Business Analytics Portfolio

Jafar Pashami

APRIL 2022

Table of Contents

Executive Summary 3

Chapter 1 – Microsoft Excel Pivot Table 4

About Excel Pivot Table 4

Dataset and Research Questions 4

Applying the analytical tool and results 4

1. What customer has the highest percentage contribution to revenue? What has been the trend of that customer's percentage contribution over the years? 4

2. Which material does not display significant seasonality? 5

3. What year had the highest overall gross margin? What was the gross margin during that year? 5

4. What was Germany's gross margin (as a percentage of sales) for the year with the highest gross margin? What was the US's gross margin (as a percentage of sales) during the same year? 6

5. What is the revenue trend over the years by sales Organizations? Is there any market loss in any branches? 6

Analysis and Critique of the tool 7

Conclusion 7

Chapter 2 – SAP Lumira Discovery 8

About SAP Lumira Discovery 8

Dataset and Research Questions 8

a) FIFA WorldCup 2014 8

b) Canadian alcohol purchasing preferences 8

Applying the analytical tool and results 8

a) FIFA WorldCup 2014 8

b) Canadian alcohol purchasing preferences 10

Analysis and Critique of the tool 12

Conclusion 12

Chapter 3- Tableau public visualization 13

About Tableau 13

Dataset and Research Question 13

Applying analytical tools and results 13

Analysis and critique of the tool 16

Conclusion 16

Chapter 4 – SAP Analytics Cloud (SAC) 17

About SAP Analytics Cloud 17

Dataset and Research Question 17
Research Questions 17
Applying analytical tools and results 17
Analysis and critique of the tool 20
Conclusion 20
Chapter 5 – Microsoft Power BI 21
About Power BI 21
Dataset and Research Question 21
Dataset 21
Research Questions 21
Applying analytical tools and results 22
Analysis and critique of the tool 24
Conclusion 24
Chapter 6: SAP Predictive Analytics 25
About SAP Predictive Analytics 25
Data Set and Research Questions 25
Applying the analytical Tool and Results 25
Data preparation 25
Data Analyzing 26
Conclusion 29
Chapter 7: SAP Lumira's Geospatial Analysis 30
About SAP Lumira's geospatial features 30
Data Set and Research Questions 30
Applying the analytical Tool and Results 30
1. What is the trend of crime incidents and their categories in Calgary between 2012 and 2017? 30
2. Which communities have more crime incidents? 31
.3 What types of crimes are most common in each community? 32
4. Are the police stations well distributed to respond quickly to high crime areas? Are there any communities that may not be well served? 33
Conclusion 34
References 35

Executive Summary

This portfolio includes my experience in implementing seven practical tools for business analytics and data visualization. These tools are mainly one of the most widely used tools. Each software and its capabilities are explained in separate chapters in this portfolio.

In each chapter, a brief description of the tool is provided, and then the datasets used are explained, followed by questions that are answered using the charts and diagrams of that chapter. In the end, a summary of each tool's capabilities and limitations compared to other software is provided.

These tools are as below:

1. Microsoft Excel Pivot Table
2. SAP Lumira Discovery
3. Tableau public visualization
4. SAP Analytics Cloud (SAC)
5. Microsoft Power BI
6. SAP Predictive Analytics
7. SAP Lumira's Geospatial Analysis

Chapter 1 – Microsoft Excel Pivot Table

About Excel Pivot Table

The most common tool used to slice and dice data is a spreadsheet. Typically slicing and dicing is done by creating a **pivot table** which creates what is called a **cross tabulated structure or crosstab.** The crosstab displays aggregated and summarized data based on which way you sort the columns and rows. The pivot table can be easily transposed by swapping the axes; hence the name, pivot.

Dataset and Research Questions

GBI dataset is available in *GBI_E5_1.xlsx*

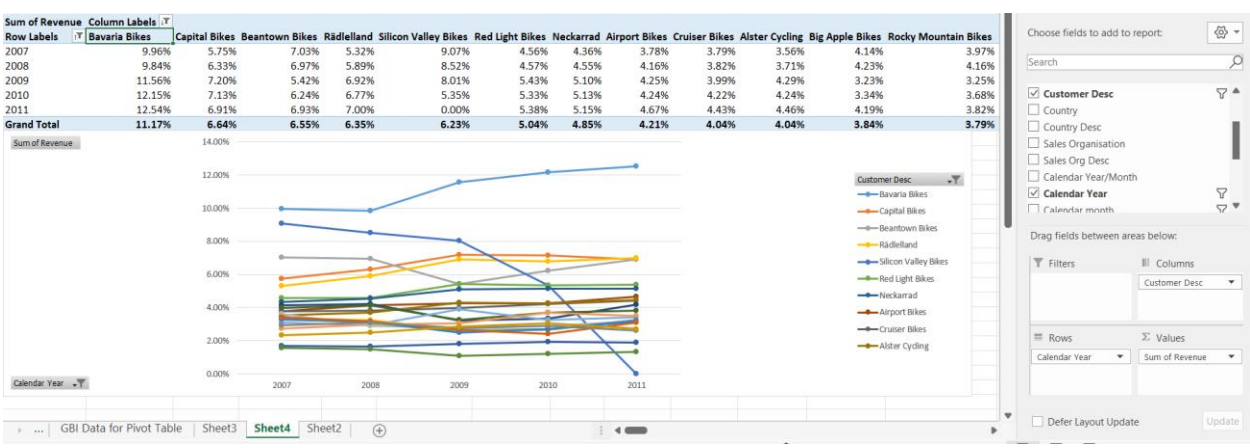
It is a dataset about a fictitious company's sales; Global Bike (GB). It provides sales information of this company from 2007 to 2011 monthly and for different sales organizations, products, and customers. It also contains sales quantity as well as sales amount and discount. No data cleaning is required. Using Pivot Table, we will answer these questions:

1. What customer has the highest percentage contribution to revenue? What has been the trend of that customer's percentage contribution over the years?
2. Which material does not display significant seasonality?
3. What year had the highest overall gross margin? What was the gross margin during that year?
4. What was Germany's gross margin (as a percentage of sales) for the year with the highest gross margin? What was the US's gross margin (as a percentage of sales) during the same year?
5. What is the revenue trend over the years by sales Organizations? Is there any market loss in any branches?

Applying the analytical tool and results

1. What customer has the highest percentage contribution to revenue? What has been the trend of that customer's percentage contribution over the years?

In Pivot table builder, Customer Desc was added to Columns while Revenue and Calendar Year remained in Value and Rows respectively. This created a crosstab from which a line chart was inserted. For each customer, to show the sales amount as a percentage of total sales in each year, the option of **as % of Row Total** is selected from the Value Field Setting. Then Customers data was sorted in descending order of Revenue.



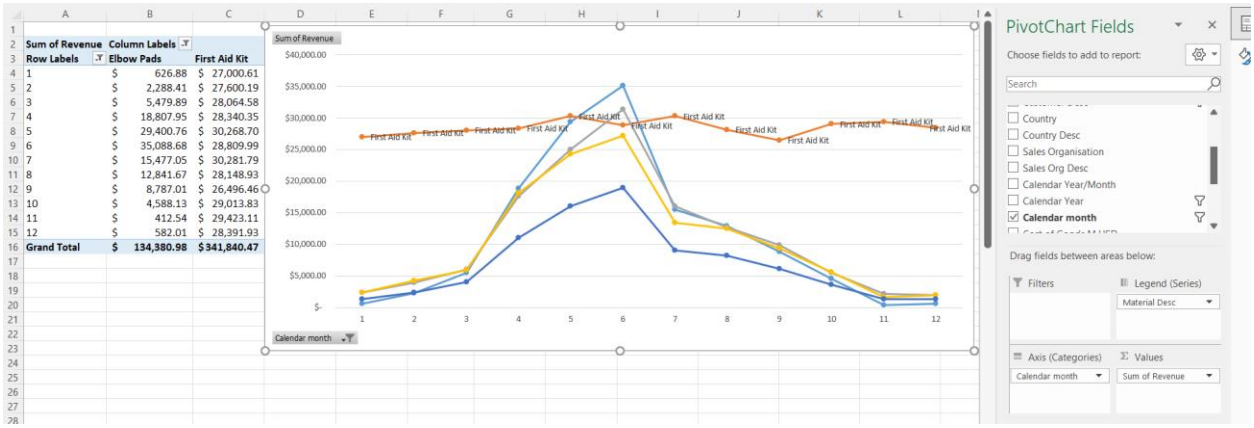
Sum of Revenue / Row Labels	Column Labels: Bavaria Bikes	Capital Bikes	Beantown Bikes	Rädlelland	Silicon Valley Bikes	Red Light Bikes	Neckarrad	Airport Bikes	Cruiser Bikes	Alster Cycling	Big Apple Bikes	Rocky Mountain Bikes
2007	9.96%	5.75%	7.03%	5.32%	9.07%	4.56%	4.36%	3.78%	3.79%	3.56%	4.14%	3.97%
2008	9.84%	6.33%	6.97%	5.89%	8.52%	4.57%	4.55%	4.16%	3.82%	3.71%	4.23%	4.16%
2009	11.56%	7.20%	5.42%	6.92%	8.01%	5.43%	5.10%	4.25%	3.99%	4.29%	3.23%	3.25%
2010	12.15%	7.13%	6.24%	6.77%	5.35%	5.33%	5.13%	4.24%	4.22%	4.24%	3.34%	3.68%
2011	12.54%	6.91%	6.93%	7.00%	0.00%	5.38%	5.15%	4.67%	4.43%	4.46%	4.19%	3.82%
Grand Total	**11.17%**	**6.64%**	**6.55%**	**6.35%**	**6.23%**	**5.04%**	**4.85%**	**4.21%**	**4.04%**	**4.04%**	**3.84%**	**3.79%**

As can be seen in the chart, there have been many changes in the sales share of some customers. Bavaria Bike's sales share increased from 2007 to 2011, while Silicon Valley Bike sales share decreased significantly

during these years, reaching zero in 2011. Bavaria Bike has the largest share among Global Bike customers with about 12.5%.

2. Which material does not display significant seasonality?

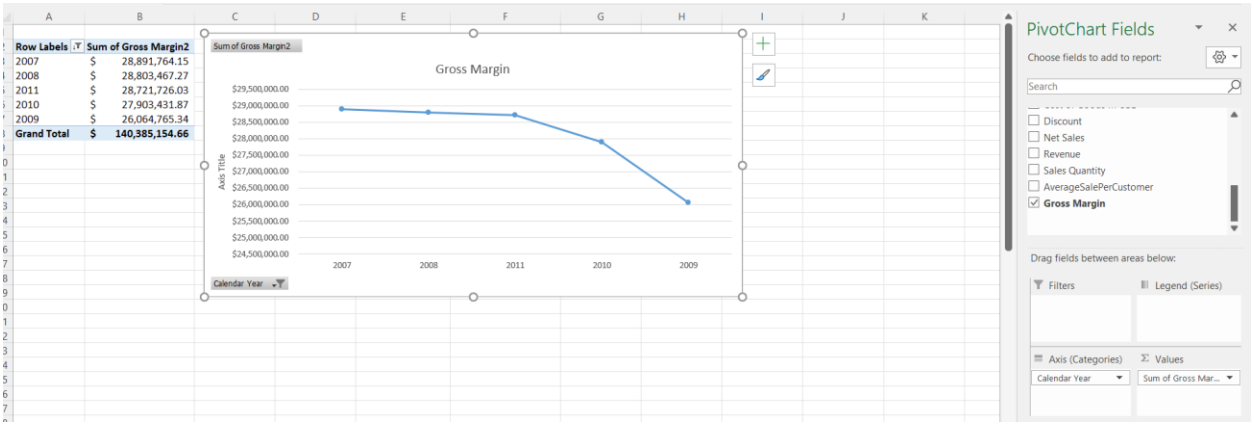
In the PivotTable Builder, Material Desc was included in Columns. Calendar Month in Rows. Revenue in Values. After forming a cross table, a Pivot Chart was selected, and a Line chart was chosen. Then for the links in the PivotChart, the highest peak became hidden by Right click à Filter à Hide Selected Item. By removing the next highest material until we see a low seasonality material (relatively flat line), we can find a product (material) that does not show significant seasonality.



According to the above diagram, it can be understood that the first aid kit has had the least seasonal fluctuations during different months.

3. What year had the highest overall gross margin? What was the gross margin during that year?

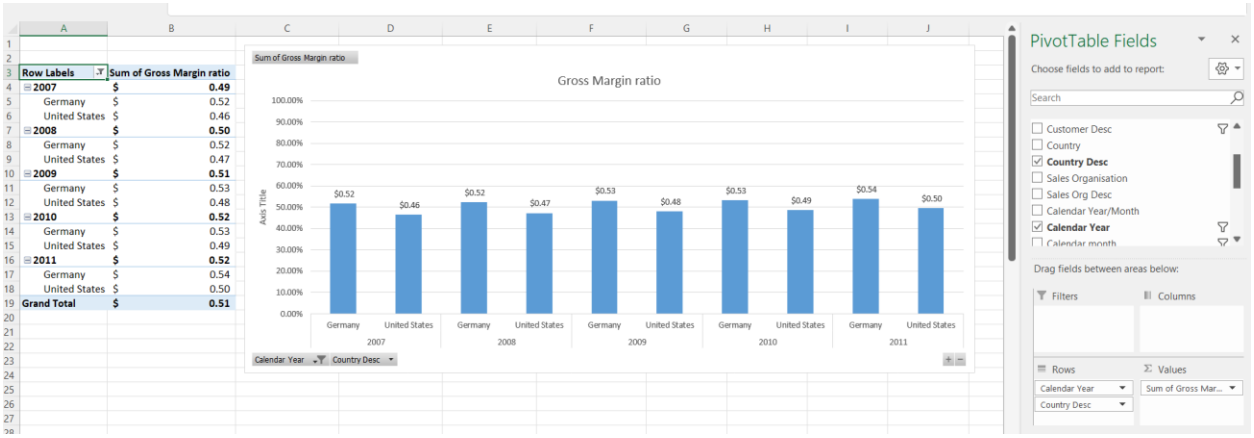
To answer this question, we need to define a new calculated measure called gross margin. To do this, the menu **fields, items and sets** is used in the PivotTable Analysis tab. Then, in the PivotTable Builder, Calendar Year was included in Rows and Gross Margin in Values. Then the Calendar Year has been sorted Decendingly based on Gross Margin.



As you can see in above diagram, 2007 is the year with the highest Gross Margin. The amount of Gross Margin in this year is about 28891764 USD.

4. What was Germany's gross margin (as a percentage of sales) for the year with the highest gross margin? What was the US's gross margin (as a percentage of sales) during the same year?

Again, to answer this question, we need to define a new calculated measure. It is the gross Margin over Net Sales. We called it **Gross Margin Ratio**. We can demonstrate the following table and chart by adding Calendar Year and Country Desc in Rows and Gross Margin Ratio in Value.

