

ANALYSIS OF DATA VISUALIZATION IN MOUNTAINEERING EQUIPMENT SALES

Ganesh kumar

X16130961

MSc in Data Analytics

Abstract:

In Current world, the data has been generated in last amount from various sources. Analyzing of those vast amount of data is difficult task. Exploring of huge amount of data require data mining techniques and may not be understood by most of the users. Hence Visualization is powerful tool to analyze and explore data. It helps even new user to understand the data and to explore the data. This paper deals with visualizing ang analyzing of mountaineering product equipment sales in various stores such as outdoors shop, sports store, department store, golf shop, rental store and warehouse store in the year 2012 to 2014.

1.Introduction:

In New world, youngsters like doing different kind of adventures in their life. Mountaineering is one of the adventure which attracts most of the people towards it. Outdoor industry became the boon for youngsters to get required equipment for the mountaineering from outdoor stores and retail stores. Industries in outdoor industry was categorized by their ownership type. Privately owned companies, and Publicly owned companies.

This paper mainly deals with visualization of sales in mountaineering equipment from these companies even though it was briefly explained outdoor industry at first. Sales contains profit, Revenue, unit cost of product, Product type, Product line, Gross margin, Unit cost, Unit sales price as their attributes. These attributes plays an important role in making effective visualizations.

This paper further describes the related work in section 2 and about the dataset in section 3. In section 4, it describes the data cleansing and finally it describes the implementation and visualization in section 5.

2.Literature Review:

Visual exploration makes humans to explore the dataset much more efficient and empowers perceptual capabilities for data navigation. [Keim, 2002] The idea of interlinking multiple sheets to overcome shortcoming in single technique is called linking and brushing. Linking and Brushing are the techniques in interact with multiple sheets at a time. Change in single sheet should change the visualization in other techniques as well. [Keim, 2002]

Difference between linking and brushing is very little. Brushing is likely pointing out the item with mouse. When a mouse is placed on certain point, it will highlight the certain point. Most of the cases we delete by using same method. Linking is like when you point out a mouse on certain point, it will change the same item in other sheets. [voigt,2002]

Most common color vision disease(CVD) is dueteranamoly in which peak sensitivity of the medium-wavelength photopigment is “shifted” abnormally close to the peak sensitivity of the long-wavelength photopigment. Frane, A. (2015).

Interactive data visualization provides the users to navigate data and to select and display the data which easy to use and interact. It is often used as a Part of Data Analytics which helps in visualizing and exploring the data. [Heer, 2008]

Color blindness Simulator is a tool for people who suffers with color deficiency to simulate them with appropriate colors.

3. Background of Dataset:

Data set is taken from IBM Watson analytics which contains 20000 rows and 11 attributes containing sales of mountaineering equipment from year 2012 to 2014. It contains attributes such as Revenue, planned revenue, gross profits, quantity, unit cost, product cost, retailer country, order method type, product line, product type, product and unit sale price.

4. Preprocessing of Dataset:

Tableau 10.3 – To visualize the data in effective manner.

Rapid miner – To know the missing values and to clean the dataset.

Dataset which is taken from IBM Watson analysis contain many missing values and hence there is need of missing value of treatment inserting in data. For better understanding of data, rapid miner is used. It shows the statistics of each attribute in a dataset and missing values contain in each attribute.

Name	Type	Missing	Statistics		
Product line	Polynominal	0	Outdoor Protection (0)	Camping Equipment (6027), Camping Equipment (6027), Personal Accessories (3973), ...[3 more]	
Product type	Polynominal	0	Least Woods (0)	Most Eyewear (2062)	Values Eyewear (2062), Watches (1911), ...[19 more]
Product	Polynominal	0	Least Trail Star (0)	Most Bella (147)	Values Bella (147), Canyon Mule Carryall (147), ...[142 more]
Order method type	Polynominal	0	Least Web (1428)	Most Telephone (1432)	Values Telephone (1432), E-mail (1428), ...[5 more]
Retailer country	Polynominal	0	Least United Kingdom (476)	Most Brazil (477)	Values Brazil (477), Canada (477), ...[19 more]
Revenue	Real	6125	Min 0	Max 5606597.900	Average 155541.548
Planned revenue	Real	6125	Min 240	Max 5606597.900	Average 163248.482
Product cost	Real	6125	Min 133.600	Max 3734223.490	Average 96331.818
Quantity	Integer	6125	Min 1	Max 183110	Average 2809.432
Unit cost	Real	6125	Min 0.850	Max 490	Average 76.074
Unit price	Real	6125	Min 2.060	Max 831.880	Average 132.379
Gross profit	Real	6125	Min -1119.040	Max 2262679.560	Average 59209.730

