BUSINESS ANALYTICS TOOLKIT PORTFOLIO

BUSI6513 – Business Analytics

B00792299
ADITYA VAISHNAVI VEMIREDDY
Dalhousie University

DALHOUSIE UNIVERSITY

Table of Contents

Chapter 1: SAP Business Object Explorer

1.1 Tool Description and Usage:

SAP Business Object Explorer is a self-service data exploration tool that not only creates visualizations leading to insights but also provides an easier way to share the insights across the organization. It is mainly used in a business context where there is a need to search and explore massive datasets. This tool is updated for a transition so that it can run on SAP Netweaver Business Warehouse (SAP BW) info-providers (Sevillano, 2014).

1.2 Data Set Exploration:

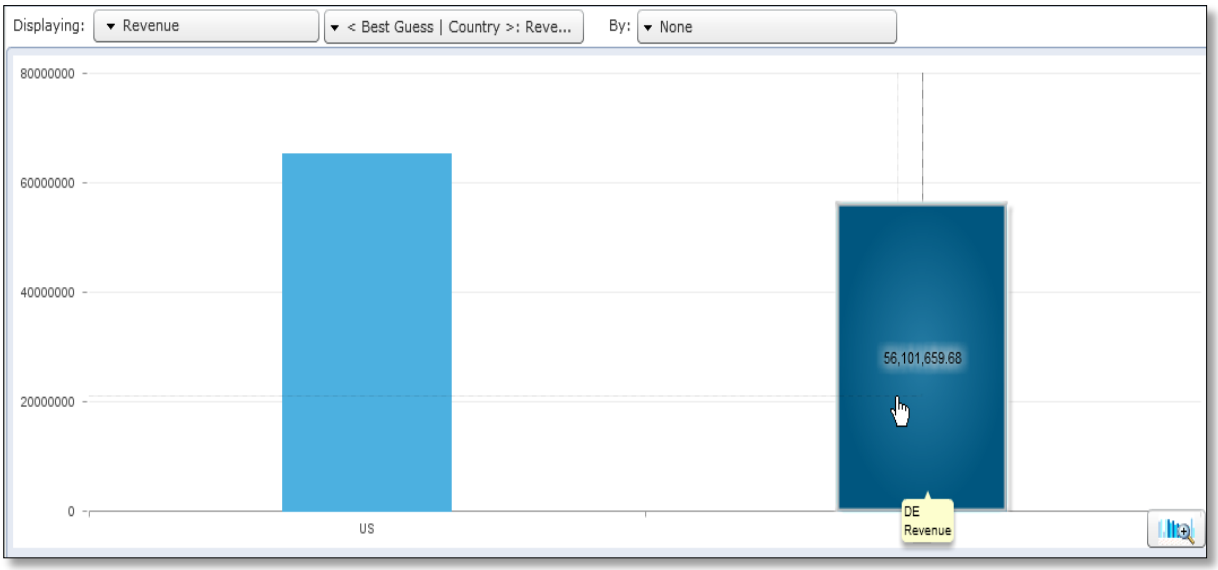
The historical sales data, from 2008 to 2011, of The Global Bike Inc. (GBI) is downloaded from the Brightspace. This data in GBI_Data_E1_1.xlsx with 36,967 recorded transactions and columns: revenue, sales organization, sales quantity, discount, currency, country and date, is explored to understand the sales, customers and products of GBI. The data is well maintained, consistent and all the columns appears to be significant for the analysis, due to which preprocessing is not required. The various trends hidden in the data are identified by using SAP Business Object Explorer.

1.3 Analytics and Insights:

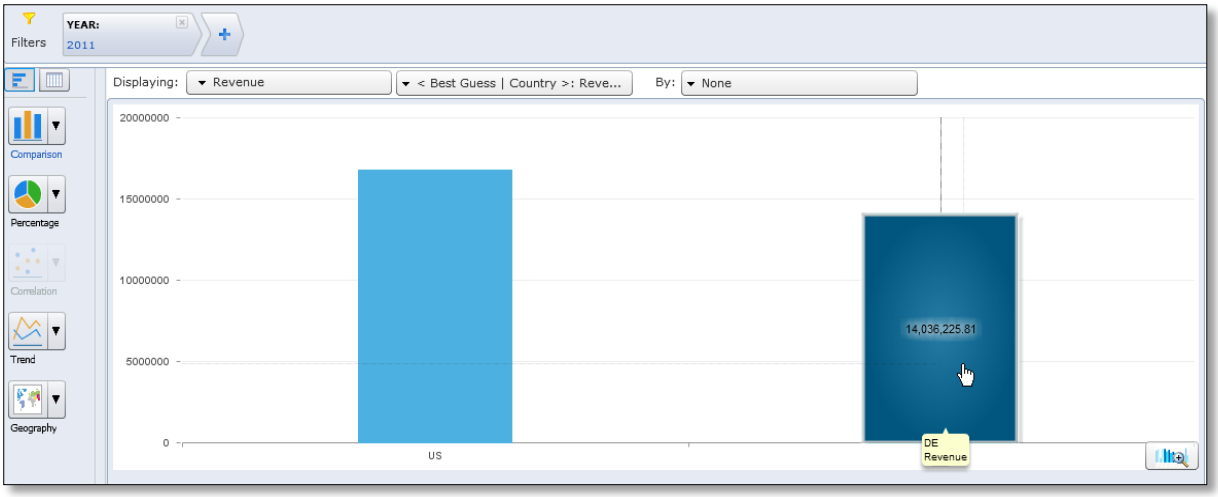
The insights generated from the analysis is described by answering following research questions derived from the respective case study from Brightspace (Kalé, Jones, Bliemel, & Lee, 2018):

1.3.1 What is the revenue of Germany?

Using a bar chart, the revenues of the Germany and US are compared over all the years from 2007 to 2011. The graph clearly indicated a slightly higher revenue for the US division of Global Bike Inc. The aggregated revenue of Germany is 56,101,659 USD.

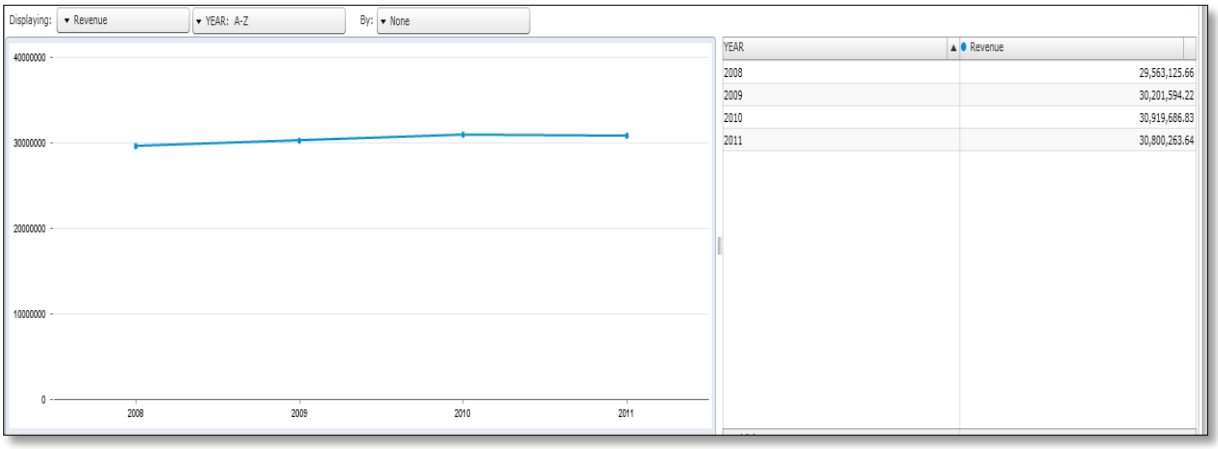


In the event of looking at the recent (2011) sales revenue across both US and Germany, it is easy to add filter on the Year attribute of the given sales data. The chart illustrating the comparison of revenue in 2011 between US and Germany is shown below, saying a higher sales revenue in the US. The revenue of Germany in 2011 is 14,036,225USD.

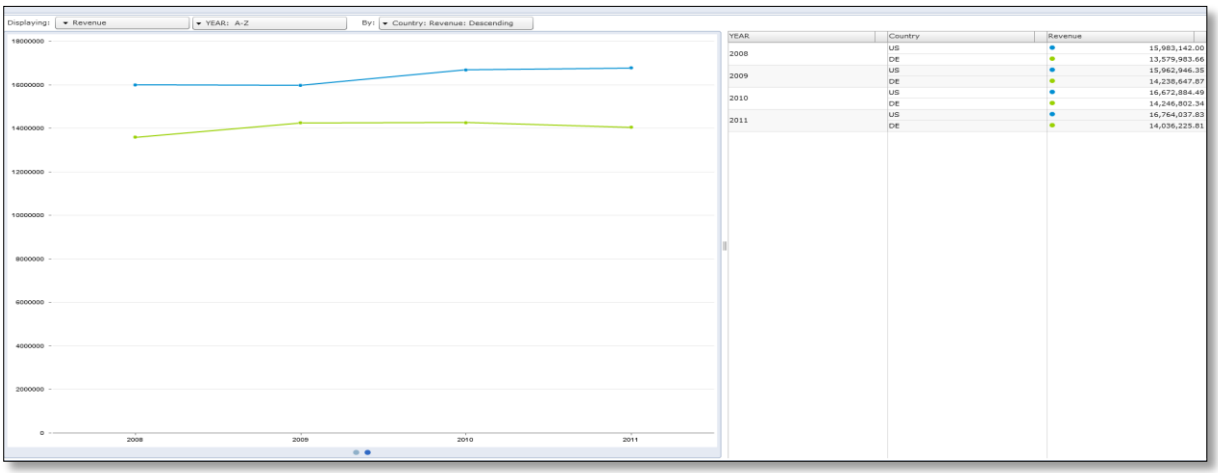


1.3.2 Illustrate the revenue trends over years?

The overall sales revenue for the Global Bike Inc. has slightly increased from 2008 to 2010, however, stayed almost the same as previous year for 2011 which is evident from the following line chart depicting the trends of revenue over time from 2008 to 2011.



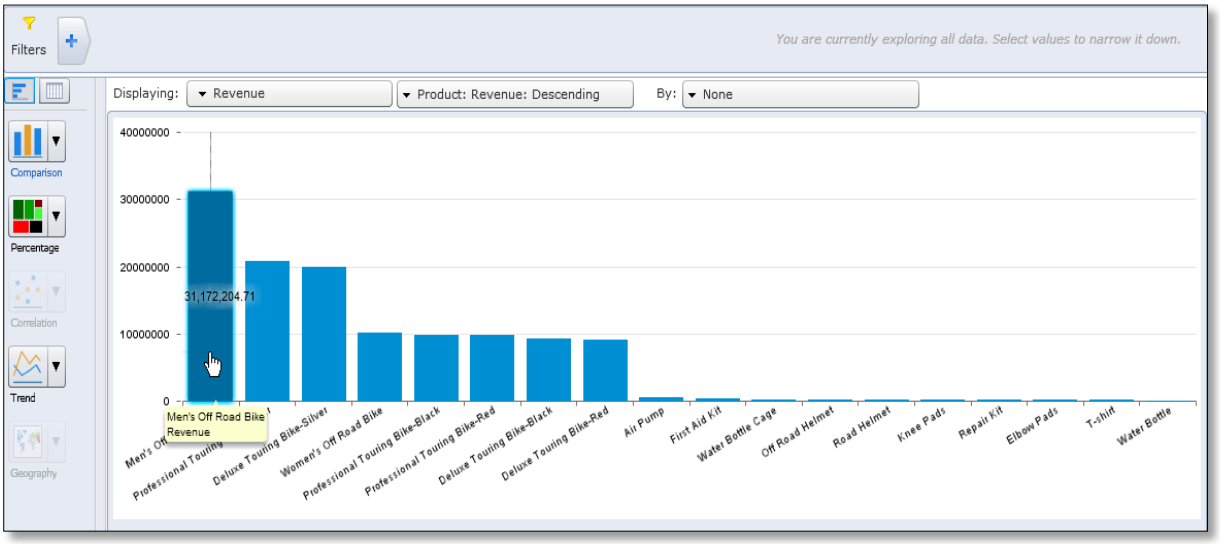
To get a better understanding and to formulate a customized strategy, the sales trends across each divisional area; US and Germany, should be analyzed. The distinct revenue trends for US and Germany are shown below.



1.3.3 What product has the highest revenue?

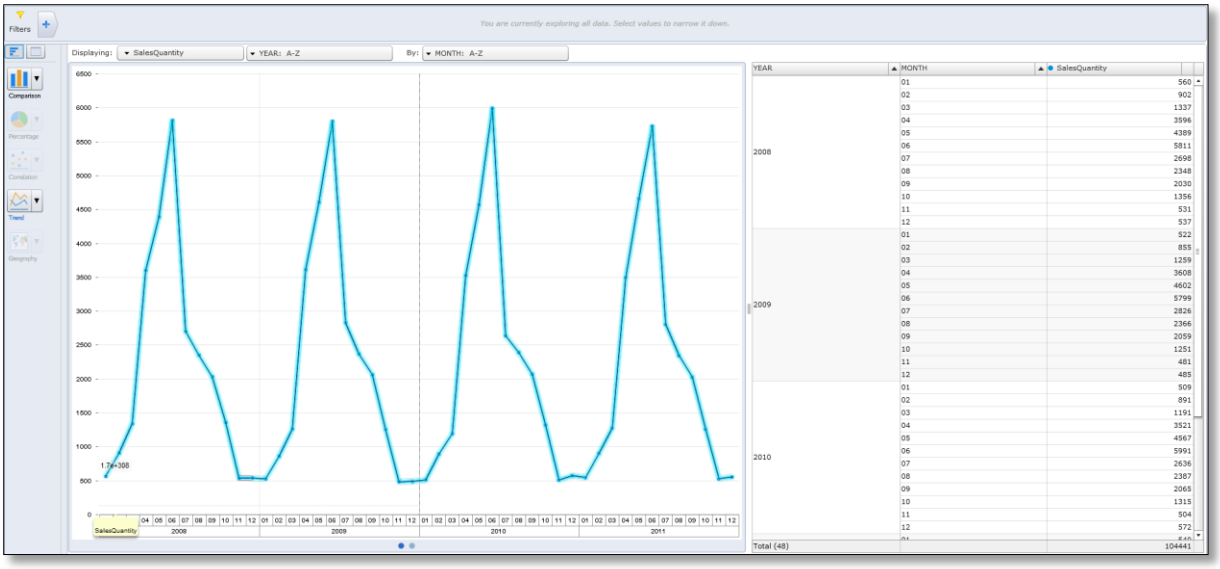
Men's Off Road Bike is the product that has the highest revenue so far among all the other product sold by Global Bike Inc. This is effectively visualized using the Tag cloud chart in SAP Business Object explorer.

One can also use the comparison option available in the SAP Business Object Explorer to make a visual comparing the revenues of different products. The visual shown below is sorted by revenues in the descending order. This can help the user to know the actual revenue generated by a product by hovering on to the corresponding bar associated with each product. Men's Off Road Bike has a revenue of $31,172,204 followed by Professional Touring Bike – Silver.



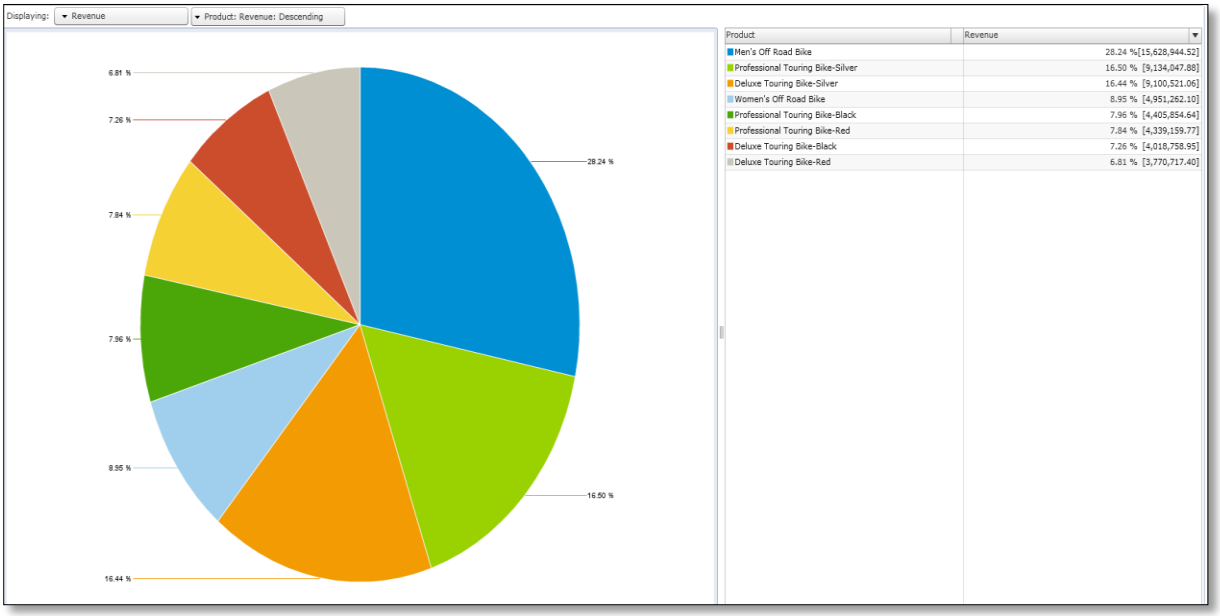
1.3.4 Is there any seasonality in Sales quantity? Which month has the highest sales quantity?

Dividing the revenue trends overs years by monthly granularity, seasonality in the sales quantity of the products of Global Bike Inc. is noticed. And it is visually presented using the trend option and further tinkering by month which is available in the SAP Business Object Explorer. Global Bike Inc. has been experiencing highest sales by quantity on every June (6 in the X axis of the visual). Here is the visual telling the existing seasonality in sales and the month with highest sales quantity.



1.3.5 What is the highest percentage of revenue provided by a bike in Germany?

The percentage section of the SAP Business Object Explorer can help develop calculated percentages based on the available data without having the actual percentage values in the data. Men's Off Road Bike has highest percentage revenue of 28%. The Pie chart depicting the percentages of revenues generated by various products under the category of bikes is further sorted by the percentage values and is shown below.



1.4 Conclusion:

In my experience working with SAP Business Object Explorer, it is a nice tool to generate visualizations and to explore data. However, this tool might be lacking a wide variety of visualization chart options as it is available with other latest BI tools and since it is a stand-alone application, its performance is highly dependent on the computer configurations. It requires some pre-knowledge on the usage and report generation indicating it is moderately user friendly.

Chapter 2: Microsoft Excel Pivot Tables

2.1 Tool Description and Usage:

One of the Microsoft Excel's features that helps to summarize and organize huge dataset into informative report is **Pivot Table**. Pivot table is a data aggregation tool that used to summarize, sort, group, count, total, average and so on in the context of data preprocessing (Tremblay, 2019). It is a structure that allows users to change the values and fields based on their requirement in just drag and drop manner. It has capabilities to create calculated columns and other features that make pivot table a more meaningful Business Intelligence. It is specifically used in a setting where the users are more familiar with Microsoft Suite of products.

2.2 Data Set Exploration:

The transactional (sales) data set from 2011 to 2014 of The Global Bike Inc., GBI_E5_2.xlsx, which is similar to data in previous chapter in structure, is downloaded from the Brightspace. After looking at the data, it appears to be consistent and no data preprocessing is necessary. The insights related to the profitability analysis of the wholesale division is generated by using **Microsoft Excel Pivot Tables** and are submitted to strategic planning team to make further decisions.

2.3 Analytics and Insights:

The insights generated from the analysis is described by answering following research questions derived from the respective case study from Brightspace (Kalé, & Jones, 2016):

2.3.1 Which year has the highest revenue ?

2014 has the highest sales revenue and the revenue was 142,498,014USD. Though the revenue had been constantly increasing from 2007, a declining revenue was witnessed after 2014. So, the company must rethink about its business strategy to improve revenue and profitability of its products.

Row Label	Sum of Revenue USD
2007	115,993,582
2008	118,763,940
2009	101,326,904
2010	109,622,920
2011	136,951,338
2012	138,086,539
2013	136,186,869
2014	142,498,014
2015	131,866,087
2016	135,151,873
Grand Total	1266448065

2.3.2 Which sales organization has the highest revenue?

East division of United States has the highest revenue of 47,179,561USD in the year of 2014. The West division of the United States has the overwhelmingly least revenue. So, the company should stay relatively more focused on the West Division of US sales by implementing good marketing methodologies, improving the number of customers by building loyalty, increasing brand value, executing appropriate pricing strategies, increasing the frequency of customer transactions and efficiently managing the resources.

YEAR	2014
Row Labels	Sum of Revenue USD
DN00	43,418,543
DS00	35,716,759
UE00	47,179,561
UW00	16,183,152
Grand Total	142,498,014

2.3.3 Compare sales discounts by country for all years. Which country has a higher sales discount as a percentage of sales?

The comparison of sales discounts by country over years is shown in the following pivot table and US has the highest sales discount (3.4%) as percentage of sales as compared to Germany which is 2.78%. US is the country with highest discount as a percentage of sales with accumulated discount amount of $19,259,464.

Row Labels	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016 Grand Total
DE	1,615,304	1,687,189	1,676,939	1,732,134	2,124,286	2,009,939	2,026,183	2,193,756	2,131,543	2,283,830	19,481,103
US	1,985,833	1,976,271	1,376,944	1,612,339	2,171,286	2,345,762	2,174,085	2,019,078	1,801,718	1,796,147	19,259,464
Grand Total	3601137.52	3663459.93	3053883.09	3344473.1	4295572.13	4355700.94	4200267.9	4212833.34	3933260.67	4079977.77	38740566.39

Row Labels	Average of % Discount
DE	2.78%
US	3.40%
Grand Total	3.06%

2.3.4 In 2014, what are the products with highest and least sales quantity?

In the year 2014, Air Pump is the highest sold product with sales quantity of 8,001 items whereas Edel Fixie is the least sold product with sales quantity of 177 units.