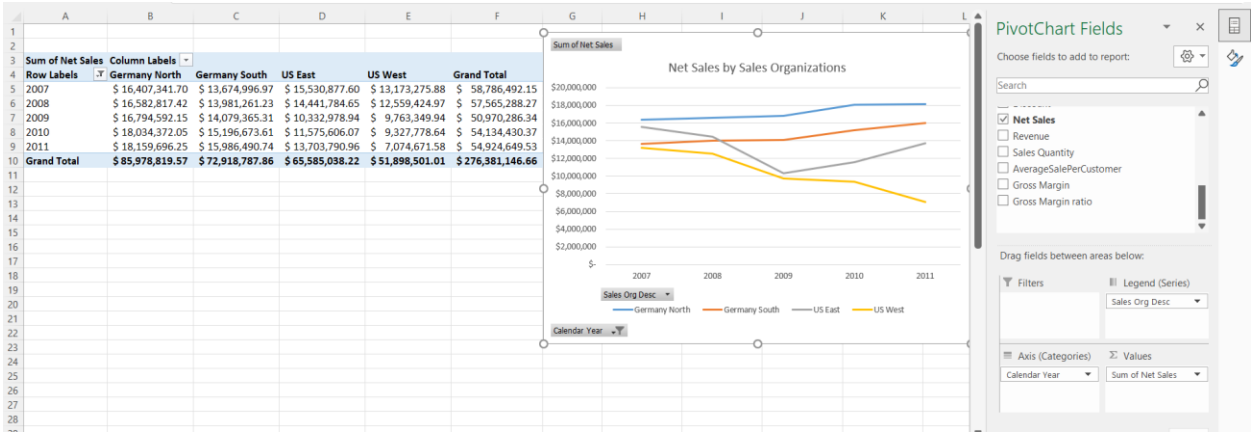


Row Labels	Sum of Gross Margin ratio
2007	$ 0.49
Germany	$ 0.52
United States	$ 0.46
2008	$ 0.50
Germany	$ 0.52
United States	$ 0.47
2009	$ 0.51
Germany	$ 0.53
United States	$ 0.48
2010	$ 0.52
Germany	$ 0.53
United States	$ 0.49
2011	$ 0.52
Germany	$ 0.54
United States	$ 0.50
Grand Total	$ 0.51

The above table (& chart) illustrate the Sales gross margin ratio for Germany and United States. We can compare this data for different years.

5. What is the revenue trend over the years by sales Organizations? Is there any market loss in any branches?

In Pivot table builder, Sales Organizations Desc was added to Columns while Net Sales and Calendar Year remained in Value and Rows respectively. This created a crosstab from which a line chart was inserted.



Sum of Net Sales	Column Labels				
Row Labels	Germany North	Germany South	US East	US West	Grand Total
2007	$ 16,407,341.70	$ 13,674,996.97	$ 15,530,877.60	$ 13,173,275.88	$ 58,786,492.15
2008	$ 16,582,817.42	$ 13,981,261.23	$ 14,441,784.65	$ 12,559,424.97	$ 57,565,288.27
2009	$ 16,794,592.15	$ 14,079,365.31	$ 10,332,978.94	$ 9,763,349.94	$ 50,970,286.34
2010	$ 18,034,372.05	$ 15,196,673.61	$ 11,575,606.07	$ 9,327,778.64	$ 54,134,430.37
2011	$ 18,159,696.25	$ 15,986,490.74	$ 13,703,790.96	$ 7,074,671.58	$ 54,924,649.53
Grand Total	$ 85,978,819.57	$ 72,918,787.86	$ 65,585,038.22	$ 51,898,501.01	$ 276,381,146.66

It seems that while the company's sales in Germany are growing, there are problems in the United States, and the company is losing market share in the western United States. The company's market share in the eastern United States has been declining since the beginning of 2007 to 2009, but since 2009, the company's resale in the eastern United States has increased, while the company has not been able to increase its market share in the western United States. The company's sales in the western United States have been declining since 2007.

Analysis and Critique of the tool

Excel is a software available to everyone and because of its ease of use is present in almost all personal computers and therefore many people are familiar with it. Also, a lot of small data is stored in the structure of Excel or CSV files that Excel and PivotTable can analyze.

Performing various mathematical and statistical calculations is easily possible in the Excel spreadsheet. In addition, the pivot table is straightforward to understand and simple, and it is also possible to use a drop and drop.

The most important features of this software are its simplicity and availability, and you can efficiently perform various analyzes of your available data using PivotTable Excel.

Running this software also does not require hardware with high capabilities and runs quickly.

The limitations of this software include the following:

1. There are limitations on the amount of data stored and processed in an Excel file.

2. Some charts are not active in PivotTable and cannot be worked with them. In general, Excel data visualization capabilities are less than other software in this field.

3. Calculated measures are not easily generated and used, and applying DAX formulas is not user-friendly when you create measures.

Conclusion

Excel is always simple and reliable, and I always use it as the first in-hand tool to perform quick analyzes.

Chapter 2 – SAP Lumira Discovery

About SAP Lumira Discovery

SAP Lumira Discovery is used by key business users who need flexibility to connect to data (usually multiple data sources), shape that data, and create ad-hoc stories with visualizations from all types of data, which others can leverage, build upon, and share. Accelerate decisions by creating impactful interactive maps, charts, infographics, dashboards, and business intelligence (BI) applications with ease. (SAP.com)

Dataset and Research Questions

I have provided to source data that I have worked on them:

a) FIFA WorldCup 2014

This data is taken from the FIFA website to look at the statistics generated during the last world cup in Brazil. It provides some statistics about performance of some football players in Fifa WorldCup 2014 in Brazil.

The question is Trying to come up with a formula for a calculated dimension that you can defend as the way to evaluate who is the Best attacker that you can visualize and defend?

b) Canadian alcohol purchasing preferences

Our business problem is "how do alcohol purchasing preferences vary by province, over the last several years?"

The sales are through provincial retailers and there is high-quality open data on annual sales by product type, as well as data on demographics by province reported by Statistics Canada.

Data source for Value of sales of alcoholic beverages:

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1010001101

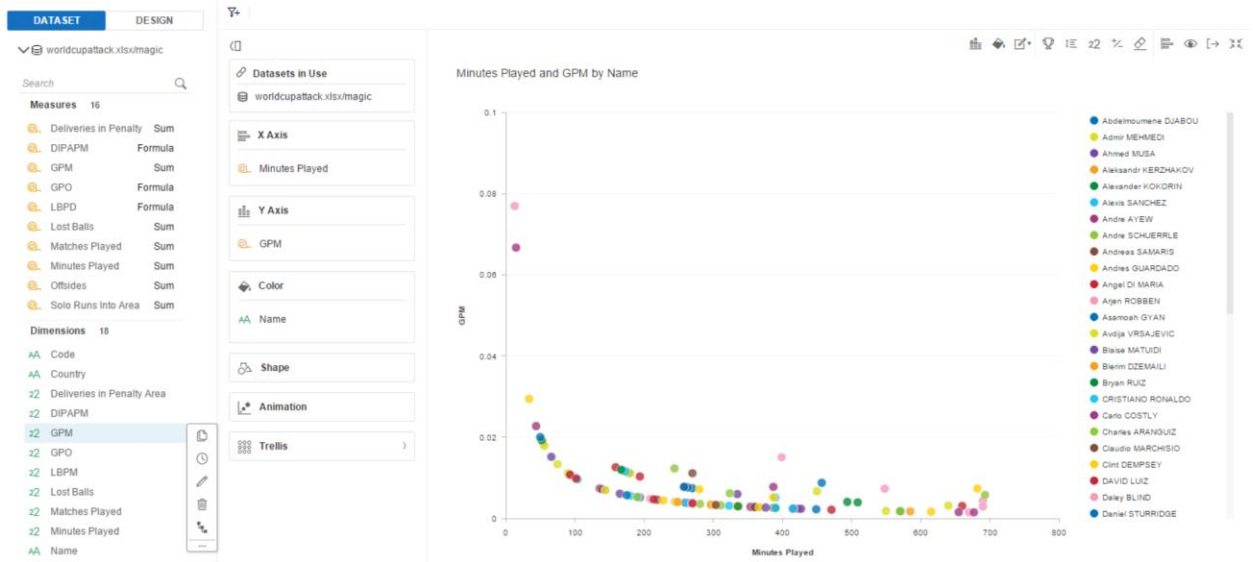
For Creating per capita measure we need to add population data also. We can apply these facts when working with the data so that we can compare how much of which product type is purchased divided by the population of over 19-year-olds (18 in Alberta, Manitoba and Quebec) to get per capita comparisons.

Data source for population: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501

Applying the analytical tool and results

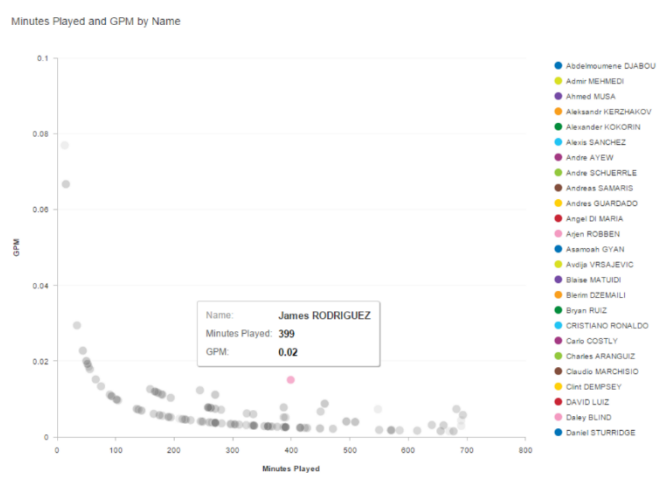
a) FIFA WorldCup 2014

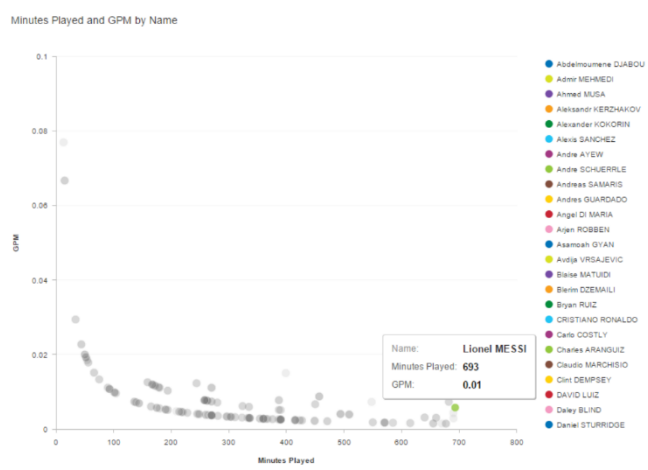
We added a Calculated dimension called GPM (Goals-per minutes). Then for creating this diagram, we put the GPM measure on Y-axis and Minute Played on the X-axis. We chose a scattered chart for this analysis.



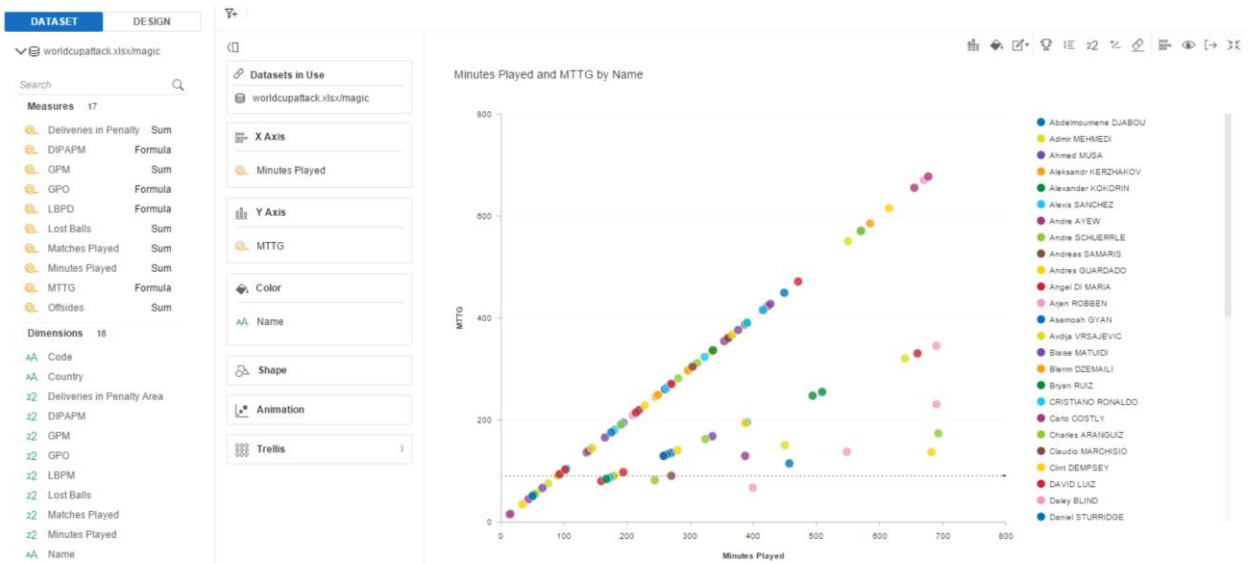
As you can see in the above chart, there is some **multiplicative inverse** curve fit on the data. Alignment curves can be seen in this diagram. These alignment curves are practically an indicator for ranking attackers. This chart is helpful because it excludes the effect of the minutes that players have compared.

You can see the names of Lionel Messi and Tomas Moler are located at the right of the chart, upper than other payers. And also, James Roudrigoez, who is much younger, recorded a fantastic performance based on this diagram.



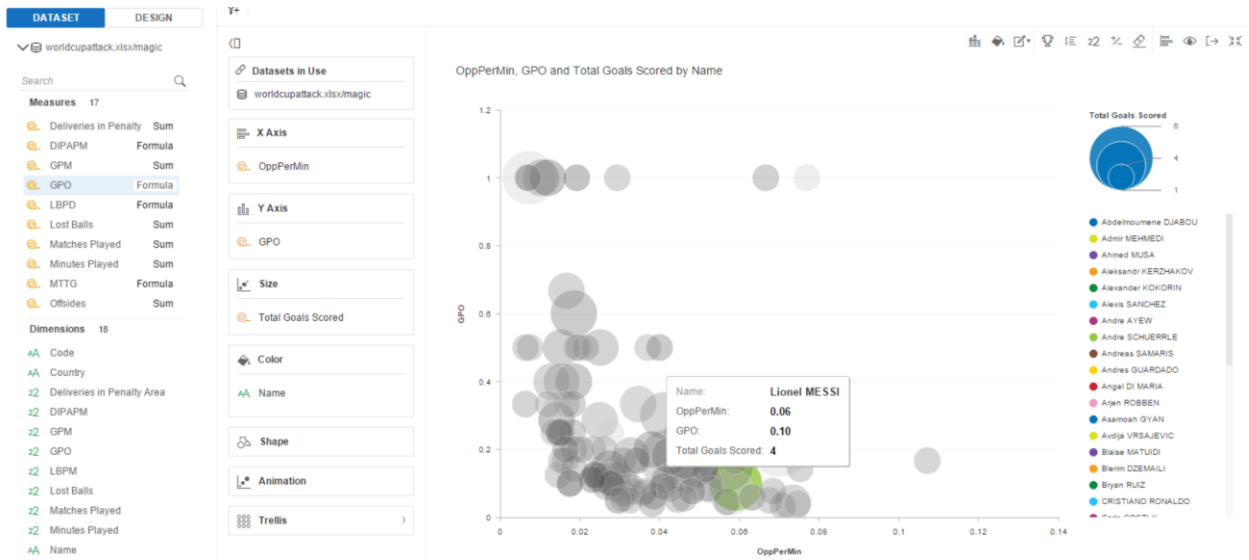


Then I changed my view and defined a new calculator instead of GPM. I named this measure MTTG (Mean time to Goal). MTTG = Minutes played / Goals scored. You can see this chart:



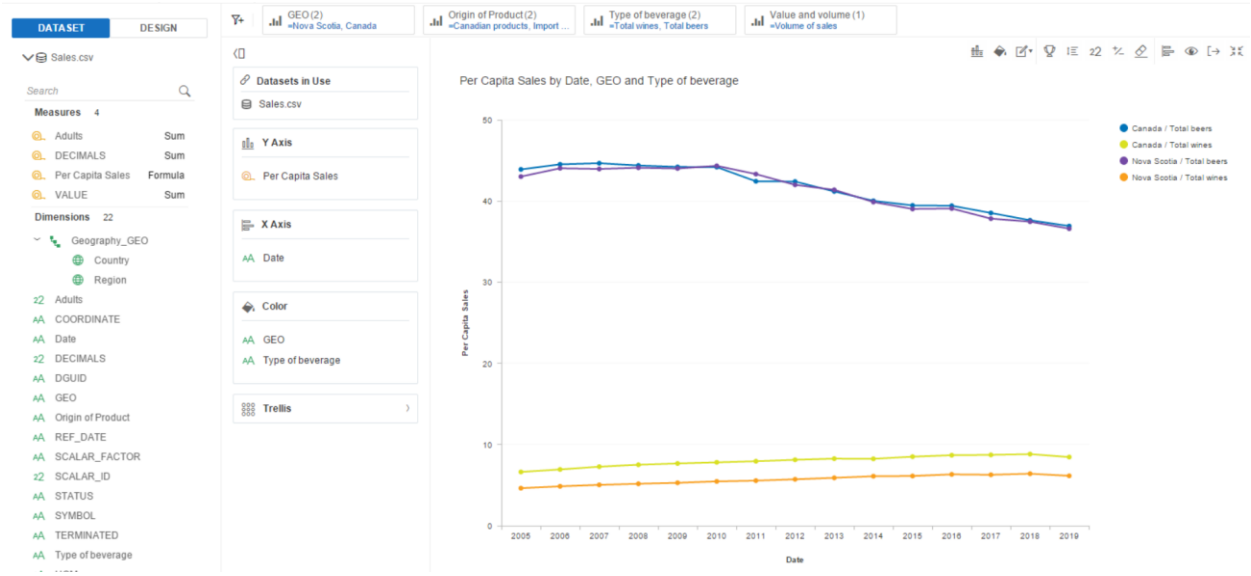
In this chart, to any player who is on a level playing field with a lower slope, it means that he has a higher prominence in scoring goals.