Fig 1. Cleaning the Dataset

There are two methods in handling missing values. List wise technique and Pair wise technique. It is important to analyze data that contains missing values at random(MCAR).

List wise technique:

It is a technique to delete the case which has one or more missing values. It has more disadvantages and it is not most considerable by Researchers due to its disadvantages. It produces bias and estimates in the dataset which made it not to be considered much. It doesn't support the assumptions of MCAR.

Pair wise technique:

It is a technique which minimizes the loss that caused in list wise technique. It finds the relations between the variables. The advantage of this method is that it will keep everything to analysis. It is preferred technique compared to List wise technique.

After completing the data cleansing, data is uploaded to Tableau in which measures and dimensions are formed based on attributes.

5. Implementation and Visualization:

To visualize the data, Tableau is used because of its advantages. Tableau is the best tool to visualize and its friendly to user to interact with it. Visualization with tableau is easy and there is no limit to data size. It is flexible in analyzing the data and does not need to be recreate reports. A fresh user can easily understand and interpret the results.

Why tableau?

- It Analyses data in real time and encourages experimentation.
- Supports all types of formats such .csv, .xml, .xls etc.
- User interface is very clear and rich and does not take much time in producing the results.
- Extraction of data can be done in tableau. It also supports joining of one or more datasets at a time.
- It has options of highlights which is more effective in visualizing and also filters the bad data.

Case studies:

The following are case studies to analyze and forecast the dataset.

- 1.Which Country is receiving more Gross profits while spending low investment on products?
- 2.Which is the most common interface used by people to order products? (Order type)
- 3.To forecast the year which has more and least revenue/ profit and What are the products that causing high investment in producing them?
4. Which product type is mostly liked by people in France and China and which products can be produced in markets of both countries?
- 5.To measure the performance of the "Ordering types" and to predict the which "Order type" will gain more profits.
- 6.Which product is generating more revenue and more gross margin?

Case study 1:

Which country is receiving more Gross profits while spending low investment on products?

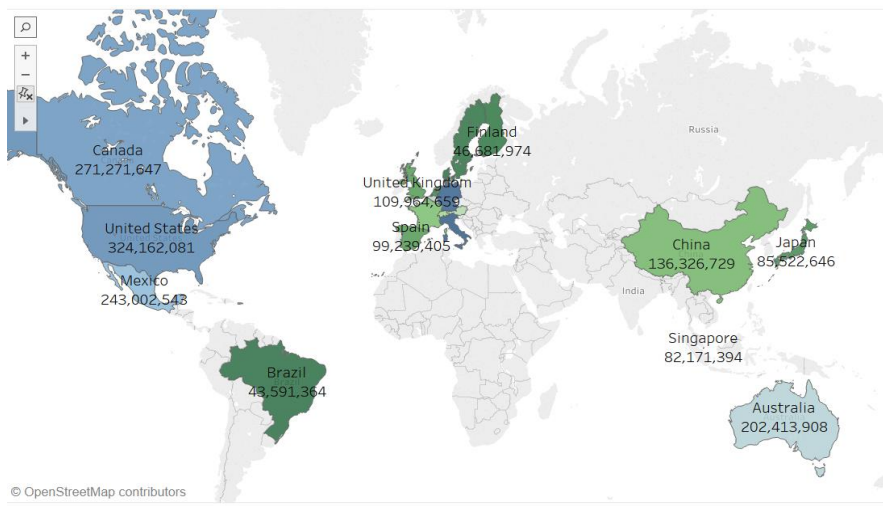


Fig 2 Planned Revenue

Revenue is the total income gaining by certain country including profits and investment. From the above figure, we can see that United states has more revenue with an income of \$324162081 followed by Canada with \$271271647. Brazil has low revenue with \$48591.364.