YEAR	2014
Row Labels	Sum of SalesQuantity
Air Pump	8,001
City Max	265
Deluxe Touring Bike-Black	2,751
Deluxe Touring Bike-Silver	4,952
E-Bike Tailwind	1,732
Edel Fixie	177
Elbow Pads	683
First Aid Kit	3,215
Knee Pads	776
Men's Off Road Bike Fully	4,623
Men's Off Road Bike Hard Tail Shimano	3,462
Men's Off Road Bike Hard Tail SRAM	4,933
Off Road Helmet	1,646
Professional Touring Bike-Black	2,506
Professional Touring Bike-Silver	5,328
Repair Kit	1,576
Road Bike Alu Shimano	4,961
Road Bike Alu SRAM	1,652
Road Bike Carbon Campagnolo	962
Road Bike Carbon Shimano	3,473
Road Bike Carbon SRAM	851
Road Helmet	1,540
T-shirt	1,584
Water Bottle	1,754
Water Bottle Cage	4,769
Women's Off Road Bike Fully	2,368
Grand Total	70,540

2.3.5 Who are five customers with lowest sales revenue overall?

Top 10 filter option is used to identify the customers with lowest sales. The five customers with lowest sales are Drahtesel, Furniture City Bikes, Ostseerad, Peachtree Bikes and Velodrom. Drahtesel has the least revenue of $40,317,175. The total revenues of these five customers are shown in the pivot table.

Row Labels	Sum of Revenue USD
Drahtesel	40317175
Furniture City Bikes	22032176
Ostseerad	26295360
Peachtree Bikes	39726335
Velodrom	34595893
Grand Total	**162966939**

Overall, the lowest four customers' sales had been increasing until 2014 and then started to decline in next years until 2016, which is presented by the conditional formatting of the following pivot table. Whereas fifth customer Velodrom has fluctuating sales revenue with a noticeable high revenue in 2013, then a decline until 2015 and increased sales in 2016.

Sum of Revenue USD	Column Labels										
Row Labels	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Grand Total
Drahtesel	2954269	3826426	3758994	3370627	4812623	3646574	4142643	5441225	4627623	3736172	40317175
Furniture City Bikes	1802090	1817196	1468034	1894074	1939877	2542848	2450742	2821124	2963748	2332442	22032176
Ostseerad	2210931	2349489	2738394	2283083	2330322	2852053	3119253	3424821	2433935	2553081	26295360
Peachtree Bikes	3686400	3292956	3203897	2941721	3980581	4173886	3472115	6211319	4645601	4117859	39726335
Velodrom	3633596	3469676	3577585	2995986	3044015	3250551	4223544	3810799	2986538	3603604	95893
Grand Total	14287285	14755744	14746902	13485492	16107419	16465913	17408297	21709287	17657444	16343158	162966939

2.4 Conclusion:

Pivot table is used to collect the summary from any given set of data, also useful in consolidating large quantity of data set present in Microsoft Excel. It helps in drawing conclusions very quickly from a given set of data. Pivot Tables consists of rows, columns and data fields which can be moved around and assists in summarizing, expanding and grouping the data. It can help any business or organization in finding any recurring patterns in the data. This in turn helps in forecasting precisely. In this fast-paced modern world, decision making must be very quick especially for business leaders and managers, pivot tables come to rescue for the managers and leaders to aid in fast, reliable and credible decision making. It appears to be providing more numeric information which should be further visualized using other chart options of Microsoft Excel. From my working with pivot tables, they are not visually pleasing and challenging for the business audience to consume the information faster. Pivot tables require some prior training on their creation, manipulation and usage.

Chapter 3: SAP Lumira Discovery

3.1 Tool Description and Usage:

SAP Lumira Discovery is an on-premise, self-service BI tool and all-in-one platform that is simultaneously useful in data discovery, visualizing, and analyzing the underlying dataset using interactive dashboards and analytical application (BI, 2018). It is a more recent tool that has covered the drawbacks of SAP BOE's performance. It is mainly used where the acquisition, preparation, visualization, exploration, and collaboration have to be done efficiently using the same platform.

3.2 Data Set Exploration:

The task of understanding the alcohol sales across the various region of Canada started with downloading the publicly available data: population.xlsx with columns: Year, Location, Sex, Age Range, and Population Count, showing the population distribution across provinces, and sales.xlsx with columns: Year, Location, Type of Beverages, Origin of Beverage, and Sales Value, showing the sales of various alcoholic beverages originated from Canada or imported across all the provinces of Canada. Most of the publicly available data is not consistent and it needs to be cleaned/prepared to further analyze the question in hand.

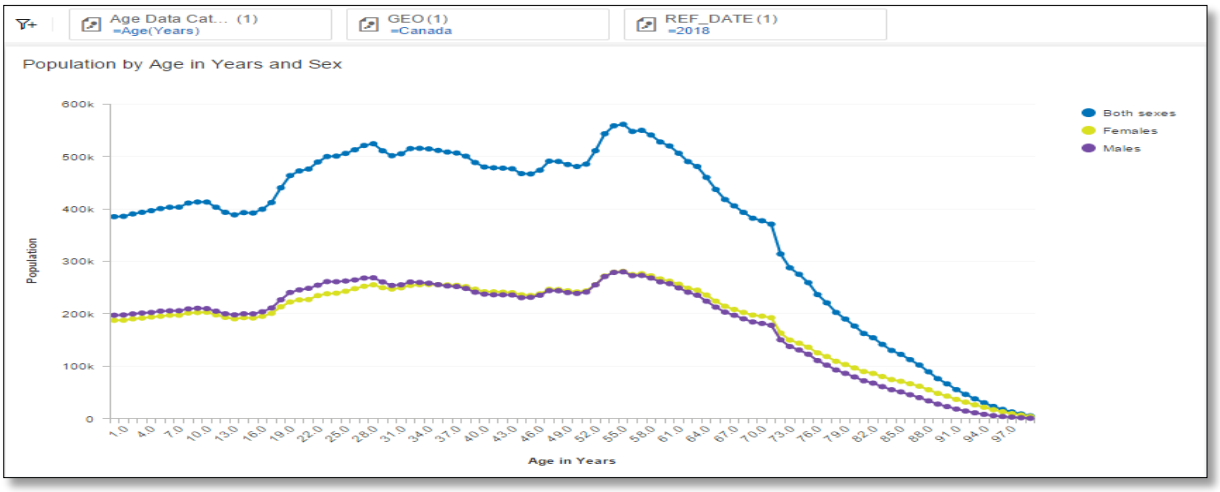
The required columns selected from the extracted data is correctly categorized into measures and dimensions. Appropriate naming is given, grouped data and new dimensions are created if necessary. For example, After the Age column with age in years and age ranges is further grouped accordingly to ease the analysis, now age in years is added as new dimension. This dataset is later merged with alcohol sales data to analyze the purchasing preferences of population.

3.3 Analytics and Insights:

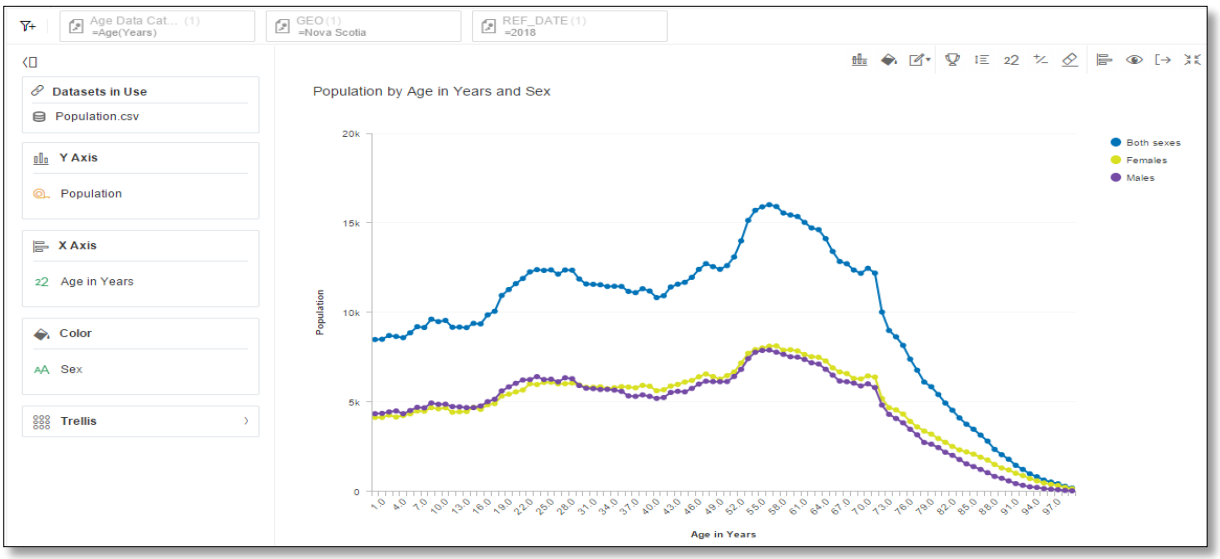
The insights generated from the analysis is described by answering following research questions derived from the respective case study from Brightspace (Bliemel, & Lee, 2018):

3.3.1 What is the population distribution of Canada?

The population count of Canada based on their age in 2018, which is further categorized into males, females and both sexes, is clearly depicted in the following line chart. Majority of population in Canada are in the age range of (53-60) years in 2018.

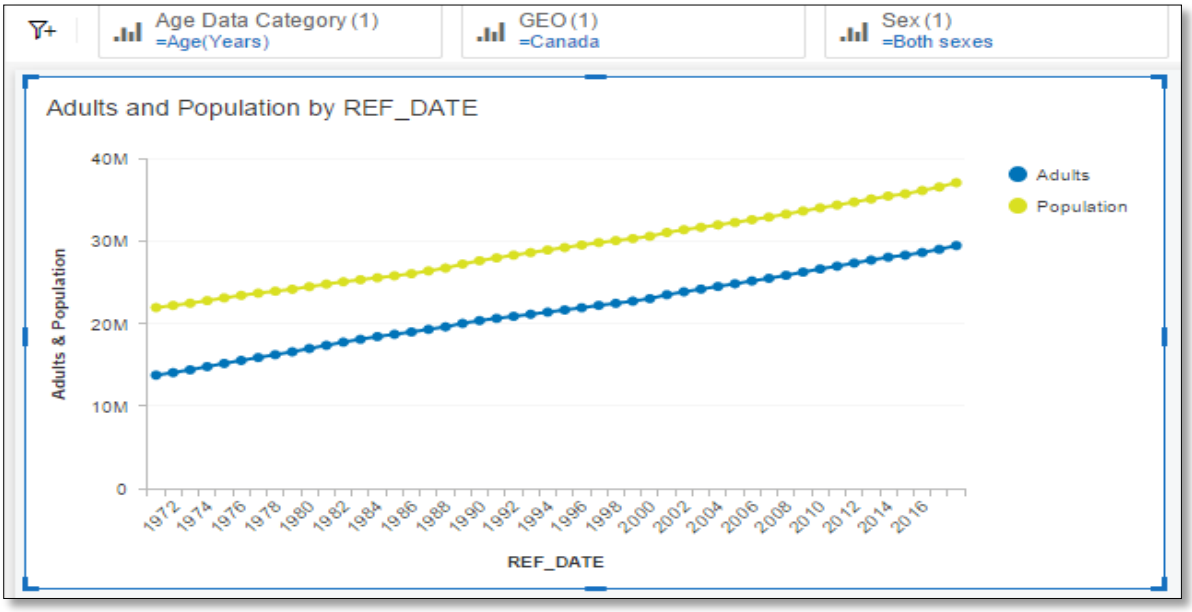


To realize the curiosity of checking the population distribution of Nova Scotia in same year as 2018, a filter on the GEO column is executed in SAP Lumira Discovery. Majority of population in Nova Scotia are in the age range of (53-65) years.

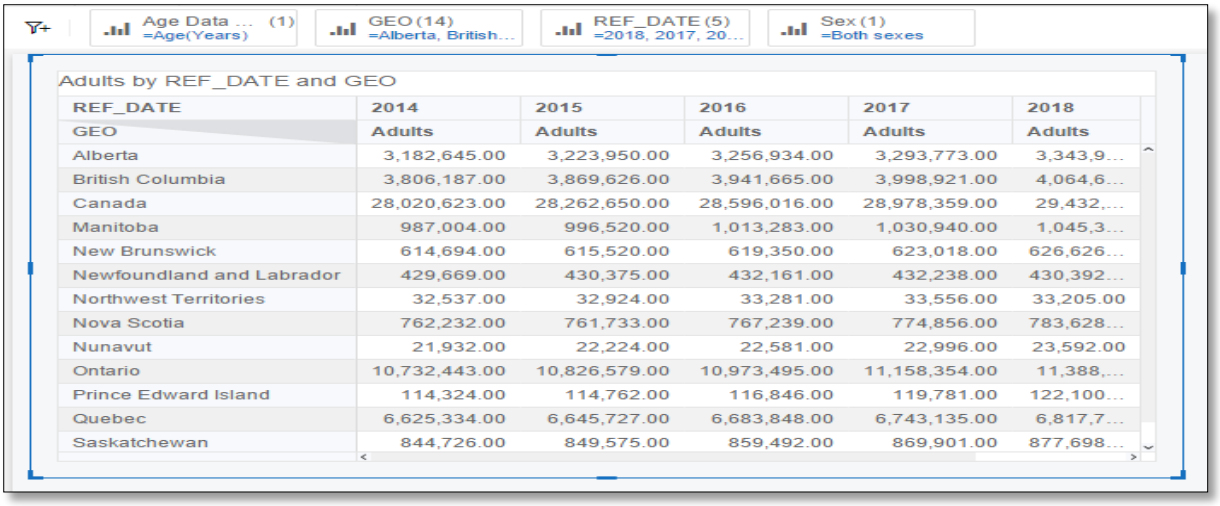


3.3.2 Depict adult population with eligibility to purchase alcohol? Compare the same for each province?

This line chart shows the number of adults in the population who are legally eligible (>=18 years in Alberta, Quebec and Manitoba and >18 years for rest) to purchase alcohol as it is necessary to answer the business question on understanding the alcohol preferences of people in Canada from 1972 to 2018.



Now the eligible population across all the Canadian regions are presented in the following cross table which is then exported as Adults.xlsx and merged with alcohol sales data to answer our business question.

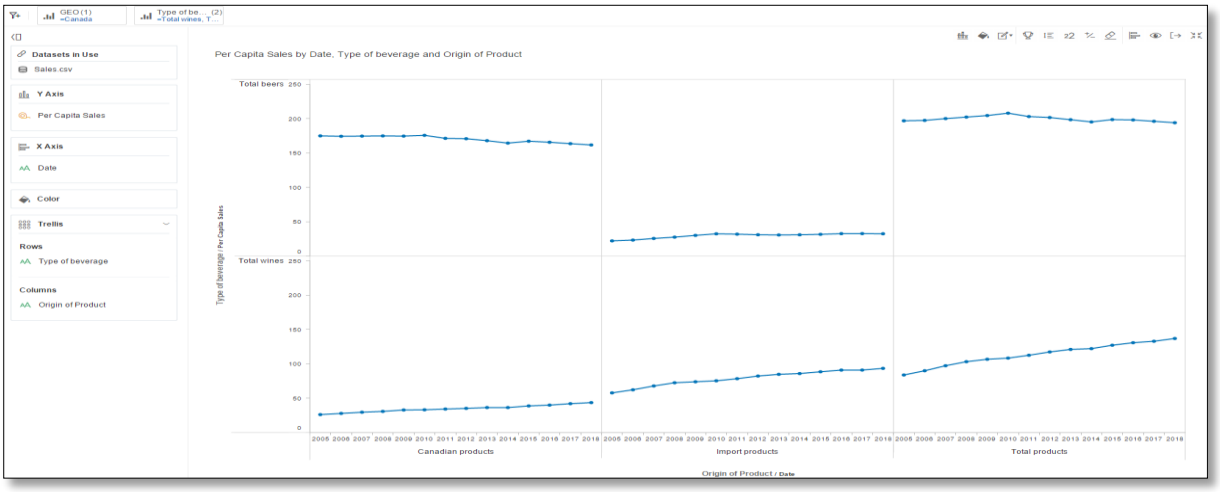


Adults by REF_DATE and GEO

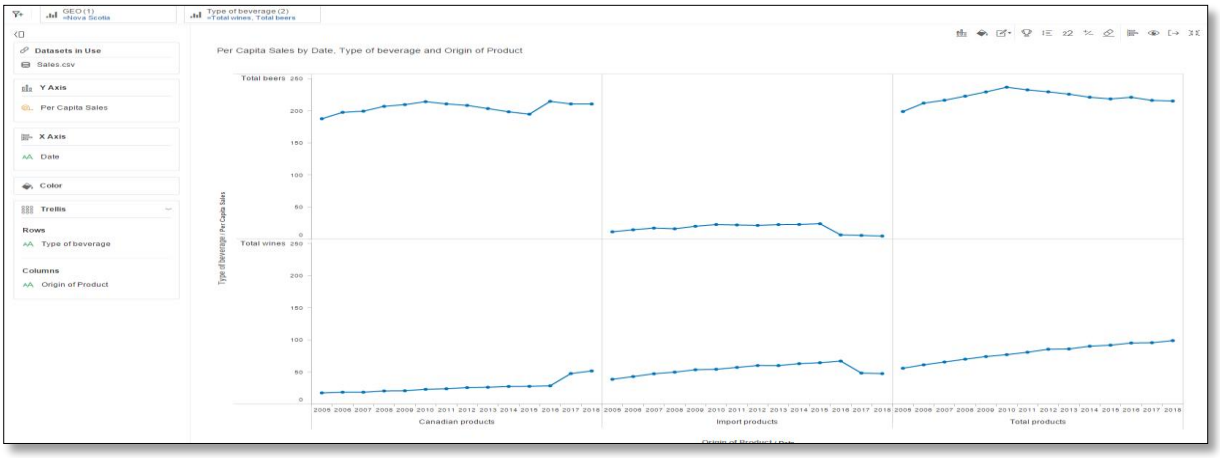
REF_DATE	2014	2015	2016	2017	2018
GEO	Adults	Adults	Adults	Adults	Adults
Alberta	3,182,645.00	3,223,950.00	3,256,934.00	3,293,773.00	3,343,9...
British Columbia	3,806,187.00	3,869,626.00	3,941,665.00	3,998,921.00	4,064,6...
Canada	28,020,623.00	28,262,650.00	28,596,016.00	28,978,359.00	29,432,...
Manitoba	987,004.00	996,520.00	1,013,283.00	1,030,940.00	1,045,3...
New Brunswick	614,694.00	615,520.00	619,350.00	623,018.00	626,626...
Newfoundland and Labrador	429,669.00	430,375.00	432,161.00	432,238.00	430,392...
Northwest Territories	32,537.00	32,924.00	33,281.00	33,556.00	33,205.00
Nova Scotia	762,232.00	761,733.00	767,239.00	774,856.00	783,628...
Nunavut	21,932.00	22,224.00	22,581.00	22,996.00	23,592.00
Ontario	10,732,443.00	10,826,579.00	10,973,495.00	11,158,354.00	11,388,...
Prince Edward Island	114,324.00	114,762.00	116,846.00	119,781.00	122,100...
Quebec	6,625,334.00	6,645,727.00	6,683,848.00	6,743,135.00	6,817,7...
Saskatchewan	844,726.00	849,575.00	859,492.00	869,901.00	877,698...

3.3.3 How alcohol purchasing preferences vary for beers and wines?

The Sales.xlsx and Adults.xlsx are merged using left outer join. The per-capita sales of total beers and total wines in Canada from 2005 to 2018 by the origin of product (Canada, Imported or Total) is shown in the following figure.



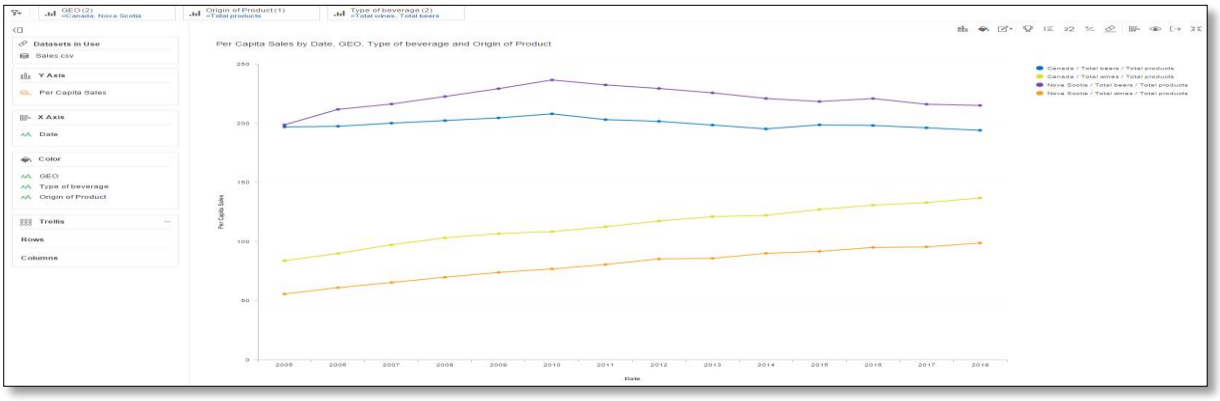
The following diagram represent the distribution per-capita sales by origin of product from 2005 to 2018 specific to the province of Nova Scotia.



The major insight from the above visualization is that Canadians are consuming more wines that are imported than home-made; whereas they are consuming less imported beers than home-made beers. Overall, the analysis shows they consume more beers than wines.

