analytics

7.1 About SAP Predictive Analytics

SAP PA is a data mining and data analytic tool used to explore the hidden insights and the relationships within the data. It enables an organization to build prediction models, identify hidden risks and foresee future events. The analysis process, traditionally done manually by developing scripts and algorithms by analysts, has been Automated in SAP PA. The SAP PA has various features such as Data manager (for preparation of analytical data), Automated Analytics (for Classification/ Regression Model, Clustering model, Time Series Analysis, Association Analysis, Social Network Analysis, and Recommendation system), and Expert Analytics (for complex analysis using statistical analysis and open-source code language R). It connects with most data sources from flat files, unstructured text files, other formats like SAS and SPSS and can integrate with SAP HANA for real-time analytics. It is a universal tool for knowledge discovery, processing, analyzing, forecasting, and visualizing.[23][24][25]

7.2 **Dataset & Research Questions**

The 'Heart Attack Possibility' dataset for this chapter was taken from Kaggle Website.[26] Originally the data was extracted from the UCI website and had 76 attributes [27] but most of the data analysis by Machine Learning researchers used only14 attributes out of them to predict the presence of heart disease with certain factors.

The data used is in CSV format and has 303 lines of records described by 14 attributes, namely '**age**', '**sex**' (1 = male; 0 = female), '**cp**'- chest pain type having 4 values (0: typical angina, 1: atypical angina, 2: non-anginal pain, 3: asymptomatic), '**trestbps**' - resting blood pressure (in mm Hg), '**chol**' - serum cholesterol (in mg/dl), '**fbs**' - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false), '**restecg**' - resting electrocardiographic results (0: normal, 1: having ST-T wave abnormality, 2: showing probable or definite left ventricular hypertrophy by Estes' criteria), 't**halach**' - maximum heart rate achieved, '**exang**' - exercise induced angina (1 = yes; 0 = no), '**oldpeak**' - ST depression induced by exercise relative to rest, '**slope**' - the slope of the peak exercise ST segment (1: upsloping, 2: flat, 3: downsloping), '**ca**' - number of major vessels (0-3) colored by flourosopy, 'thal' - 0 = normal; 1 = fixed defect; 2 = reversable defect. The outcome variable is discrete – valued, where '**target**' - 0= less chance of heart attack 1= more chance of heart attack. There are no missing values. All the variables are of numeric data type.

The objective of this chapter was to explore the data to extract insights and build a classification model to be able to predict heart attack.

7.3 **Applying Analytical tools and Results**.

7.3.1 Explore the 'heart' dataset and provide insights.

For initial data exploration SAP PA **Expert Analytics** is used. The data is loaded using the File -> New -> Text Load a text file (*.csv;*. -> NEXT -> and selecting the 'heart' from local computer.

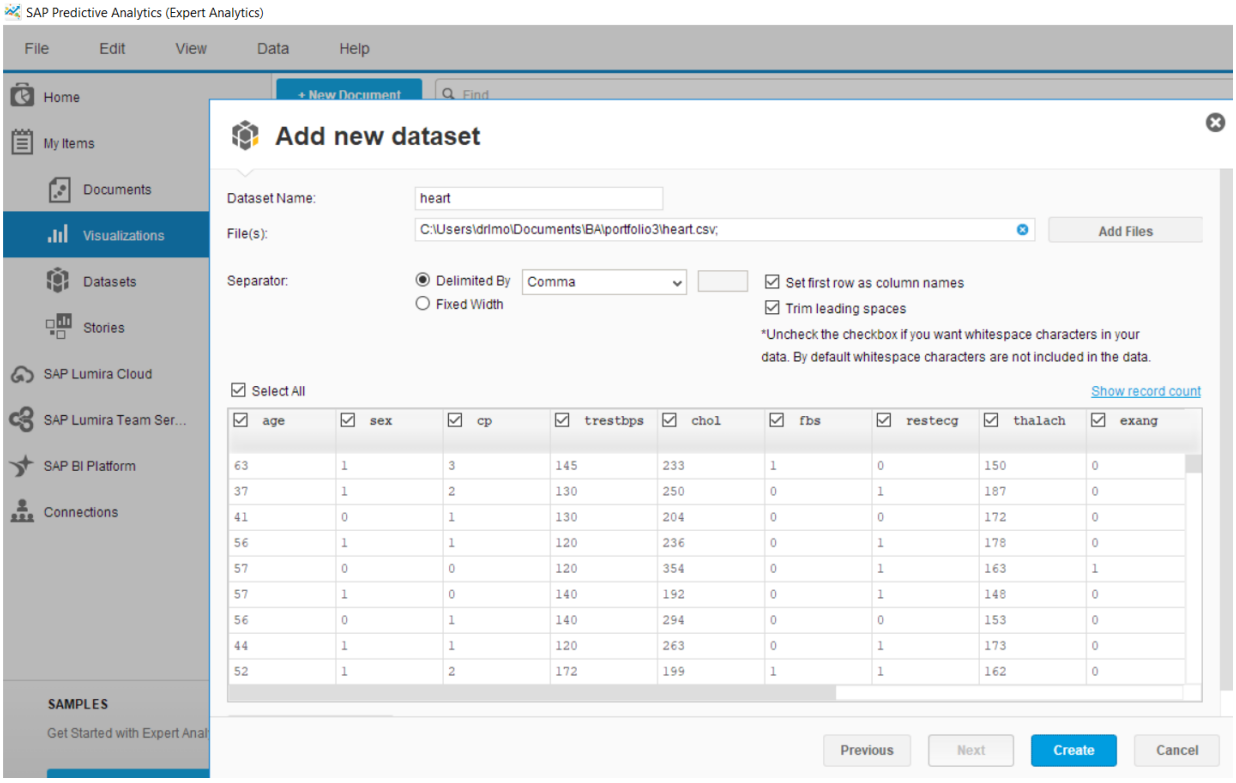


Figure 39

The first visualization was a column chart to see the age with the maximum number of heart cases. The 'target' measure was added in Y-axis and the 'age' in X-axis.
To see the occurrence of heart disease in males and females, 'sex' was added in Y-axis and the 'target' measure in X-axis and bar chart was chosen as the choice of visualization.
Lastly, a bar chart was plotted to visualize the presence of heart disease across various types of chest pain. The 'target' measure was added in Y-axis and the 'cp' in X-axis.

