Trent University

A Report on

Sales Data Analysis in PowerBI

Under the subject of
Big Data
M.E., Semester-I
(Big Data Analytics)

Submitted By

Sr.	Name of Student	Enrolment No.
1.	Nisarg Patel	0692153

Guided By
Prof. Brian Srivastava
Instructor

Academic Year
2021-2022

PROJECT WRITE UP

Abstract :

Business Intelligence term can be defined as an answer that companies need to make a bold decision in business processes. It is used to very effectively organized and transform the data into great values and this is used as knowledge in decision making. The company can easily monitor business performance and respond to another external environment with the help of business intelligence solutions.

Power BI is developed by Microsoft and used as a business intelligence software. Most of the companies are focusing on improving the growth of their product. So, they look for interesting hidden patterns from the data. For this reason, in this project, I have developed the dashboard for the commissioning sales department, a vendor of adventure cycle works which required to monitor the yearly sales performance. So, the purpose of this project is to make a sales displaying report which displays the yearly sales by visualizing the data in the Power BI. Along with that, I was also focused to use maximum functionalities.

In this project, I have set up the Power BI software and performed much useful functionality of power BI. I have explained each and every functionality that I have used to develop the dashboard. Furthermore, the empirical steps that I have specified in this project can provide interesting results to follow and improve the sale in the future.

Keywords: Power BI, Data visualization, Business Intelligence, Sales Dashboard,

Problem Statement :

Adventure Works Cycles is a large, fictitious, and multinational manufacturing company that manufactures and sells metal and composite bicycles, clothes, accessories, and components to the commercial market of six different countries named Australia, Canada, France, Germany, UK, and the US. It has a total of 9 different database files in which consist of the information of currency, customer, geography, product category, sales, etc. and these are all messy data. So, they are several problems like the user reports are ambiguously targeted and it is not easily read which data and information to display. Secondly, we have different files for different information, so it is required to build a report that has all the necessary information.

Overview of Power BI :

Power Bi is powerful cloud-based business analytics and graphical representation tool for business users to monitor their business performance and gain useful insight from the data. Microsoft developed power bi allows users to connect data from various data sources like local files, Microsoft share points, data warehouse. It also provides additional service for excel users like, they can also publish the insights that created in excel to power BI like pivot tables, dash-boards, etc. in other words, power bi provides four traditional excel which are power query, power view, power pivot and power map which are user friendly who use excel.

There is a total of three main services available in the power bi which are power bi desktop, power bi service, and power bi mobile application. Among all three, the power bi desktop has the strongest abilities to connect the data, visualize the data and transform the data. After making a connection with the data sources, we can transform or clean the data into a query editor. Moreover, users build a data model and visualize the data through a report, as with ought limitation of ad hoc and time dashboard, and then we can publish. It is also easy for users to make changes and manage visualization output based on requirements. In the power bi application, it can be downloaded and easily installed on android and iOS devices in which users can view and interact with published data without limitations of time and locations.

The usage of power bi is also depended on the user's role like if the user mainly works in data processing and creation of a report environment then that user will prefer power bi desktop while on the other hand sales executive whose job is mainly to monitor the process of sales work and get useful insight into new sales opportunity then that user primarily chooses the mobile application.

How power BI will scale up large databases ?