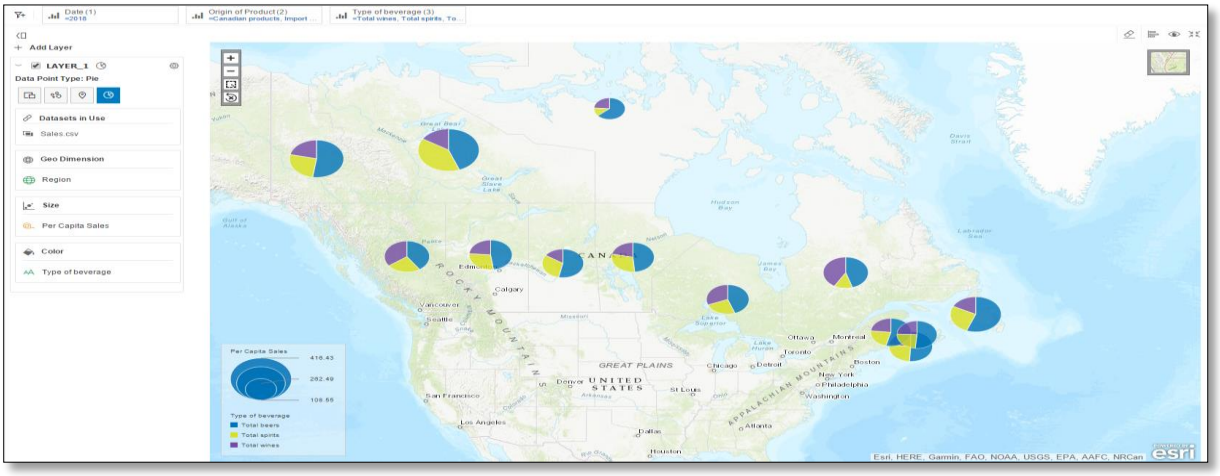
3.3.4 Compare per capita sales of wines and beers?

The per-capita sales of total wines and total beers are compared and further compared between Nova Scotia and Canada, which is shown in the following line chart showing a higher consumption of beers than that of wines.



3.3.5 Geographically represent per capita sales of beers, wines and spirits?

SAP Lumira Discovery has the functionality to tie the geographical data with working dataset and then visualizing using geo-map chart. The per-capita sales of total beers, total spirits and total wines across all the provinces of Canada is appealingly visualized in the following geo-map. The size of the pie sections helps to compares the per-capita sales on beers, wines and spirits while also being consistent with the location.



3.4 Conclusion:

SAP Lumira Discovery appears to be an intuitive tool that helps to analyze the complex business problems in a way without much IT professional involvement. It is easy to use for anyone and prior knowledge is not necessary. The performance of the SAP Lumira Discovery can be further improved. The additional options of the certain charts can be implemented. It is best known for its visualization features.

Chapter 4: SAP Analytics Cloud

4.1 Tool Description and Usage:

SAP Analytics Cloud is an integrated analytics tool built on the reliable and high-performance cloud which makes it accessible through web browser ("SAP Analytics Cloud: End-to-end," n.d.). This tool is accessible from different locations with any browser installed, preferably with Google Chrome supporting a mobile workforce. SAP Analytics Cloud supports a strong and fast decision making using its fast information processing system. The data from multiple sources can be easily merged and utilized to form insights. The features of this tool can be further extended by using some API and applications in its ecosystem.

4.2 Data Set Exploration:

The data from the ERPSIM game is collected which shows the sales and revenue generated by playing the virtual business game in terms of teams. The ERPSIM excel data has 10 columns consisting of details related to sales orders of Muesli cereal, including the team sold, area of operation, on which round and day, quantity sold and through which distribution channel, the product, its price and the revenue generated from that sale. The data set is explored and it is clean so that no further data preprocessing is required.

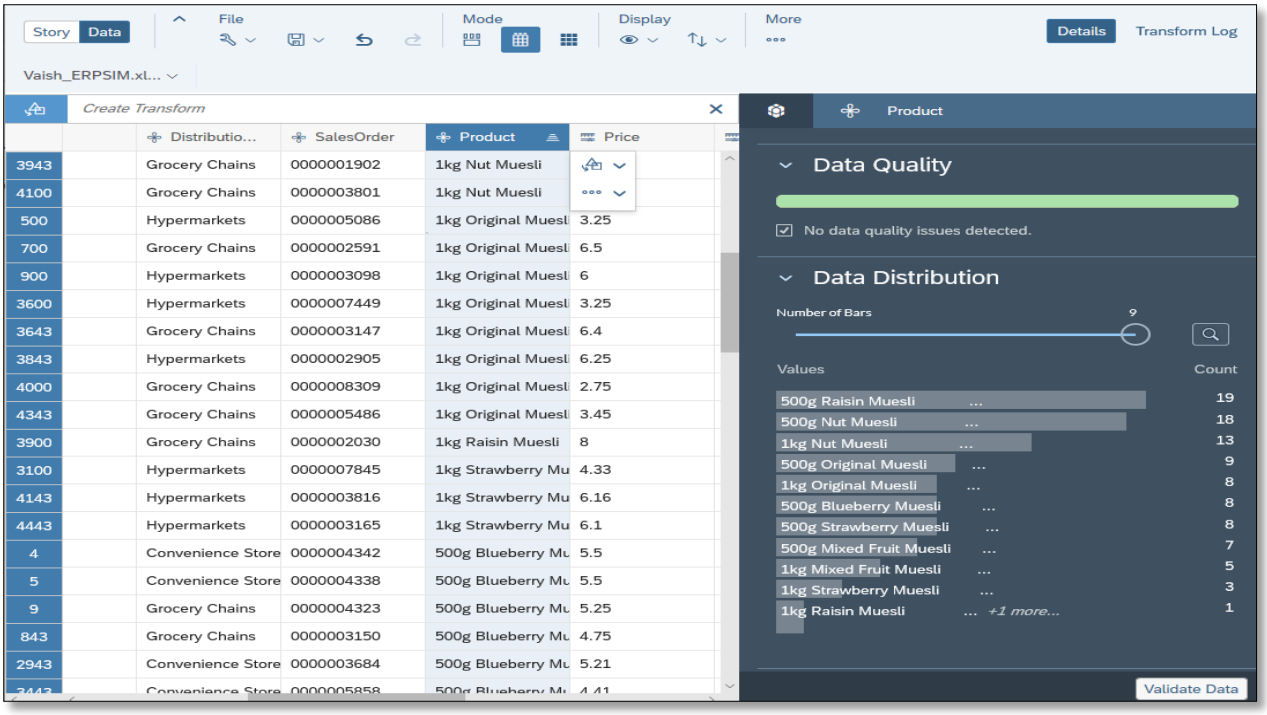
4.3 Analytics and Insights:

The insights generated from the analysis is described by answering following research questions derived from the respective case study from Brightspace (Kalé, Jones, & Lee, n.d.):

4.3.1 Which Distribution Channel and what are the top 3 products in terms of the number of sales orders?

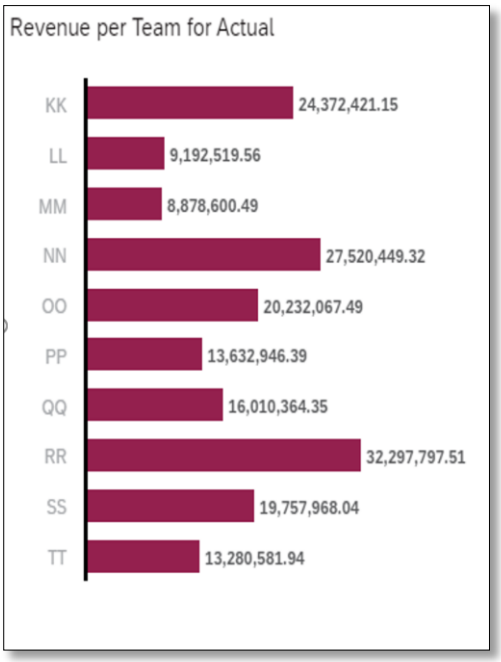
Convenience stores is the distribution channel that has the highest number of sales which is shown below which is not part of the story created from the ERPSIM data, but it is the Data Distribution section of the properties panel in the SAP Analytics Cloud platform. An initial analysis along with some visuals underneath the data distribution is shown in the properties panel.

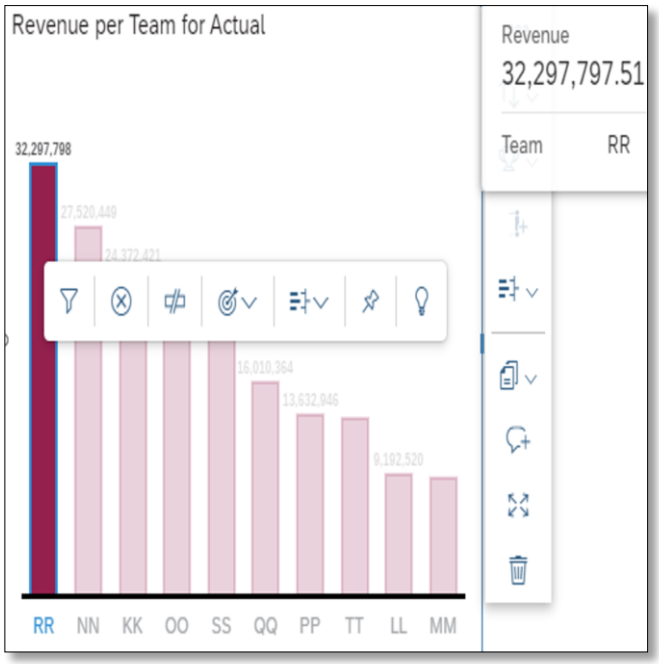
500g Raisin Muesli, 500g Nut Muesli and 1kg Nut Muesli are the top three products with frequent sales orders.



4.3.2 Which team had the highest revenue? What was the revenue?

Team RR had the highest revenue across all the rounds and all the days. The overall revenue of team RR was 32,297,797 USD. The same visual can be shown in descending way by sorting by revenue which is relatively more informative and provide quick solution to the underlying question here. This tool provides the feature to click on a certain bar and highlight the numerical details which is the revenue amount here.

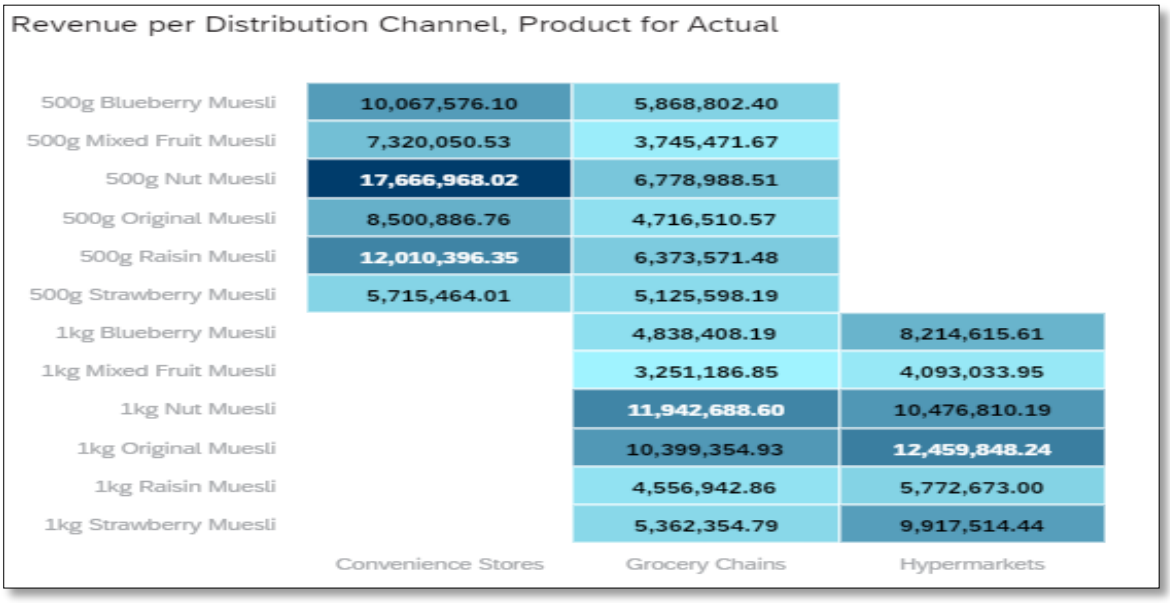


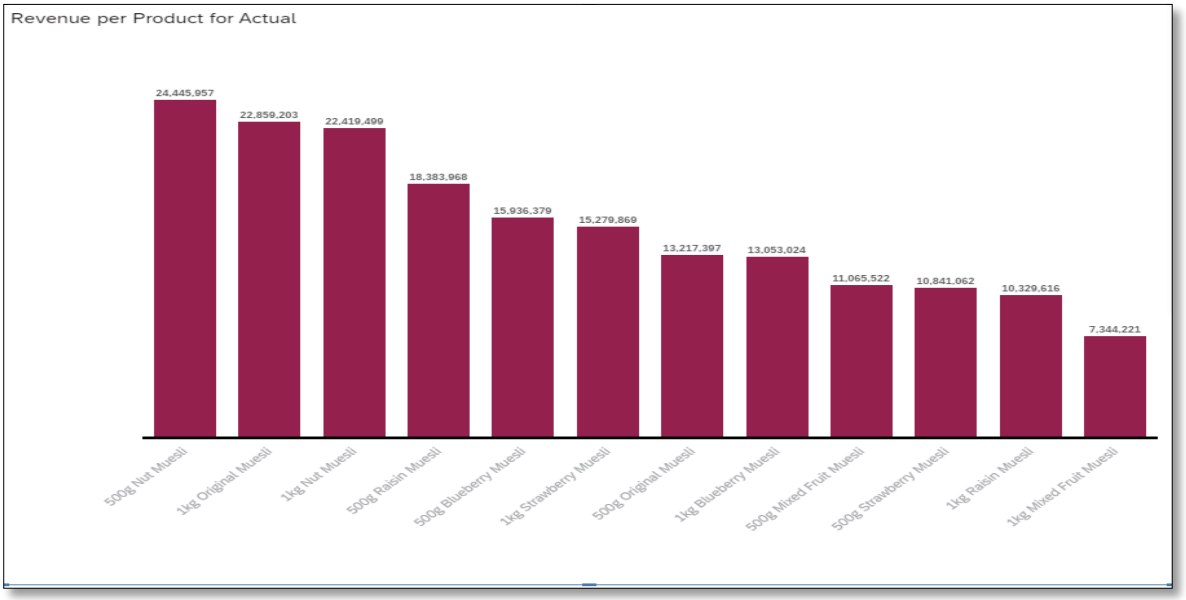


4.3.3 Are there any products that do not sell in specific distribution channels? Which product had the highest revenue?

Yes, there are few products that did not sell in few distribution channels which is clearly shown in the picture below. All the 1kg products (1kg Blueberry Muesli, 1kg Mixed Fruit Muesli, 1kg Nut Muesli, 1kg Original Muesli, 1kg Raisin Muesli, 1kg Strawberry Muesli) are not sold in the Convenience Stores. All the 500g products (500g Blueberry Muesli, 500g Mixed Fruit Muesli, 500g Nut Muesli, 500g Original Muesli, 500g Raisin Muesli, 500g Strawberry Muesli) are not sold in Hypermarkets. However, Grocery Chains are selling both 500g and 1kg Muesli products.

This heatmap also gives more information on the most selling and least selling products at different distribution channels which is shown in tune with the color density. For example, 500g Nut Muesli had brought in the highest revenue so far which is conveyed by both bar chart and the heatmap below.



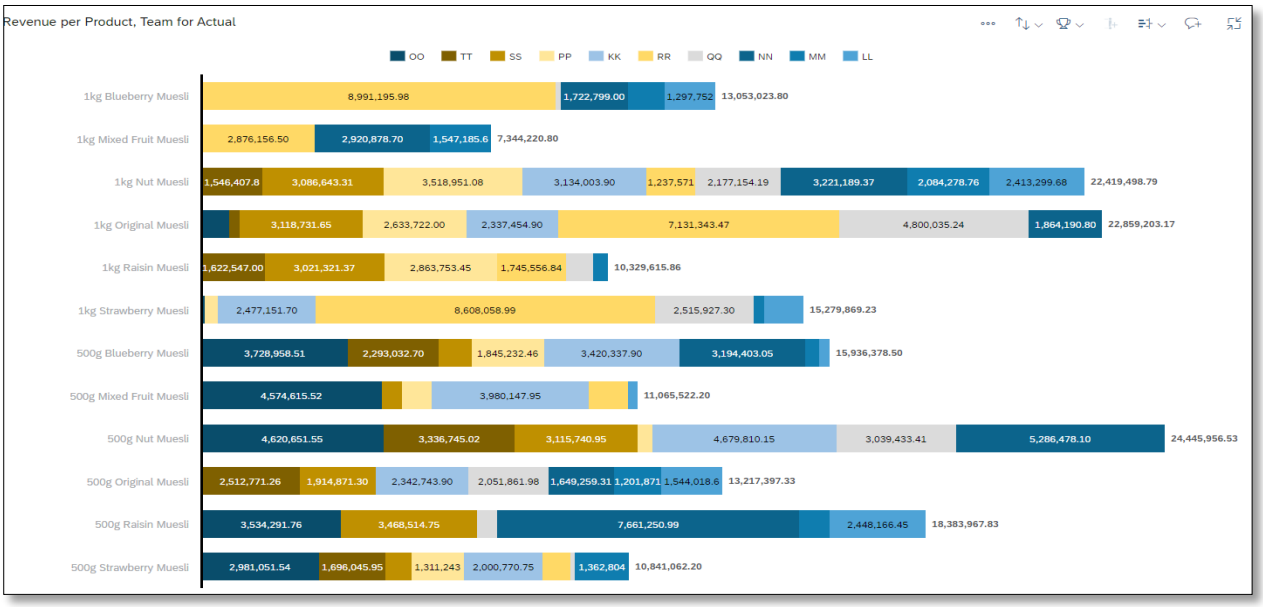


4.3.4 What was the highest priced muesli on the day 18 of round 1? What team sold it? What was the price & sales order number?

Team PP sold 1kg Raisin Muesli at 7USD which is the highest priced muesli in the day 18 of round 1. The sales order number is 0000002050. The following bar chart describes the team, product and corresponding sales order number along with the product and the corresponding prices the certain product is sold for. It is easy to determine secondary information along with the main solution.

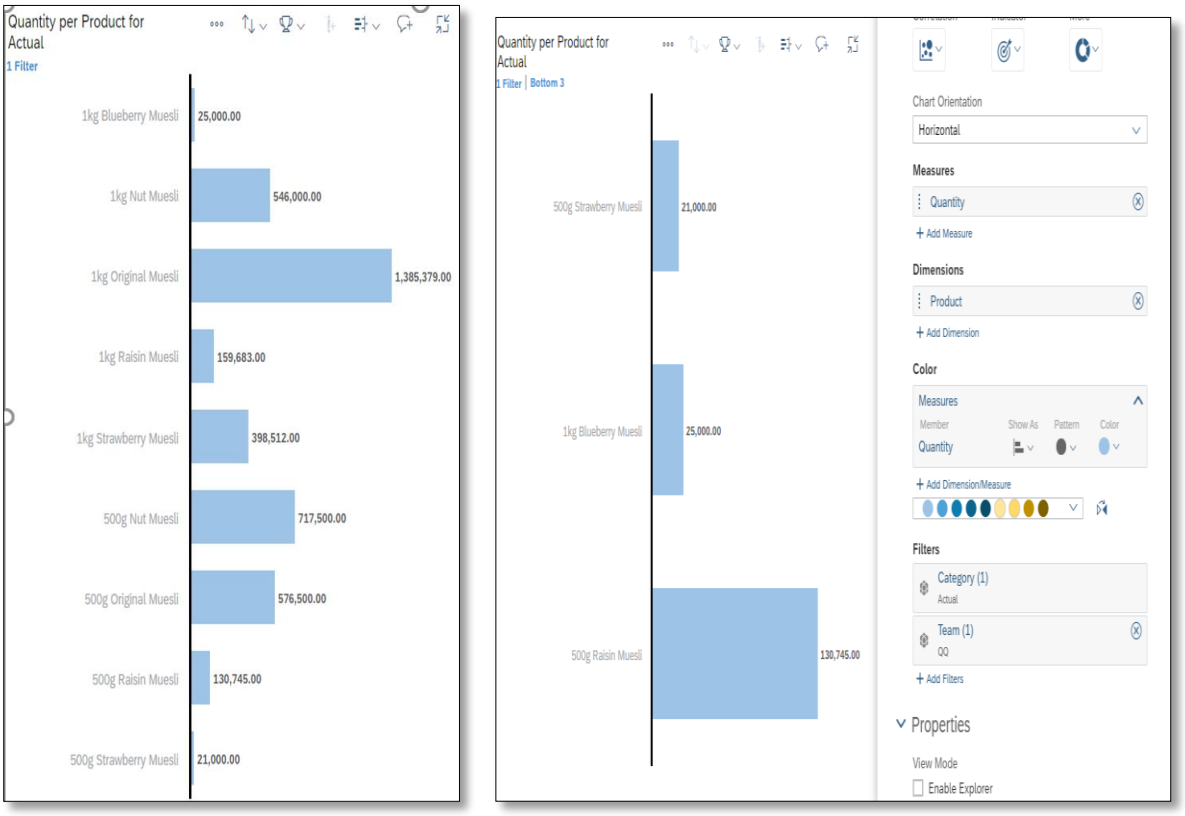
4.3.5 Show the market share (in terms of revenue) of each team by product?

The following horizontal stacked bar chart shows the market share in terms of revenue of each team by product. This clearly explains which product is the cash cow for each team and which team is not reasonably generating revenue by a certain product. For example, team RR accounted for the major portion of revenue generated by 1kg Blueberry Muesli.



4.3.6 Please identify three products that generate the lowest quantity (i.e., bottom 3 quantity) by the team QQ.

By using the available filter option on the dimension Team let the team QQ is chosen. The first chart shows all the products sold by the team QQ in terms of the quantity sold. However, the goal here is to find the subset of products which is three products that generate the lowest quantity. So using the Top N feature to further filter helps to identify the bottom 3 products by applying the condition on the measure quantity sold. The second chart shows only the bottom 3 products with lowest number of quantities sold.



4.4 Conclusion:

In my experience working with SAP Analytics Cloud, it is a nice tool to generate visualizations and to process data quickly supporting fast decision making. However, being a cloud-based tool, any down time would dramatically affect the company in terms of financial performance, in the critical times. Its smart-assist capabilities along with live data connectivity makes it a more appealing tool ("SAP Analytics Cloud for newbies," n.d.). It might have some wait time after creating an SAP analytics cloud account to get it activated and to use the tool. It is more interesting to be able to create dynamic stories through SAP Analytics Cloud application.