To know the Country which is gaining more profit?

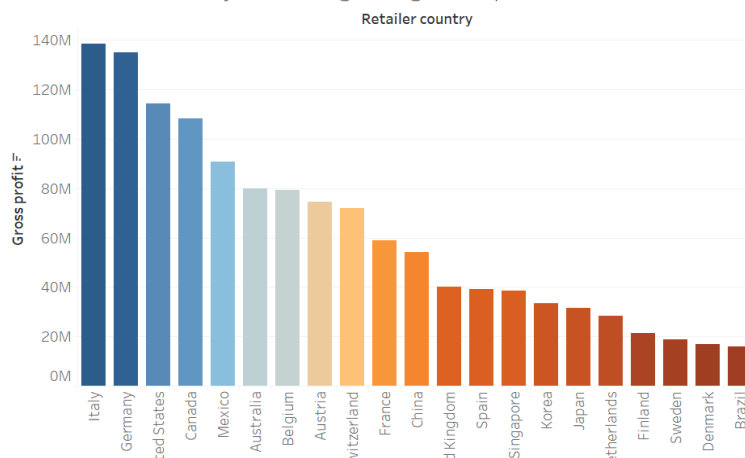


Fig 3 Gross profit

Gross profit is calculated by excluding investment in revenue. It is calculated as follows.

Gross profit = Revenue - Investment

From above figure, Italy has more gross profit followed by Germany with \$140 approximately. Brazil has least gross profits with \$20M.

Countries such as Italy and Germany are receiving more profits with low investments on products.

Case study 2:

Which is the most common interface used by people to order products?

order type

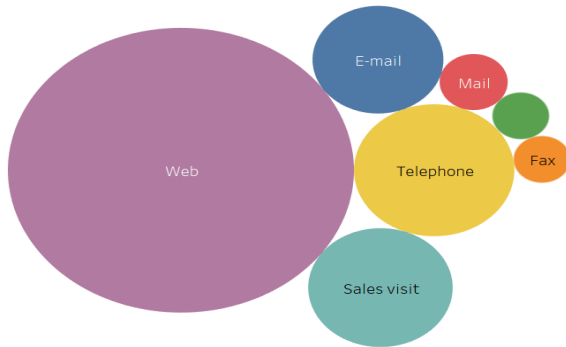


Fig 4 Order type

From the figure, it clearly indicates that **Web** is the most commonly used interface to order their products. **Telephone** and **sales visit** stands followed by one with very slight difference. **Fax** is the least term used by the customers.

Case study 3:

To forecast the year which has more and least revenue/ profit and What are the products that causing high investment in producing them?

Revenue and profit

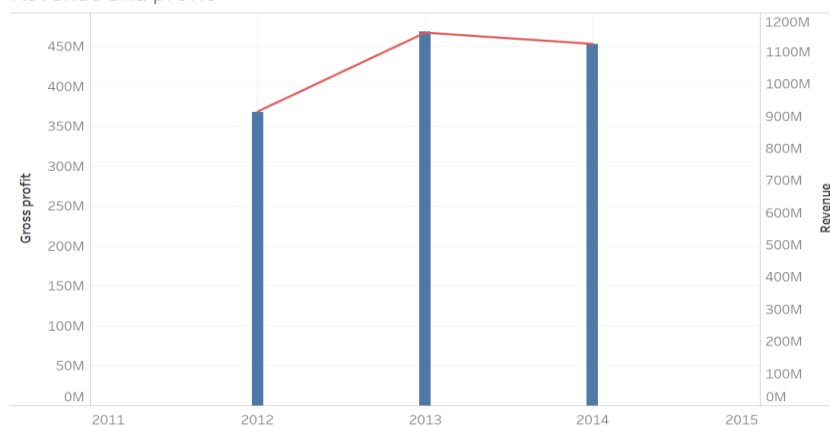


Fig 5 Revenue/ profit

From the above figure, Blue line indicates Revenue and red line indicates profits. We can depict results such as **2013 has more** profits and revenue in both. **Year 2012** had very hard times in which Revenue and Gross profits has least points in graph. But they gradually increased to 2013.