This is another graph that shows all the goals scored in terms of all the positions of receiving the ball in the vertical axis and also shows the number of positions of receiving the ball in terms of the minutes of the game in the horizontal axis. All positions for receiving the ball include all receptions in the penalty area, all tackles and all individual moves.



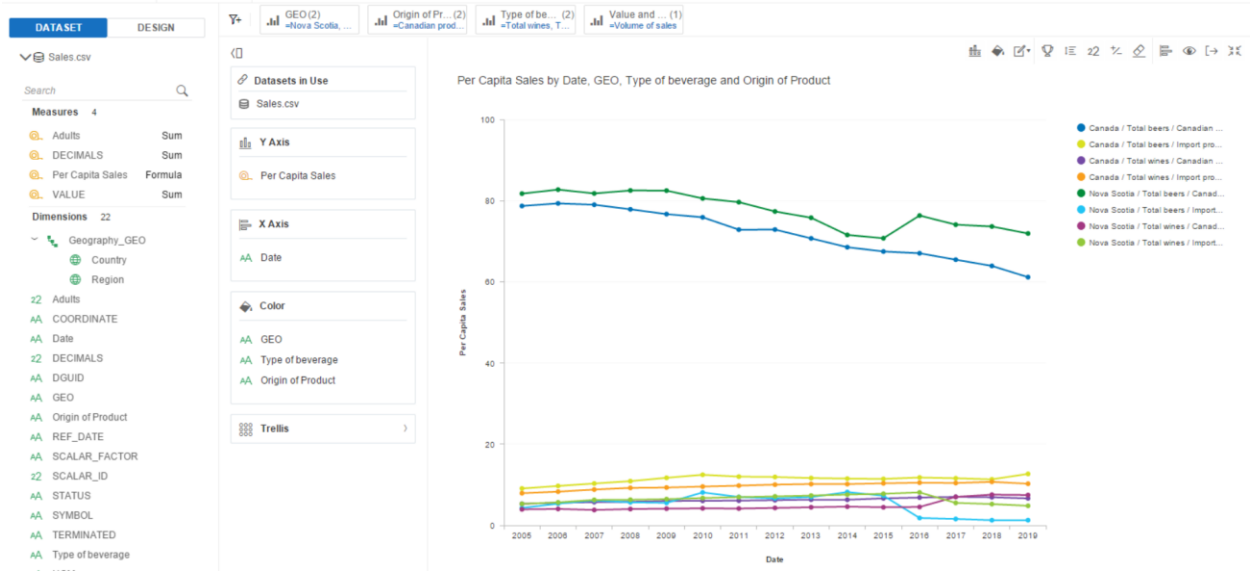
b) Canadian alcohol purchasing preferences

Using Lumira Discovery we can draw some different charts that show alcohol purchasing preferences:



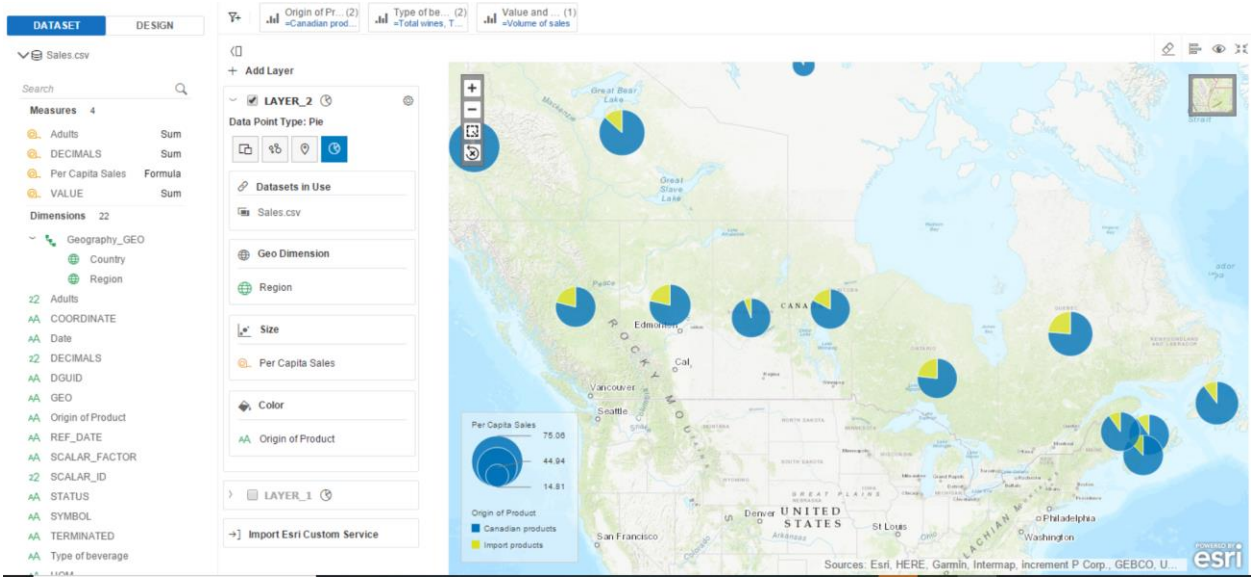
The above figure shows that beer consumption per capita in Nova Scotia is the same as Canada's Beer consumption per capita. But Canada's wine consumption per capita is much bigger than nova scotia.

But in the following figure, it is shown that Nova Scotians consume more local beer rather than the Canadian average.

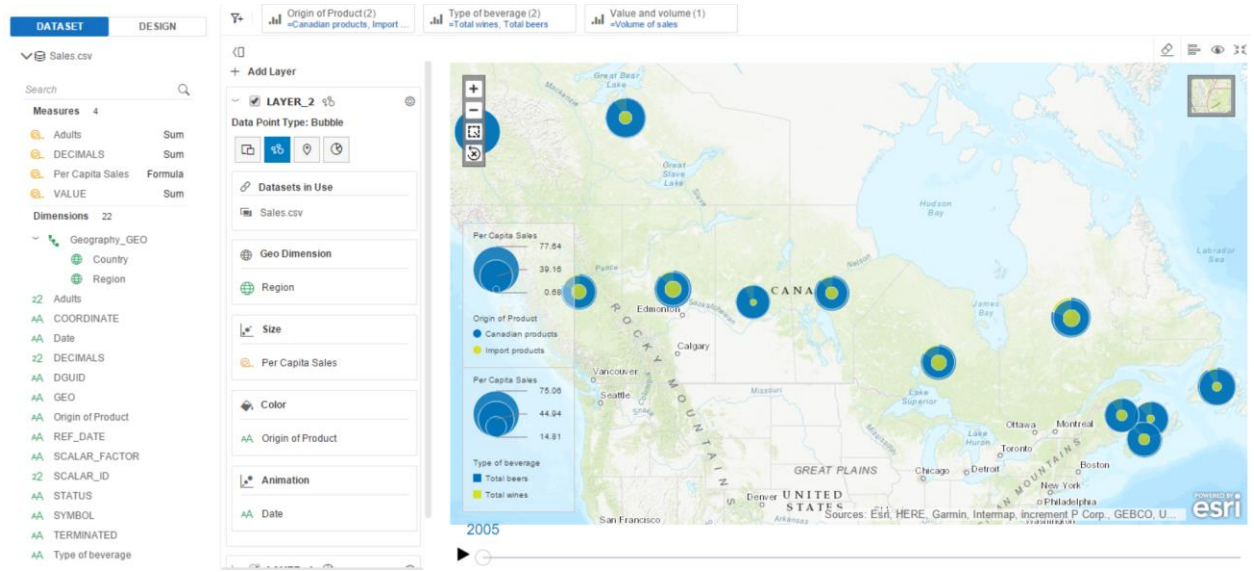


After 2016 Noca Scotia bear per capita from imported products decreased dramatically.

In the following figure, the combination of imported and local products for all provinces of Canada has been shown on the map.



Lumira Discovery can animate data using a bubble chart type. In the following picture, an animation attribute based on date dimension has been added to the chart. Moreover, you can add layers to your map.



Analysis and Critique of the tool

SAP Lumira software combines self-service BI discovery and visualization with interactive dashboards and analytic applications. It can support multiple story-telling of data visualization as well as different chart types. It has a feeding panel that easily supports drag and drop for fast configuring and changing charts. You can add calculated dimensions and measures straight forward.

On the other have, I found that this software is not fast as it should be in loading figures or changing the filters.

Conclusion

SAP Lumira Discovery is a tool that has all capabilities needed for a visualization tool.

Chapter 3- Tableau public visualization

About Tableau

Tableau is a data visualization software that was founded in 2003 as a result of a computer science project at Stanford. Its popularity is increasing because of its ease-of-use interface and its powerful and stable environment as well as its cloud facilities for saving and publishing the visualization reports. The goal of the creators of this software is to create insight using analysis and display of data for everyone. Therefore, the audience of this software is not only people with technical skills and different users of a business with different specialties can also benefit from its capabilities. After a successful IPO in 2013, Tableau was acquired by Salesforce for $15.7B on Jun 10, 2019 (Crunchbase, n.d.). The total investment for building Tableau before acquisition has been $15M. (Crunchbase, n.d.)

Dataset and Research Question

I downloaded the updated data set from The World Bank website (CO2 emissions (kt), 2022)

The data set is CO2 emission data provided by World Bank from 1960 to 2018.

Link to the data set: https://data.worldbank.org/indicator/EN.ATM.CO2E.KT

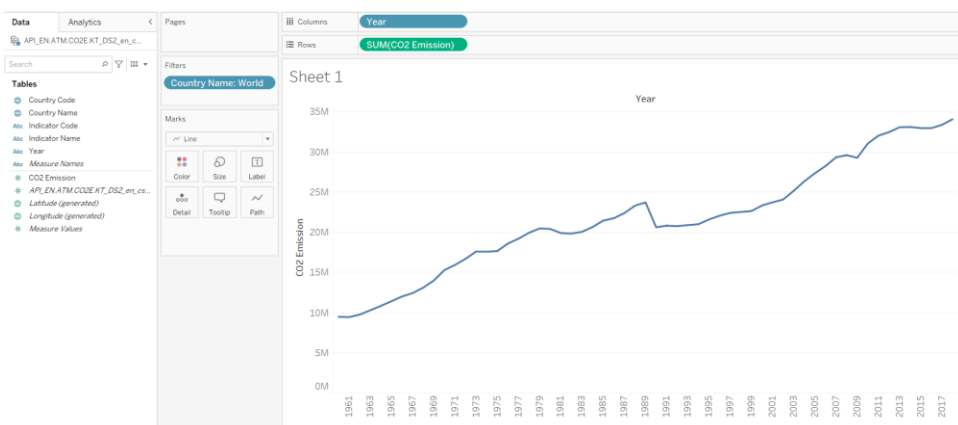
Research questions are:

1. What is the trend of CO2 emission level by years for selected countries?
2. What is the accumulative CO2 emission since 1960 in different countries?
3. What were the top ten countries with the highest amount of CO2 emission in 2018?
4. Are there any relations between income and CO2 emission?

Applying analytical tools and results

First, after importing data, we need to do data preparation and cleansing. For doing that Tableau provides a Data Interpreter to do this task. It recognizes table headers and data structure. Then we must transform the data set from Wide form to Long-form. To do that after selecting the columns related to years, we can apply pivot to create the years' column for proceeding with our analytical tasks.