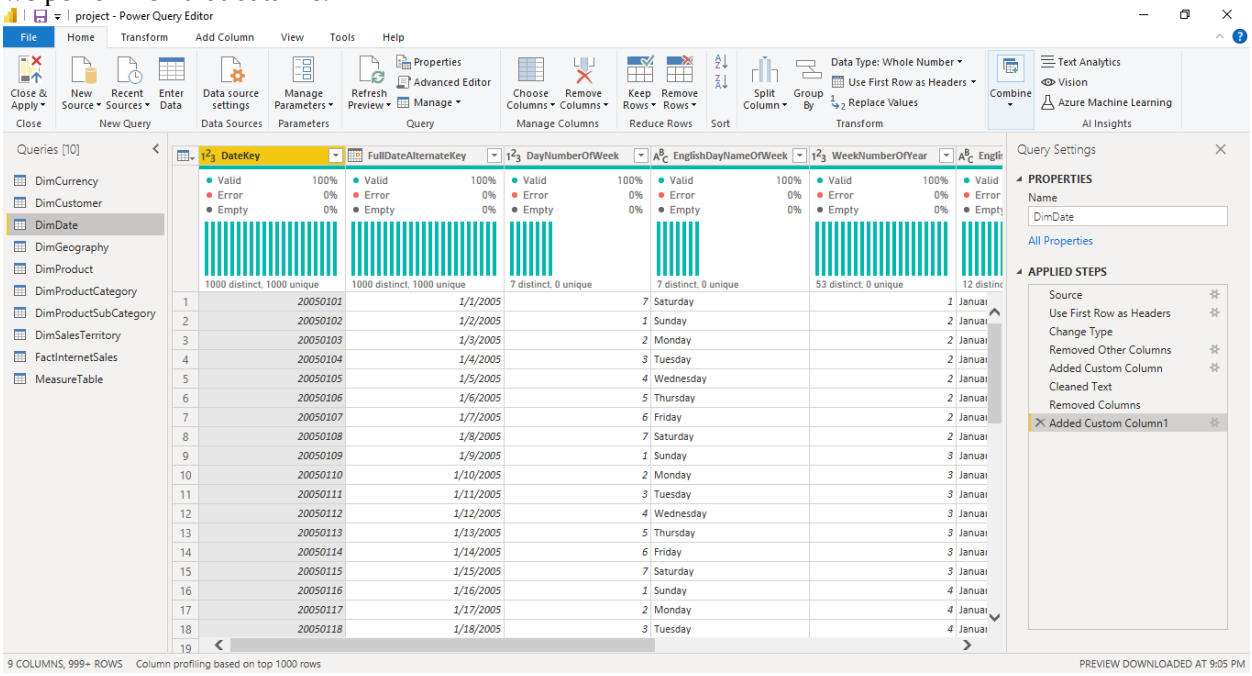
Power BI datasets are able to store the data into the highly compressed in-memory cache for optimized query performance and enabling fast user interactivity. For example, for this project total size of the database is around 30MB but after loading into the power bi, it only takes around 2 MB to store the whole data which is extremely great data compression.
Microsoft also provides premium capacities for those which datasets are beyond the default size, which can be enabled with the large dataset storage format setting. And when it is enabled, dataset size is limited by the premium capacity size by the administrator. A large database that contains single files of more than 10 GB, enabling large dataset storage has extra benefits.
Users can enable the large dataset for all the premium P SKUs, PPU and A SKU, etc. we can compare the model size of this with Microsoft azure analysis services.

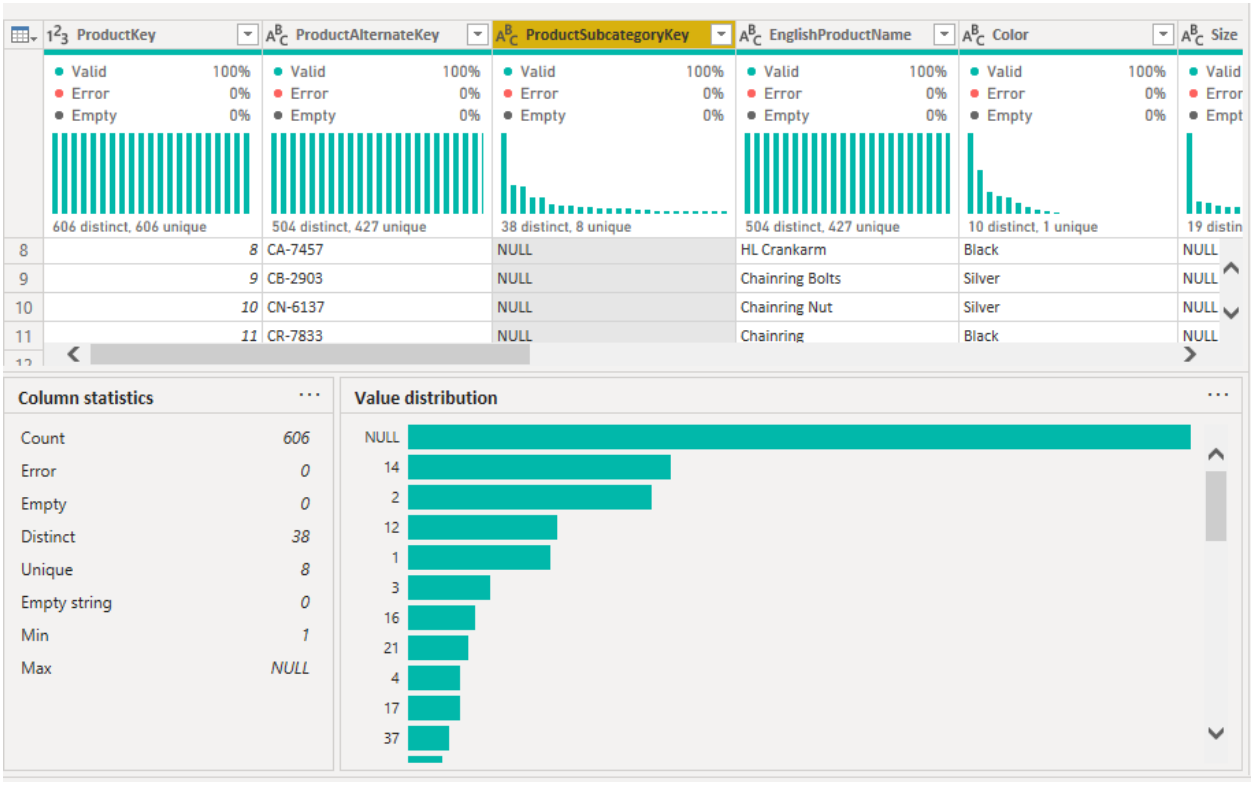
Method: What did you do that worked towards solving the problem
Power Query Data Cleansing