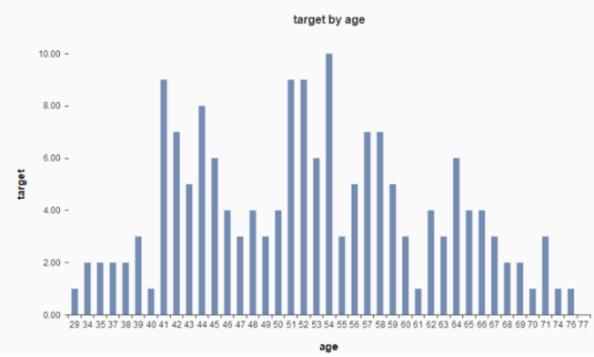


Figure 40

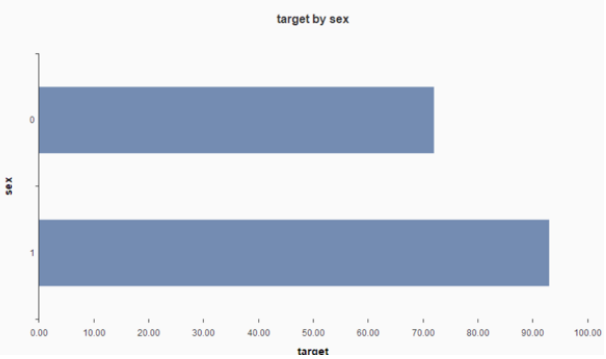


Figure 41

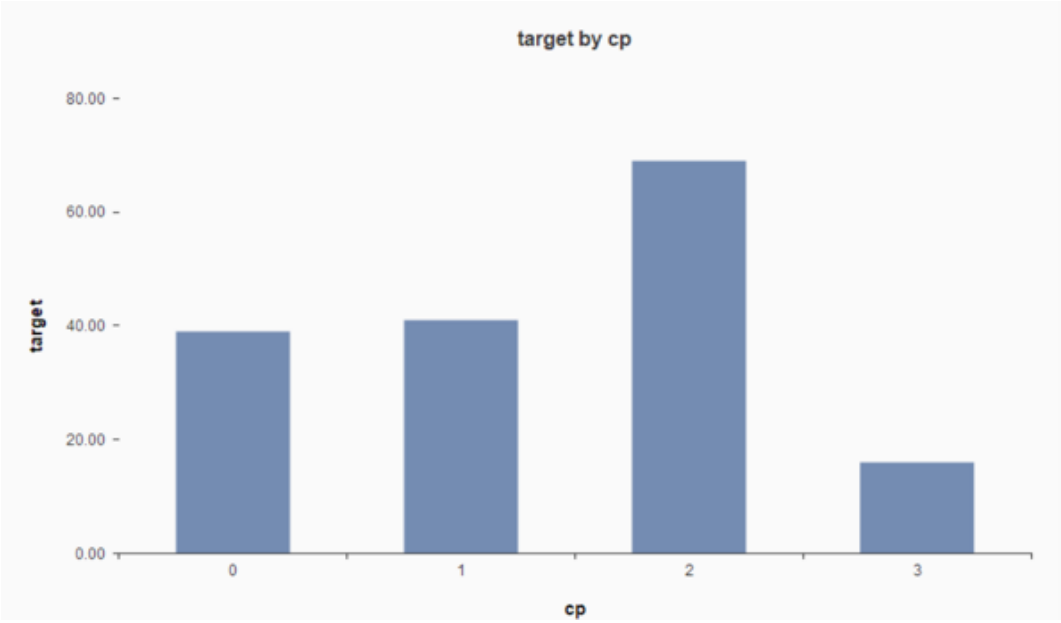
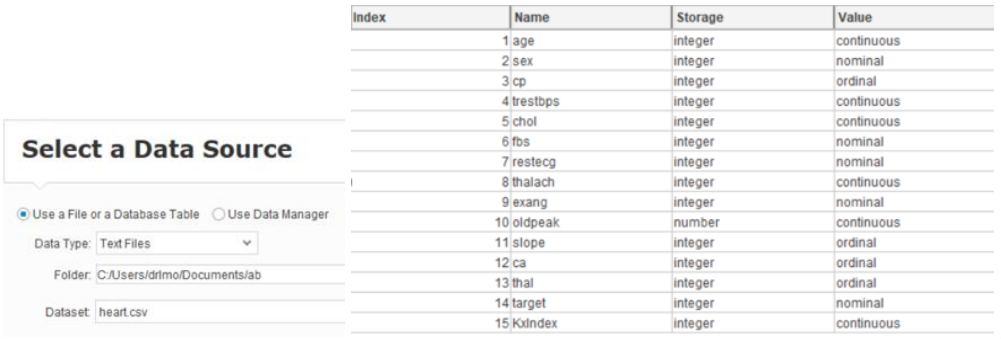


Figure 42

Figure 40, 'target by age' shows that in the above dataset the maximum of heart diseases is seen at age 54, where in the second image (Figure 41) it is seen males (1) are more prone to suffer from heart disease compared to females (0). From figure 42 it can deduce that individuals with non-anginal pain cp value – 2 show having the maximum number of heart disease.