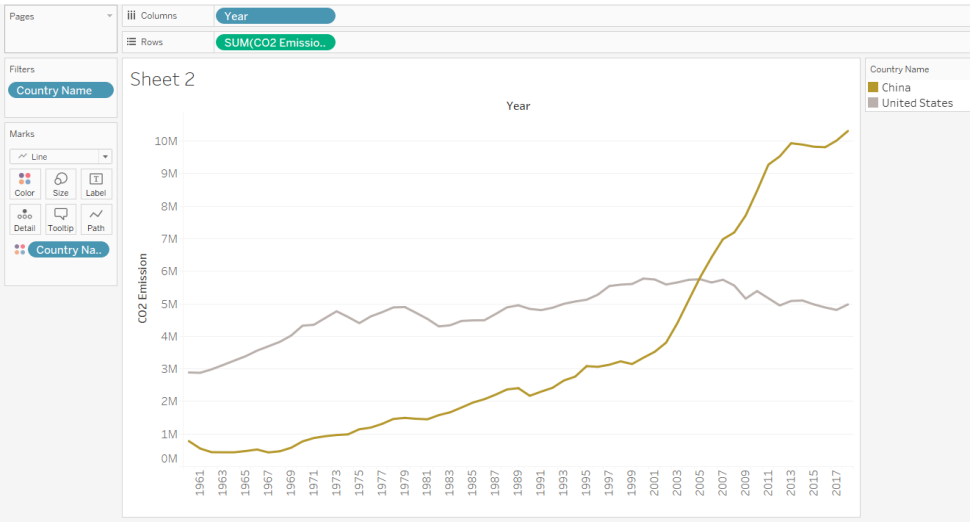
By locating CO2 Emission measure and year dimension and filtering country name as World we can have the below diagram:



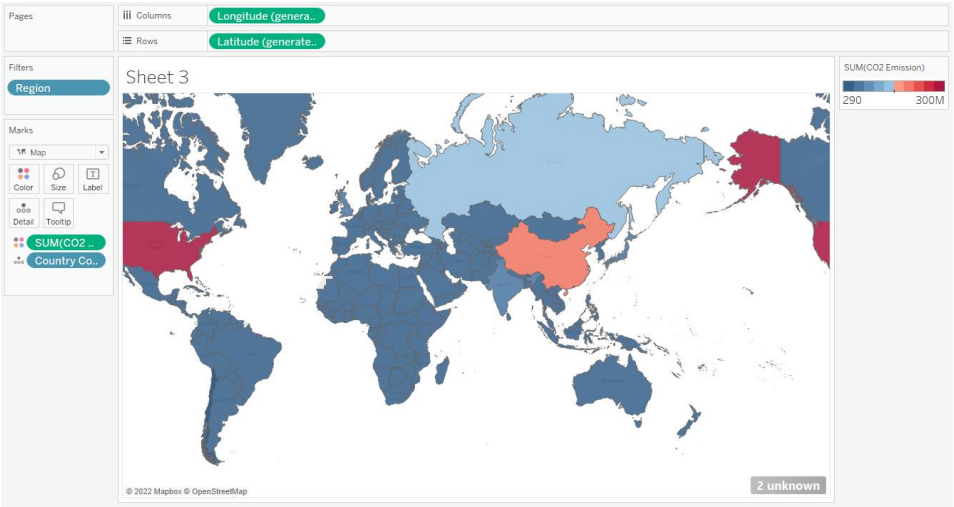
As you see here, there is a drop in 1990, it backs to changing in data sources that is mentioned in metadata file: "Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data:

Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions." (CO2 emissions (kt), 2022)

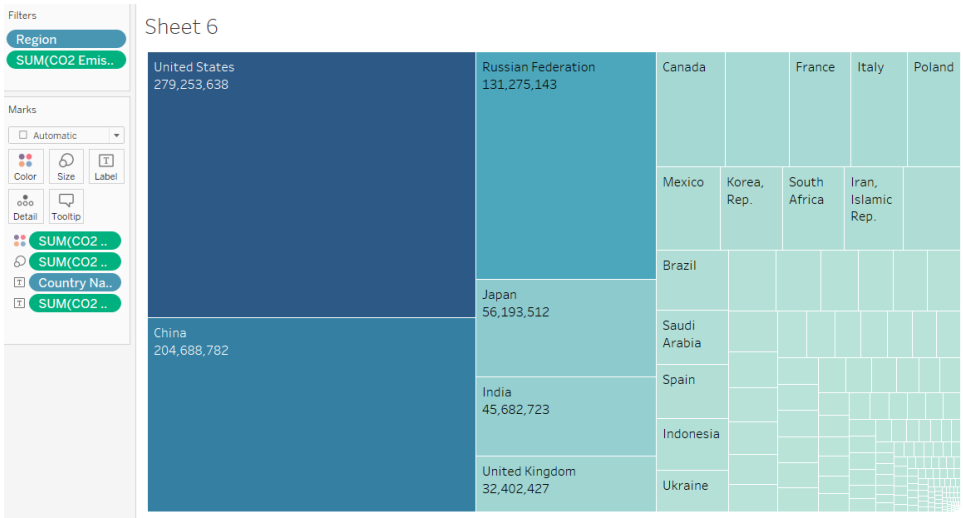
Then I filtered country names to the United States and China.



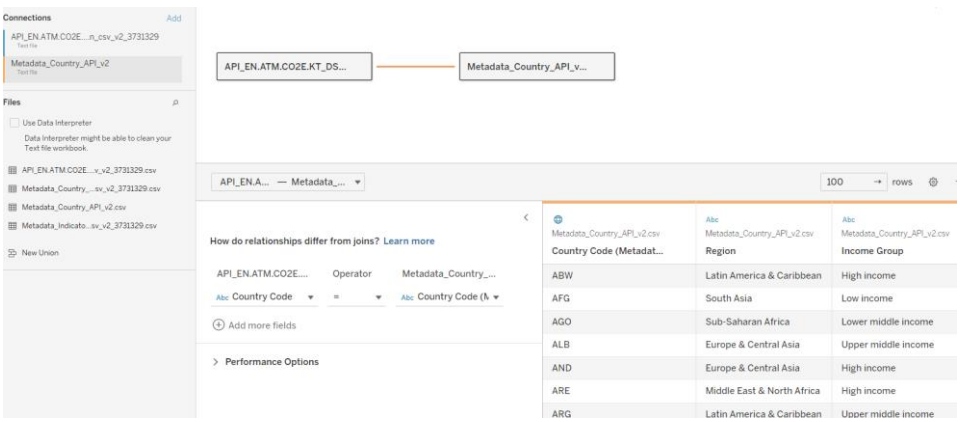
China's CO2 emission has surpassed the USA and It reaches more than double rather than United States CO2 emission. We can see the accumulative emission on the map:



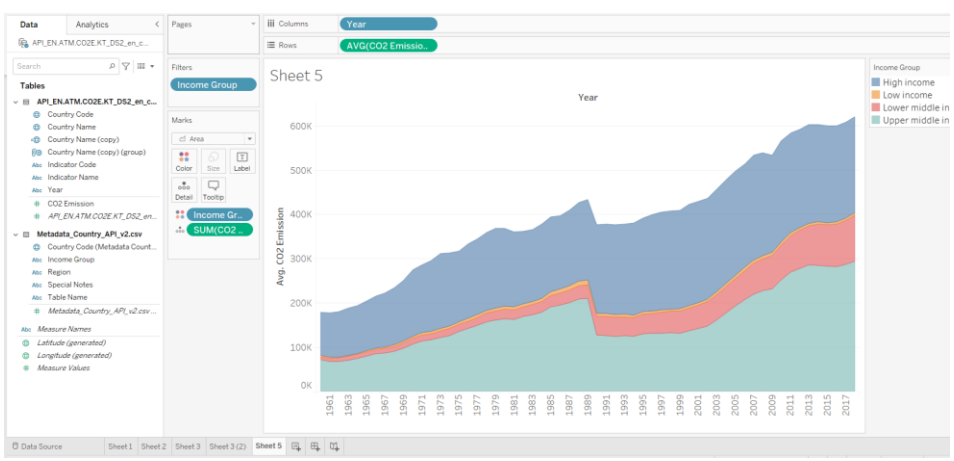
From historical prespective, the united states by emitting about 280 Billion tons of CO2 emission is the top of the countries in this list and china has the second position by about 200 Billion tons of CO2 emission.



This Treemaps shows the name of the countries how have the most emission. For having more dimensions, I added Metadata about countries' regions and income to my data source:



Now:



We can see that middle-income countries, especially high-middle-income countries, have less control over CO2 emissions, and therefore it seems that the participation of these countries is as important as the participation of other high-income countries.

Analysis and critique of the tool

Based on what I did using Tableau, the following points seemed to me to be its pros and cons:

pros:

- Ease of use: for most of the applications it provides a straightforward way to create a visualization.
- Easy to share in Tableau public is a crucial element when you are working as a team, or you want to have a repository for your data.
- Compatibility with Mac: It can be used on MacBook in addition to windows and from the web.
- Separating the worksheets from the dashboard was also one of the advantages.

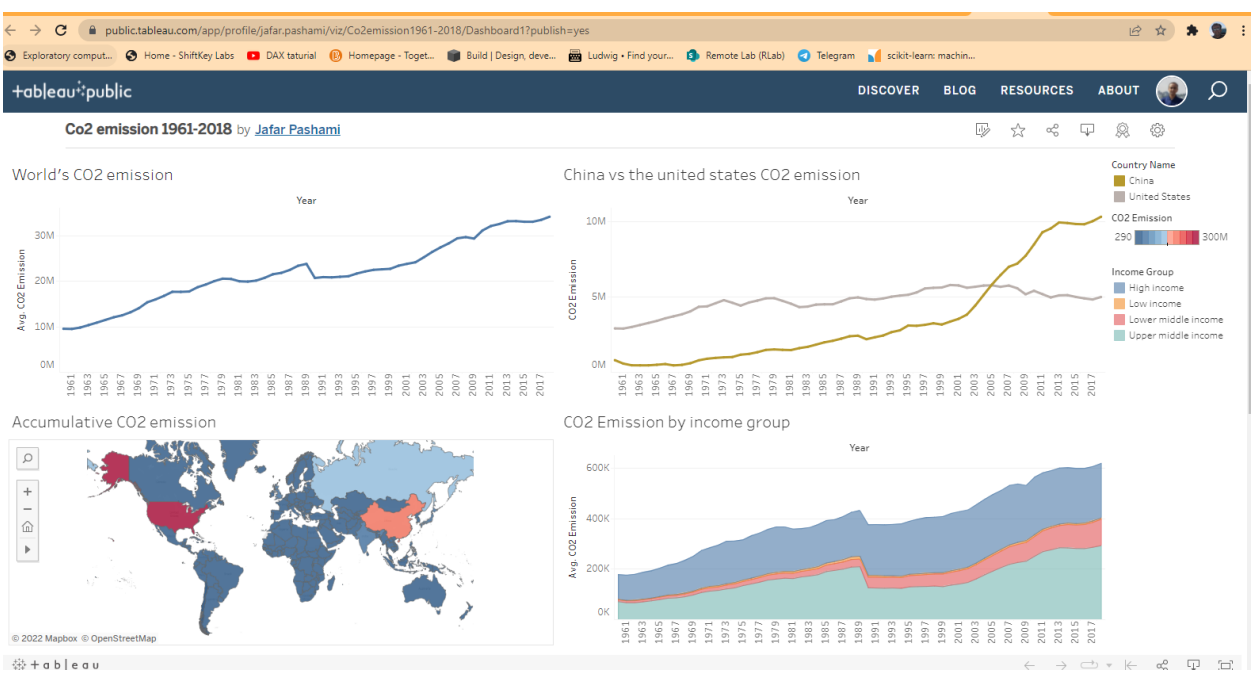
Cons

- It has its terminology that you need to learn.
- Tableau provides a high price per user.
- In my opinion, Microsoft power BI is more powerful for data wrangeling.

Conclusion

In this section, we downloaded the data related to gas emissions from the beginning and added additional databases to it using Tableau and tried to answer the questions that were given at the beginning of the report by using different diagrams. To draw the charts, we used the features of Tableau in all stages of analysis and visualization, including data preparation and the use of various charts, and we also published the report in Tableau Public.

Link: https://public.tableau.com/app/profile/jafar.pashami/viz/Co2emission1961-2018/Dashboard1?publish=yes



Chapter 4 – SAP Analytics Cloud (SAC)

About SAP Analytics Cloud

With the introduction of this tool, SAP Technology has integrated all the tools related to Data visualization, Business Intelligence, and Predictive analytics in the cloud environment. SAC is the analytical front layer of all SAP technologies and supports advanced analytics for organizations.

Dataset and Research Question

The goal of this report is to use SAC to analyze ERP simulation game data. ERPSIM is an ERP Simulation Game to introduce the importance of having integrated business processes for SAP University Alliance member schools. To learn more, you can browse this link: https://erpsim.hec.ca/

The Data set has been provided from several rounds of the game (up to 12 rounds of 20 virtual days each).

The link for downloading the data set is:

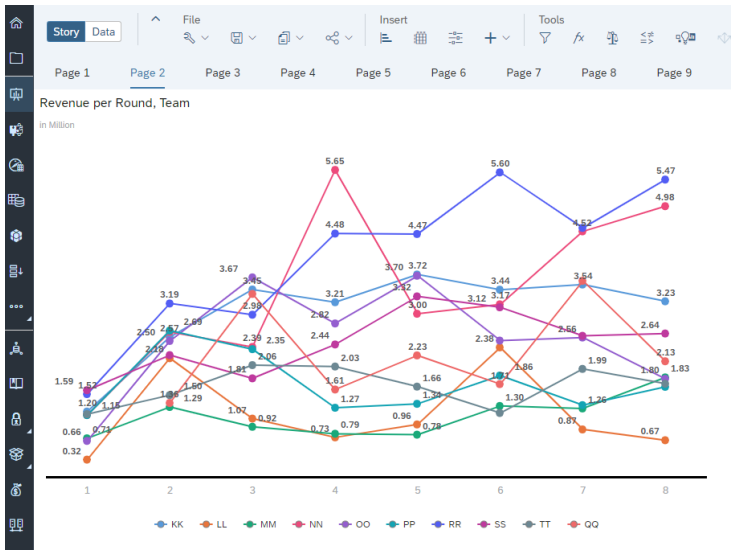
https://dal.brightspace.com/d2l/le/content/202874/viewContent/2815082/View

Research Questions

1. Display the trend of revenue over rounds for each team
2. What is the revenue per team and revenue per product
3. What is the market share (in terms of revenue) of each team by product?
4. Which team sold the most quantity of muesli? For that team, what was the most sold product?
5. What product on what day and round brought the highest revenue (for which team)?

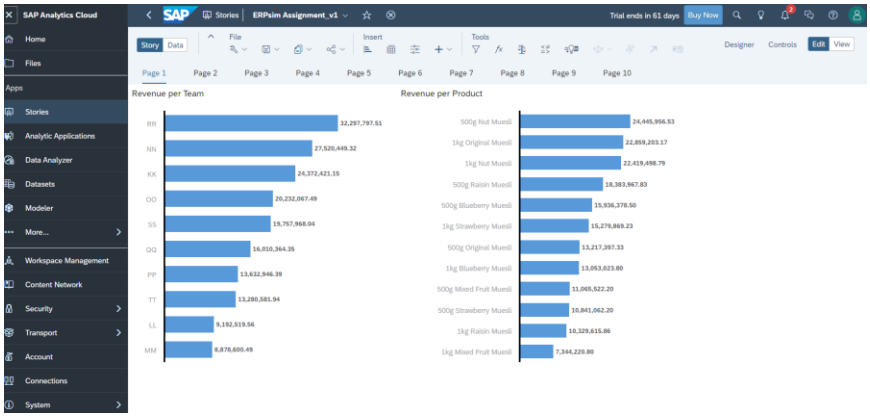
Applying analytical tools and results

After uploading the data to SAC, SAC detects automatically the dimensions and measures. We should check them or change their category. In another step, we did some minor enhancements to the data. Then for having a day-round investigation we need to create a column by concatenating day column and round column. Now the data set is ready for analysis and answering the questions. In order to display the revenue trend, we use a line chart and put Revenue as a measure and Round as Dimension and Team for different Color.

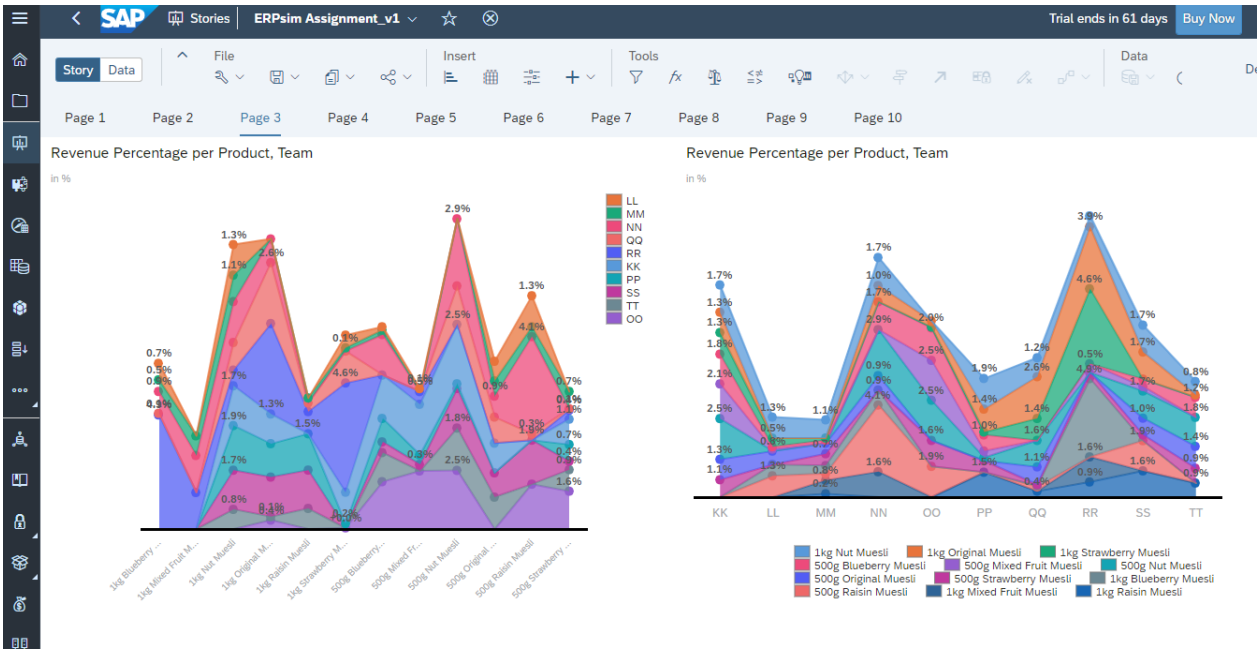


As you can see in the chart above, it seems that the teams with the highest scores have been among the top teams in all rounds of the game, and the teams with the lowest scores have always been among the weakest teams.