Chapter 5: Bex Query Designer and SAP Business Object Analysis for Excel

5.1 Tool Description and Usage:

Business intelligence should not only able to be derived from a two-dimensional data but also from a multidimensional data sources with star or snowflake schemas, efficiently by building queries on data cubes, which is the SAP BW Infocube. The SAP Business Object Analysis module can, then, be plugged into Microsoft Excel to analyze the results from the query which is designed using **Bex Query Designer**. Multidimensional analysis for the Online Analytical Processing (OLAP) data sources can be well performed using Bex Query Designer and SAP Business Object Analysis for Excel.

5.2 Data Set Exploration:

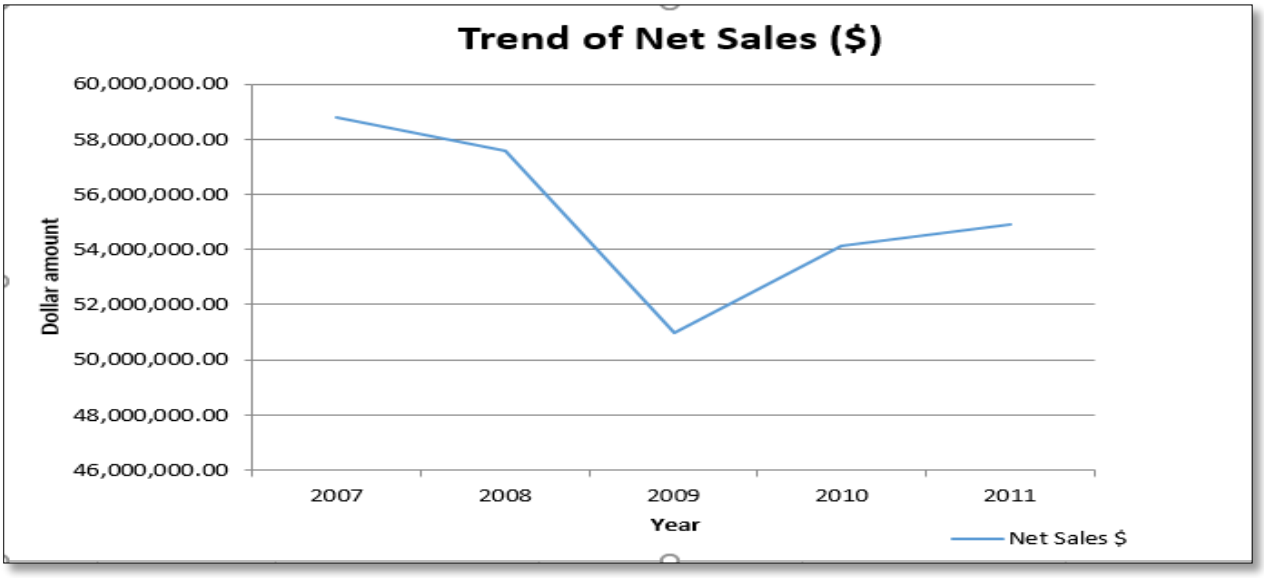
The data warehouse of Global Bike Inc. is explored which contains the sales data from 2007 to 2011. In contrary to traditional relational databases, data warehouses are associated with data cubes with facts and dimensions are related through star or snowflake schema. The InfoProvider (Infocube) for GBI Inc. called GBI Infocube is accessed and a query is executed in it. The GBI Infocube of historical sales data is readily available in the SAP Netweaver Business Warehouse (BW) and no further data processing is necessary.

5.3 Analytics and Insights:

The insights generated from the analysis is described by answering following research questions derived from the respective case study from Brightspace (Kalé, Jones, Bliemel, & Lee, 2016):

5.3.1 Show the trend of Net Sales over years?

Overall, the net sales of the GBI is decreasing from 2007 to 2011. It went dramaically down during 2009, which might be due to the financial recession at the end of 2008. However, the net sales started to increase form 2009, perhaps not at the same rate as desired, but showing the recouping from the financial distress which is a positive sign. The company must consider formulating strategies to increase the quanitities sold by appropriately marketing there by gaining a major market share.



5.3.2 What is the overall Revenue in USD?

The total revenue for the Global Bike from the sale of all variety of products is 285,250,874USD. Just by dragging revenue from data source to columns component of design panel and material to rows component of design panel, the revenue generated from the sale of each product with a grand total is displayed in the crosstab.

	Revenue
Material	$
Air Pump	698,167.68
City Bike Max	587,920.44
Deluxe Road Bike (Shimano)	24,624,777.19
Deluxe Road Bike (SRAM)	7,097,986.00
Deluxe Touring Bike (black)	12,558,803.79
Deluxe Touring Bike (red)	12,856,625.33
Deluxe Touring Bike (silver)	27,170,368.38
E-Bike Tailwind	4,023,195.23
Elbow Pads	134,381.12
First Aid Kit	341,840.35
Fixed Gear Bike Plus	93,723.34
Knee Pads	134,799.28
Men's Off Road Bike Fully	30,613,418.25
Men's Off Road Bike Hard Tail (Shiman	14,819,077.64
Men's Off Road Bike Hard Tail (SRAM)	24,602,305.16
Off Road Helmet	213,027.14
Professional Road Bike (Campagnolo)	8,000,167.10
Professional Road Bike (Shimano)	36,763,975.87
Professional Road Bike (SRAM)	7,276,109.83
Professional Touring Bike (black)	14,197,132.88
Professional Touring Bike (red)	13,443,587.37
Professional Touring Bike (silver)	30,298,155.26
Repair Kit	134,317.12
Road Helmet	214,624.39
T-shirt	126,444.92
Water Bottle	83,314.79
Water Bottle Cage	259,198.17
Women's Off Road Bike Fully	13,883,430.28
Overall Result	**285,250,874.29**

5.3.3 What is the Revenue for E-Bike Trailwind in 2010?

The revenue for E-Bike Trailwind in 2010 is 2,024,435 USD. Material in the Rows tab can be filtered to E-Bike Tailwind and Calendar year is dragged to rows component followed by Material dimension to show the following crosstab answering the question in hand. It is import to understand the ordering of dimensions accordingly.

E-Bike Tailwind	2010	2,024,435.82
	2011	1,998,759.41
	Result	**4,023,195.23**

5.3.4 What year had the highest Sales Quantity and the lowest Revenue? In the year with the lowest Revenue, which Material had the lowest revenue?

2007 has the highest sales quantity and so does the revenue of 60,715,832 USD. Whereas 2009 has the lowest revenue of 52,610,815 USD and lowest sales quantity. There is a constant decrease in the revenue from 2007. This is a serious issue to be investigated further deeply to understand the reason for the decreased sales quantity and revenue.

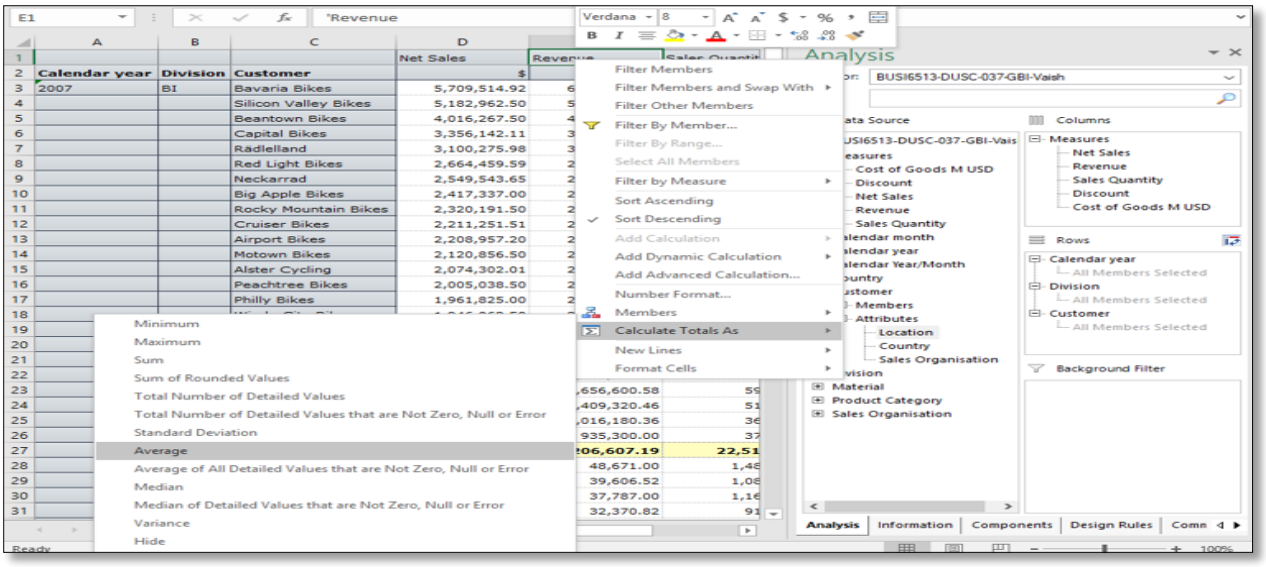
	A	B	C
1		Sales Quantity	Revenue
2	**Calendar year**	PC	$
3	2007	37,537	60,715,832.76
4	2008	36,088	59,444,067.04
5	2010	32,237	55,854,416.28
6	2011	31,880	56,625,742.97
7	2009	31,112	52,610,815.24
8	**Overall Result**	**168,854**	**285,250,874.29**

The year with the lowest revenue is 2009. In 2009, only 10 units of Fixed Gear Bike Plus was sold generating lowest revenue, whereas Professional Road Bike contributed for the highest revenue in all the major revenue generating products likes Bikes. The revenue is conditionally formatted which allows the audience to grasp the information quickly about the trend, maximum and minimum revenue generating products without much effort. This can help to make a decision to probably keep few profitable products and remove few products without growing sales.

	A	B	C	D	E
1				Sales Quantity	Revenue
117	2009	PUMP1000	Air Pump	4,116	125,485
118		ORHT2000	Men's Off Road Bike Hard Tail (SRAM)	2,572	4,756,749
119		CAGE1000	Water Bottle Cage	2,570	50,422
120		DXRD1000	Deluxe Road Bike (Shimano)	2,498	4,623,118
121		ORMN1000	Men's Off Road Bike Fully	2,223	5,836,680
122		PRTR2000	Professional Touring Bike (silver)	1,598	5,565,854
123		PRRD1000	Professional Road Bike (Shimano)	1,597	6,960,614
124		DXTR2000	Deluxe Touring Bike (silver)	1,565	5,111,357
125		ORHT1000	Men's Off Road Bike Hard Tail (Shiman	1,473	2,567,157
126		FAID1000	First Aid Kit	1,455	63,333
127		ORWN1000	Women's Off Road Bike Fully	974	2,667,508
128		PRTR1000	Professional Touring Bike (black)	776	2,710,564
129		DXRD2000	Deluxe Road Bike (SRAM)	749	1,348,337
130		DXTR3000	Deluxe Touring Bike (red)	737	2,412,271
131		DXTR1000	Deluxe Touring Bike (black)	727	2,372,711
132		SHRT1000	T-shirt	719	23,584
133		RHMT1000	Road Helmet	718	38,983
134		BOTL1000	Water Bottle	710	15,506
135		PRTR3000	Professional Touring Bike (red)	710	2,471,552
136		OHMT1000	Off Road Helmet	704	38,375
137		RKIT1000	Repair Kit	702	24,435
138		KPAD1000	Knee Pads	297	24,293
139		EPAD1000	Elbow Pads	287	23,474
140		PRRD3000	Professional Road Bike (Campagnolo)	281	1,380,604
141		PRRD2000	Professional Road Bike (SRAM)	279	1,277,747
142		CITY1000	City Bike Max	65	106,629
143		FXGR1000	Fixed Gear Bike Plus	10	13,475
144		**Result**		**31,112**	**52,610,815**

5.3.5 What is the average Revenue per Customer for the top 5 customers?

By following the path shown in the following screenshot, it is easy to change the aggregation method of a measure, for example the aggregation method to the measure Revenue is changed to average. So, the subtotal represents the average instead of default “sum”.



Sorting is applied on the Calendar year and the Top N (Top 5 here) functionality is applied on the Customer measure based on the revenue. All these data manipulations can be easily executed using the SAP Business Object Analysis for MS excel by just right clicking on the corresponding dimension and selecting the functionality.

		Revenue
Customer	**Calendar year**	$
Bavaria Bikes	2011	7,102,274.17
	2010	6,788,681.24
	2009	6,083,127.61
	2007	6,049,622.58
	2008	5,851,914.83
	Result	**6,375,124.09**
Capital Bikes	2010	3,983,787.32
	2011	3,913,352.04
	2009	3,789,976.56
	2008	3,760,612.81
	2007	3,489,737.97
	Result	**3,787,493.34**
Beantown Bikes	2007	4,265,437.00
	2008	4,141,214.49
	2011	3,922,743.91
	2010	3,486,673.19
	2009	2,853,740.15
	Result	**3,733,961.75**
Rädlelland	2011	3,962,873.12
	2010	3,781,113.71
	2009	3,638,600.04
	2008	3,504,132.27
	2007	3,228,531.52
	Result	**3,623,050.13**
Silicon Valley Bikes	2007	5,504,421.00
	2008	5,062,580.88
	2009	4,216,175.73
	2010	2,989,108.38
	Result	**4,443,071.50**
Overall Result		**4,390,434.69**

5.4 Conclusion:

SAP Business Object Analysis supports enterprise customers to large extent. The visualizations are mostly done in excel which is almost similar to excel charts. However, it is easy to use; just by dragging dimensions or measures into corresponding areas in the design panel, the crosstab is modified accordingly. The tool might not be well integrated with other external platforms. It appears to be outdated when compared with the existing BI tools in the current era of technological advancements. This might not support quick decision making. It might not be well suited in the context of predictive analytics. It requires a prior knowledge on creating and working on data cubes and queries with a pre-installation of SAP Business Object Analysis plugin for Excel. It is not a very useful tool to visualize and might be challenging to get insights faster.

Chapter 6: Tableau (Desktop)

6.1 Tool Description and Usage:

Tableau desktop is a business intelligence application that virtually visualizes all kinds of structured data in the form of interactive charts, dashboards and reports (Intellipaat, 2019). It can be connected to an extensive list of data sources from simple spreadsheets to complex data warehouses. It is one of the sought-after tools to effectively tell stories to audience by converting raw data into meaningful information in an easily understandable and visually elegant ways. Real-time analytics and data blending are significant features of Tableau that makes it widely used in the industry (Intellipaat, 2019).

6.2 Data Set Exploration:

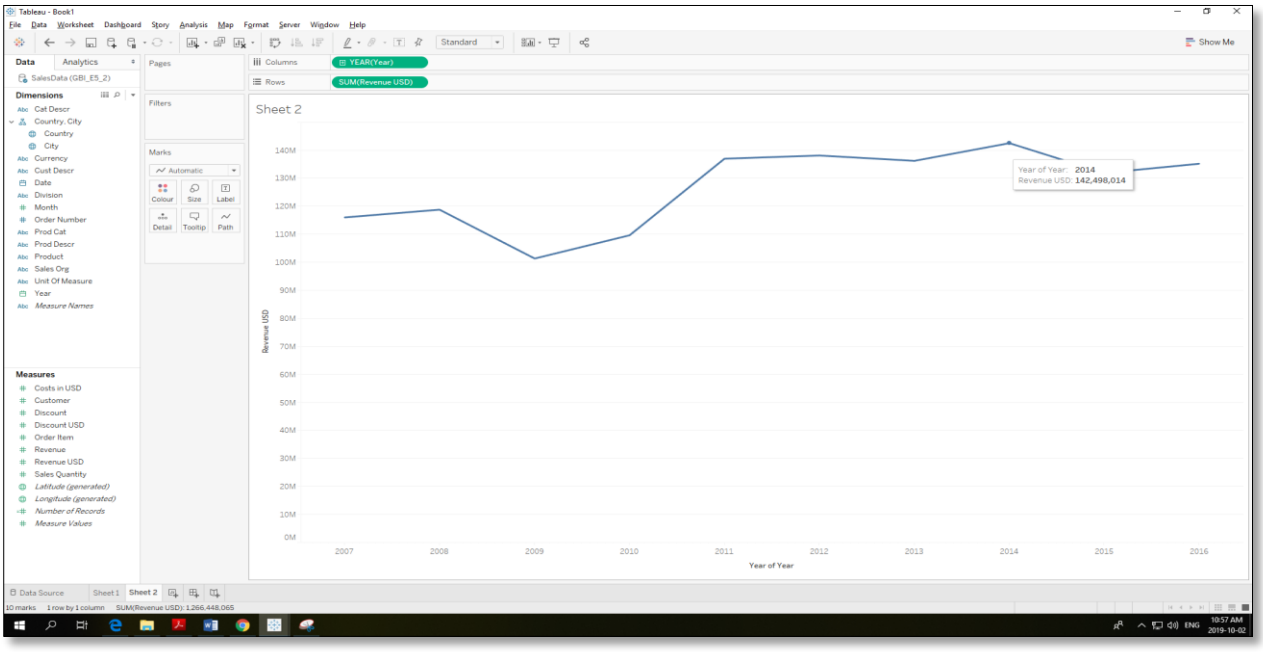
The transactional (sales) data set from 2011 to 2014 of The Global Bike Inc., GBI_E5_2.xlsx is downloaded from the Brightspace. The data has 23 columns consisting the details of the purchased product, customer, sales organization, sales order and other financial information and 132760 rows of sales orders. After exploring at the data, it appears to be consistent and no data preprocessing is necessary. This data from Excel is loaded into tableau desktop for further analysis.

6.3 Analytics and Insights:

The insights generated from the analysis is described by answering following research questions derived from the respective case study from Brightspace (Kalé, Jones, & Lee, 2017):

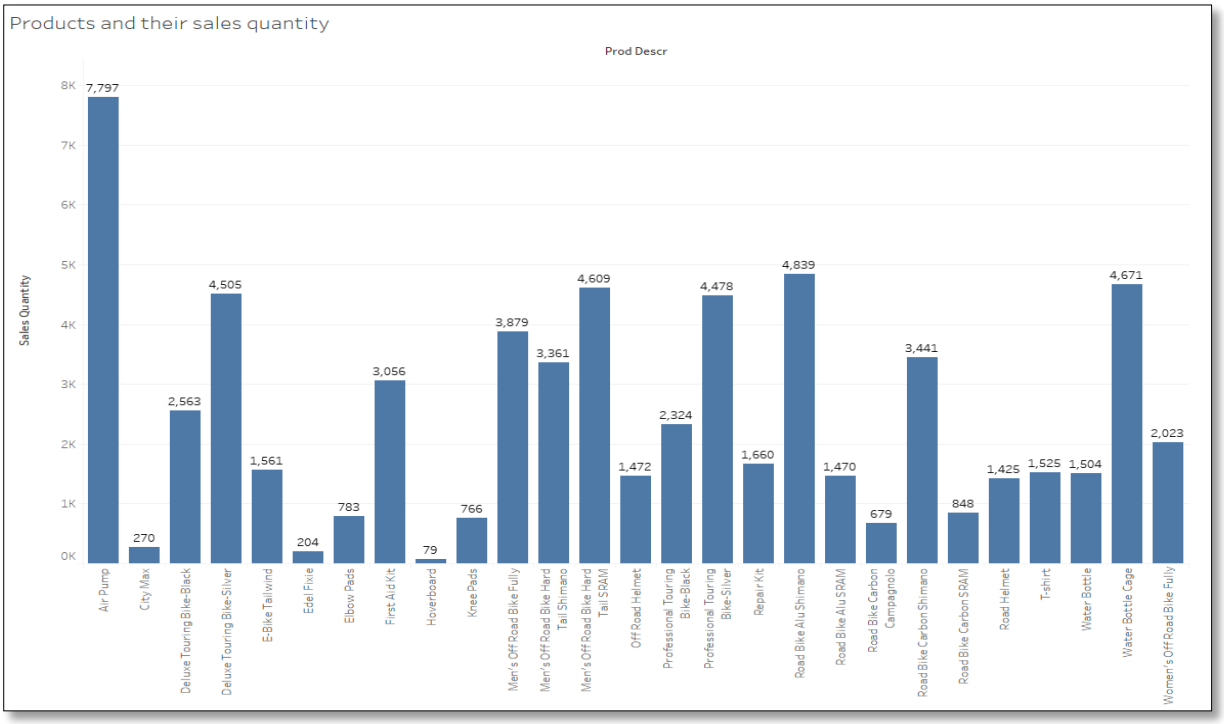
6.3.1 Which year had the highest revenue overall?

2014 has the highest revenue of 142,498,014 USD. Tableau allows the trend to be shown easily by just dragging revenue from measures into rows and year from dimensions to columns. The default aggregation for all the measures is sum which is relevant in this context to find the sum of revenues generated from all the products in a certain year. The actual numerical data i.e., total revenue in a year is shown by hovering on to that data point in the line chart shown below.