Step 1: Import the data
Datatype detection: based on the first 200 rows, based on the entire database, do not detect any data type. I will recommend choosing based on the entire database because sometimes it may get errors if there is more than one type of data. I have selected Datatype detection based on the first 200 rows because I have fewer data.

Step 2: clean the data in which I removed some columns which are not necessary for the analysis. In case of need, we can also add those columns easily. Apart from this, I have added some new columns.
From the figure, it is clearly seen that Power BI gives much great functionality through power query editor in which we can see all the descriptions for every specific column like Valid, Error, and Empty.

Apart from that, we also have a special tab for applied steps through which we can look up all the steps at we perform on that datafile.





In above figure, we can see that by clicking on any specific column powerbi provides all the statistics about that column like Count, Distinct values , unique values, any empty string, min and max. it also provides value distribution of column.

I have loaded all the columns from csv files that required for the analysis.

Dimcurrancy: it consist only 3 columns and we need all columns: CurrencyKey, CurrencyAlternateKey, CurrencyName

DimCustomer : I have removed irrelevant columns like address, number of children , middle name etc.

Then I combined 2 columns firstname and lastname with customcolumn and created new column named customer full name and set type of that column as ABC(text).
So, I did same for all files by removing the columns that are not important for our analysis will reduce the space complexity and also increase the speed of analysis.

DimDate: datekey, fulldatealternativekey, Daynumberofweek, englishdaynumberofweek, weeknumberofyear, englishmonthname, monthnumberofyear, calenderyear, month-year

Dimgeography: GeographyKey, city, StateProvinceCode, StateProvinceName, CountryRegionCode, EnglishCountryRegionName, SalesTerritoryKey

DimProduct: ProductKey, ProductAlternateKey, ProductSubcategoryKey, EnglishProductName, Color, EnglishDescription, StartDate, EndDate, Status

DimProductCategory : ProductCategoryKey, ProductCategoryAlternateKey, EnglishProductCategoryName

DimProductSUBCategory : ProductSubcategoryKey, ProductSubcategoryAlternateKey, EnglishProductSubcategoryName, ProductCategoryKey,

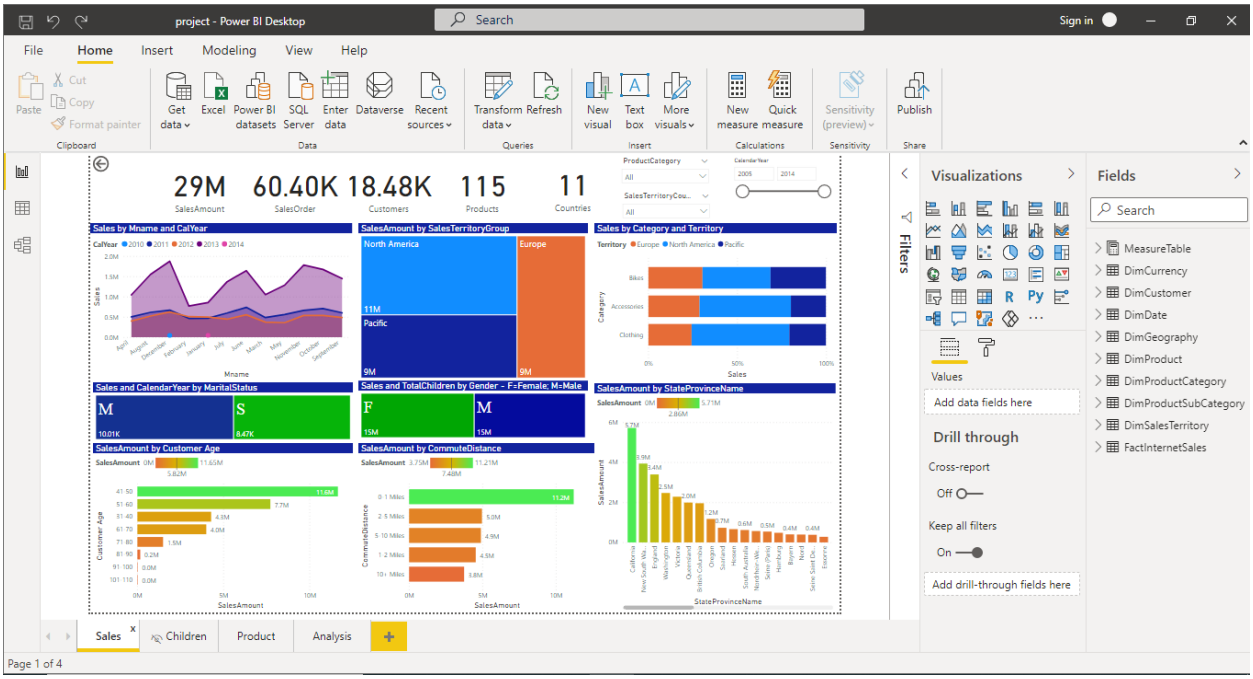
DimSalesTerritory: SalesTerritoryKey, SalesTerritoryAlternateKey, SalesTerritoryRegion, SalesTerritoryCountry, SalesTerritoryGroup