In the following diagram we see revenue per team and revenue per product:

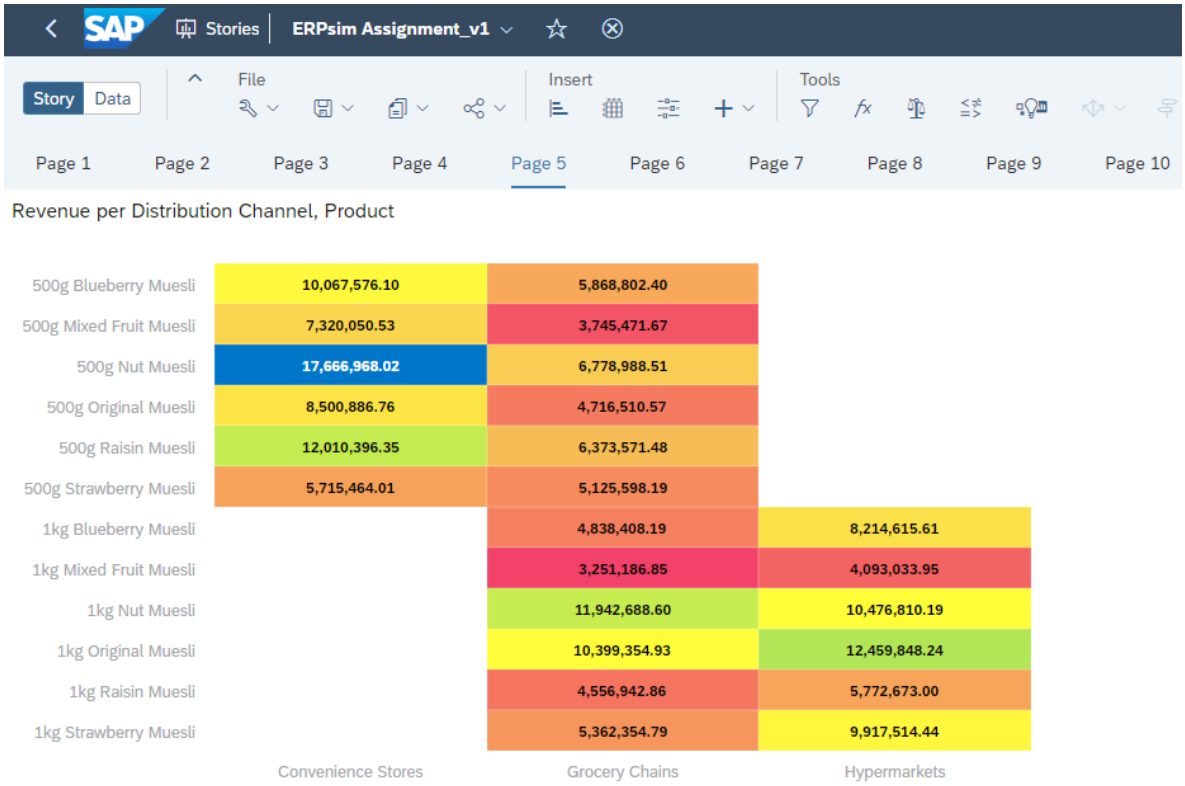


The market share of each team by product revenue has been shown in the following charts:

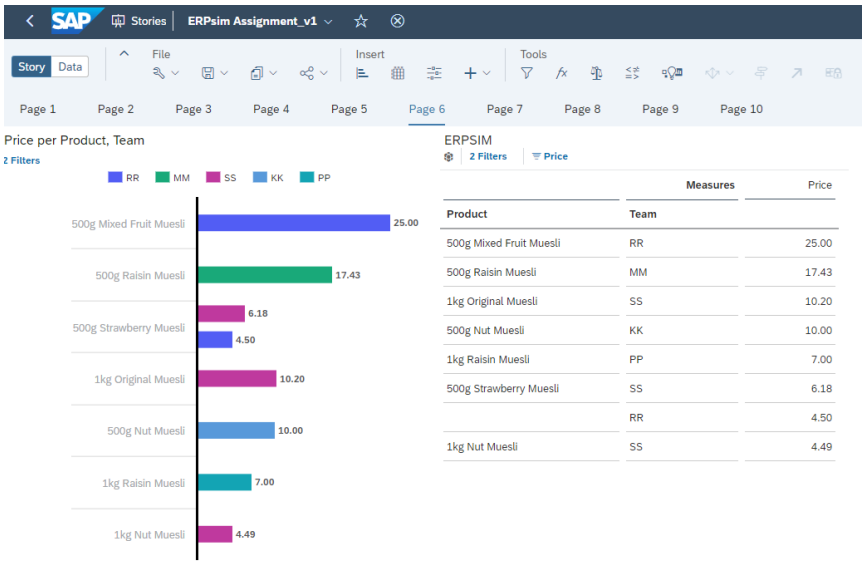


The diagram illustrates that teams have chosen different strategy for selling the products.

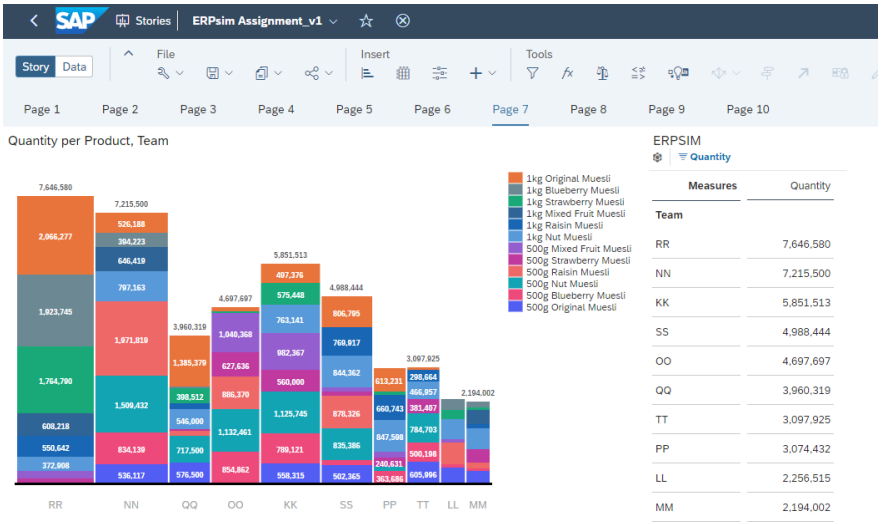
The below Heat map chart indicates revenue per distribution channels by products. As you see, small packages of the product haven't sold on convenience store and hypermarkets.



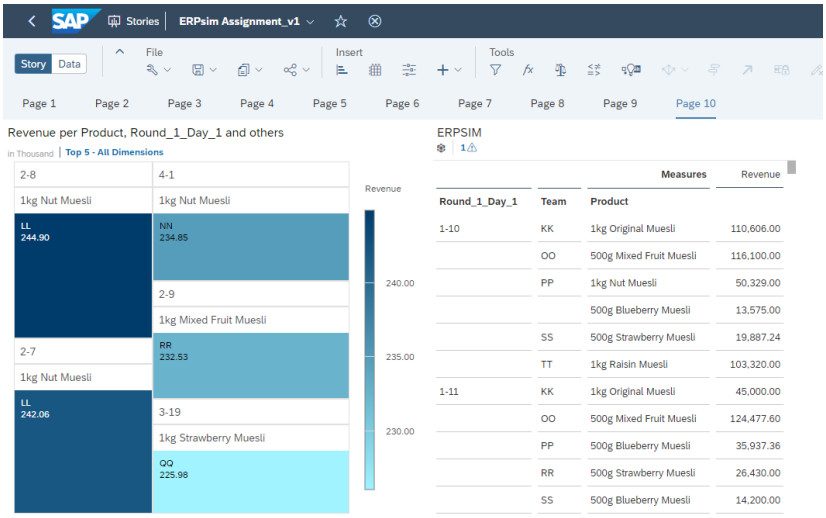
For analyzing the highest priced muesli in an specific day and round, I used price from measures and Team, Product and SalesOrder from dimensions:



The following chart is about the quantity of muesli sold by each team. I used Marimekko map.



Following chart is for finding the highest revenue product and its day and round and team:



Analysis and critique of the tool

SAC is very user friendly and contain lot of different charts. It runs on cloud so it doesn't need to install any software and from there it can be shared. It has modules for more advanced analytics such as predictive analysis. It has an intuitive design so you can learn how to work with very fast.

About the negative points, sometimes it has poor response time maybe because of network connection. This software is relatively costly.

Conclusion

Using SAP Analytics Cloud software, we examined the performance of different teams in the simulated ERP game and answered the required question. SAP Analytics Cloud is a powerful front-end layer of SAP analytical technologies and It can be substitute other SAP analytical softwares.

Chapter 5 – Microsoft Power BI

About Power BI

PowerBI provides a set of tools and technologies for preparing, editing, and displaying data. It can connect to more than 70 types of Databases. It is 14 years that Microsoft has been positioned as a Leader in the 2021 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms. (Ulagaratchagan, 2021)

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms

Source: Gartner (February 2021)

Dataset and Research Question

Dataset

The Dataset contains data from 7200 Apple Store apps that was provided by Ramanathan. (Ramanathan, 2017)

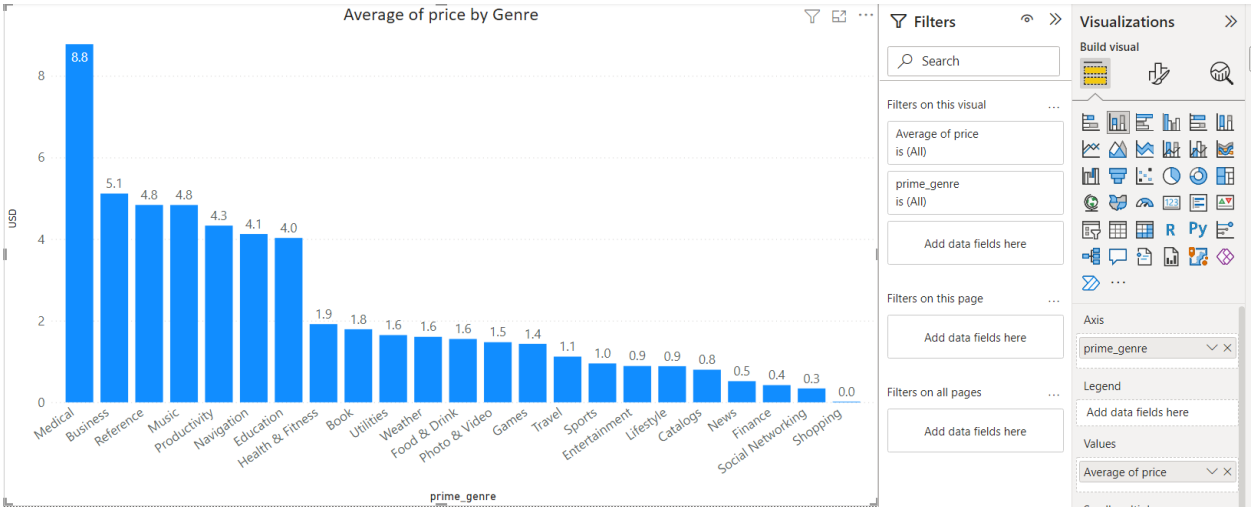
Link to dataset: https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps

Research Questions

1. What was the category with the highest average price? Why do you think this was so high?
2. Visualize the 10 apps with the largest number of ratings.
3. In the Social Networking category, what is the total count of apps with a 17+ content rating? How does this compare to the total number of Social Networking apps?
4. What is the highest-priced app in the Education genre?
5. Is there any relation between the user rating and app's price?

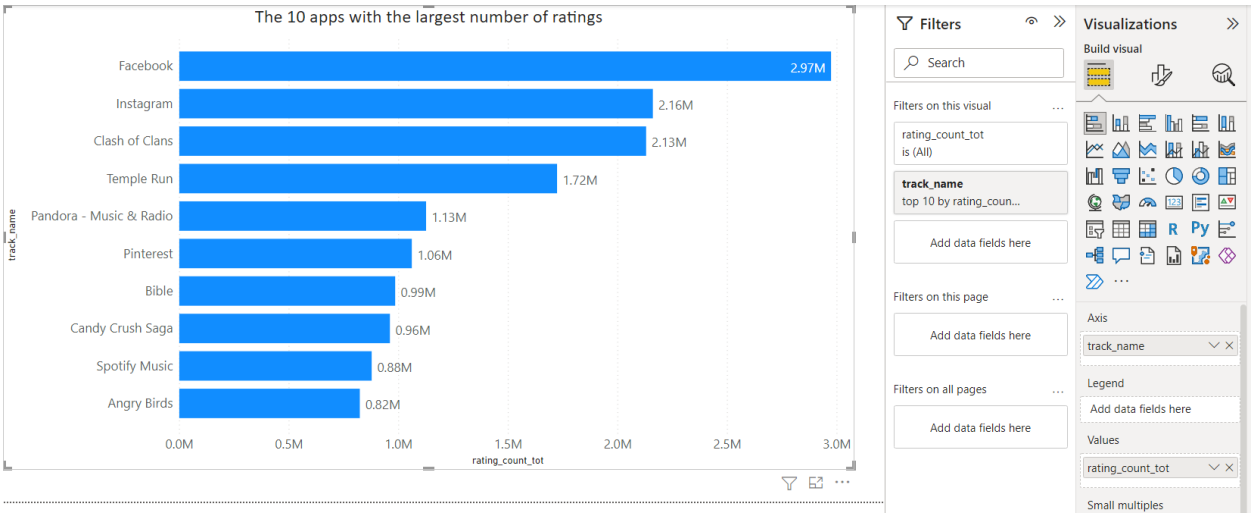
Applying analytical tools and results

For answering the first question, we use a bar chart and locate Genre and Average price as their axes:

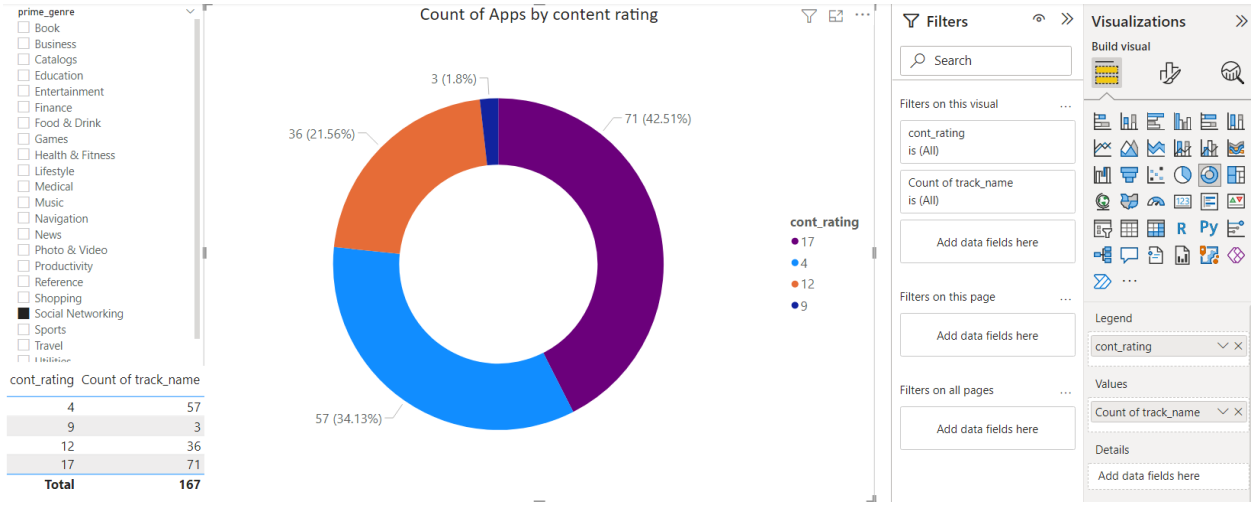


We see that medical app are much more expensive than other. May be this is because of demand-intensive nature of medical apps.