7.3.2 To evaluate the predictability of Heart disease possibility, build the classification model and explain the results.

On SAP PA, Automated Analytics was used to build a classification model by clicking on Create a Classification/Regression Model . The data was loaded by selecting the dataset after selecting Text files in Data Type on Data Source screen. In the Data Description screen, Analyze is clicked which then shows and describes all the variable names and their data type and value. Figure 43. Click Next.



Index	Name	Storage	Value
1	age	integer	continuous
2	sex	integer	nominal
3	cp	integer	ordinal
4	trestbps	integer	continuous
5	chol	integer	continuous
6	fbs	integer	nominal
7	restecg	integer	nominal
8	thalach	integer	continuous
9	exang	integer	nominal
10	oldpeak	number	continuous
11	slope	integer	ordinal
12	ca	integer	ordinal
13	thal	integer	ordinal
14	target	integer	nominal
15	KxIndex	integer	continuous

Figure 43

In Selecting Variables screen, 'target' is selected as the target variable. Click Next. In the Summary of Modeling Parameters screen, select the 'Compute Decision Tree' and 'Enable Auto-selection'. Click Generate.

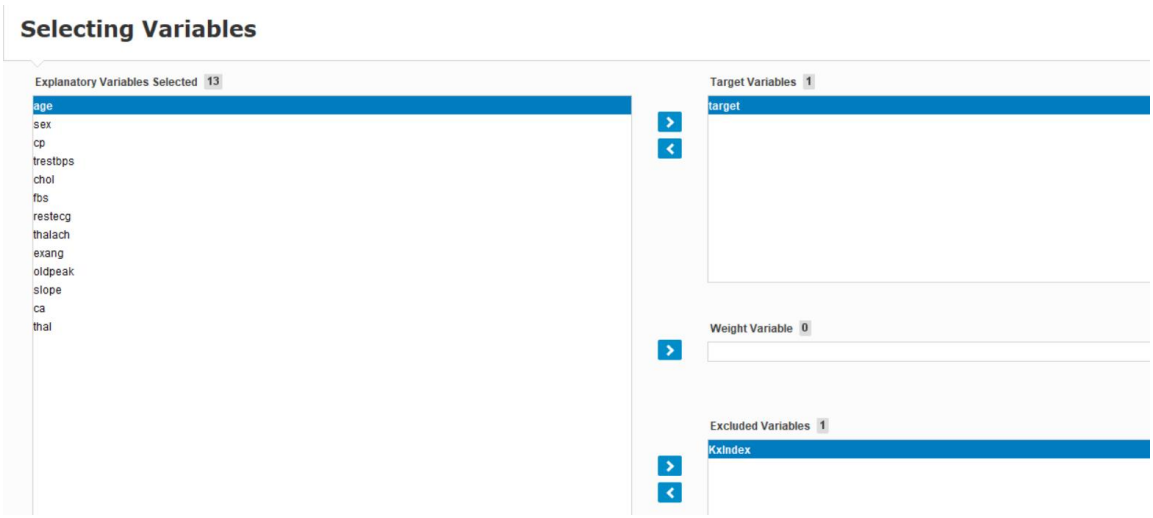


Figure 44

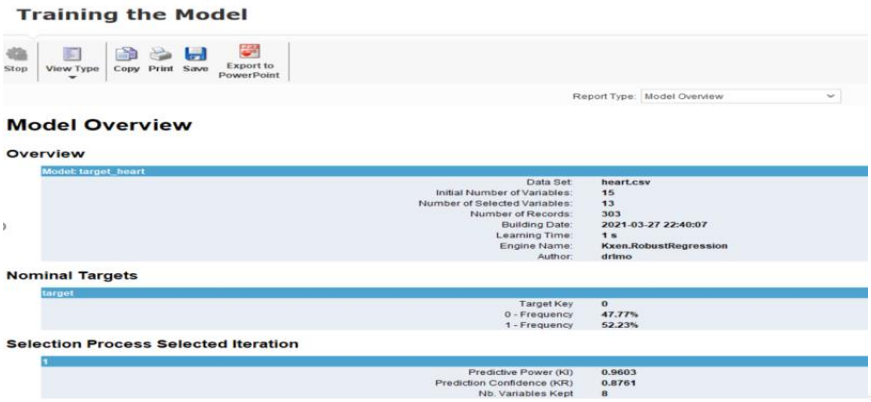


Figure 45

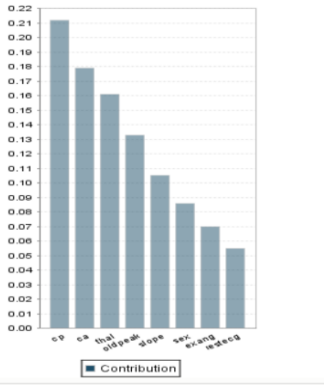
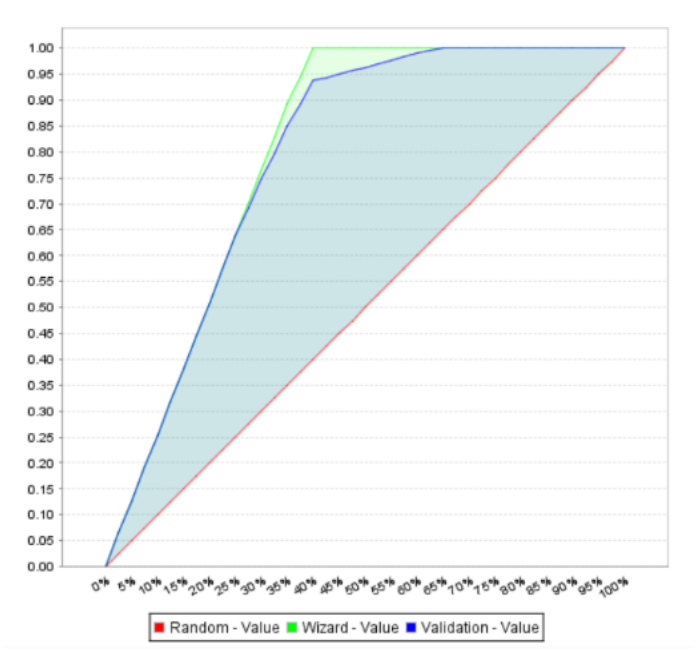