FactInternateSales: ProductKey, OrderDateKey, ShipDateKey, CustomerKey, PromotionKey, CurrencyKey, SalesTerritoryKey, SalesOrderNumber, SalesOrderLineNumber, RevisionNumber, **OrderQuantity**, UnitPrice, UnitPriceDiscountPct, SalesAmount, CarrierTrackingNumber, CustomerPONumber, OrderDate, ShipDate

Power BI has a report workspace where I have created all the visuals as well as different reports. Here I have created many useful charts in the dashboard like line graph, bar graph, area chart and many more.

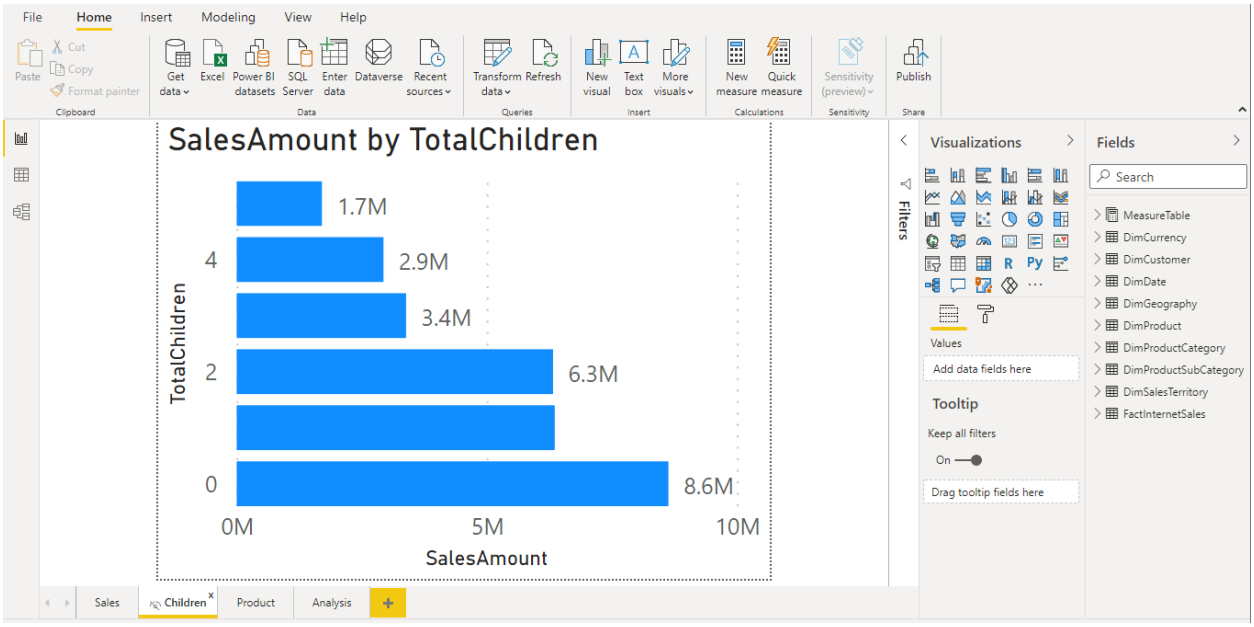
First, I have created a measure table where I have calculated all the sales amount with the help of sum function which sum all the values from sales amount field from FactInternetSales table.
SalesAmount = SUM(FactInternetSales[SalesAmount])

With the help of total salesamount I have visualized card chart of the total sale. Apart from that I have also counted salesorder, total number of customer then total product and countries.



In the dashboard, right side at the top there are three slicers and under that, there is a visualized information in a comprehensive way. With the help of a slicer, users can check filtered information in the visualization by selecting different times and products or sales territory.
Here, I have created three slicer named Product Categories through which we can see specific information regarding the different product, second slier is SalesTerritory which is used to gain information with respect to different countries like Australia, Canada, US, Uk, etc. third slicer is for the calendar year in which we can change different years according to our requirement.