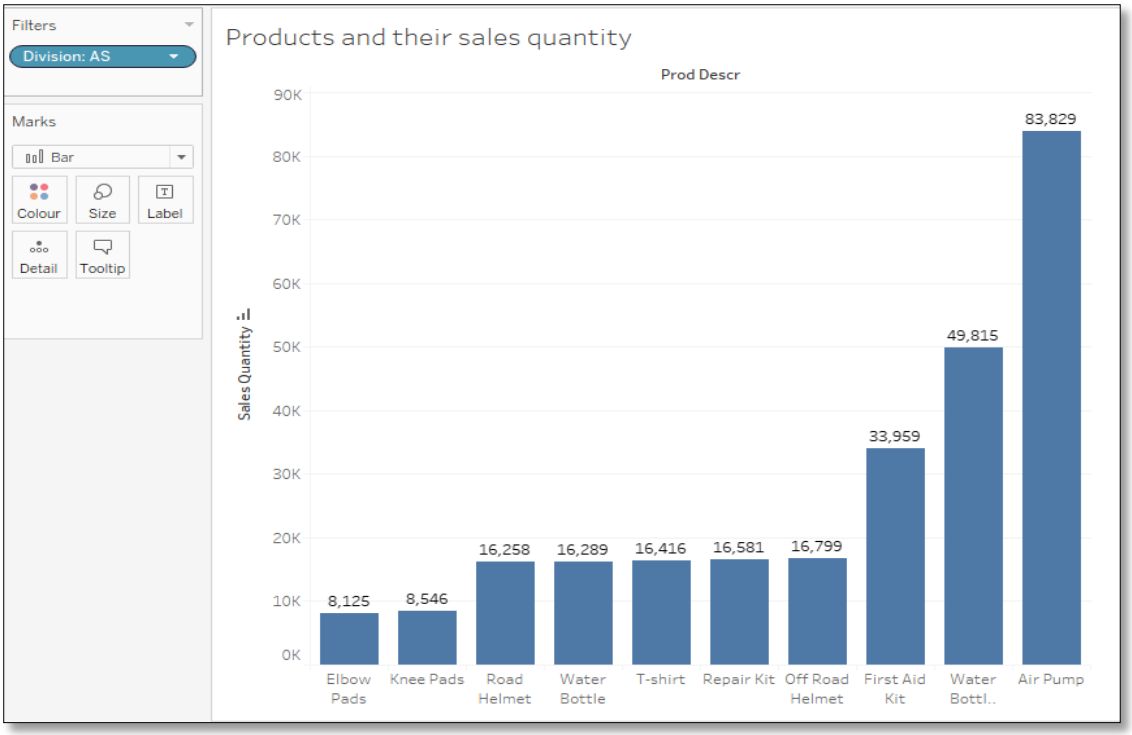


6.3.2 Show the products and corresponding sales quantity. Which product was the lowest selling accessory in 2016?

All the products and their sales quantities are shown in the visual below.



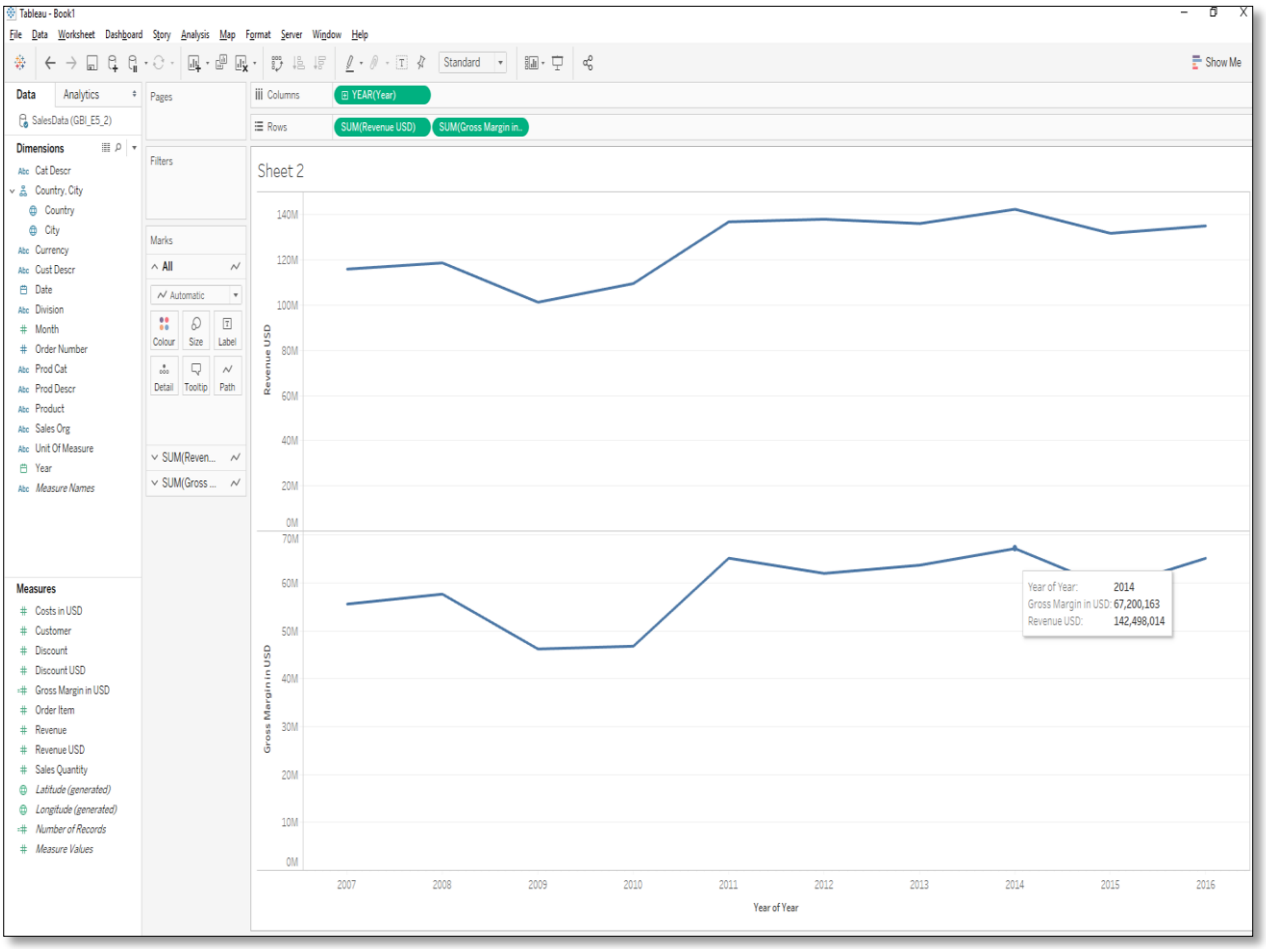
In 2016, Elbow Pads are the lowest selling accessory product in terms of sales quantities. Filters are applied on Division dimension and year to choose the sales quantities only for 2016 and only for Division AS representing accessories. The bar chart shows the sales quantity corresponding to each product, it is easy to sort the product in the chart either chronologically by product names or numerically based on sales quantity. This would enhance the understanding.



6.3.3 Compare the trend of gross margin and revenue over years?

After exploring the data set, it is understood that gross margin value is not provided. However, it is easy to calculate the gross margin using calculated field by using the formula representing the subtraction of discounts and costs from the revenue which is shown below.

It does make sense to have revenues and gross margin a similar trend as they are almost directly proportional to each other. It is obvious from the chart below that revenue and gross margin is high for the year 2014.



6.3.4 Compare gross margin ratios by division for all years. What is the gross margin ratio for each division?

Gross margin ratios are calculated appropriately using respective formula using the calculated measures and dragged into the row component. The aggregate measures for the gross margin ratio is changed to average which is sum by default. By placing country and division is same order, the data is first divided based on country and then each segment of data is further divided based on the division. So, order of the dimensions is important to get relevant results. Overall, Germany appears to have higher gross margin ratio. In both Germany and United States AS Division is doing better than BI Division which is obvious from higher gross margin ratios.

6.4 Conclusion:

Tableau can perform various kinds of data manipulation operations like sorting, filtering, ranking or aggregating at the same level as many other tools. However, it is rich in terms of producing visualizations which are perfect for analysis. There are options to optimize the operations in tableau thereby enhancing the performance and speed of analysis. It appears to be user friendly and a prior coding knowledge or experience may not be required. The rich feature set of Tableau allows to customize report based on the needs of audience. It is more of a visualization tool than a data preprocessing or preparation tool.

Chapter 7: IBM Cognos Insight

7.1 Tool Description and Usage:

IBM Cognos Insight is one of the tools under IBM Cognos Business Intelligence umbrella that is majorly used to work with complicated data sources in easy to use workspaces. This tool provides flexibility to import and merge diverse data like historical business intelligence, databases and other files (Application Consulting Group, Inc., n.d.). This tool can be easily integrated with other tools of IBM family than external platforms. It is used to analyze data and share those workspaces with the managers in supporting decision making, without much reliance on Information Technology (IT) team. This tool appears to work well in simpler business scenarios than complicated ones.

7.2 Data Set Exploration:

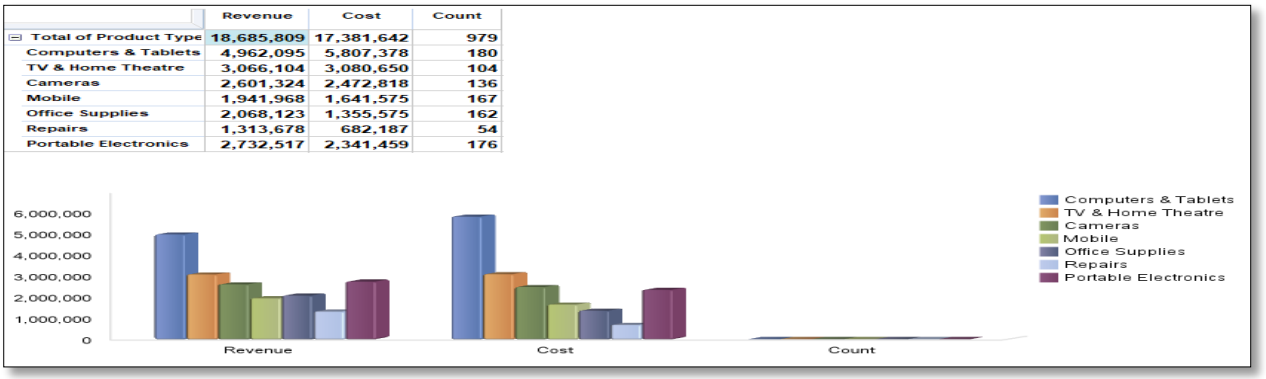
The dataset "Sales.csv" related to the sales transactions of electronic products is downloaded from the Brightspace. It consists of eight columns with details related to 979 sales transactions of electronics: Computers & Tablets, TV & Home Theatre, Cameras, Mobile, Office Supplies, Portable Electronics and repairs related to them, sales channel: Direct, Internet, Retail, customer type: Entertainment Venues, Retail, Internet Direct, Branded Stores, revenue, cost, quarters, and months (YM) of transactions across various regions: North America, Europe, Asia Pacific and South America. The observed data is inconsistent as the quarter column reflected 2012, but YM column reflected 2019 which is loss of integrity in data and the transaction takes place only on 12th of every month in 2019.

7.3 Analytics and Insights:

The insights generated from the analysis is described by answering following research questions derived from the respective case study from Brightspace ("IBM Workshop," n.d.):

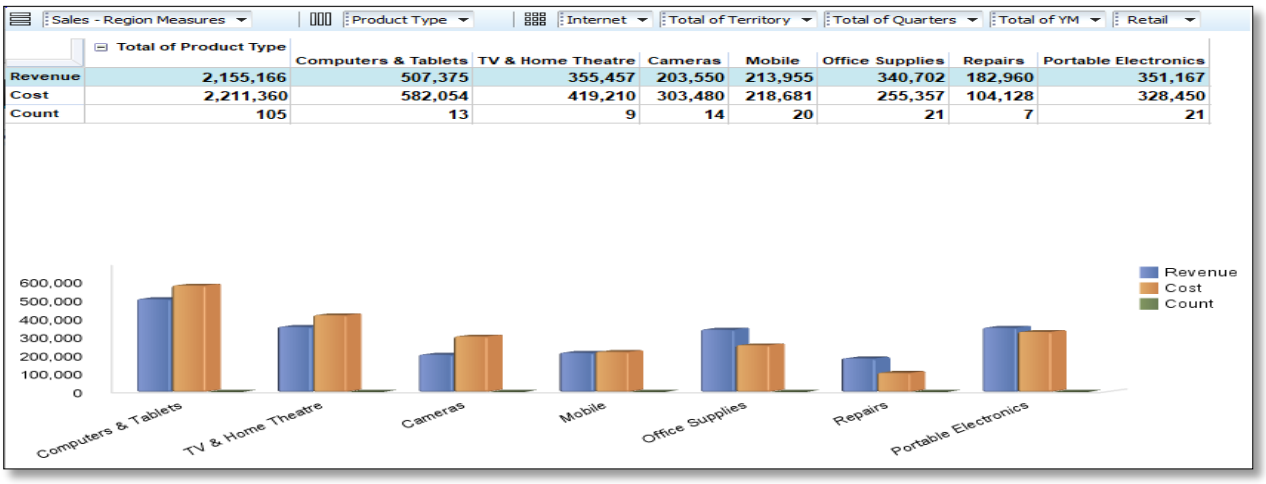
7.3.1 Project the revenue, cost and count of products from the sales transactions data?

IBM Cognos, by default, converts all the numeric values are created as measures and each column is created as a separate dimension. The following crosstab is generated immediately after the data is loaded into the platform with default measures and dimensions. Out of the available columns, revenue and cost are converted as measures as they are numeric and displayed in the crosstab. In addition, IBM Cognos has Count as measure by default. Computers & Tablets has the highest sales and the highest costs; however, it appears the cost is slightly higher than the revenue. And the revenue is significantly more than cost for the office supplies indicating efficient use of resources.

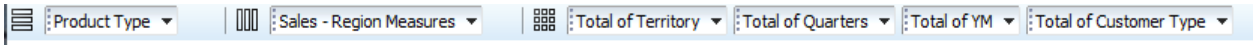


	Revenue	Cost	Count
Total of Product Type	18,685,809	17,381,642	979
Computers & Tablets	4,962,095	5,807,378	180
TV & Home Theatre	3,066,104	3,080,650	104
Cameras	2,601,324	2,472,818	136
Mobile	1,941,968	1,641,575	167
Office Supplies	2,068,123	1,355,575	162
Repairs	1,313,678	682,187	54
Portable Electronics	2,732,517	2,341,459	176

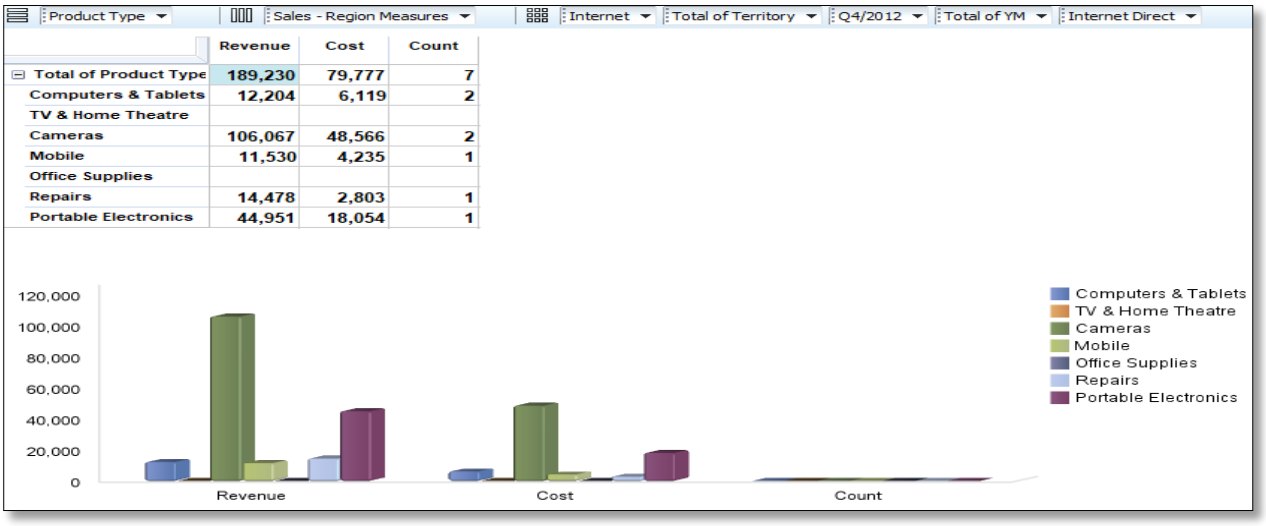
It is easy to switch rows and columns projecting that the tool is user interactive which would be helpful for a business user. The same projection, by flipping the rows and columns i.e., by each product is shown in the below bar chart.



7.3.2 What products are profitable that are sold through internet with Internet Direct as customer type for Quarter 4 of 2012?



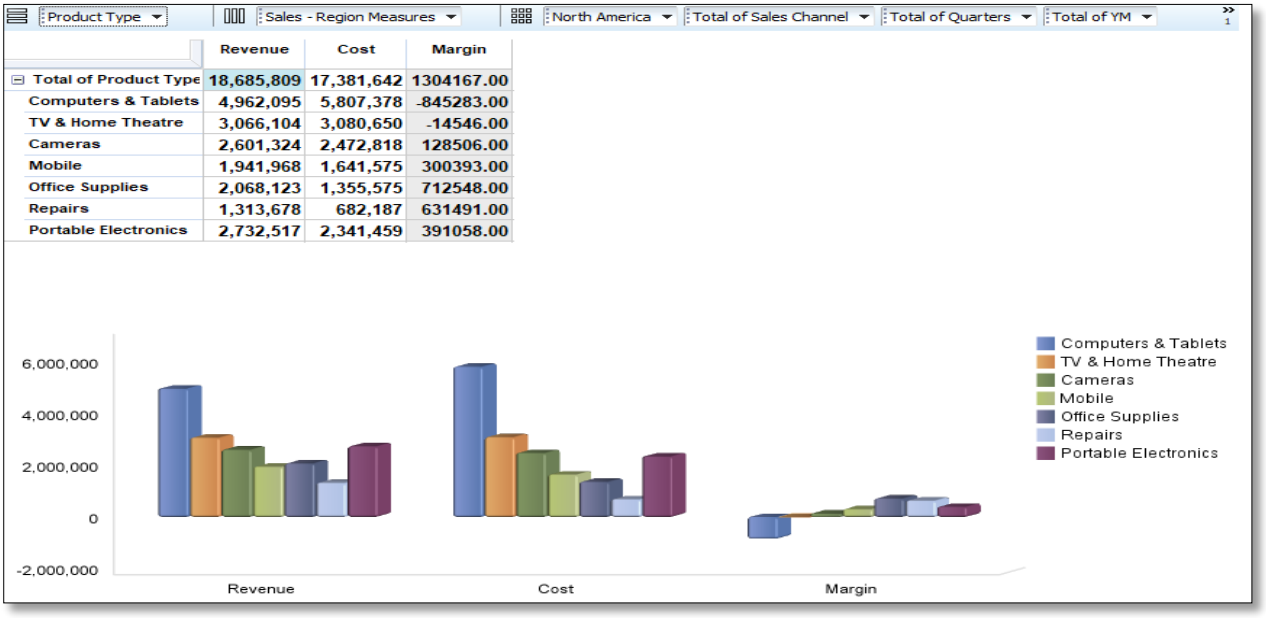
The above question can be answered just by choosing the right value from the corresponding dropdown options for each category (dimension) in the overview area of Cognos Insight user interface. The overview area is shown above. TV & Home Theatre and Office Supplies do not have any sales through internet to internet direct customer. However, cameras are sold more through internet and to internet direct customer. Overall, there is significant profit related to sales performed through internet channel and sold to internet direct customer.



7.3.3 Compare profit margin for all the product types in the North America?

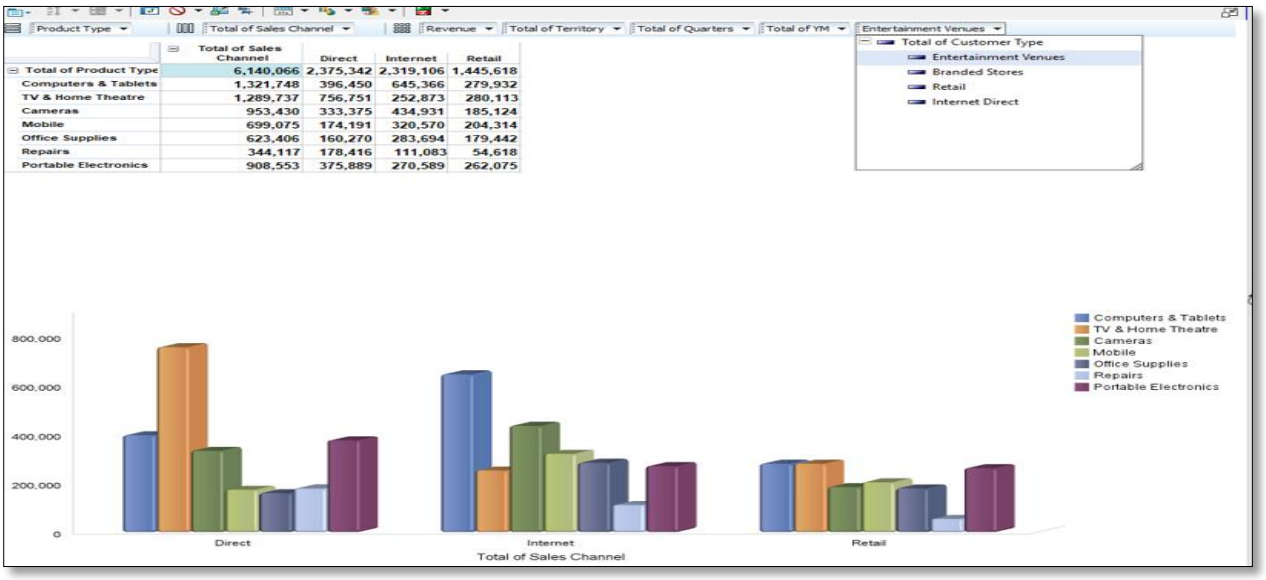
Though the profit margin is not created as a measure, one can easily create margin as a calculated measure by clicking on calculation icon in the menu tab. Though the top selling line is Computer & tablets, due to higher costs, margin is very low and less profitable in North America. Similarly, TV & Home Theatre also has negative margin indicating higher costs. This

necessitates the company to focus and investigate its value chain activities striving for cost leadership strategy to remain profitable. All the remaining products shows a positive margin which is shown below. Despite having a positive aggregated margin, a focused approach to reduce the higher costs is needed to sustain longer in the market.