Figure 46

In 'Training the Model', the overview shows out of all variables Nb. Variables kept are 8. The model's Predictive model = 0.97 and Prediction Confidence = 0.88 is appreciated.
The contribution of the predictors towards the prediction of heart disease can be visualized on the column chart, where 'cp' – chest pain type shows maximum contribution.



Profit Curve

The Profit Curve shows the model performance. The blue line shows the training data performance while the green line shows the validation data performance. It is seen that the performance of the validation data is better than the training data inferring it to be a good prediction model.

Figure 47

7.3.3 What are the highest contributing factors for Heart attack?

SAP PA Automated Analytics provides various displays of the classification model such as the Executive Report, Decision tree, Confusion Matrix, Scorecard, Variable Contribution, etc.

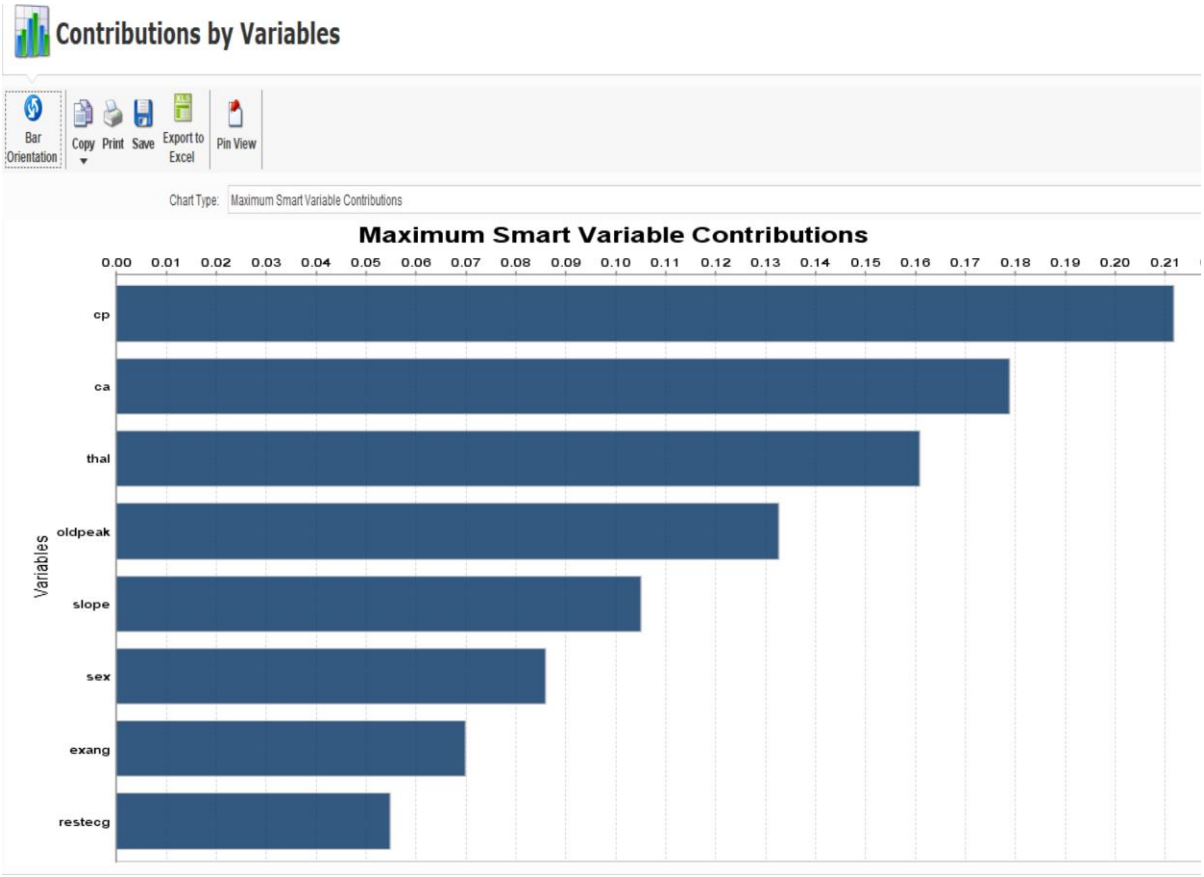


Figure 48

The Smart Variable Contribution bar chart shows the 8 influencing variables sorted in descending order. The 'cp' - Chest Pain type is the highest contributor followed by 'ca' – the major blood vessel involved, in the prediction of heart disease.

7.3.4 Build a Decision Tree of heart attack prediction.

A decision tree is the visualization of nodes and branches where each node represents the test on the attribute node and each branch represents the result of the test on that node. While the topmost node is known as 'root node', the most terminal node is known as 'leaf node'. The decision tree can be built by selecting on Decision Tree on the model screen. By clicking on the nodes, we can expand various nodes and explore their branches and the contribution of each node towards the

heart disease prediction. Here again, it is the most influencing factor is the type of chest pain ‘cp’.