In the visualization at the top tile, an area chart is used to visualize the trends in the market with values. An area chart is very useful at displaying the trends in a set time interval and as for displaying sales performance, it also compares two or more values through which we can get an idea about the sale. Here, we can see a line for the sales amount for every year with respect to other year and easily compare them. All the charts in the PowerBI are interactive like we can check to get the specific information regarding the graph by just hovering on the lines.



Power BI provides the treemap which displays the data in rectangle boxes and I have used it for sales amount by different sales territory groups. Treemaps are very useful to display the largest data in a small region. We can see from the treemap that North America had the highest sale among all the countries which is accounted for around 11M while Europe and the pacific share the same amount of sales.

Besides the treemap, I have created a 100% stacked bar chart which is used to show the relative percentage of different data series in the stacked bars in which the total of every stacked bar equals 100%. Here, I have used this chart for displaying the data for different categories with respect to the territory. Then, with the help of a treemap, I have created two visualization charts such as sales by marital status and sales with respect to gender. In which I can see that overall we cannot see any sale difference by gender as they have same sale amount which is around 15M. I also added a drill-down facility to explore more information in the graph.

At the bottom, I have used three stacked bar graph in which category is divide into small categories and we can see the relationship of each part on the total amount. It is also used to show the comparison between categories of data which can break down the amount into smaller parts according to categories.

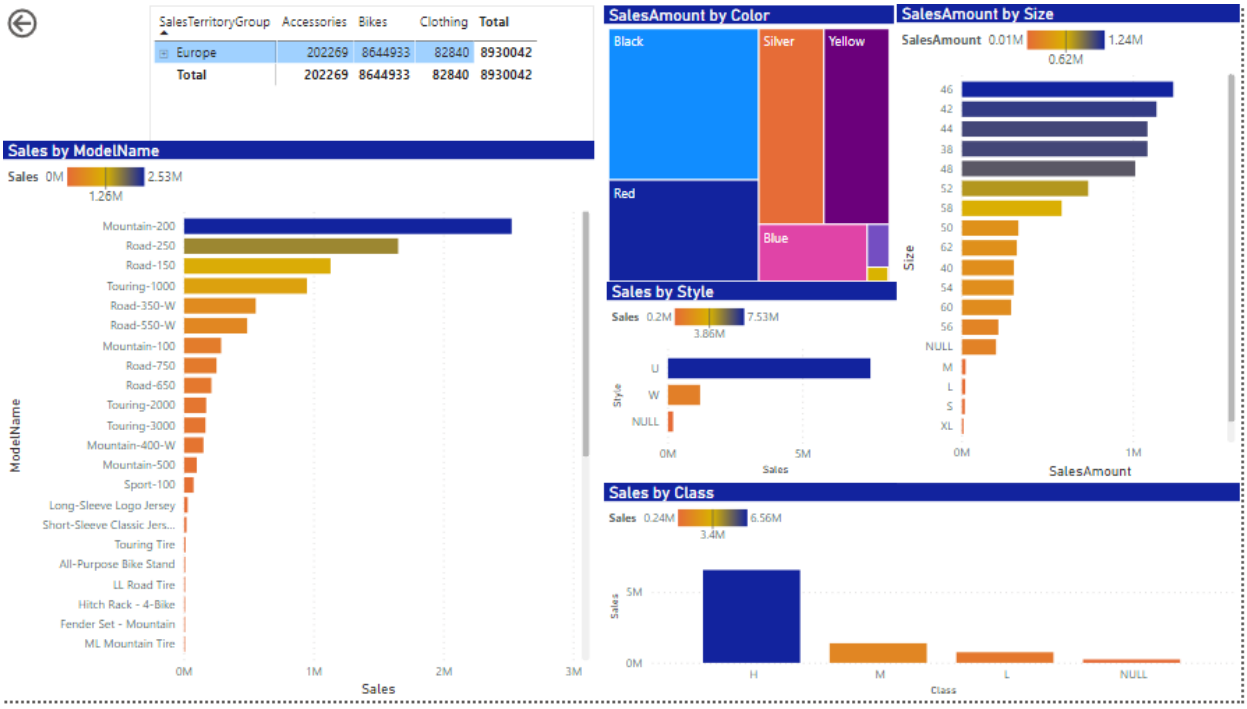
```
Customer Age =
VAR __AGE = DATEDIFF(DimCustomer[BirthDate],TODAY(),YEAR)
RETURN
SWITCH(TRUE(),
__AGE > 100, "101-110",
__AGE > 90, "91-100",
__AGE > 80, "81-90",
__AGE > 70, "71-80",
__AGE > 60, "61-70",
__AGE > 50, "51-60",
__AGE > 40, "41-50",
__AGE > 30, "31-40",
__AGE > 20, "21-30",
```

__AGE > 10, "11-20","01-10")

In the first graph, customer age is divided into smaller categories and according to that, we got the result of sales amount. I have counted the difference between the present date and birth year of the customer and allocate it into various age categories. We can see age group of 41-50 has generated more revenue while there are close to 0M sales by the customer who has an age greater than 91.