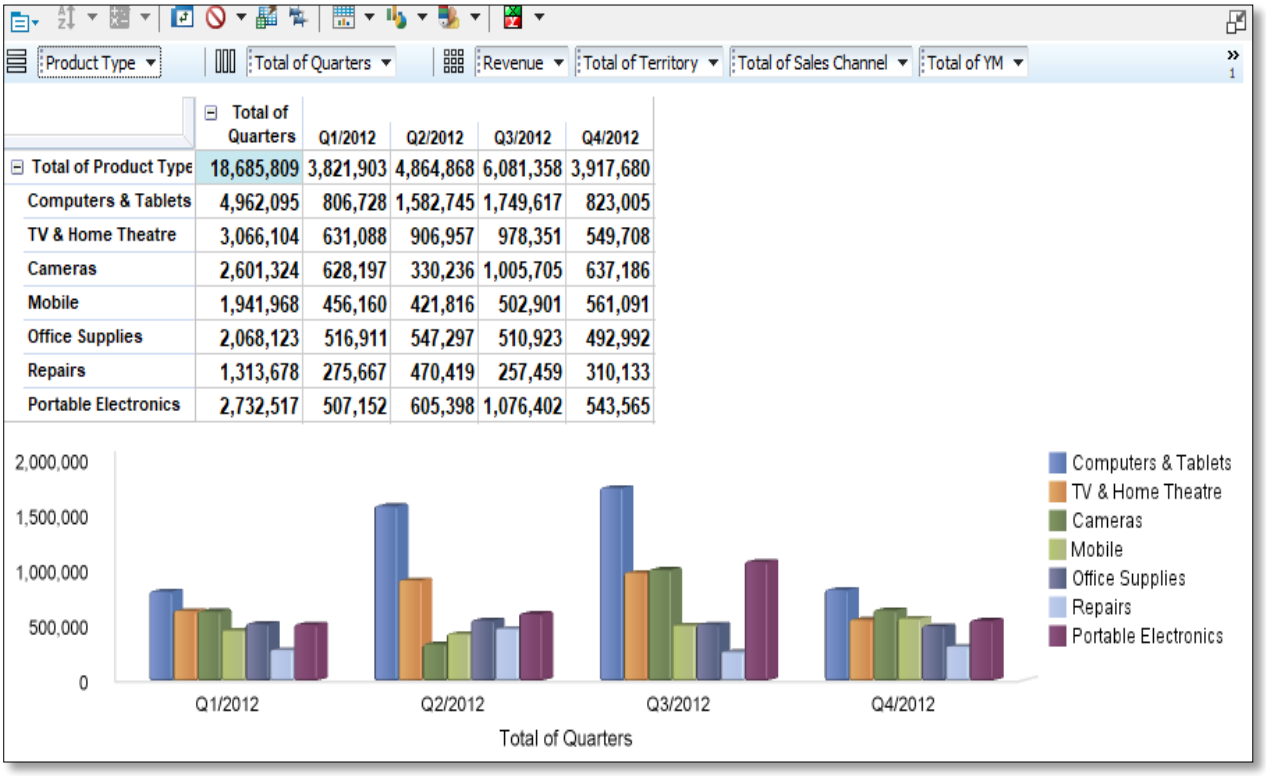
7.3.4 Compare revenues generated across various sales channels for the Entertainment Venues customer?

By changing revenue as the measure, changing columns to sales channel and rows to product type, applying filter to customer type by selecting Entertainment Venues from the corresponding dropdown list in the overview menu leads to answering the question. By comparing revenues for each product across various sales channels, the chart below conveys: Direct channel is associated with higher sales of TV & Home Theatre, Internet channel has higher computer & Tablets sales and lower repairs, and retail channel has higher and consistent revenues for Computer & Tablets, TV & Home Theatre and Portable Electronics without much variation among all products as in other sales channels.



7.3.5 What is the trend for all products across all the quarters?

The trend of sales for all the products across all the quarters is shown below. One of the major insights is Cameras are selling significantly low in second quarter compared to other quarters. The firm should make a strategy to not to manufacture many Cameras for the season quarter 2 to avoid overstock and extra inventory costs. Computers are selling very well in second and third quarter. Portable Electronics are selling better in third quarter. Mobiles sales are consistent across all the quarters. So, a business can avoid the overstock, inventory costs or capacity problems by choosing the right quality of products to manufacture or procure.



7.4 Conclusion:

From my experience working with IBM Cognos Insight, I understand it is very quick with good computational performance than many other tools. However, it appears to be lacking in terms of visualizations and predictive analytics. There could have been many diverse chart options for the visualizations. This tool might fit well for a business user who needs quick results than a technical data analysts or data scientists.

Chapter 8: SAP Predictive Analytics (Clustering)

8.1 Tool Description and Usage:

SAP Predictive Analytics is a business intelligence software from SAP designed to allow users to manage and analyze large volumes of data (big data) and predict future outcomes and behaviors (Rouse, & O'Donnell, n.d.). It appears to have significant predictive power through the application of predictive models (machine learning/AI) onto the data sets to predict future by analyzing the historical data. It works in synchrony with other tools of SAP's Business Objects Suite, especially SAP Lumira Discovery to visualize the outcomes (Wailgum, 2019).

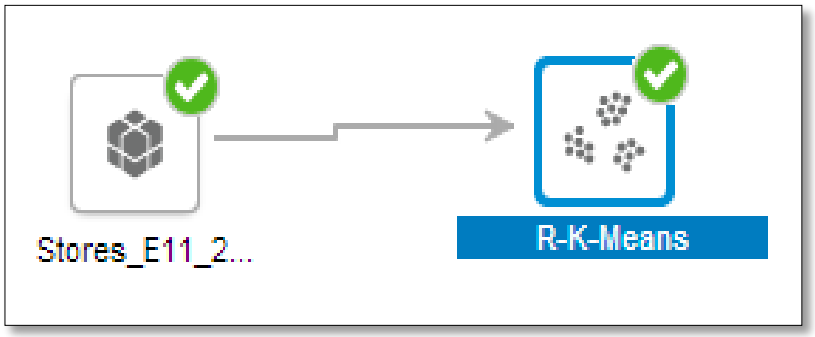
8.2 Data Set Exploration:

The data set used to understand the concept of clustering under unsupervised models of predictive analytics is Stores.csv. The data set has stores information at 150 locations in United States. The information related to each store is available in four columns: Sales Turnover, Store Size, Staff Size and Profit Margin. These stores data are segmented using the K-means clustering algorithm, where k represents the number of clusters into which the data is segmented. When the entire data is considered as a single set, the average sales turnover is 5.84, the average store size is 3.0, the average staff size is 3.75, and the average profit margin is 1.2. However, given the differences among the stores, the average values differ when the similar stores are clustered into groups, which would provide a better understanding of the information to finalize business plans for customized sales promotion strategies.

8.3 Analytics and Insights:

The insights generated from the analysis is described by answering following research questions derived from the respective case study from Brightspace (Kalé, & Jones, 2016):

8.3.1 Group the stores into 3 clusters using K-Means algorithm and analyze the summary?

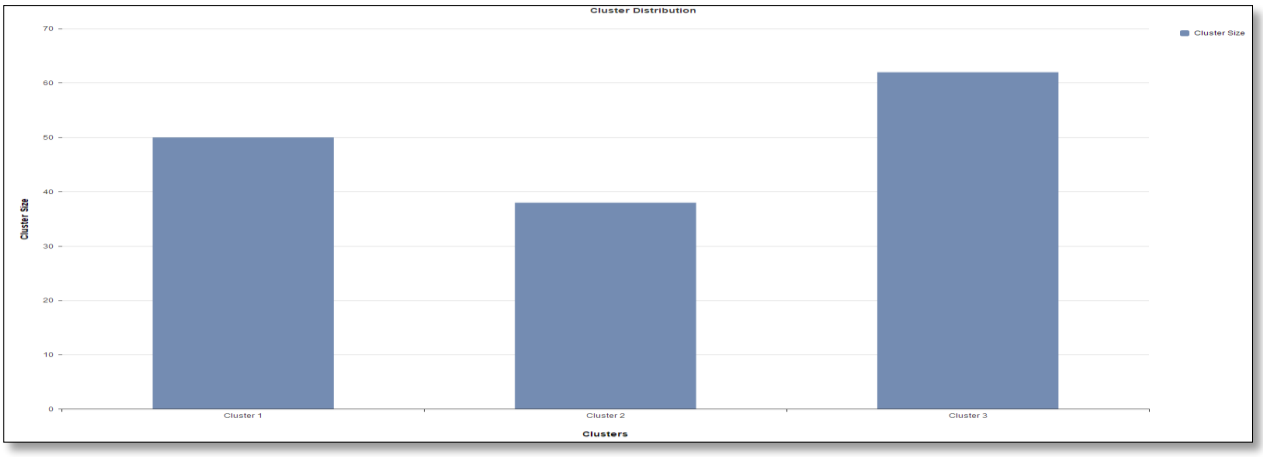


There is an algorithm section in the designer tab of predict panel in the Expert Analytics package of SAP Predictive Analytics. It is simple to cluster the data in SAP Predictive Analytics, which takes the form of dragging the R-K-Means from algorithms tab to the designer tab, which is shown in the diagram above. Since the task under hand is to cluster the stores into *3* segments to support the country manager, we must make the value of K to 3 in the configuration setting tab of the R-K-Means algorithm. Once the algorithm is successfully executed, the summary tab below shows the details about the clusters. For example, the columns on which K-Means is executed, the cluster centers (3) and their details, sum of squared errors (SSE), and the size of each cluster.

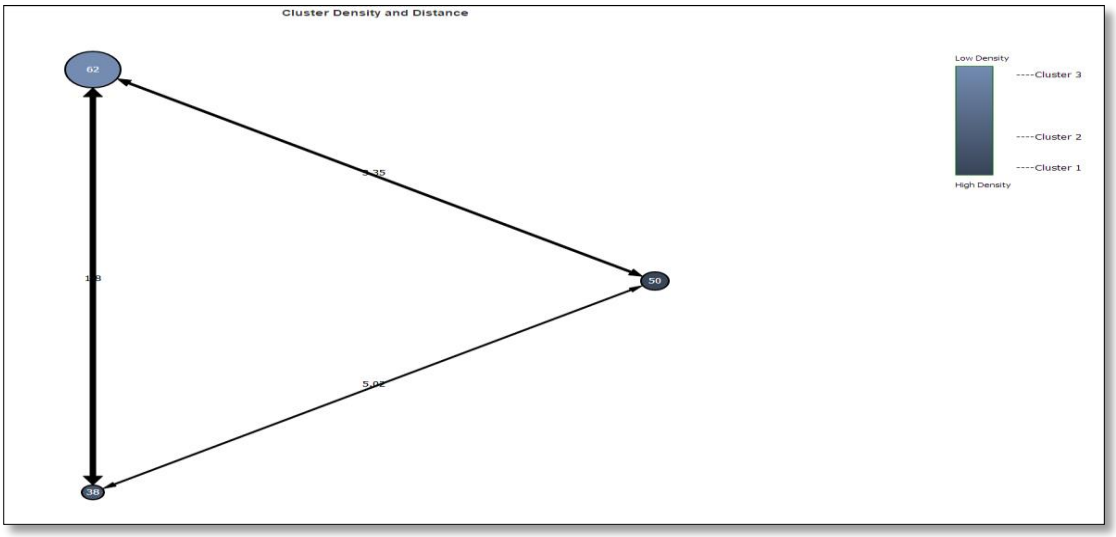
```
Summary of the model from R Scripts

Information of the columns used in the algorithm
-----------------------------------------------------
Independent Columns
 Sales Turnover : Double
 Store Size : Double
 Staff Size : Double
 Profit Margin : Double

Summary of the Model
             Length Class  Mode
cluster      150    -none- numeric
centers       12    -none- numeric
totss          1    -none- numeric
withinss       3    -none- numeric
tot.withinss   1    -none- numeric
betweenss      1    -none- numeric
size           3    -none- numeric
iter           1    -none- numeric
ifault         1    -none- numeric


Centers
  Sales_Turnover Store_Size Staff_Size Profit_Margin
1       5.006000   3.418000   1.464000      0.244000
2       6.850000   3.073684   5.742105      2.071053
3       5.901613   2.748387   4.393548      1.433871


Within cluster sum of squares
[1] 15.24040 23.87947 39.82097


The size of each cluster
[1] 50 38 62
```

8.3.2 Explain the size distribution and the density of the generated clusters?

Though, the summary option have the numeric values of the size of each cluster, a bar chart is used for an effective comparision of the sizes. As shown in the bar chart below, cluster3 has most number of stores (62) and cluster 2 has the least number of stores (38) and the number of stores in cluster1 (50) is between that of cluster2 and cluster3.

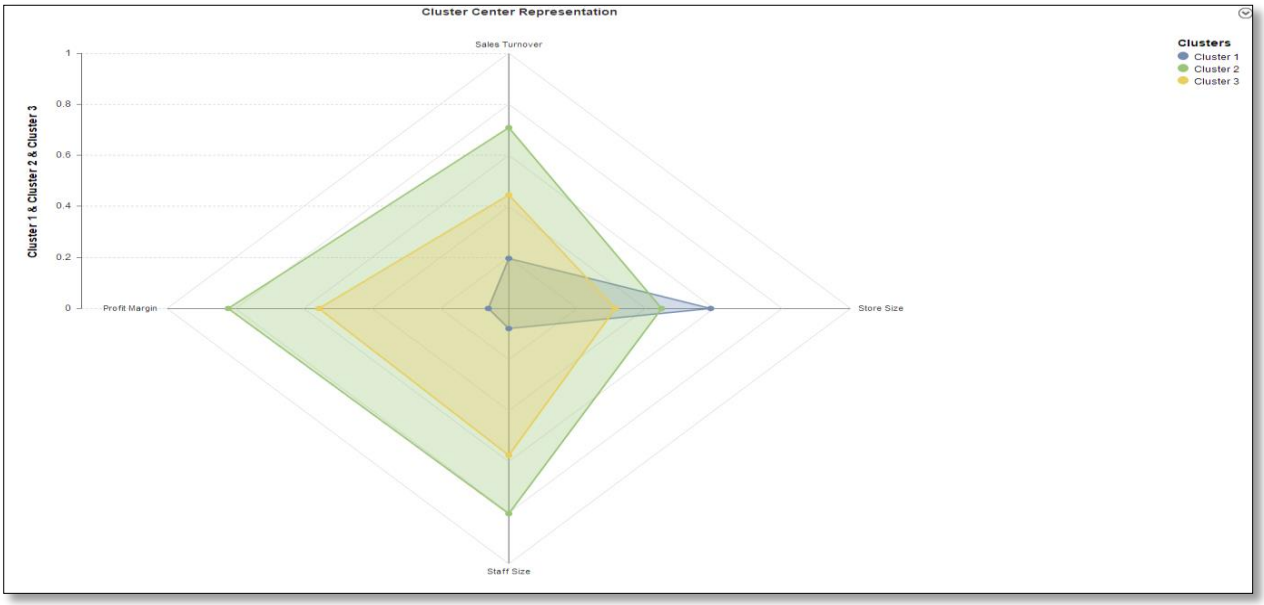


The distance (inter-cluster) and the density (intra-cluster) are depicted in the following figure. Though the number of stores (data points) in cluster 3 is lesser, the similarity among them is higher, which the density of cluster describes. Cluster 3 has low density which indicates that the data points in cluster3 are sparsely located, and loosely associated with the chance of noise and outliers i.e., stores in cluster3 are moderately similar.



8.3.3 Compare the various features across the clusters?

The following radar chart with axes as Sales Turnover, Profit Margin, Store Size and Staff Size, compares the cluster averages. Despite the fact that cluster1 has higher store size, it has relatively lower profit margin, sales turnover and staff size. This might indicate that the stores in cluster1 are a kind of warehouses with huge store sizes but with least direct selling activities. On the other hand, cluster2 has relatively higher profit margin, sales turnover, staff size and store size, which might be a chain of popular, high performing retail stores.

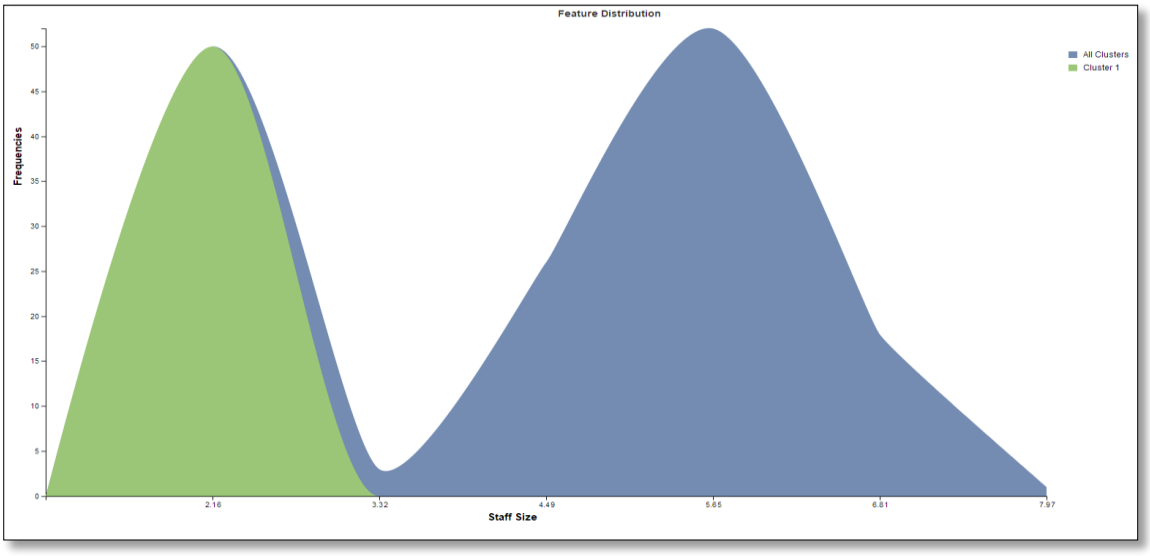


The dimensions of the stores can also be compared by considering them in pairs and analyzing the relationship between them using individual scatterplot for each pair of attributes. This pair plot or scatter matrix chart for all the four attributes across three clusters is shown below.

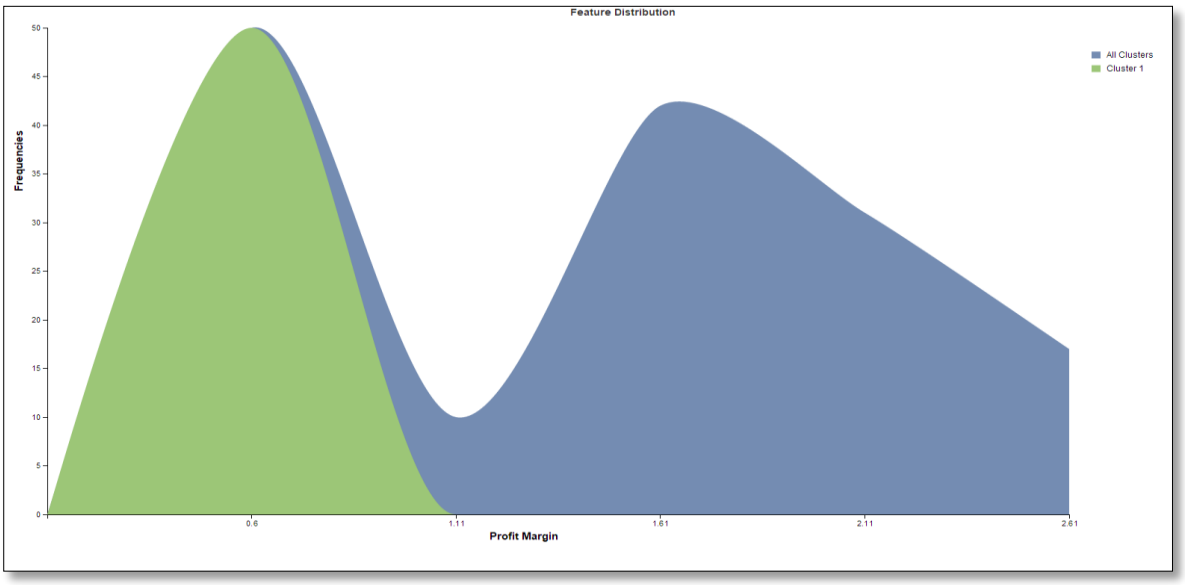
The data points in cluster1 reside closely together but far from other cluster points when compared in terms of all possible pairs of dimensions, which differentiate cluster1 from other clusters. For example, Staff Size – Sales Turnover scatter plot shows that cluster1 has lower staff size followed by cluster3 and cluster2.

8.3.4 Choose a cluster and position it against all the stores in terms of all the attributes?

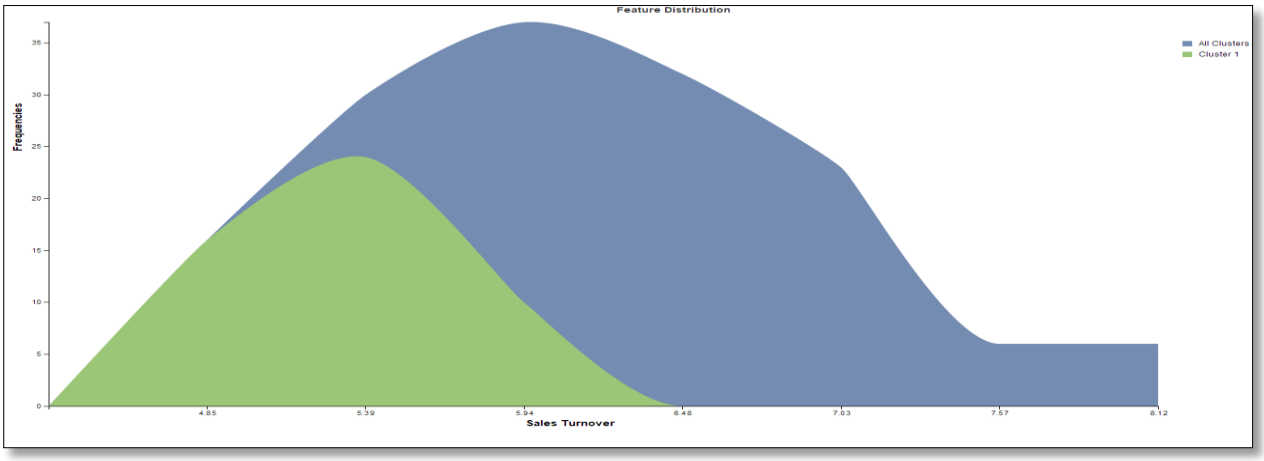
I have chosen cluster1 and the distribution of each variable in cluster1 is compared against the entire data set using the area charts representing feature distribution in cluster representations pane. The measure is changed to each dimension to identify the distribution of that dimension and compare across clusters.