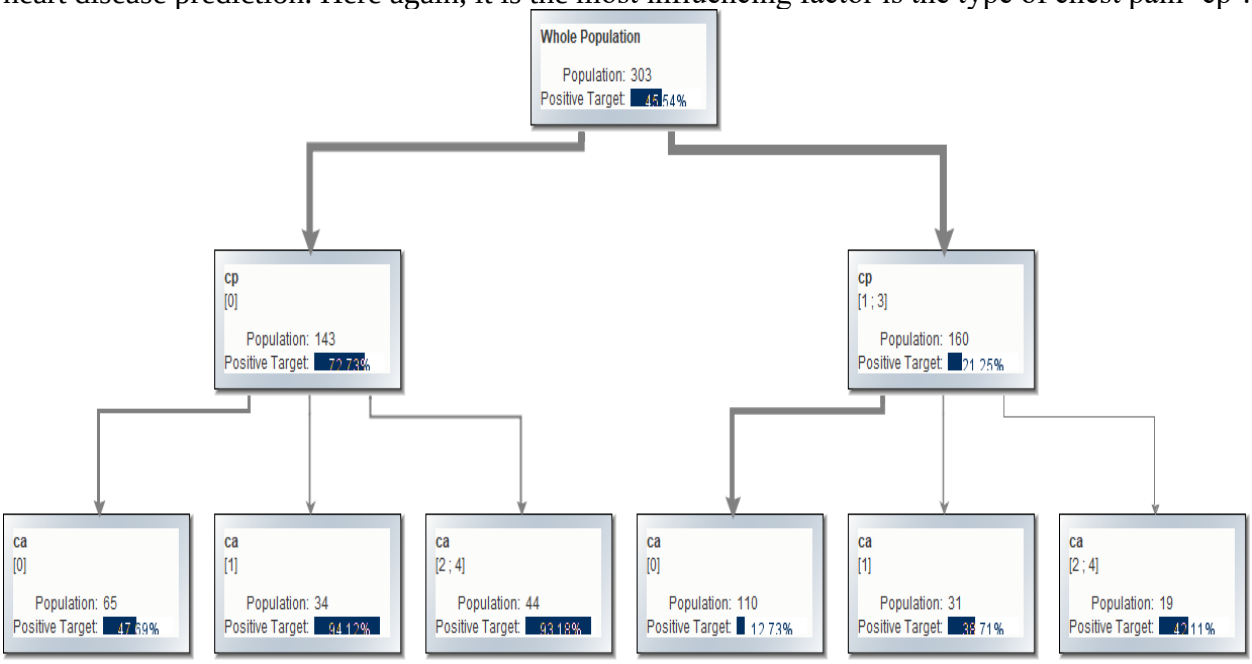


Figure 49

The Decision tree of the model built on the heart dataset illustrates that in the total population of 303 i.e., 45% of positive target (individuals with heart disease) at the root node, only 21% - 160 individuals of cp value 1- 3 have heart disease. Thus, inferring that individuals with atypical angina, non-anginal pain, and asymptomatic chest type of pain have a 21% chance of having a heart problem. While 73% of individuals with ‘cp’ value 0 (typical angina) have a heart problem. Similarly, the contribution of the other 7 variables can be easily visualized by expanding each node.

7.3.5 Explain the performance of the model.

The model’s performance is evaluated with confusion matrix and receiver operating characteristic curve, or ROC / Profit curve. The confusion matrix display is selected from the model screen. The matrix shows the number of rightly predicted true and false cases, known as True positive and True negative, and wrongly predicted true and false cases known as False-positive and False-negative.

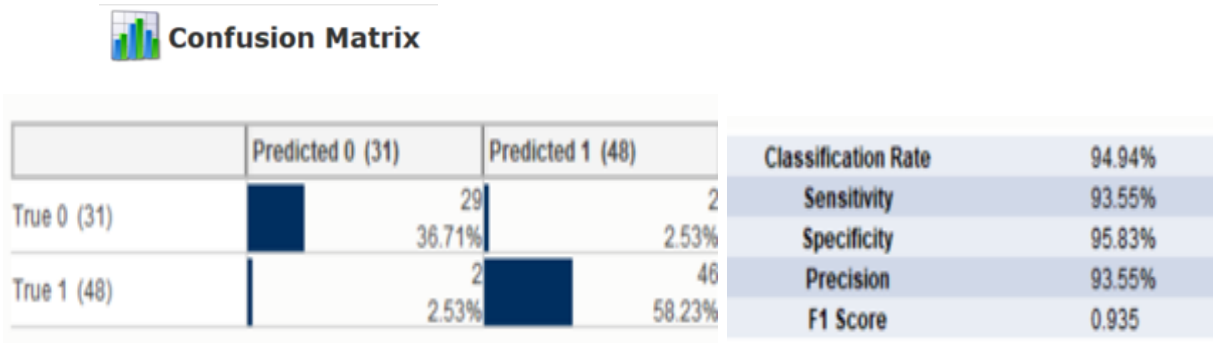


Figure 50

The heart dataset classification model shows 58% of TP and 36% TN. The F1 Score is 0.9, showing the model performance is of high quality.
The ROC or Profit Curve (Figure 47) explained in Question 7.3.2 also represented that the model is of good performance.

7.4 Analysis & Self-reflection of the tool

SAP PA has a clear and intuitive interface. It allows to explore the data and create predictive models by connecting to data of small to large size. Its extensive data visualization capabilities demonstrate forecasted results that are easier to comprehend. The results, charts, diagrams, and tables can be exported and be integrated into reports. The automated functionality has made it easier for even business users and data analysts to create a prediction model. The most troublesome part while using the SAP PA as a BI tool, is its long loading time (probably because it tries to analyze the data before loading), and frequent hanging issues make it difficult to perform the data analysis job at a stretch. This could be because of the free version used for the project. Secondly, the tool did not show an option for data cleaning such as to remove outliers or null values, which are very common in medical data.