For finding top 10 apps with highest rating, again we use bar char and then use filtering to show top 10 apps:

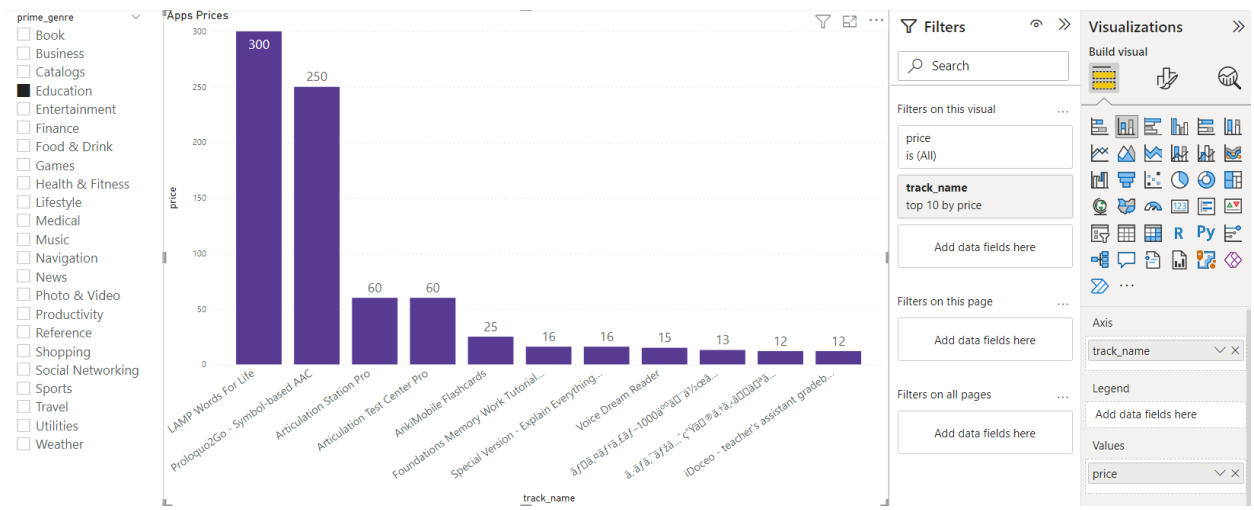


For content rating in social networking app we use a donut chart and an slicer for choosing social networking:

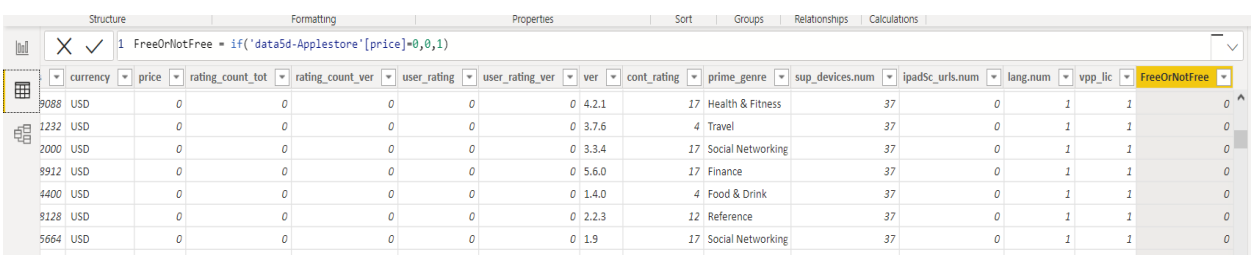


You see that 42.5% Social networking Apps has been rate as +17 years old.

In order to find the highest-priced app in the Education genre we use a bar chart and an slicer.

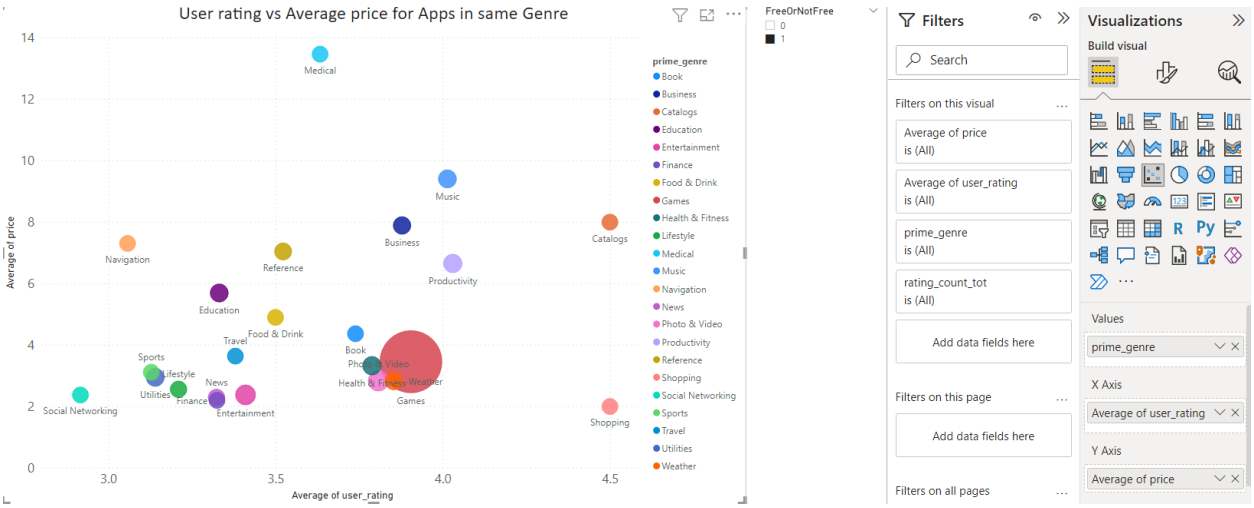


For finding a relation between price and user rating at first we need to exclude the free apps because they have different quality with price equal to ziro. For that we define a new column in our dataset. This column is a binary column with value 0 for free apps and 1 for non-free apps. I used a simple conditional DAX formula for creating this column:



	currency	price	rating_count_tot	rating_count_ver	user_rating	user_rating_ver	ver	cont_rating	prime_genre	sup_devices.num	ipadSc_urls.num	lang.num	vpp_lic	FreeOrNotFree
9088	USD	0	0	0	0	0	4.2.1	17	Health & Fitness	37	0	1	1	0
1232	USD	0	0	0	0	0	3.7.6	4	Travel	37	0	1	1	0
2000	USD	0	0	0	0	0	3.3.4	17	Social Networking	37	0	1	1	0
8912	USD	0	0	0	0	0	5.6.0	17	Finance	37	0	1	1	0
4400	USD	0	0	0	0	0	1.4.0	4	Food & Drink	37	0	1	1	0
8128	USD	0	0	0	0	0	2.2.3	12	Reference	37	0	1	1	0
5664	USD	0	0	0	0	0	1.9	17	Social Networking	37	0	1	1	0

After that I use a slicer and a scatter chart for this data visualization:



As you can see here we can study about a kind of correlation between the price and user rating in payed-Apps.

Analysis and critique of the tool

Power BI is powerful data visualization tool that integrate with all Microsoft software. It is more economical rather than other visualization tools because Power BI desktop is free and power BI Pro is available for a low monthly price per user. In my opinion, Power BI is much powerful in preparing data it uses power query for working with raw data and it has DAX formulation language which is a coding language for working with data. Power BI has an embedded R and Python module. It has a market for creating more visuals with 3rd parties.

I think the most weakness of Power BI is that it runs just on Windows OS and it doesn't work on Mac computers.

Conclusion

In this chapter, we uploaded data related to applications in the Apple Store to PowerBI software, answered a question related to this data, and reviewed the various capabilities of the software.

Chapter 6: SAP Predictive Analytics

About SAP Predictive Analytics

SAP Predictive Analytics is one of the SAP tools designed for statistical analysis and data mining solutions. So, you can use it for both visualizing data as well as predicting data using some machine learning techniques such as classification, regression, clustering, time series, and association rules. (SAP Predictive Analytics Overview, n.d.)

Data Set and Research Questions

I am going to investigate the supply side of seniors' home-sharing potential market in Halifax as well as in Nova Scotia and Canada. To do that, I have downloaded the whole data set about housing indicators of senior people for statistics Canada. This Dataset is a part of Census 2016 in Canada. It is about the Family Characteristics of Seniors (above 65 years old). It has about 4.9 million rows and 30 columns. (Data tables, 2016 Census, 2019)

Therefore, the research questions are as below:

1. What is the number of living houses for senior people living alone in Halifax? And in Canada?
2. What is the number of convenient housing the senior people owned?
3. How many bedrooms do they have? And what is the percentage of the alone seniors' houses with more than one bedroom?
4. Estimate the potential market of seniors for home-sharing.

Applying the analytical Tool and Results

Data preparation

At first, because the amount of data fetching data has been taken a few minutes. After fetching, I need to know about the structure of the data, so I read the metafile that is prepared for a better understanding of the data set and its dimensions.

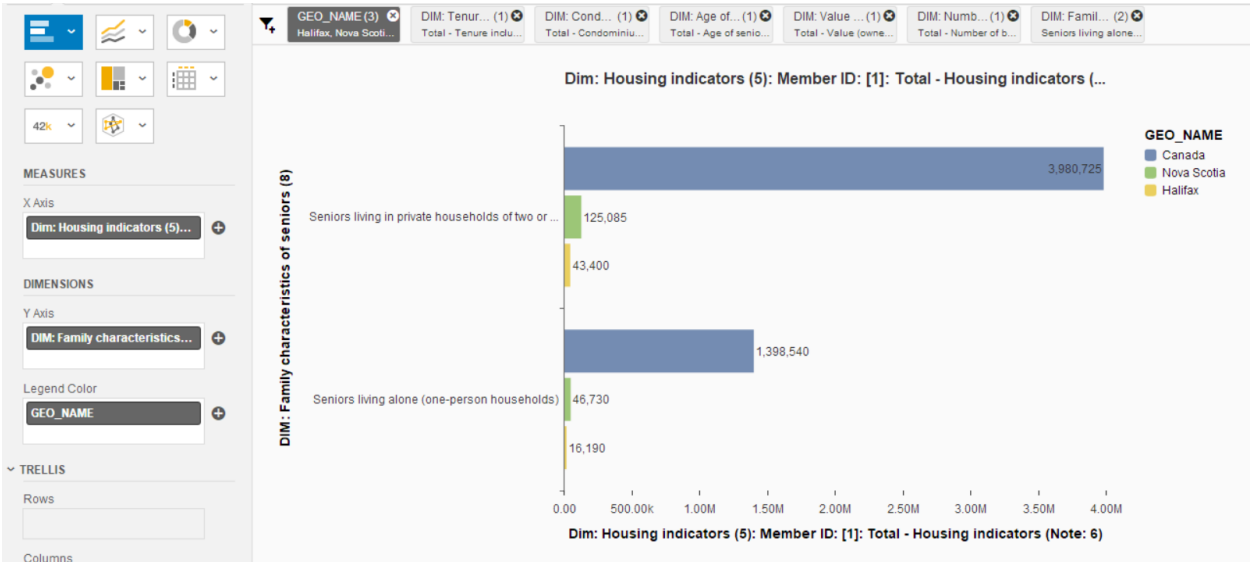
We have the value of aggregation of data and single dimension values with each other in the dimensions because of that if we don't filter data, we will count them multiple times. Therefore, I defined at least 7 filters for all the visualizations. One of these filters was global it was GEO_NAME filter that filters data to contain just Halifax, Nova Scotia, and Canada data. Other filters are local filters that use especially in each chart.

In addition to filtering, I defined a calculated measure that I describe it later in this report.

Data Analyzing

1. *What is the number of living houses for senior people living alone in Halifax? And in Canada?*

As you see in the following chart there are 16190 houses in Halifax where a single senior is living there. This number for Nova Scotia is about 47000 houses and for Canada is 1.4 million houses.