In the second, stacked graph I have divided the sales amount with respect to the cumulate distance that traveled by customer to purchase the item and it has interesting facts that 11.2M sale that company created is within the circle of 0 to 1 miles.

In the third graph, I have checked the distribution of sales according to state province and I found out that California had the highest sales amount while there are 11 countries that have total sale amount less than 1M.

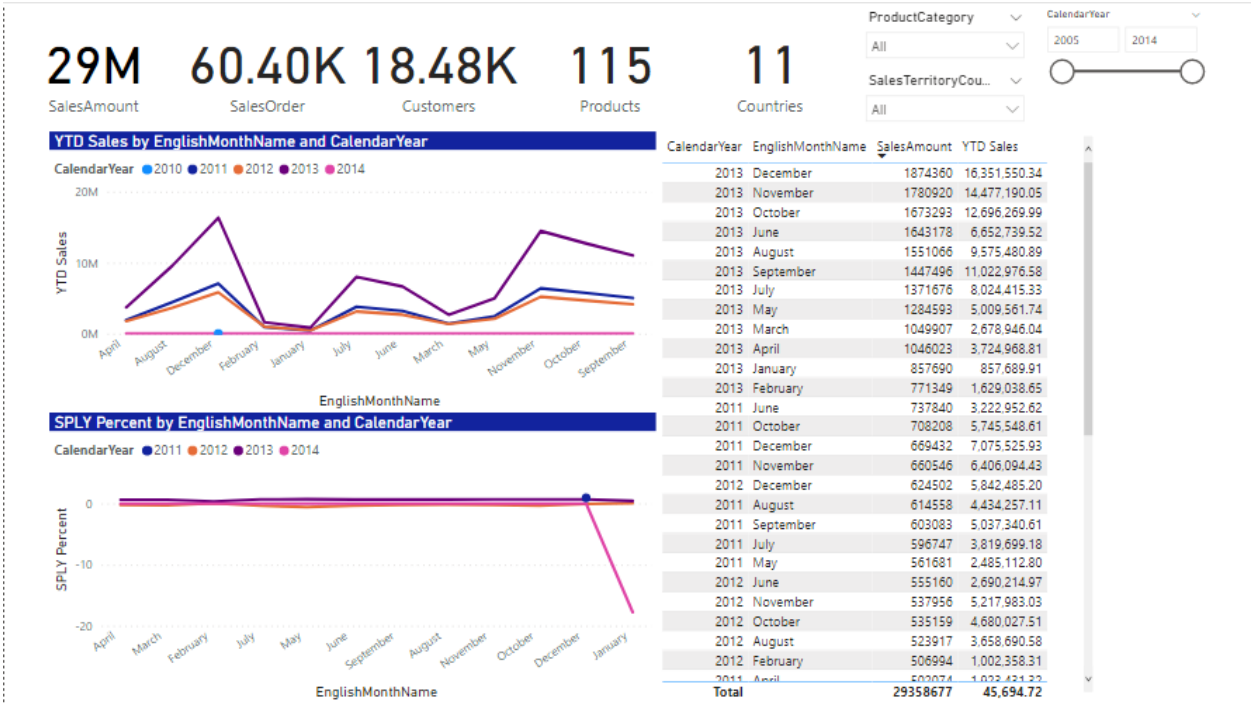


I have created a separate page for the details of the product. As we can see I have started the page by creating a table of salesterritorygroup, accessories, bikes, and clothing. Through the table, we can check the count of each item by their country. The next chart that I created is a treemap to check is there any hidden pattern in the purchase. I have created the treemap with different colors and saw that product which has black color are sold more than any other colors.

Power BI has a chart called clustered bar chart which displays the values in which the length of bar or column is proportional to the data. The clustered chart can also be used to juxtapose categories in one label item categories. Clustered bar charts are means to show the measure against a single dimension. I have used a cluster bar diagram for three different purposes. In the first diagram, I compared the sales with respect to model names, in the second graph, I have checked the sales amount by a style which has main two styles called: U and W. and in the third graph sales amount with respect to size. From all the

graph, I got the information that Mountain-200 is very popular model which is followed by road-250. So, from this data company can focus on improving lower sale model.
In the last, to check the sales by class I have created bar graph. We have total 4 class which are: H, M, L and last one is null. Among all, H class had the highest sales then followed by M which is around 0.1M.