7.5 Conclusion

The SAP PA is an excellent platform for data mining, text mining, data analysis, prediction and visualization, and creating reports. The built-in algorithms make it easy for even a casual user to perform all the above functions. When handling noisy data, other tools need to be used to remove outliers or NA values before uploading on SAP PA. However, it stands as an ‘all-in-one’ tool for easy and efficient data analysis for Business users.

Works Cited

1. D. Bruns, “Excel Pivot Tables | Exceljet.” [Online]. Available: https://exceljet.net/excelpivot-tables. [Accessed: 02-FEB-2021].

2. M. Večeřa, “What is a Pivot Table & How to Create It? Complete 2020 Guide – Lumeer.” [Online]. Available: https://www.lumeer.io/pivot-table-complete-guide/. [Accessed: 02-FEB-2021]

3. Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Pivot_table [Accessed: 02-FEB-2021].

4. M.L. Ivan, “Improving Business Intelligence Application by Using New Generation of Web and Mobile technologies.” Informatica Economic Journal, Vol. 19, No. 4,2015, pp. 81-89.

5. Blog www.sapinsideronline.com/blogs/sap-lumira-an-introduction/ accessed 02-02-2021

6. BI, V. (2018, March 08). SAP lumira Discovery: An overview by Visual BI SOLUTIONS. Retrieved April 08, 2021, from https://visualbi.com/blogs/sap/sap-businessobjects/sap-lumira-discovery/sap-lumira-discovery-overview/

7. Government of Canada, S. (2021, April 08). Data. Retrieved April 08, 2021, from https://www150.statcan.gc.ca/n1/en/type/data

8. C. (2019). Common challenges, shared priorities - cihi. Retrieved April 8, 2021, from https://www.cihi.ca/sites/default/files/document/shp-companion-report-en.pdf

9. *SAP predictive analytics in 2021 - Reviews, features, Pricing, Comparison*. (2021, March 8). https://www.predictiveanalyticstoday.com/sap-predictive-analytics/.

10. O'Donnell, J. (2016, February 29). *What is SAP Predictive Analytics? - Definition from WhatIs.com*. SearchSAP. https://searchsap.techtarget.com/definition/SAP-Predictive-Analytics.

11. *SAP help PORTAL*. https://help.sap.com/viewer/5978826050354bc89761a57cfb60473f/3.0/en-US#:~:text=SAP%20Predictive%20Analytics%20is%20a,make%20predictions%20about%20future%20events.

12. *SAP Predictive Analytics: overview*. SAP Predictive Analytics: overview | JET BI. https://jetbi.com/blog/sap-predictive-analytics-overview.

13. Kokemuller, N. (2016, September 29). What is the 80/20 rule of marketing? Retrieved April 08, 2021, from https://yourbusiness.azcentral.com/80-20-rule-marketing-6554.html

14. *IBM Cognos Insight Version 10.2.2: Tutorial*. Office of Technology Services (OTS) | Illinois institute of Technology. https://ots.iit.edu/.

15. *IBM Cognos INSIGHT delivers outstanding data Interactivity, analysis and visualization capabilities to individual desktop users*. https://www.iconresources.com/cognos-insight/.

16. *IBM Cognos Insight – Overview and Use Case*. IBM Cognos Proven Practices. (2013, February 28). https://www.ibm.com/developerworks/data/library/cognos/infrastructure/cognos_specific/page627.html#:~:text=IBM%20Cognos%20Insight%20allows%20you,ODBC%20data%20sources.

17. SAP analytics CLOUD: BI, planning, and predictive analysis tools. (n.d.). Retrieved April 08, 2021, from https://www.sap.com/canada/products/cloud-analytics.html#:~:text=The%20SAP%20Analytics%20Cloud%20solution,supports%20advanced%20analytics%20enterprise-wid

18. *What is SAP analytics Cloud (sac)?* https://www.createch.ca/blog/what-is-sap-analytics-cloud-sac#:~:text=SAP%20Analytics%20Cloud%2C%20commonly%20referred,visualization%20tool%20from%20the%20SAP.&text=SAP%20Analytics%20Cloud%20combines%20the,Planning.

19. Anurag. (2020, September 17). What is Tableau, its uses and applications. Retrieved April 08, 2021, from https://www.newgenapps.com/blog/what-is-tableau-its-uses-and-applications/

20. What is Tableau? (n.d.). Retrieved April 08, 2021, from https://www.tableau.com/why-tableau/what-is-tableau

21. Types of Tableau data sources with connection Establishment Process. (2019, November 15). Retrieved April 08, 2021, from https://data-flair.training/blogs/tableau-data-sources/

22. Access data and reports. (n.d.). Retrieved April 08, 2021, from https://www.cihi.ca/en/access-data-reports/results?fs3%5B0%5D=content_format%3A806&fs3%5B1%5D=primary_theme%3A684

23. SAP help PORTAL. (n.d.). Retrieved April 08, 2021, from https://help.sap.com/viewer/product/SAP_PREDICTIVE_ANALYTICS/3.3/en-US

24. O'Donnell, J. (2016, February 29). What is SAP PREDICTIVE Analytics? - definition from whatis.com. Retrieved April 08, 2021, from https://searchsap.techtarget.com/definition/SAP-Predictive-Analytics

25. SAP predictive ANALYTICS: Overview. (n.d.). Retrieved April 08, 2021, from https://jetbi.com/blog/sap-predictive-analytics-overview

26. Your machine learning and data science community. (n.d.). Retrieved April 08, 2021, from https://www.kaggle.com/

27. UCI machine Learning REPOSITORY: Heart disease data set. (n.d.). Retrieved April 08, 2021, from https://archive.ics.uci.edu/ml/datasets/Heart+Disease