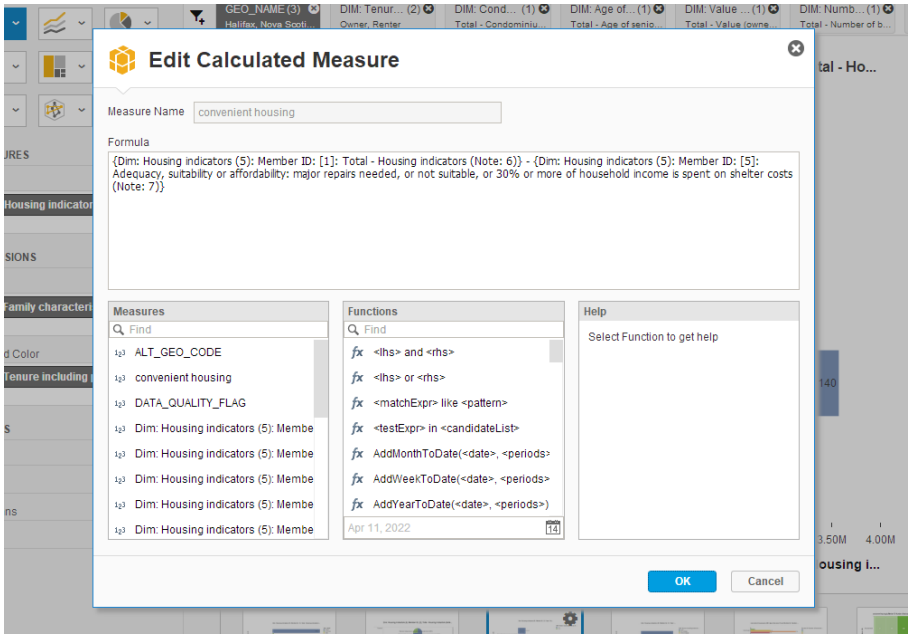


There following figure will show the share of this housing indicator based on living alone or with others:

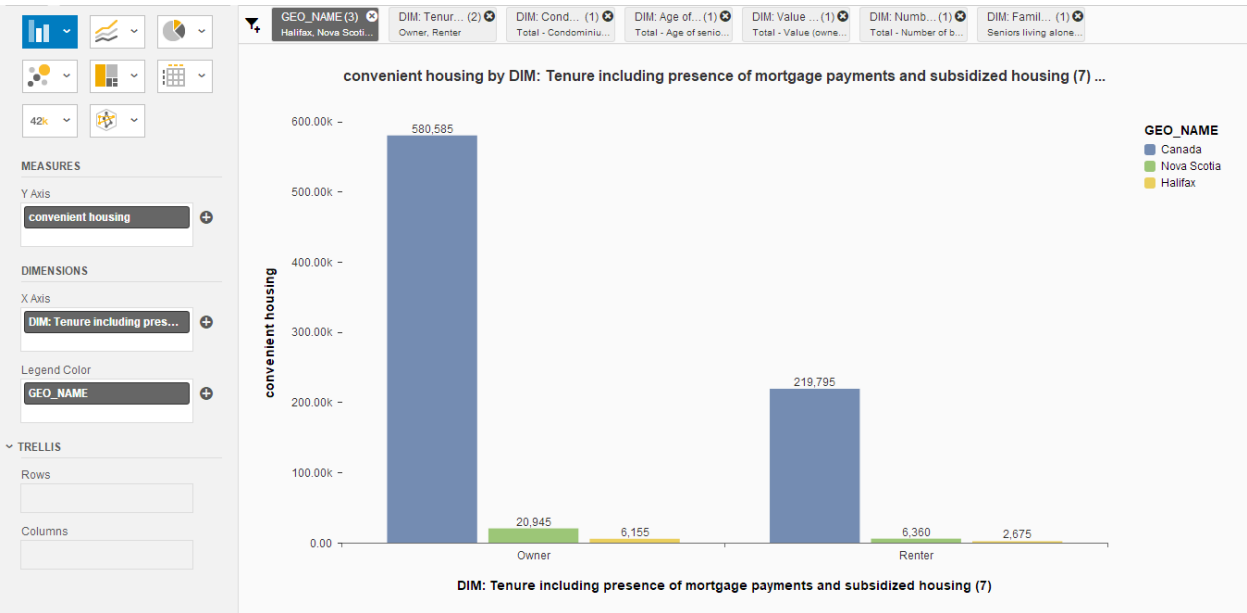
As you see, we have approximately the same pattern in Halifax, Nova Scotia, and Canada in all of them about 26%-27% of seniors (65 years old and higher) are living alone.

2. *What is the number of convenient housing the senior people owned?*

To answer this question first we should define a new measure that is the result of subtracting non-convenient housing from the total housing provided in this data. The following picture shows the formula:

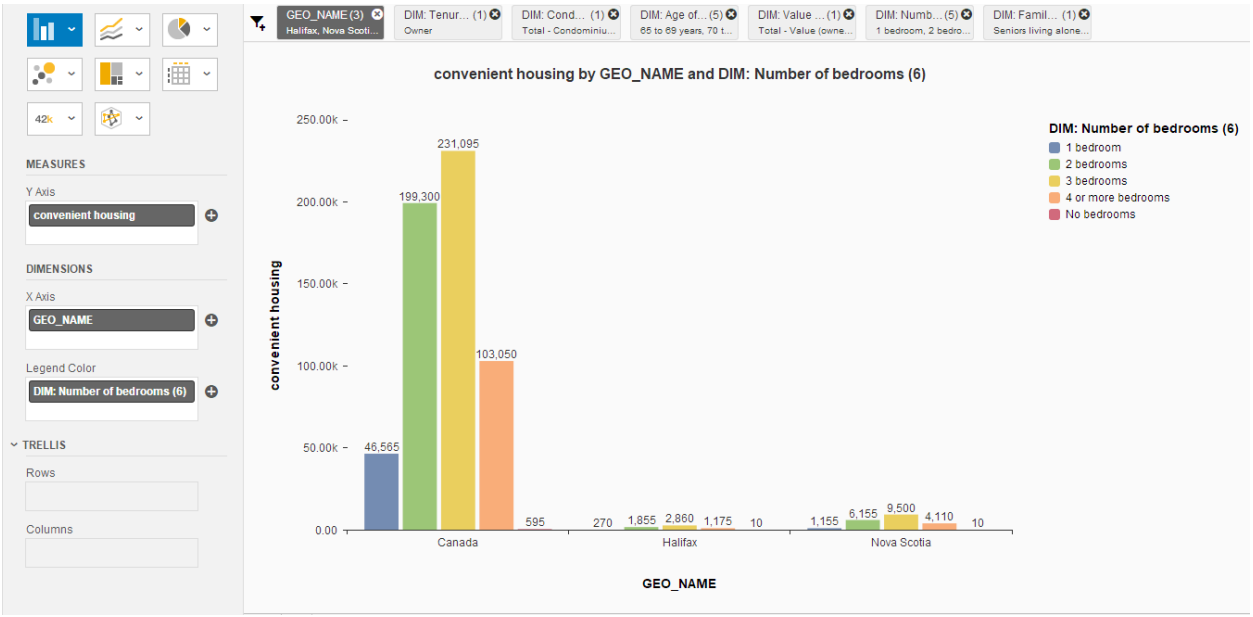


After creating this measure, we apply it to all the charts as the number of convenient houses. In addition, we need to know how many of these houses are owned by seniors. You can see the results in the following figure:

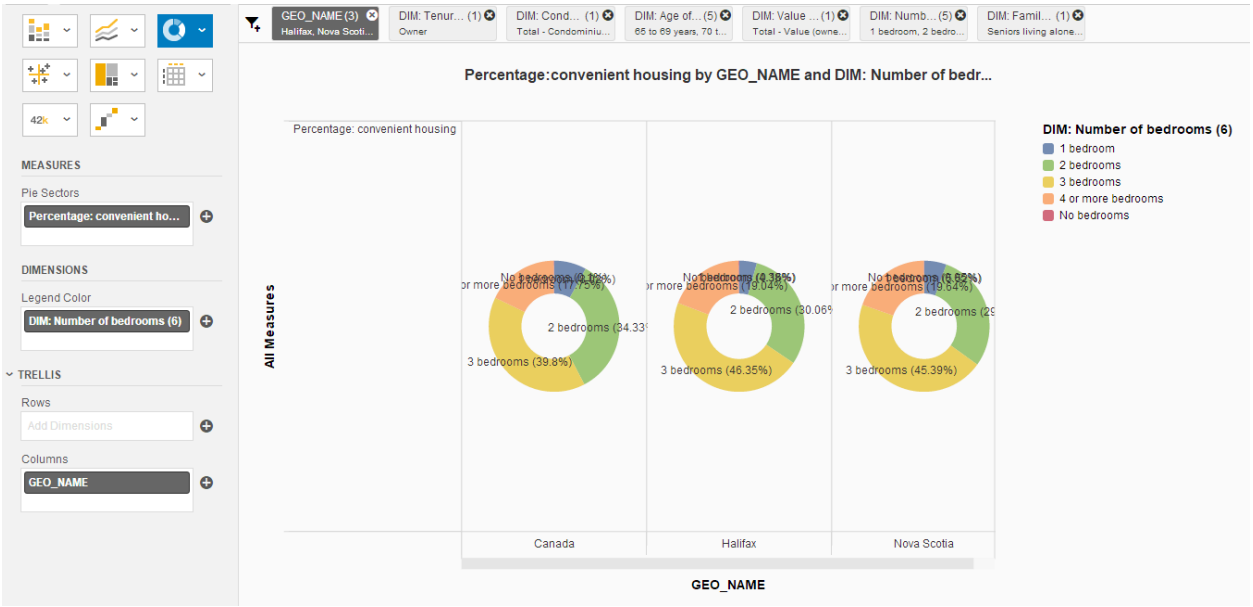


So, about 6155 houses in Halifax are owned by alone seniors and have appropriate conditions for living. This indicator for Nova Scotia is about 21000 houses and for Canada is about 581000 houses.

3. *How many bedrooms do they have? And what is the percentage of the alone seniors' houses with more than one bedroom?*

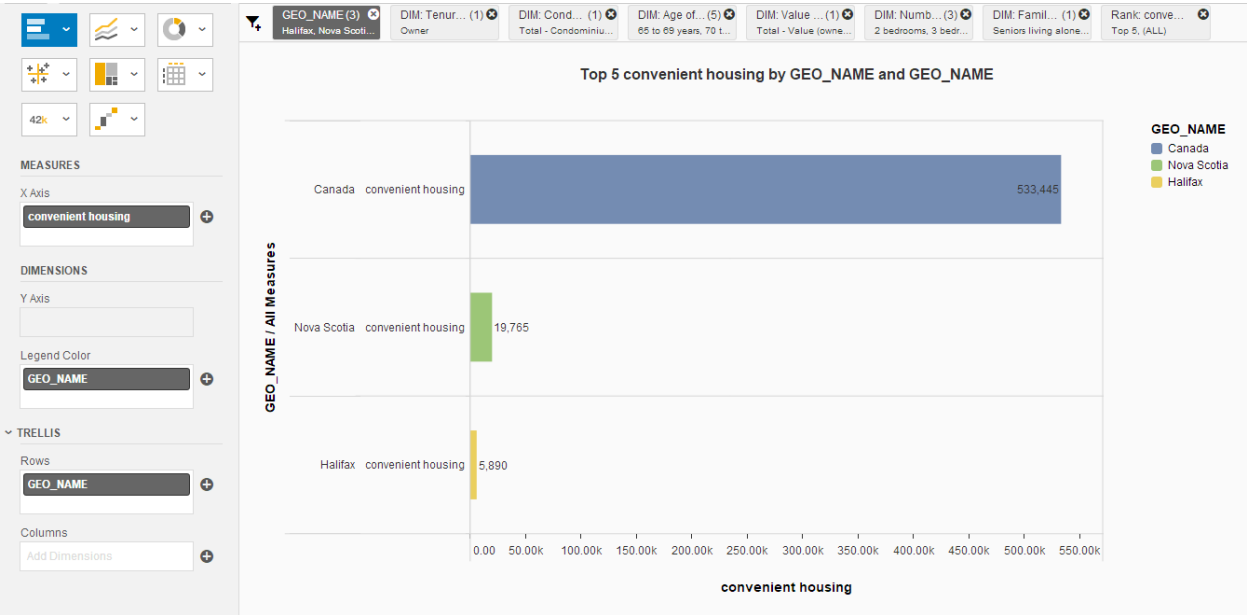


In Halifax, 280 houses out of 6155 have less than 2 bedrooms. It is about 4.5% So about 95% of convenience houses that are owned by alone seniors have more than one bedroom. We see a similar pattern in Nova Scotia (94%) and Canada (92%) as well:



4. *Estimate the potential market of seniors for home-sharing.*

Based on filtering one-bedroom and no bedroom values, we have this chart as our potential market that is the number of houses owned by alone seniors with more than one bedroom:



Conclusion

Like any other tool, this tool has strengths and limitations that must be considered when it comes to data analysis tools.

One of the strengths was that this data, which contained about 5 million records and was about 1.2 GB in size, was quickly loaded into this tool, although there were occasional errors when working with the tool.

Another strength, in my opinion, is having different sections of preparation, visualization, and forecasting that help you work with better mastery of your data set and do not need tools such as Excel when working with data.

Regarding limitations, first, some errors occur during the work. This is common behavior of most SAP tools in comparison to Microsoft Power BI and Tableau.

In addition, another limitation of this tool is its integration with other products outside of SAP.

Moreover, I did not see the ability to add new charts.

Finally, I did not find proper capabilities in setting fonts and the page layout of reports in this tool.

Chapter 7: SAP Lumira's Geospatial Analysis

About SAP Lumira's geospatial features

Business Analytics tools are emphasizing heavily easy-to-understand visualizations. One way is using Geospatial Analysis. SAP Lumira Discovery uses Esri embedded maps to provide an updated map with four base maps for its visualization. There are several capabilities in Lumira Discovery such as creating a Geo Hierarchy, Choropleth layer, Bubble layer, and Pie chart layer, …. In this chapter, we will analyze Crime Data using SAP Lumira Discovery.

Data Set and Research Questions

The Data Set is the crime data for the city of Calgary between 2012 and 2017. We have for data set in CSV formats that will merge for further analysis.

1. The first one is about the number of incidents of crime category that happened in communities.
2. The second one is the geographical location of Calgary's communities.
3. The third one is Calgary's communities' population between 2012 and 2017
4. The fourth one is Police Locations in Calgary

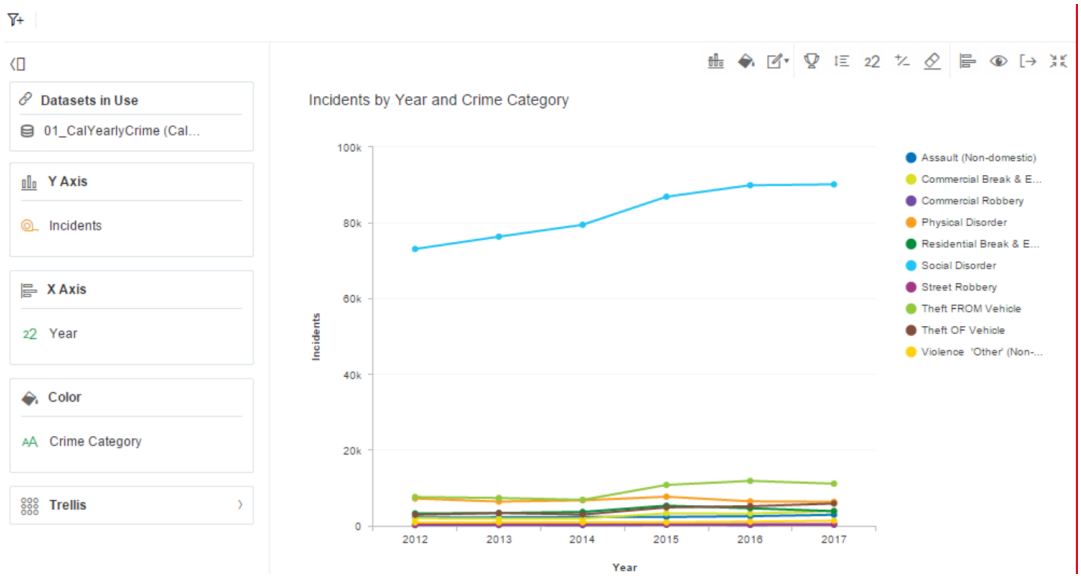
The questions we are going to answer are:

1. What is the trend of crime incidents and their categories in Calgary between 2012 and 2017?
2. Which communities have more crime incidents?
3. What types of crimes are most common in each community?
4. Are the police stations well distributed to respond quickly to high crime areas? Are there any communities that may not be well served?