In year 2014, both profit and revenue gradually decreased 450M to 420M and 1190M to 1100M respectively.

Product cost and unit cost

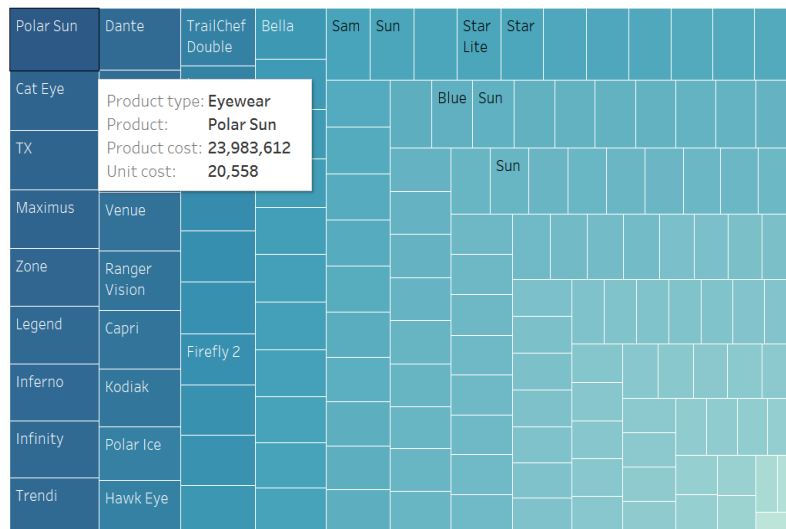


Fig 6 Product and unit cost

Product cost and unit cost tells about the investment of product. From above picture, unit cost of polar sun is high with \$20558 and product cost is \$23983612. Followed by, Products such as Cat eye, Tx, maximus, Zone etc., have most unit and product cost.

Polar eye, cat eye, Tx, maximus, Zone are demands high investments.

Case study 4:

Which product type is mostly liked by people in France and China and which products can be produced in markets of both countries?

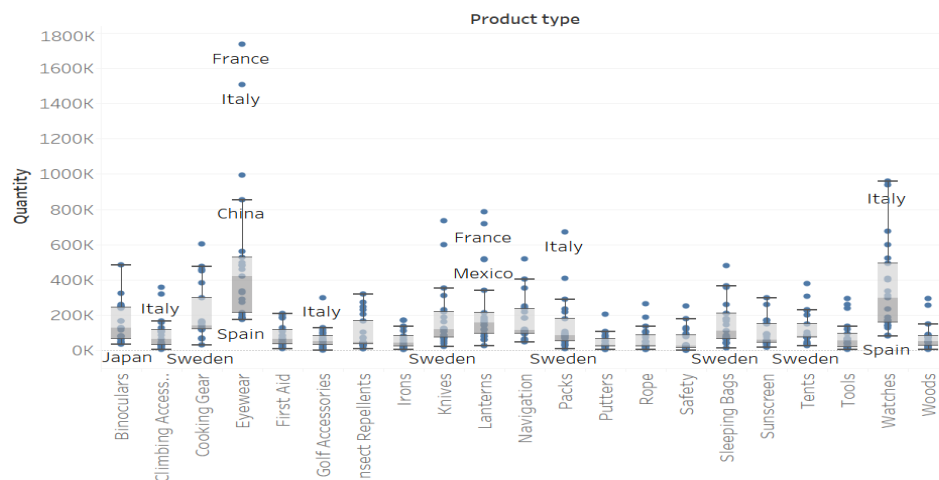


Fig 7 Quantity of Product type

Eye wear is product type which attracted most of the people in France and China which can be interpreted from figure. In addition, People in France shows interest towards lanterns.

Eyewear products can be produced more in markets of **China and France**. **Lanterns products** can be produced more in **France market**.

Case study 5:

To measure the performance of the “Ordering types” and to predict the which “Order type” will gain more profits.