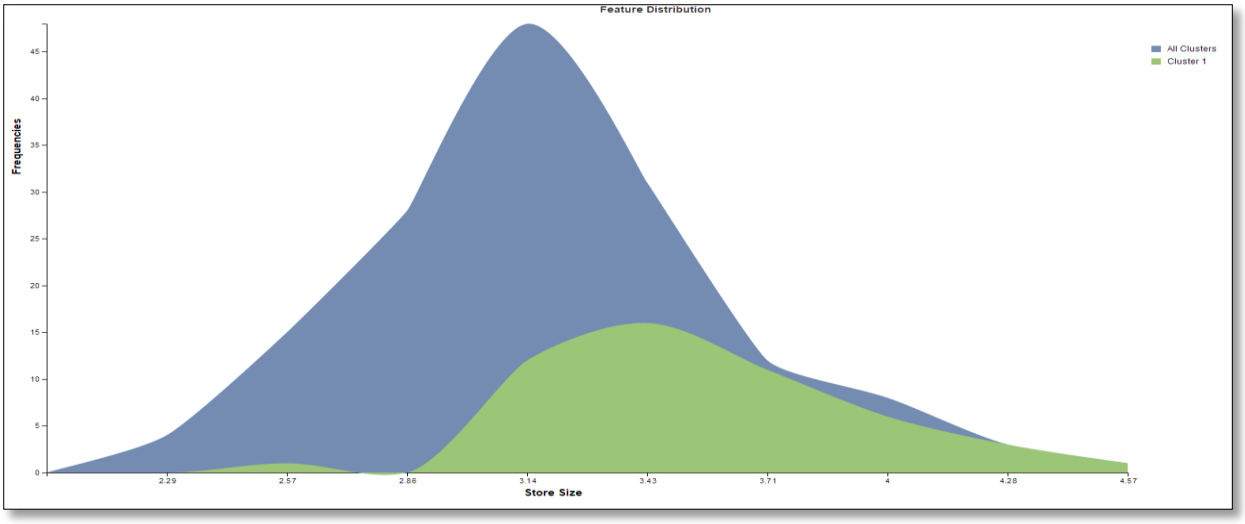
The chart above shows that cluster1 is highly differentiated in terms of staff size when compared with the staff size of overall data set. The staff size of the stores in cluster1 is always less than 3.32, whearas the staff size of all the stores spans towards 7.9.



The chart above shows that cluster1 is also differentiated in terms of profit margin when compared with that of overall data set. The profit margin of the stores in cluster1 is always less than 1.11 (lower), whearas the there are stores with profit margin of 2.6 as well. Stores in cluster 1 have very low profit margins against all the stores.



The chart above shows that cluster1 is also differentiated in terms of sales turnover when compared with that of overall data set. The sales turnover of the stores in cluster1 is always less than 6.5, whearas the there are stores with sales turnover of 8.12 as well. Stores in cluster 1 have relatively low sales turnover against all the stores. In terms of analyzing the frequencies, there are around 25 stores with sales turnover of 5.3.



The chart above shows that store size of cluster1 is little scattered with very few stores with size 2.57, most of the stores with store size between 3.1 to 4 and few stores with huge store size around 4.5. Stores in cluster 1 have relatively higher store size compared to an average store size across all the stores.

8.4 Conclusion:

SAP Predictive Analytics supports appealing visualization and predictive capabilities. There are multiple R language-based algorithms that would perform both unsupervised tasks like clustering and association analysis, and supervised tasks like classification. From my working with SAP Predictive Analytics for clustering, it supports K-Means clustering algorithm but lacks other clustering techniques like K-Medoids, DB-scan etc. However, it provides flexibility in terms of choosing attribute (measures) while analyzing clusters from various perspectives at the same time providing excellent visualizations of the clusters. This tool also accommodates the needs of Data Analysts and Data Scientists.

Chapter 9: SAP HANA (Data Modelling)

9.1 Tool Description and Usage:

SAP HANA is platform that has in-memory database technology that allows quick processing of massive volumes of historical and real-time data. This allows SAP HANA to perform both OLTP (Online Transactional Processing) and OLAP (Online Analytical Processing) operations at the same time. In terms of its usage in business, it includes many components with capabilities of data processing, data warehousing, database management system, application platform, data integration platform, and also acts as machine learning and data science platform (Jewett & , 2019).

9.2 Data Set Exploration:

The data from the Global Bike Inc., which includes the details of its customers, products and sales transactions. These data are stored in the server where SAP HANA is running. This data-modelling case study also included extracting the data from remote source, which is discussed further in the next section.

9.3 Analytics and Insights:

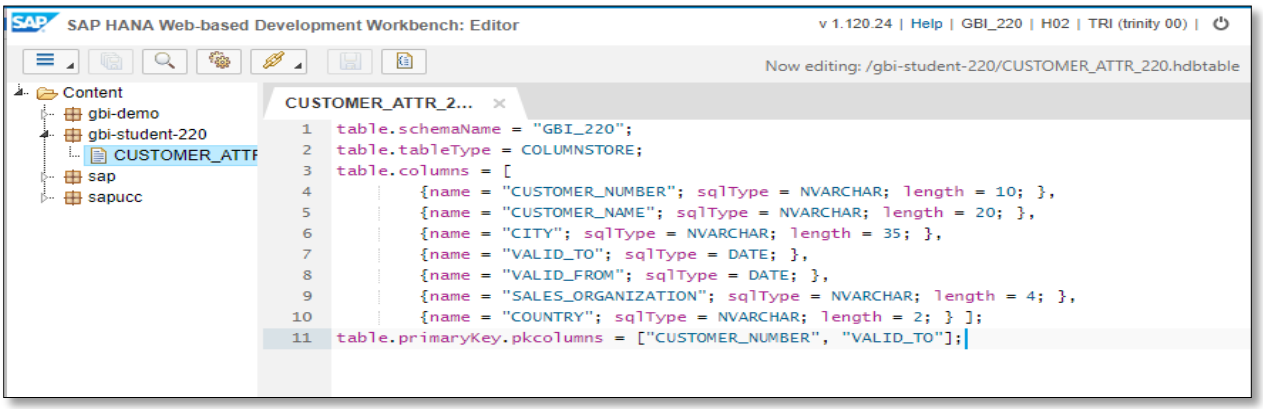
The insights generated from the analysis are described by answering following research questions derived from the respective case study from Brightspace ("SAP Hana," n.d.):

9.3.1 Create table definitions for Customer, Product and Sales data of Global Bike Inc. in SAP HANA?

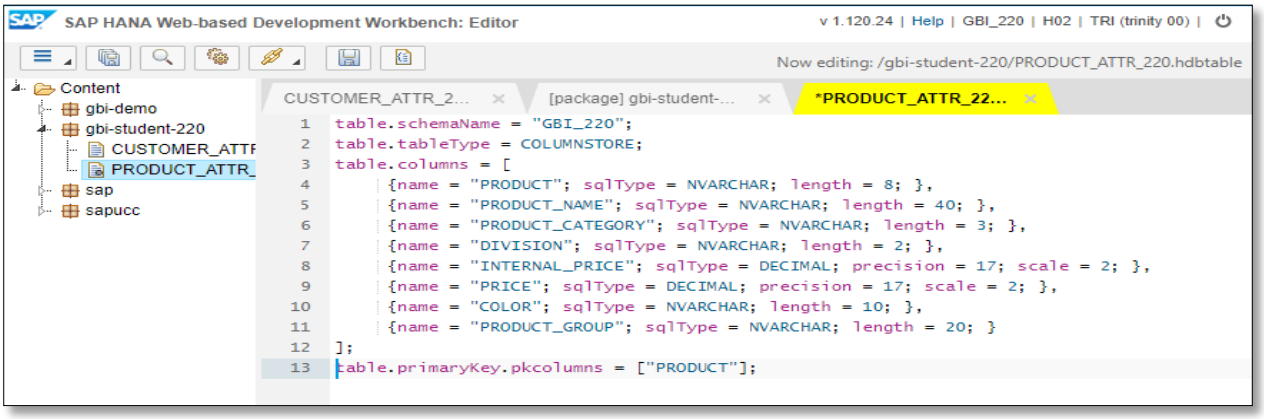
Creating a table definition means creating a schema or structure that the data follows and how data is represented after it is loaded into the SAP HANA in-memory database. One has to run an SQL query to create the table definitions. The SQL query indicates:

- The type of storage, which is columnar as SAP HANA is a columnar database in contrast to relational databases.
- Name of the schema
- Names of the columns in the structure
- Primary key columns which can uniquely identify to certain data point, foreign keys if any

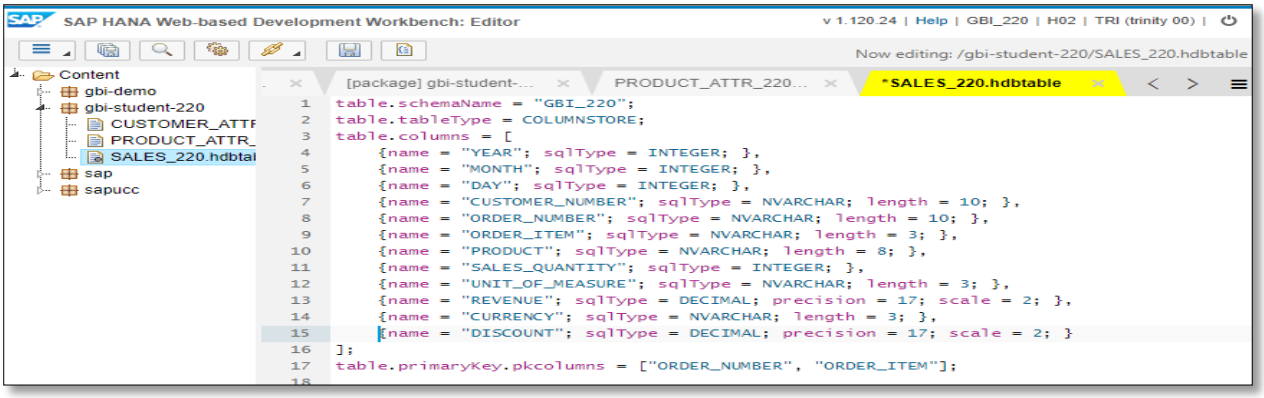
The SQL query to define the customer master data is shown below



The SQL query to define the product master data is shown below



The SQL query to define the sales transaction data is shown below



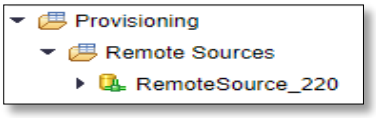
When the above queries are executed in the SAP HANA Development Workbench, the relevant schemas for the customer, product and sales data is created. However, they are empty, as they are not yet populated with the corresponding data.

9.3.2 Populate the created tables with relevant data, which are stored as flat files in the SAP Server?

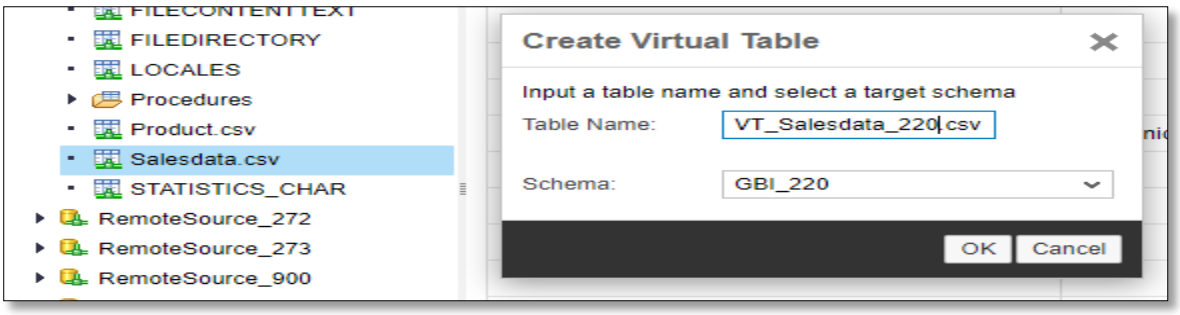
The steps involved in populating tables with the data are as follows:

- Data Provisioning
- Virtual Table creation
- Flowgraph model creation and execution

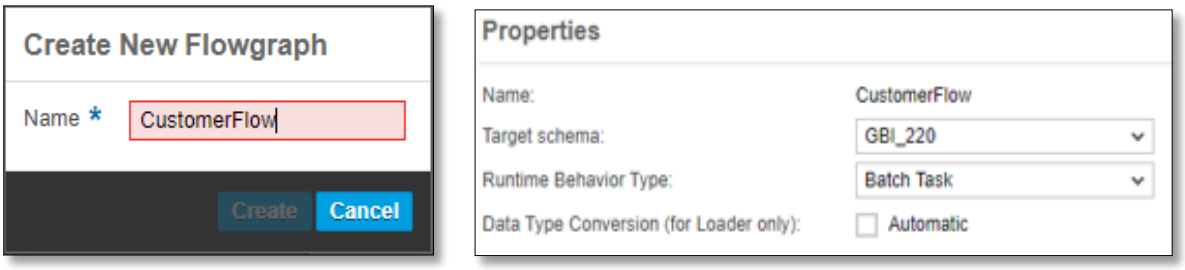
Firstly, in the catalog perspective of SAP HANA, remote sources referring the flat files on the server are added to the SAP HANA system, which is shown below.



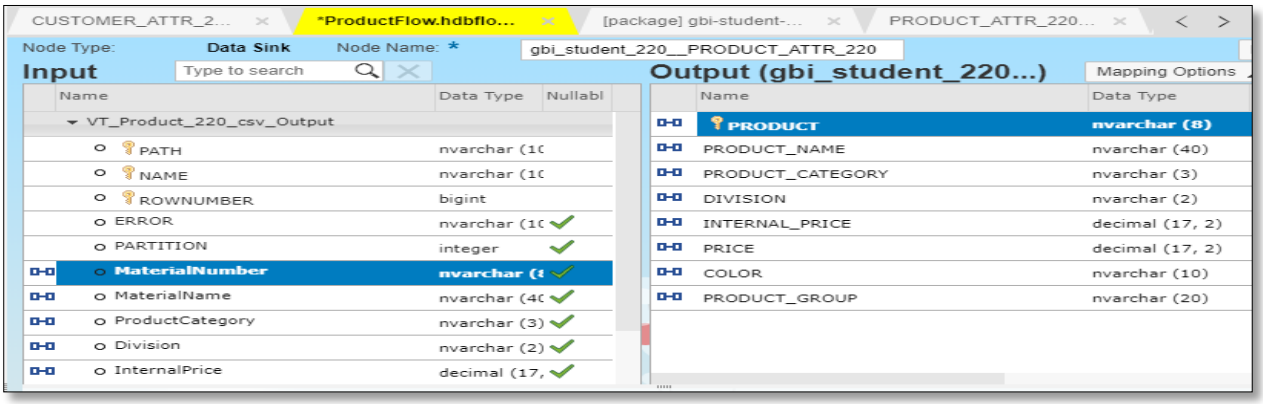
Then, the content in the server: Customer.csv, Product.csv, and Salesdata.csv, under remote source is created. This process is called data provisioning. Virtual tables for each .csv file representing Customer, Product and Salesdata are created. The virtual table creation for sales data is shown below.



A flowgraph model is used to load data from a virtual table to a table in SAP HANA database. In the editor perspective of SAP HANA, flowgraph models for each of the Customer, Product and Salesdata are created. Creation of a new flowgraph model to load the customer data is shown below.

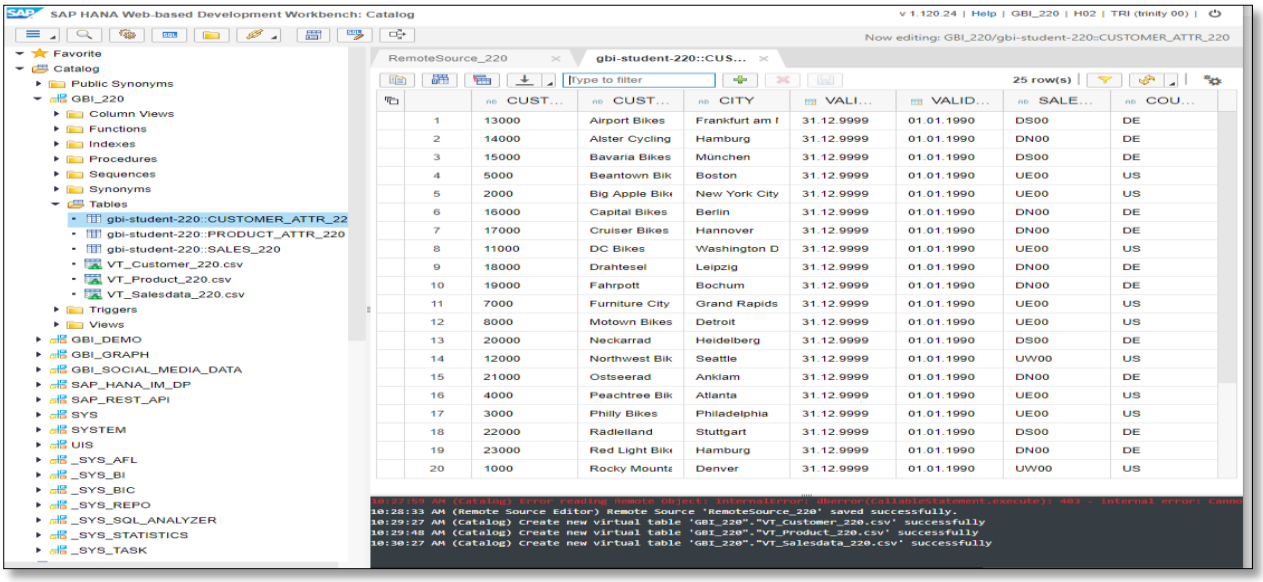


Within the created flowgraph models for Customer master data, Product master data and Sales transactional data, data connections between the each of the data tables (data sinks) with corresponding virtual tables (data sources) are established. The connection between sales product virtual table and product tables in SAP HANA and choosing the relevant field mapping is shown below



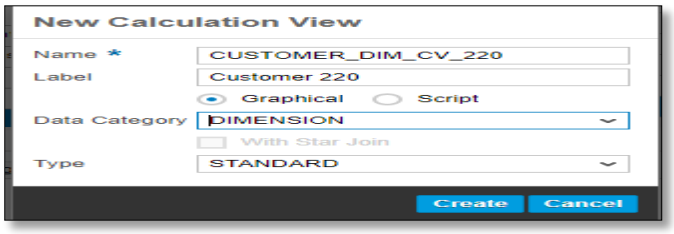
Once the flowgraph models are executed, the data is loaded into corresponding tables, which are created using SQL queries in previous question. Content is populated in each of the table

which can be seen in the catalog perspective of SAP HANA. The customer data table with populated data after the execution of customer flowgraph is shown below.

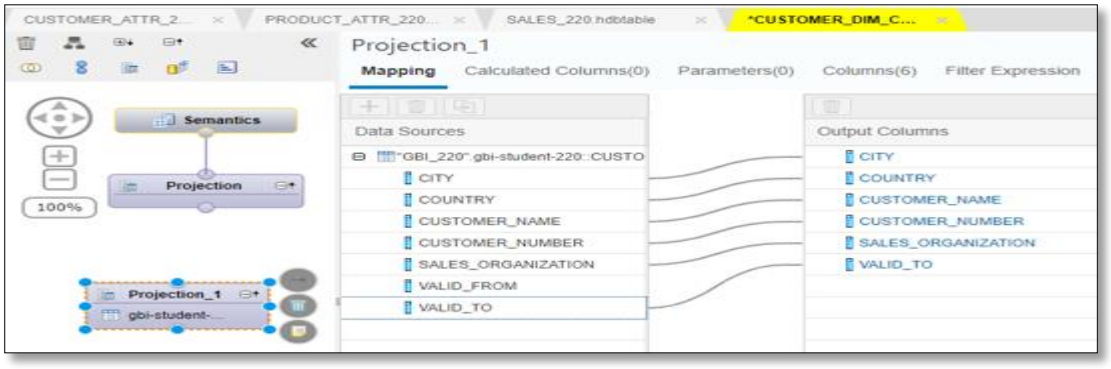


9.3.3 Create calculation view for each data table: customer, product and sales?

Calculation view perspective in SAP HANA is useful in performing few complex manipulations across tables like joins, unions, column addition, removal, and calculated column creation and so on. The creation of calculation view for customer is shown below.

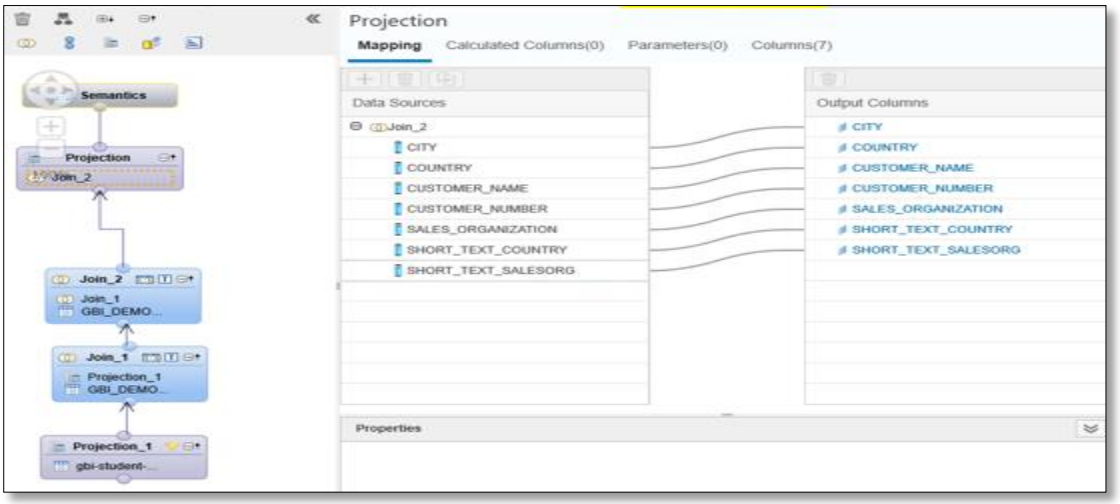


Since the country and sales organization attributes in the customer tables are short keys, the descriptions can be added to the customer table by using a projection node with customer table and valid_from column is removed.