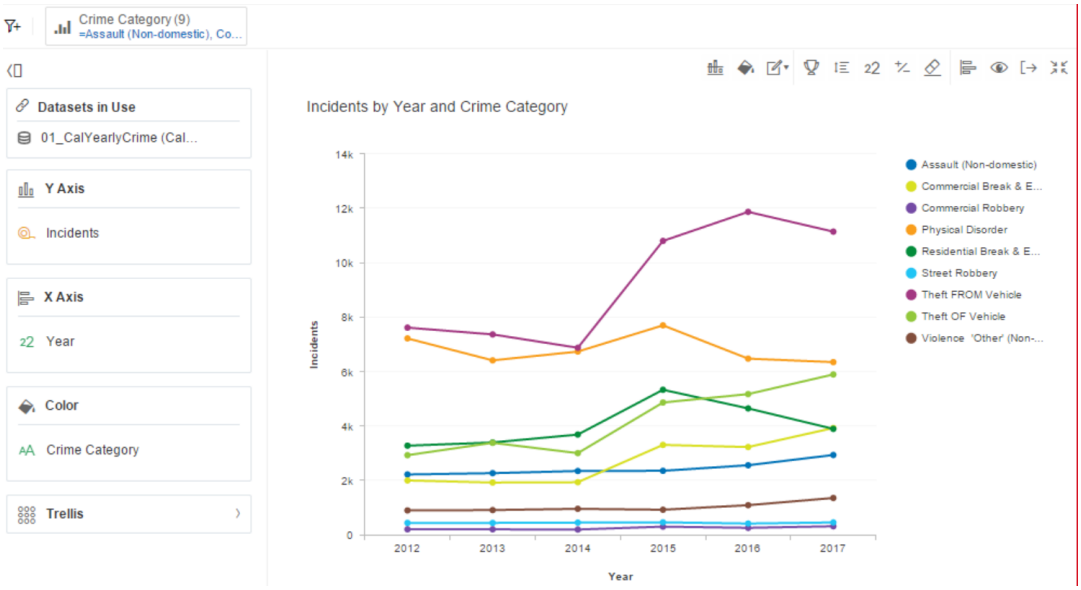
Applying the analytical Tool and Results

1. What is the trend of crime incidents and their categories in Calgary between 2012 and 2017?

First, we upload the first data set in Lumira Discovery. Then using the Design View select a linear chart as Chart type and the drag and drop incidents to Y-axis, Year to X-axis, and Crime Category to Color. The following line chart shows the crime incident trend in Calgary:



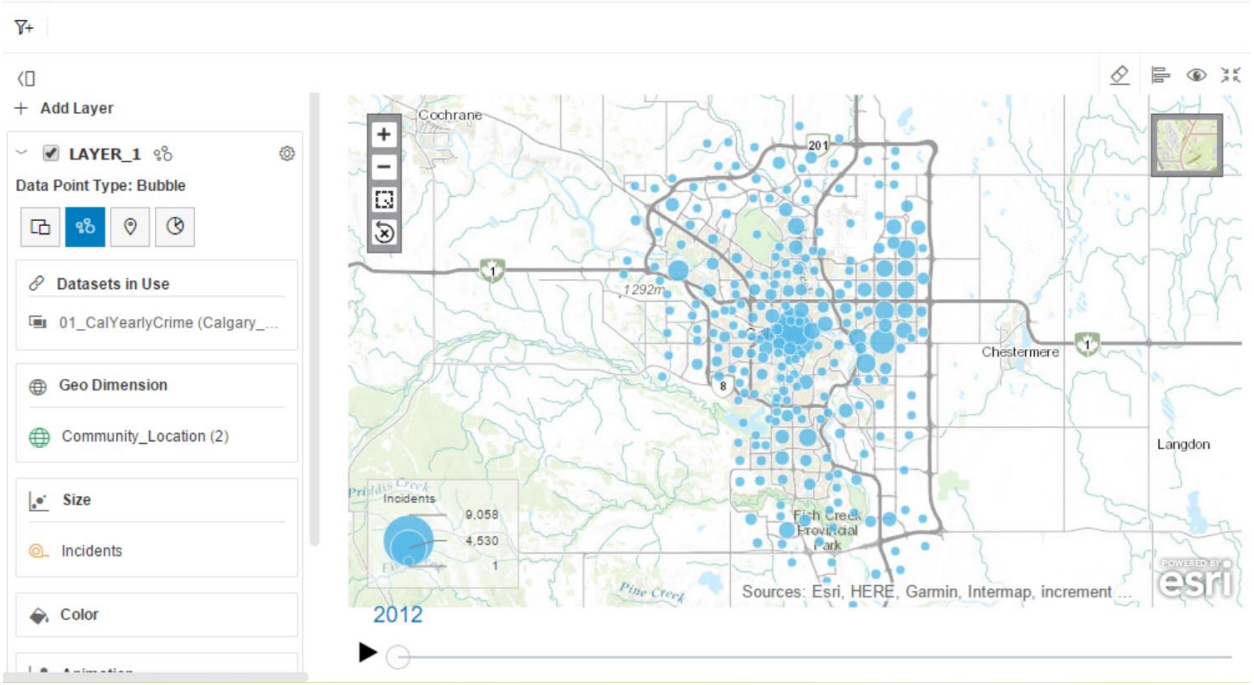
By adding a filter to remove Social Disorder that is more frequent data we can have a better view of other crimes:



As you see, we have an overall increase in crime incidents up to 2015 however after that some of them have been controlled and others are growing in 2017.

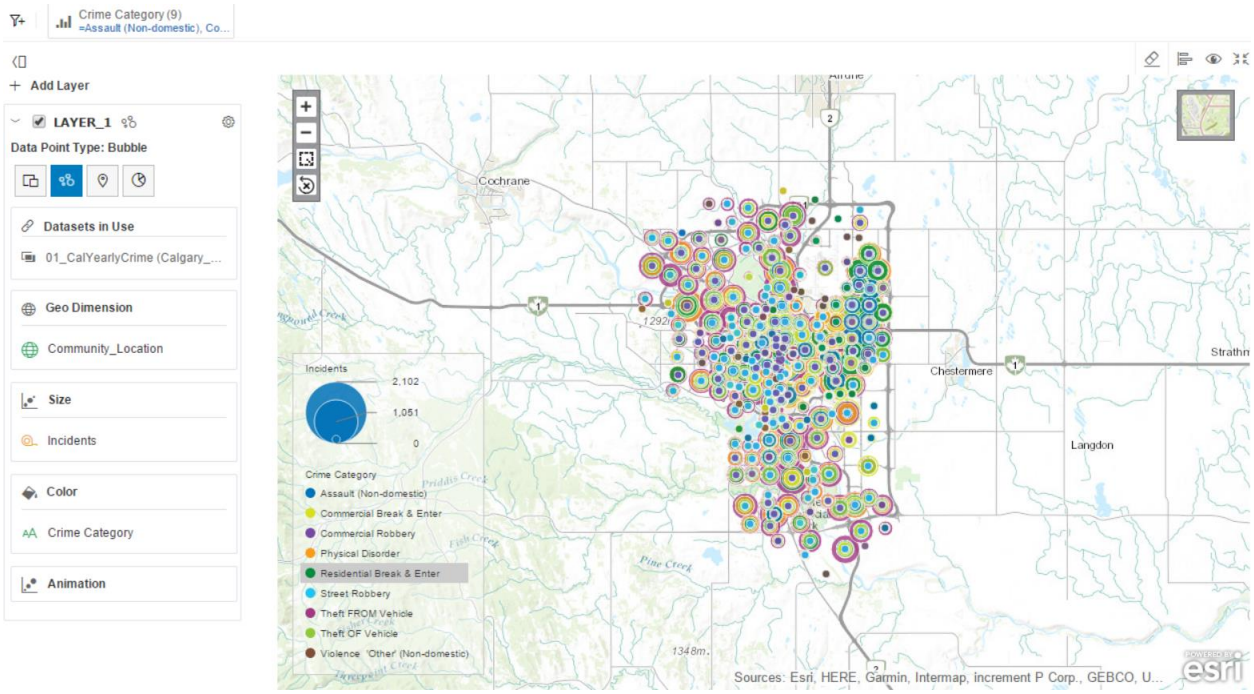
2. Which communities have more crime incidents?

To answer this question, we need to add the second data set includes geographical data of communities. We merge two data sets by the left outer join technique. Then we create a Geographical hierarchy by the longitude and latitude information added to our data set. After that, a Map chart is used to show the values of the map. We change the Datapoint type to bubble and put incidents as the size of the bubble. By looking at the chart below, we can easily see crime-prone communities.

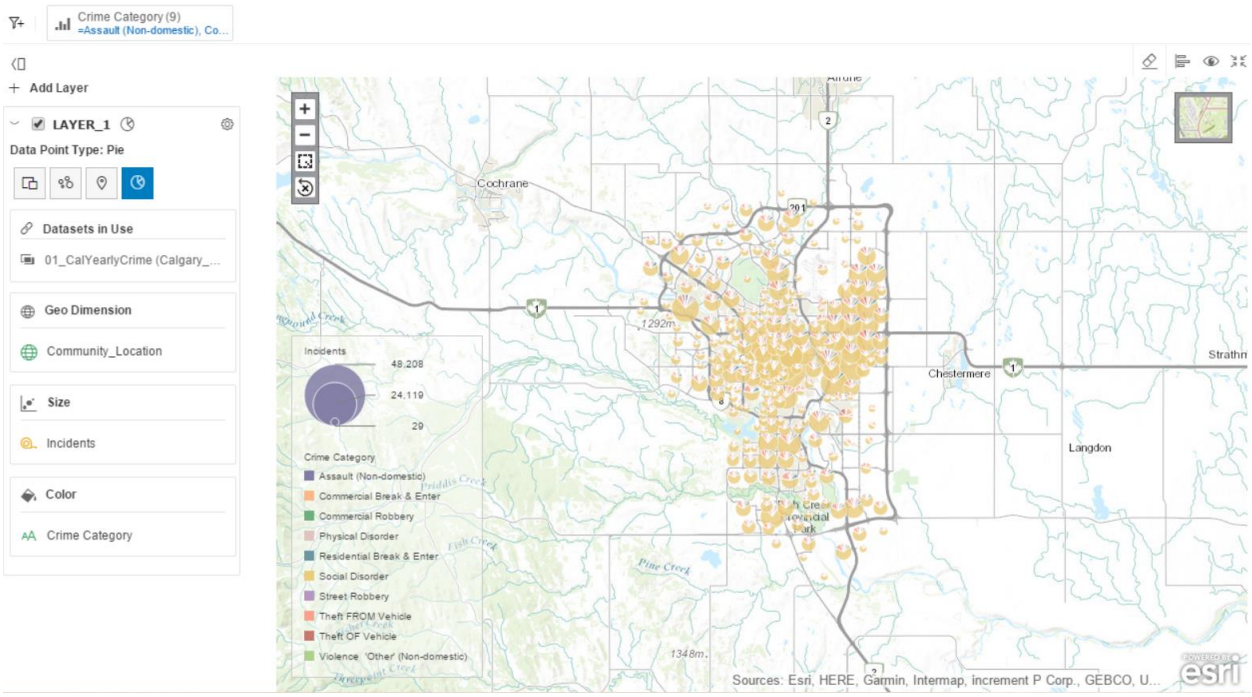


3. What types of crimes are most common in each community?

We can add crime categories to the bubble chart on the map to see the different crimes with different colors and different sizes in each community. To have a better visualization, we filter out social disorder from data.

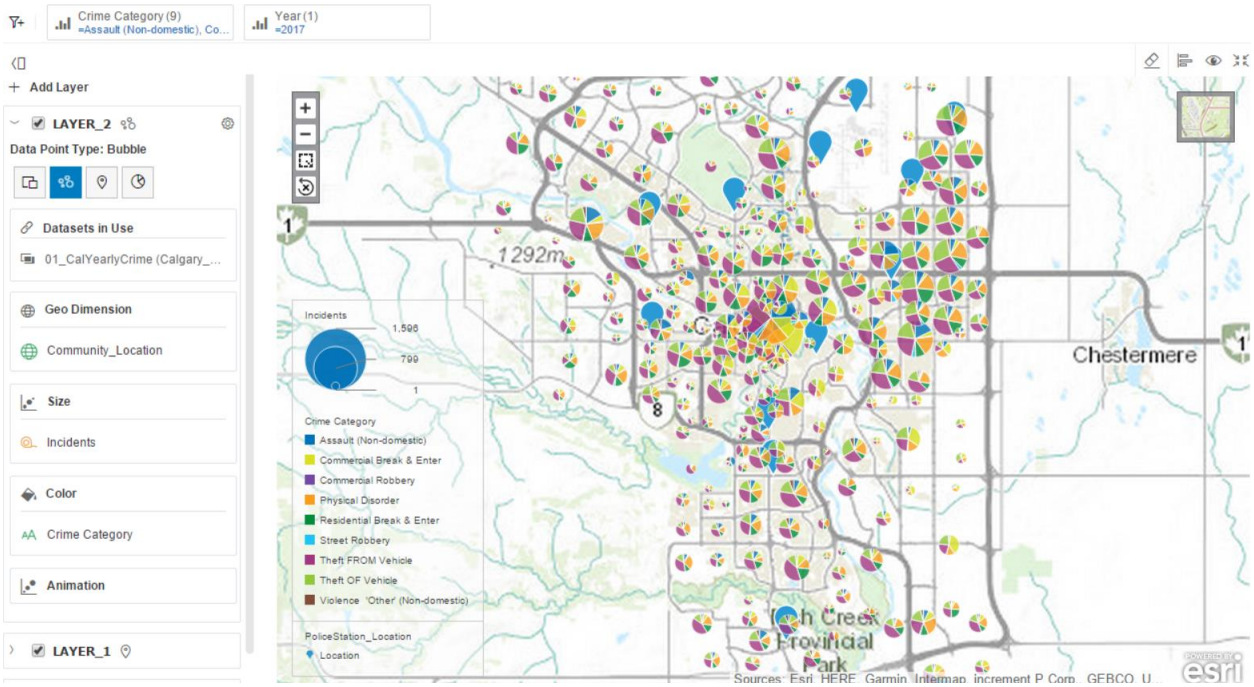


We can change the Datapoint type to pie to find a share of crime categories in each community.



4. Are the police stations well distributed to respond quickly to high crime areas? Are there any communities that may not be well served?

To answer this question, at first, we need to import police location data and then merge it with our base data set.



You see, it seems that police locations are not optimized to be close to high crime communities and it takes time for police to reach a high-crime areas

Conclusion

One of the essential advantages of this software is that it can support different map layers. These layers can be divided into three categories. The first is the base layers that are present in all maps, and the second is the layers designed. Some applications can show and hide. And the third layer is the layer of using data points, which exists in the form of bubble diagrams, chart legs, etc.

The most significant limitation of using GIS in this software is that working with this data is very slow for personal computers and is accompanied by numerous errors related to the execution of the map.

References

(n.d.). Retrieved from Crunchbase: https://www.crunchbase.com/organization/tableau

CO2 emissions (kt). (2022, 2 15). Retrieved from The Word Bank Data: https://data.worldbank.org/indicator/EN.ATM.CO2E.KT

Data tables, 2016 Census. (2019, 06 17). Retrieved from Statistics Canada: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Rp-eng.cfm?TABID=4&LANG=E&A=R&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=205&GL=-1&GID=1341700&GK=10&GRP=1&O=D&PID=110578&PRID=10&PTYPE=109445&S=0&SHOWALL=0&SUB=0&Temporal=2017&THEME=121&VID=0&VNA

Ramanathan. (2017, July). *Mobile App Store (7200 apps)*. Retrieved from Kaggle.com: https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps

SAP Predictive Analytics Overview. (n.d.). Retrieved from SAP help Portal: https://help.sap.com/docs/SAP_PREDICTIVE_ANALYTICS/5978826050354bc89761a57cfb60473f/e876b18b5e4c45308ff1a08cad909b15.html?version=3.0

Ulagaratchagan, A. (2021, February 18). *Microsoft named a Leader in the 2021 Gartner Magic Quadrant for Analytics and BI Platforms*. Retrieved from powerbi.microsoft.com: https://powerbi.microsoft.com/en-us/blog/microsoft-named-a-leader-in-2021-gartner-magic-quadrant-for-analytics-and-bi-platforms/