28. How the end of alcohol restrictions is changing kugluktuk | cbc news. (2019, February 18). Retrieved April 09, 2021, from https://www.cbc.ca/news/canada/north/nunavut-alcohol-kugluktuk-1.5012237

Appendix

Appendix 1

Total **Revenue** for **SalesOrder**

in Euro

185,175,716.24 Euro

Total number of **SalesOrder**

6,558

Minimum value

in Euro

6.00 Euro

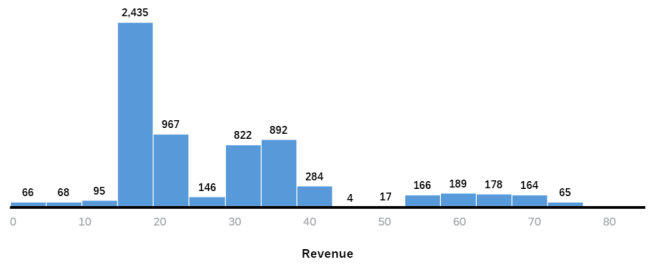
Maximum value

in Euro

76,563.20 Euro

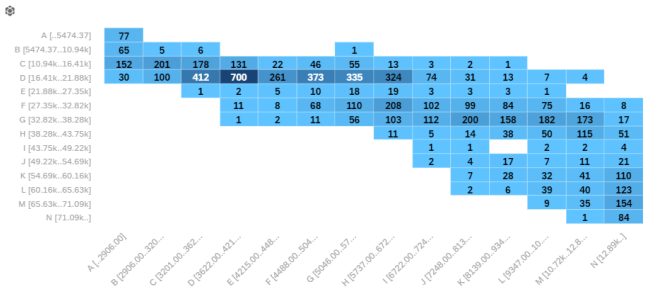
Distribution of **SalesOrder** by **Revenue**

in Thousand Euro | 16 Bins



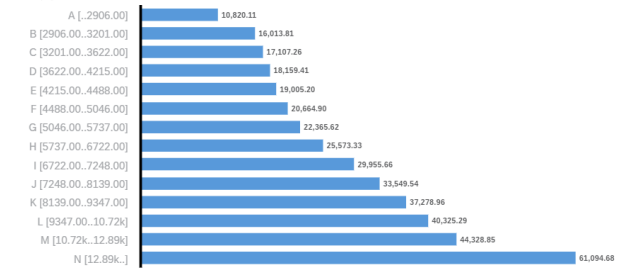
How **Quantity** influences **Revenue** for **SalesOrder**

Distribution of **SalesOrder** by **Revenue** and **Quantity**



Average Revenue for **SalesOrder** by **Quantity**

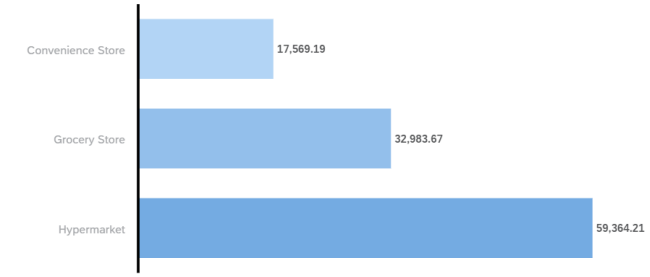
in Euro



How **Distribution Channel** influences **Revenue** for **SalesOrder**

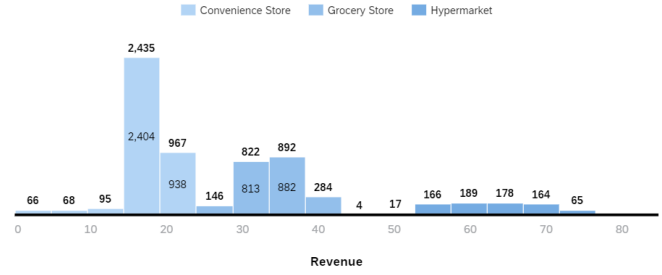
Average Revenue for **SalesOrder** by **Distribution Channel**

in Euro



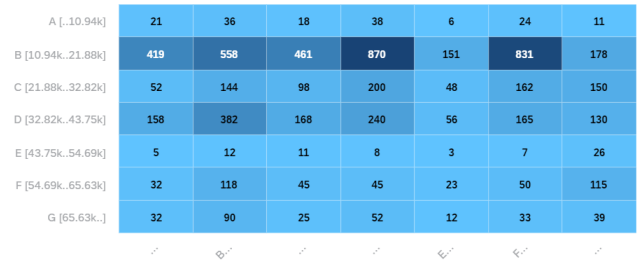
Distribution of **SalesOrder** by **Revenue** and **Distribution Channel**

in Thousand Euro | 16 Bins



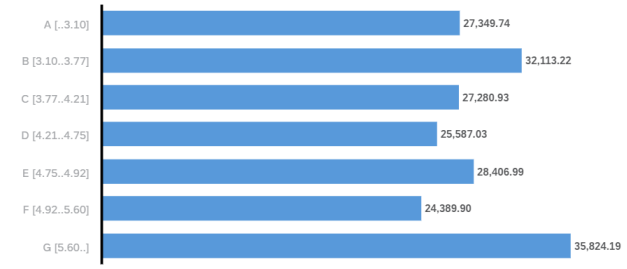
How **Price** influences **Revenue** for **SalesOrder**