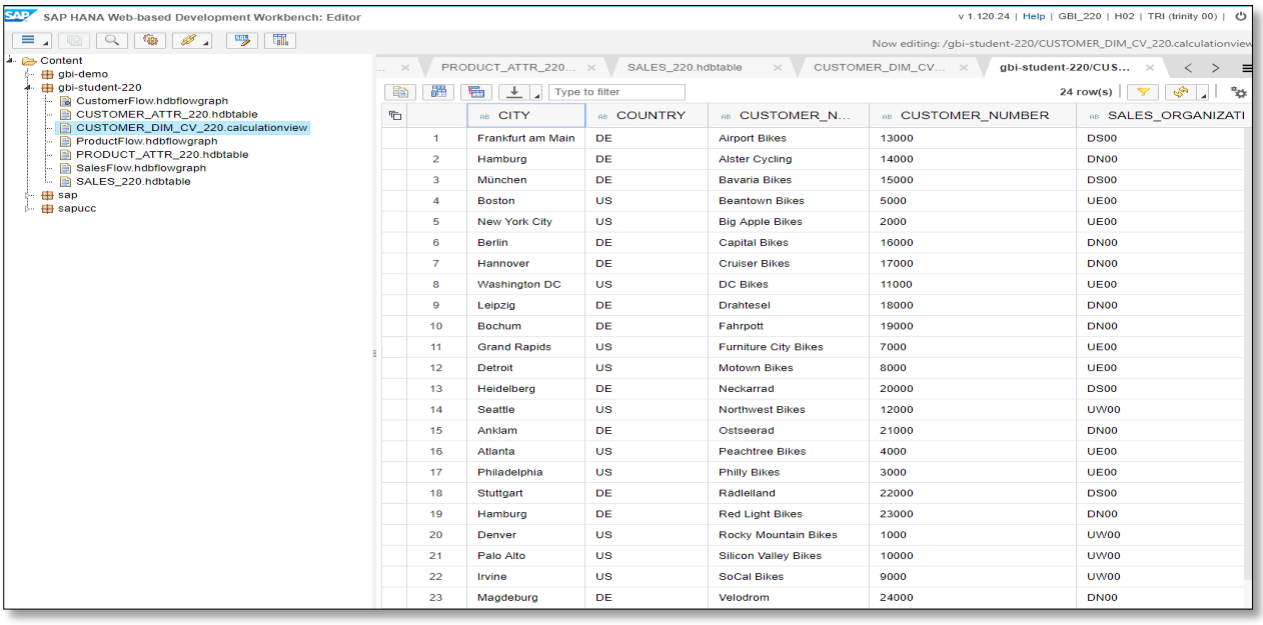


The table GBI_DEMO_COUNTRY (text data with country key and description) with the customer data using a text join in Join Definition tab and the corresponding description is added to customer data in the mapping tab which is shown below.

The same process is repeated to add the description of sales organization to the joined data i.e., customer data with country description, using another text join node. The description mapping for the fields COUNTRY and SALES_ORGANISATION in the output columns of customer data is performed which is shown below.

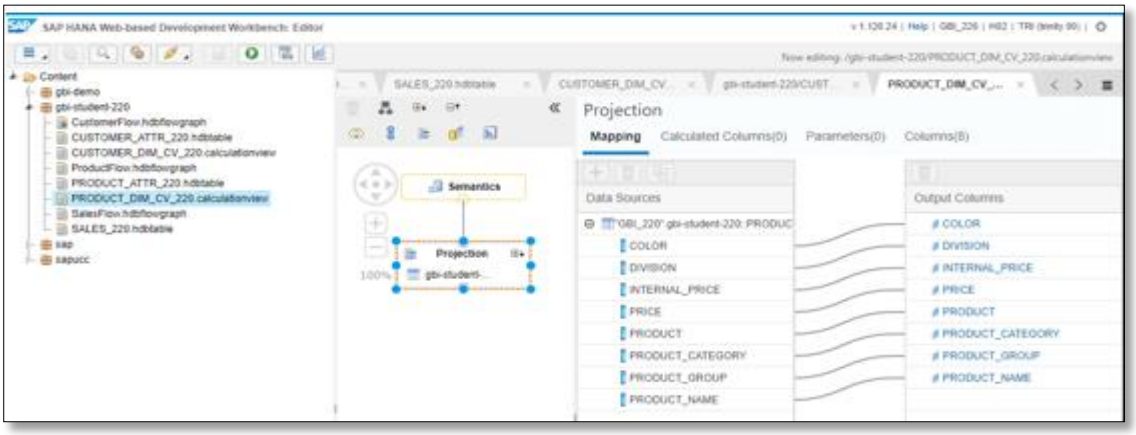


The output data can be seen in catalog perspectives with added descritpions.

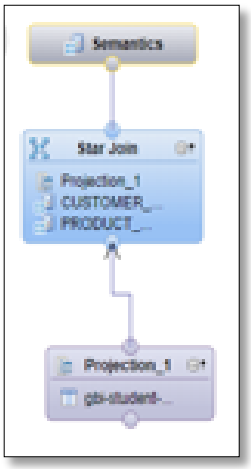


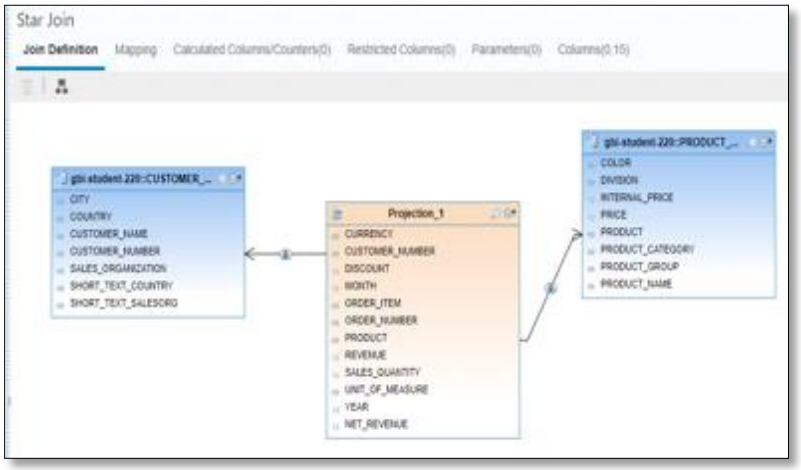
	CITY	COUNTRY	CUSTOMER_N...	CUSTOMER_NUMBER	SALES_ORGANIZATI
1	Frankfurt am Main	DE	Airport Bikes	13000	DS00
2	Hamburg	DE	Alster Cycling	14000	DN00
3	München	DE	Bavaria Bikes	15000	DS00
4	Boston	US	Beantown Bikes	5000	UE00
5	New York City	US	Big Apple Bikes	2000	UE00
6	Berlin	DE	Capital Bikes	16000	DN00
7	Hannover	DE	Cruiser Bikes	17000	DN00
8	Washington DC	US	DC Bikes	11000	UE00
9	Leipzig	DE	Drahtesel	18000	DN00
10	Bochum	DE	Fahrpott	19000	DN00
11	Grand Rapids	US	Furniture City Bikes	7000	UE00
12	Detroit	US	Motown Bikes	8000	UE00
13	Heidelberg	DE	Neckarrad	20000	DS00
14	Seattle	US	Northwest Bikes	12000	UW00
15	Anklam	DE	Ostseerad	21000	DN00
16	Atlanta	US	Peachtree Bikes	4000	UE00
17	Philadelphia	US	Philly Bikes	3000	UE00
18	Stuttgart	DE	Radlelland	22000	DS00
19	Hamburg	DE	Red Light Bikes	23000	DN00
20	Denver	US	Rocky Mountain Bikes	1000	UW00
21	Palo Alto	US	Silicon Valley Bikes	10000	UW00
22	Irvine	US	SoCal Bikes	9000	UW00
23	Magdeburg	DE	Velodrom	24000	DN00

The calculation view for product is shown below

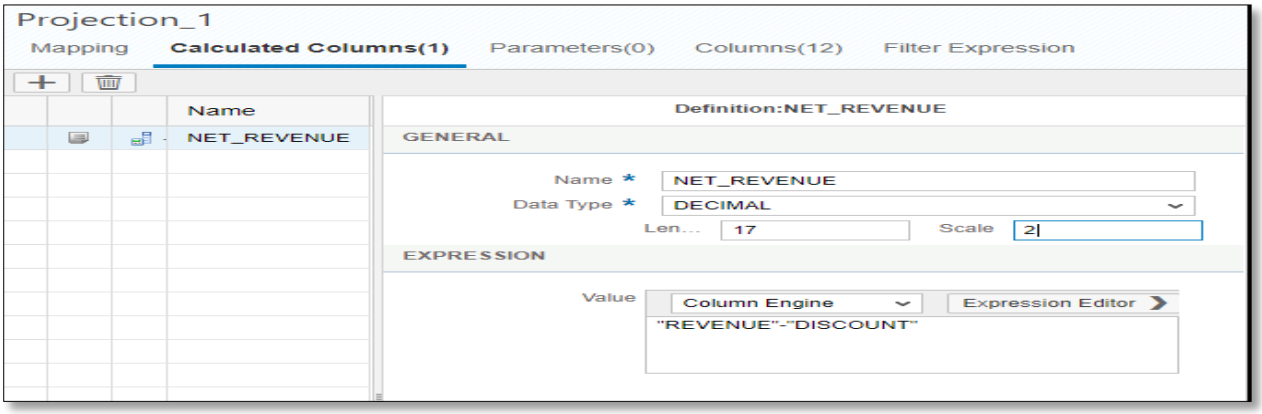


The calculation view for sales data requires a star join operation as it should also contain details of both customers and product, leading to the creation of sales data cube with transactional data which is shown below.

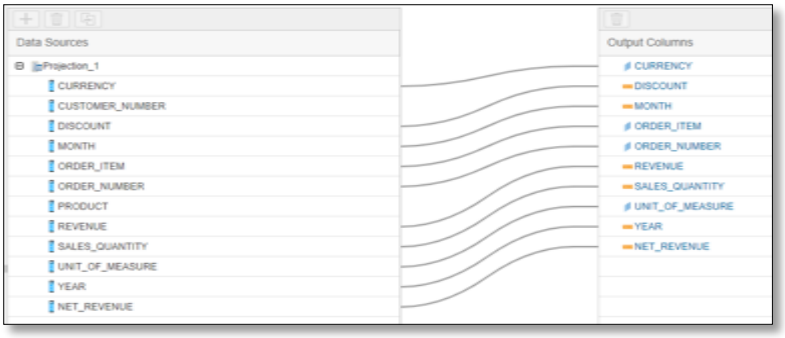




The creation of new calculated column “Net Revenue” is shown below



The necessary columns along with the calculated column “Net Revenue” are mapped from the view to the output table.



The resulted output sales data in the catalog perspective is shown below. This data is further analyzed with other business intelligence/ analytics tools to gather insights.

gbi-student-220/PROD... | SALES_CUBE_CV_22... | gbi-student-220/SAL...

Type to filter

	YEAR	PRODUCT_NAME	PRICE	DISCOUNT	NET_REVENUE	CUSTOMER_NUMBER	S...	SALE...
1	07	Deluxe Touring Bike-Red	3000	0	2450.8	19000	Germany	DN00
2	07	Professional Touring Bike-Black	3200	0	2614.18	19000	Germany	DN00
3	07	Professional Touring Bike-Red	3200	0	2614.18	19000	Germany	DN00
4	07	Men's Off Road Bike Fully	2400	0	1960.64	19000	Germany	DN00
5	07	Men's Off Road Bike Hard Tail Shima	1600	0	1307.09	19000	Germany	DN00
6	07	Road Bike Alu SRAM	1650	0	1347.94	19000	Germany	DN00
7	07	Road Bike Carbon Shimano	4000	0	3267.73	19000	Germany	DN00
8	07	Off Road Helmet	50	0	40.85	19000	Germany	DN00
9	07	Road Helmet	50	0	40.85	19000	Germany	DN00
10	07	Repair Kit	32	0	26.14	19000	Germany	DN00
11	07	Water Bottle	20	0	16.34	19000	Germany	DN00
12	07	First Aid Kit	40	0	32.68	19000	Germany	DN00
13	07	Deluxe Touring Bike-Silver	3000	0	2450.8	18000	Germany	DN00
14	07	Road Bike Alu Shimano	1700	0	1388.79	18000	Germany	DN00
15	07	Road Bike Carbon Shimano	4000	0	3267.73	18000	Germany	DN00
16	07	First Aid Kit	40	0	32.68	18000	Germany	DN00
17	07	Water Bottle Cage	18	0	14.7	24000	Germany	DN00
18	07	Deluxe Touring Bike-Black	3000	0	2450.8	24000	Germany	DN00
19	07	Professional Touring Bike-Silver	3200	0	2614.18	24000	Germany	DN00
20	07	Men's Off Road Bike Fully	2400	0	1960.64	24000	Germany	DN00
21	07	Men's Off Road Bike Hard Tail SRAM	1700	0	1388.79	24000	Germany	DN00
22	07	Road Bike Carbon Shimano	4000	0	3267.73	24000	Germany	DN00
23	07	Road Helmet	50	0	40.85	24000	Germany	DN00

9.4 Conclusion:

After working on the data modeling case study with SAP HANA system, I embraced how different views, for example calculation view, can be useful in making efficient manipulations without actually manipulating or distorting the original data stored in the data base. A prior knowledge on SQL (Structured Query Language) is expected and it demands to have a prior training to the user before using the SAP HANA system. It is a data management platform, but neither a visualization tool nor a predictive tool.

Chapter 10: Executive Summary

I have been exposed to diverse state-of-the-art business intelligence tools and data sets throughout the course of Business Analytics. From my enhanced understanding on the application of these tools, I realized, despite having some similarities, each tool is different in terms of its target users: business user, data analyst, data scientist, data processing capabilities, visualization options, predictive power, the level of user interaction, types of supported data: structured, unstructured, semi-structured, volume of data (big data) that can be handled, computational performance, accessibility, and so on.

There are wide data sources, particularly, the contemporary devices like mobile devices, location sensors, temperature sensors, social media, transportation vehicles, traffic data and so on generates data every miniscule of time. Business Analytics exposed me to how important it is to harness this massive data and translate them into actionable insights that would better the lives of people by easing their activities. Besides working with various tools with different capabilities in terms of creating story-telling dashboards, I have secured a strong foundational knowledge on geospatial and graph analytics, social analytics, text analytics, machine learning concepts of supervised learning – implementing linear regression analysis, sentiment analysis based on classification algorithms, and unsupervised learning – clustering and association analysis.

The summary of a subset of the business analytics tools that I worked in this course and which are explored in this document using the delegated case studies are as follows:

- **SAP Business Objects Explorer**

  SAP Business Objects Explorer is more like an explorative tool than a visualization tool. It is used to explore massive data sets and share insights easily. This tool lacks predictive power, but it is updated to be integrated with other tools of SAP through SAP Netweaver. This tool might serve data analysts with prior knowledge better than data scientists and might not be an excellent option for business users.

- **SAP Lumira Discovery**

  SAP Lumira Discovery is an all-in-one business intelligence platform. Though it is popular for its appealing visualization capabilities, it can be used for data discovery, exploration, data manipulations, and analysis simultaneously. SAP Lumira lacks predictive power, but can be integrated with other SAP tools easily. SAP Lumira might serve the business user and data analyst but limited to the visualization needs of data scientists.

- **SAP Analytics Cloud**

  SAP Analytics Cloud is an analytics tool solely built on the high-performance cloud for better information processing speed. It is accessible to authorized users remotely. However, the security and downtime of cloud might be the concerns. One of the key features of SAP Analytics Cloud is that it allows easier merge of data from multiple sources while maintaining live connections to the data sources. It serves the needs of business users, data analysts better than data scientists.

- **Bex (Business Explorer) Query designer and SAP Business Object Analysis for Excel**

Bex Query Designer is used to design queries to retrieve data, relevant to business question in hand, from the SAP Business Warehouse. Then SAP Business Object Analysis add-on for MS Excel is used to analyze the results of the executed query. It provides options to sort, filter, and aggregate the data. It lacks predictive and visualization capabilities. It appears to be obsolete in relation to other contemporary BI tools

- **SAP Predictive Analytics**

SAP Predictive Analytics is mainly used for its predictive capabilities while having similar features of other SAP tools discussed above. It has R-Language based machine learning algorithms that can be used for supervised and unsupervised learning tasks. However, a wide variety of algorithms can be added to further accelerate the accuracy of its predictions. This tool best serves the needs of a data scientist.

- **SAP HANA**

SAP HANA is a data management platform with in-memory data base technology that supports data storage, data discovery, data handling, data retrieval and manipulations on both transactional data and analytical (historical) data simultaneously, very quickly. It is a centralized data platform and an effective tool to serve the data needs of expanding businesses. A prior working knowledge, like SQL, is required for anyone to use SAP HANA. It might not be a BI tool but support all the business analytical tools with underlying data management, effectively.

- **Tableau**

Tableau, like SAP Lumira is widely known for its visualization capabilities. It visualizes the data with promising user interface and have easy manipulative options. Tableau support real time analytics. It stands out in terms of data processing and elegant visualization options. It also serves the needs of a data scientist.

- **IBM Cognos Insight**

IBM Cognos Insight is a business analytics tool that is not only used to analyze the data but also support easier sharing of insights across the organization without much reliance on IT. However, it does not appear to fit well with contemporary business scenarios with complex business data. Though it is very quick in presenting the result, it lacks sophisticated visualization and predictive capabilities.

- **Microsoft Excel Pivot Tables**

Pivot table is a feature of Microsoft Excel application that summarizes large data sets into informative reports. It is not a great visualization tool but has limited predictive capabilities like forecasting and can handle limited amount of data. It might serve the needs of business users and data analysts more than data scientists. MS Excel could be a better data processing tool.

In conclusion, the emergent needs for packed analytics solution containing business intelligence, data exploration, data modelling, and predictive capabilities are increasing. So, I strongly believe, being ready with prior experience working with variety of tools would help me enter and sustain in the industry with strong foundation. Hopefully, I could be strong among other job applicants for an analytics role in the industry.

References

Application Consulting Group, Inc. (n.d.). *IBM Cognos Insight*. Retrieved from https://www.acgi.com/ibm-cognos-insight.

BI, V. (2018). *SAP Lumira Discovery: An Overview by Visual BI Solutions*. Retrieved from https://visualbi.com/blogs/sap/sap-businessobjects/sap-lumira-discovery/sap-lumira-discovery-overview/.

Bliemel, M & Lee, K. Y. (2018). *Data Wrangling and Visualization case* [PDF file]. Retrieved from https://dal.brightspace.com/d2l/le/content/102286/viewContent/1371370/View

Intellipaat. (2019). *What is Tableau - Uses and Applications*. Retrieved from https://intellipaat.com/blog/what-is-tableau/.

Jewett, E. (2019). *Overview of SAP HANA features and how it's being used today*. Retrieved from https://searchsap.techtarget.com/tip/Overview-of-SAP-HANA-features-and-how-its-being-used-today.

Kalé, N., Jones, N. (2016). *Chapter 5: Exercise 1 Using excel pivot tables for analytics* [PDF file]. Retrieved from https://dal.brightspace.com/d2l/le/content/102286/viewContent/1371361/View

Kalé, N., & Jones, N. (2016). *Chapter 11: Exercise 2 Segmenting stores using clustering* [PDF file]. Retrieved from https://dal.brightspace.com/d2l/le/content/102286/viewContent/1477856/View

Kalé, N., Jones, N. & Lee, K. Y. (2017). *Chapter 5: Exercise 2(T) Data manipulation for analysis* [PDF file]. Retrieved from https://dal.brightspace.com/d2l/le/content/102286/viewContent/1456524/View

Kalé, N., Jones, N., & Lee, K. Y. (n.d.). *Data Visualization using SAP Analytics Cloud* [PDF file]. Retrieved from https://dal.brightspace.com/d2l/le/content/102286/viewContent/1451415/View

Kalé, N., Jones, N., Bliemel, M & Lee, K. Y. (2016). *Extracting multidimensional data using SAP Business Objects Analysis for Excel and Business Explorer Query Designer* [PDF file]. Retrieved from https://dal.brightspace.com/d2l/le/content/102286/viewContent/1440948/View

Kalé, N., Jones, N., Bliemel, M., & Lee, K. Y. (2018). *Chapter1: Exercise 1 Data Exploration with SAP Explorer* [PDF file]. https://dal.brightspace.com/d2l/le/content/102286/viewContent/1371366/View

IBM workshop: Personal Analytics with Cognos Insight [PDF file]. (n.d.). Retrieved from https://dal.brightspace.com/d2l/le/content/102286/viewContent/1440949/View

Rouse, M., & O'Donnell, J. (n.d.). *What is SAP Predictive Analytics?*. Retrieved from https://searchsap.techtarget.com/definition/SAP-Predictive-Analytics.

SAP Analytics Cloud for newbies. (n.d.). Retrieved from https://blogs.sap.com/2019/04/01/sap-analytics-cloud-for-newbies/.

SAP Analytics Cloud: End-to-end Analytics for the Intelligent Enterprise. (n.d.). Retrieved from https://www.sapanalytics.cloud/.

SAP HANA Data Modelling Case study Business Scenario [PDF file]. (n.d.). Retrieved from https://dal.brightspace.com/d2l/le/content/102286/viewContent/1487821/View

Sevillano, A. (2014). *Running SAP Business Objects Explorer on top of SAP BW*. Retrieved from https://www.clariba.com/blog/tech-20140218-running-sap-businessobjects-explorer-on-top-of-sap-bw-alvaro-sevillano.

Tremblay, T. (2019). *Online Database Software*. Retrieved from https://www.kohezion.com/blog/what-is-a-pivot-table-examples-and-uses/.

Wailgum, T. (2019). *SAP Predictive Analysis: What It Can and Cannot Do*. Retrieved from https://www.asug.com/news/sap-predictive-analysis-what-it-can-and-cannot-do.