I have created the separate page to see some analytics.
In the last, I have calculated growth percentage year by year so we can see our company is having profit or loss with exact amount with respect to last year. To understand the process by example, In calculating growth, sales of January 2020 will be compared with the sales of 2022. For this calculation we have DAX function called “SAME PERIOD LAST YEAR”



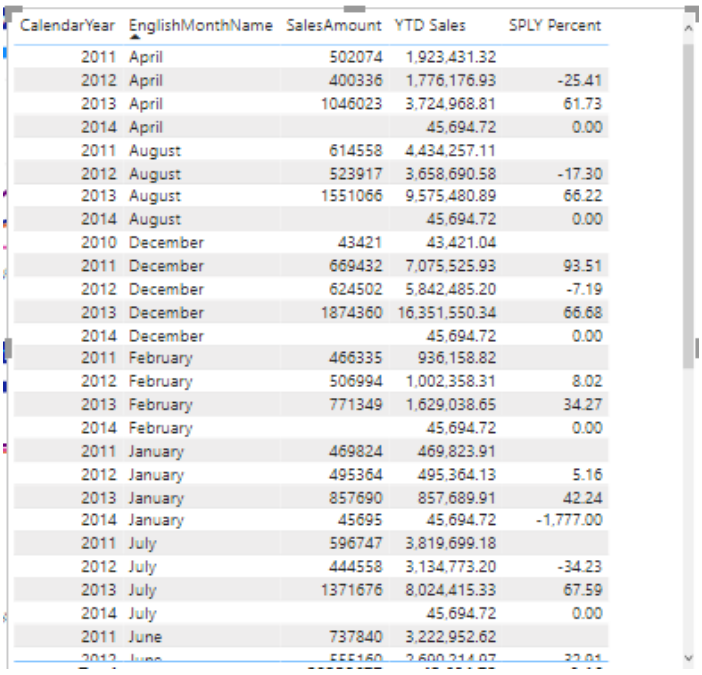
Codes :

SPLY = CALCULATE([SalesAmount], SAMEPERIODLASTYEAR(DimDate[FullDateAlternateKey]))

It will calculate the growth of sales compared to last year in percentage.

SPLY Percent =
VAR _Differ =
IF(
ISBLANK([SPLY]),BLANK(),
[SalesAmount]-[SPLY]
)
RETURN
DIVIDE(_Differ,[SalesAmount],0)

YTD Sales =
CALCULATE([SalesAmount], DATESYTD(DimDate[FullDateAlternateKey]))



CalendarYear	EnglishMonthName	SalesAmount	YTD Sales	SPLY Percent
2011	April	502074	1,923,431.32	
2012	April	400336	1,776,176.93	-25.41
2013	April	1046023	3,724,968.81	61.73
2014	April		45,694.72	0.00
2011	August	614558	4,434,257.11	
2012	August	523917	3,658,690.58	-17.30
2013	August	1551066	9,575,480.89	66.22
2014	August		45,694.72	0.00
2010	December	43421	43,421.04	
2011	December	669432	7,075,525.93	93.51
2012	December	624502	5,842,485.20	-7.19
2013	December	1874360	16,351,550.34	66.68
2014	December		45,694.72	0.00
2011	February	466335	936,158.82	
2012	February	506994	1,002,358.31	8.02
2013	February	771349	1,629,038.65	34.27
2014	February		45,694.72	0.00
2011	January	469824	469,823.91	
2012	January	495364	495,364.13	5.16
2013	January	857690	857,689.91	42.24
2014	January	45695	45,694.72	-1,777.00
2011	July	596747	3,819,699.18	
2012	July	444558	3,134,773.20	-34.23
2013	July	1371676	8,024,415.33	67.59
2014	July		45,694.72	0.00
2011	June	737840	3,222,952.62	