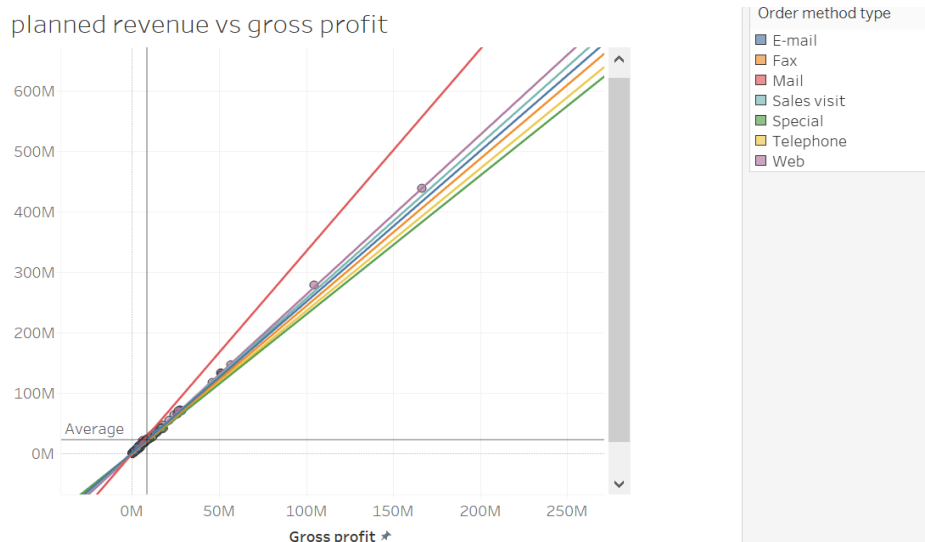


Fig 8 Planned revenue vs Gross profit

To predict the “Order type” which gains more profits, I divided the scatterplot in to four areas by taking average on both gross profit and planned revenues. By inserting trend lines, each color indicates respective Ordering type such as Email in blue, Fax in yellow Etc.

Violet colored line indicating Web expands in both profit and revenue predicts that It will bring more gross profits.

Case study 6:

Which product is generating more revenue and more gross margin?

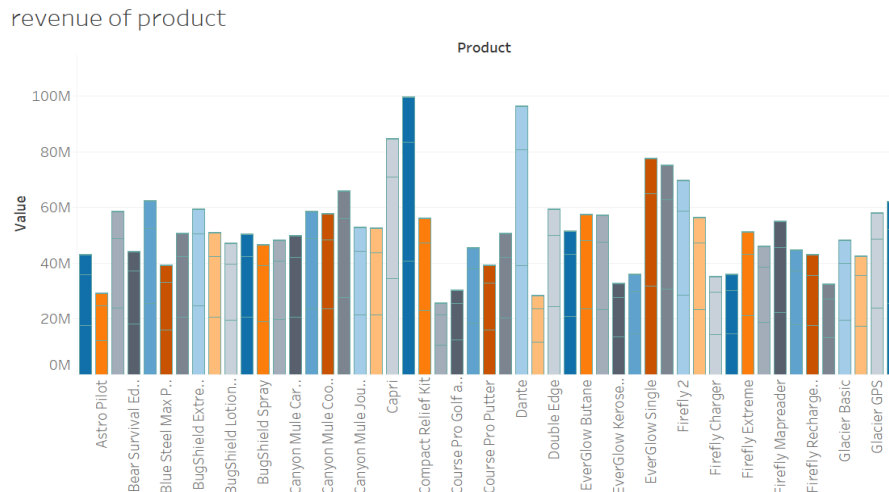


Fig 9 Revenue of product

Capri is generating more revenue and Compact retail kit is producing less revenue which can be interpreted from above figure.