Distribution of **SalesOrder** by **Revenue** and **Price**

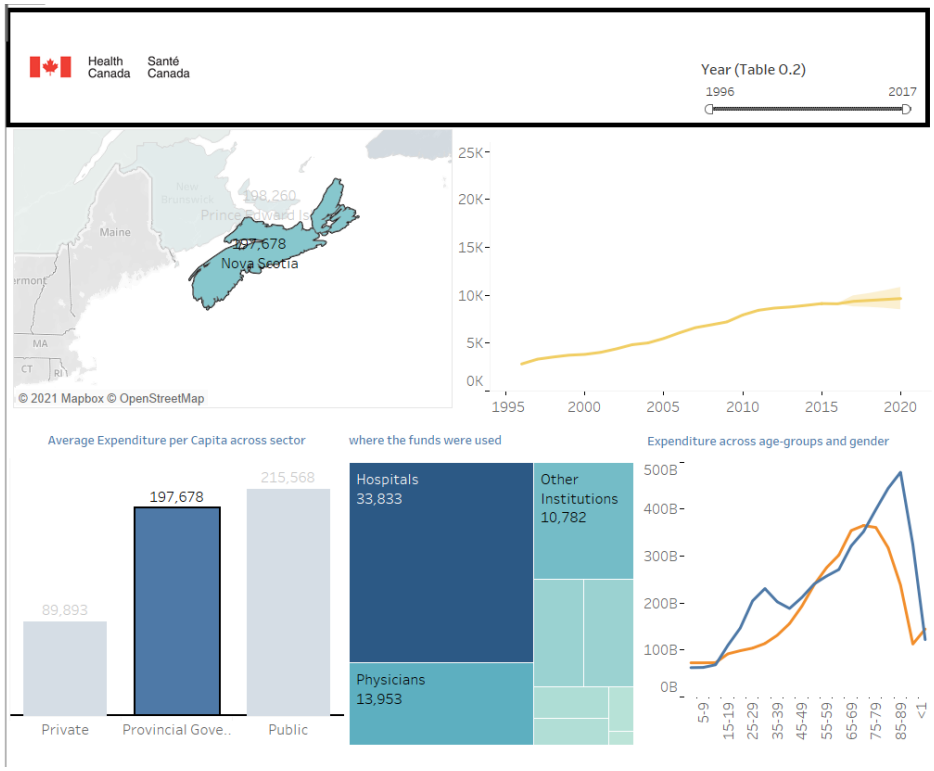


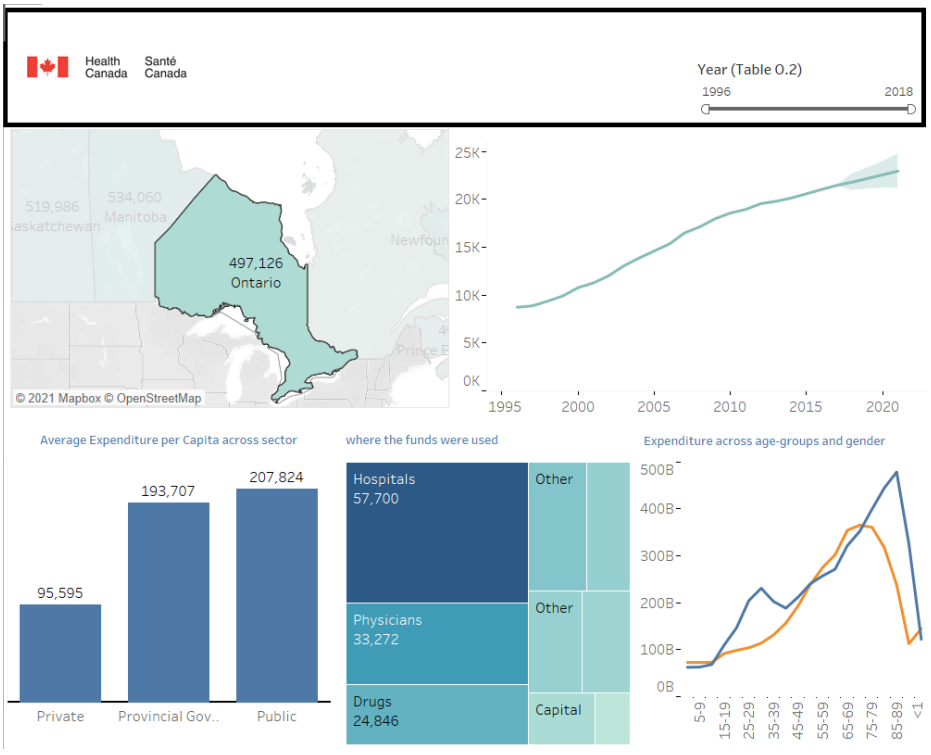
Average Revenue for **SalesOrder** by **Price**

in Euro



Appendix 2


Health Canada
Santé Canada
Year (Table 0.2)
1996
2017
198,260
Prince Edward Is.
297,678
Nova Scotia
Maine
MA
CT
RI
© 2021 Mapbox © OpenStreetMap
25K
20K
15K
10K
5K
0K
1995
2000
2005
2010
2015
2020
Average Expenditure per Capita across sector
89,893
197,678
215,568
Private
Provincial Gove..
Public
where the funds were used
Hospitals
33,833
Other Institutions
10,782
Physicians
13,953
Expenditure across age-groups and gender
500B
400B
300B
200B
100B
0B
5-9
15-19
25-29
35-39
45-49
55-59
65-69
75-79
85-89
<1


Health Canada
Santé Canada
Year (Table 0.2)
1996
2018
519,986
534,060
Manitoba
497,126
Ontario
© 2021 Mapbox © OpenStreetMap
25K
20K
15K
10K
5K
0K
1995
2000
2005
2010
2015
2020
Average Expenditure per Capita across sector
95,595
193,707
207,824
Private
Provincial Gov..
Public
where the funds were used
Hospitals
57,700
Other
Physicians
33,272
Other
Drugs
24,846
Capital
Expenditure across age-groups and gender
500B
400B
300B
200B
100B
0B
5-9
15-19
25-29
35-39
45-49
55-59
65-69
75-79
85-89
<1