top sales order number

SO43697

3578

SalesAmount

SO43702

3578

SalesAmount

SO43703

3578

SalesAmount

SO43706

3578

SalesAmount

SO43707

3578

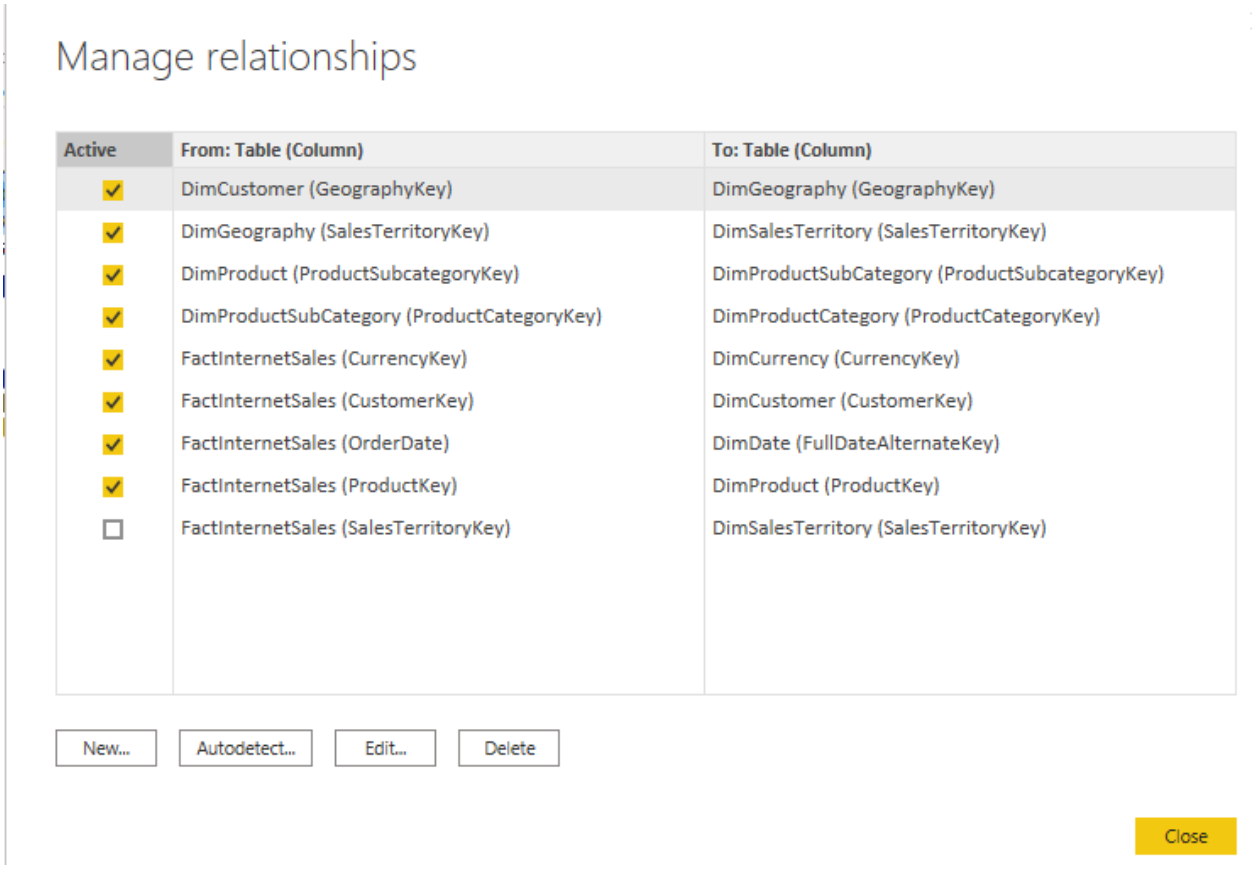
SalesAmount

SO43709

3578

SalesAmount

With the help of Dax, I found out the sale percentage of the company with respect to last year's sale which can be seen in the figure.
In the next figure, Power BI provides a powerful tool through which we can ask any query to the system and it searches for an appropriate answer. The main advantage of this is we don't need to remember the codes as we can do it with regular language. In the below image, I have created the relation between table using primary key of that columns which help us to get data from other tables too.



Results/Discussion :

This whole empirical process is a workflow to transform the raw data into information and the aim of this study is to provide valuable information to make a decision. From the dashboard, I had some useful stats about the sales like total sale is 29 million and for that total 60.40K order done which consist 18.48K customer and 115 products.
If a company is thinking about gaining more profit then there is some good suggestion derived from the data like from the commute distance graph, if they want to open new store then it should be around 5-10 miles from its other store, because as the distance increase sales decrease but if we open new store withing proper distance so people can travel to reach there then it's a good option.
Other than that, we know from the data that as the customer age increases the amount of sales decrease. It is also noted good to note that there is no difference in sales because of gender.

Conclusion:

From the analysis, I can clearly find out noticeable points from the report which was the problem earlier. Power bi used for making a bold decision for the companies like with the help of a graph, the company can work on how to increase the sales and check which item had the highest sales and lowest sales with lots of various filters like search by model name, class, colors, etc. power bi provides an easy collaboration of different types of database which is very effective. Here in this project, I have included more than one data file and it handles the files very accurately with the help of the power bi query editor tool.

References :

Clark, D. 2017. Beginning Power BI: A Practical Guide to Self-Service Data Analytics with Excel 2016 and Power BI Desktop. Book from ProQuest. Accessed 2 January 2019.
https://ebookcentral.proquest.com/lib/laurea/detail.action?docID=4815394&query=clark

History of Business Intelligence. No date. Accessed 13 December 2018.
https://www.panorama.com/blog/history-business-intelligence/

Microsoft. 2018. Intro to dashboard for power BI designers. Accessed 5 January 2018.
https://docs.microsoft.com/en-us/power-bi/service-dashboards

The History of BI: The 2000's and now. 2014. Accessed 14 December 2018.
https://dataconomy.com/2014/07/the-history-of-bi-the-2000s-and-now/