gross margin

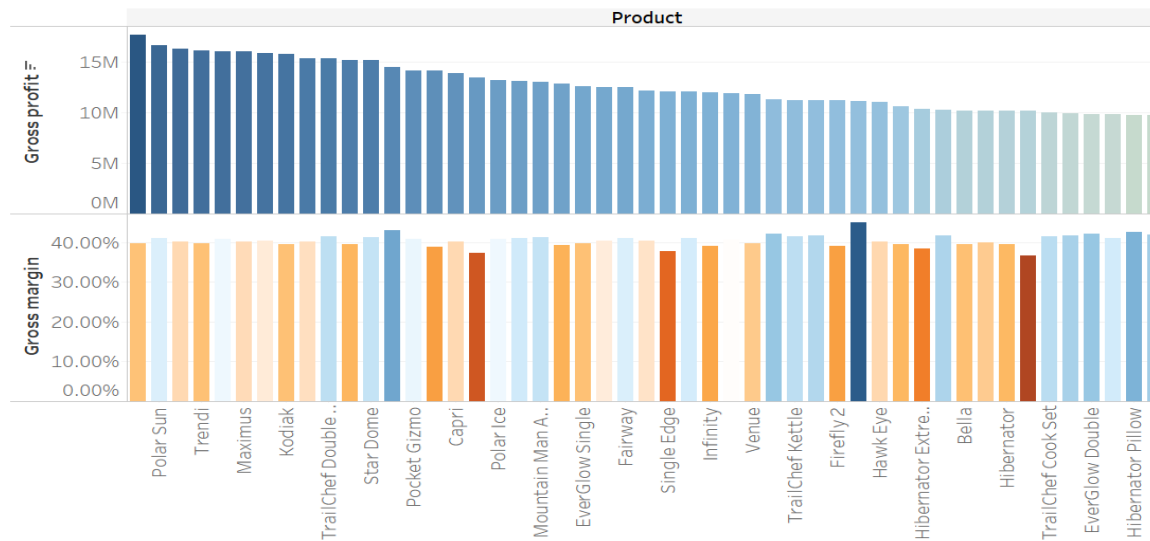


Fig 10 Gross margin and gross profit of product

Gross margin is calculated by dividing gross profit by revenue.

Gross margin = Gross profit/ Revenue.

From the figure, Firefly 2 has more gross margin with 40 percent and most of the products are little less with 37 percent approximately

Creating Dashboards:

Linking and brushing are the concepts in creating Dashboards to compare multiple sheets at a time to make effective visualization and to overcome the effects in single technique.

Dashboard 1:

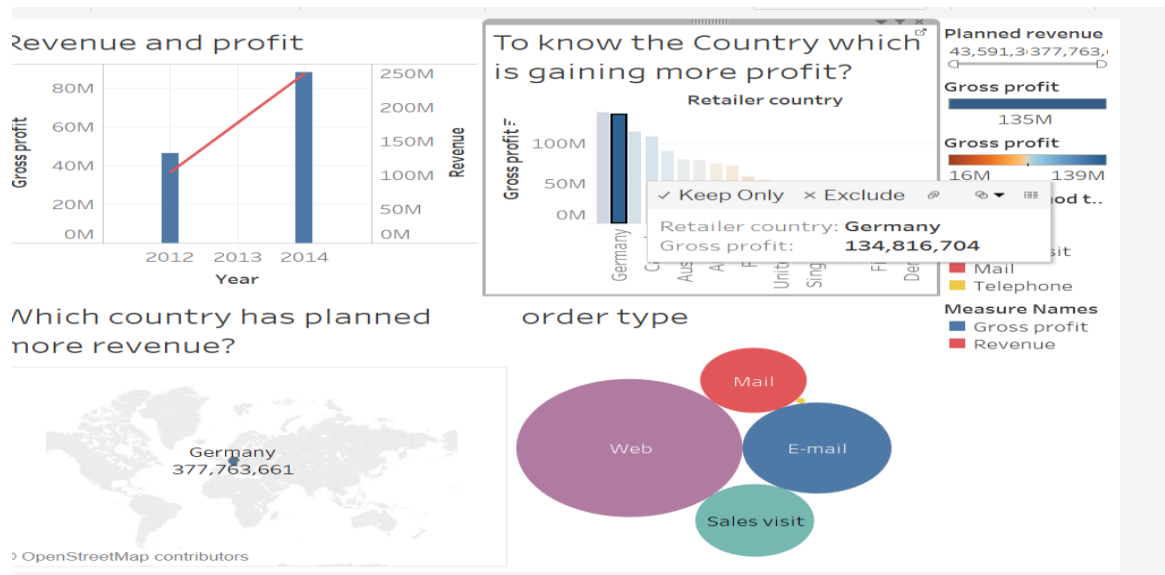


Fig 11 dashboard 1

Dashboard is created by bringing the figures in case studies 1,2 and 3. We can see from the figure that, when Germany is used as filter, It shows the gross profit ,revenue and their Order type respectively.

Dashboard 2:

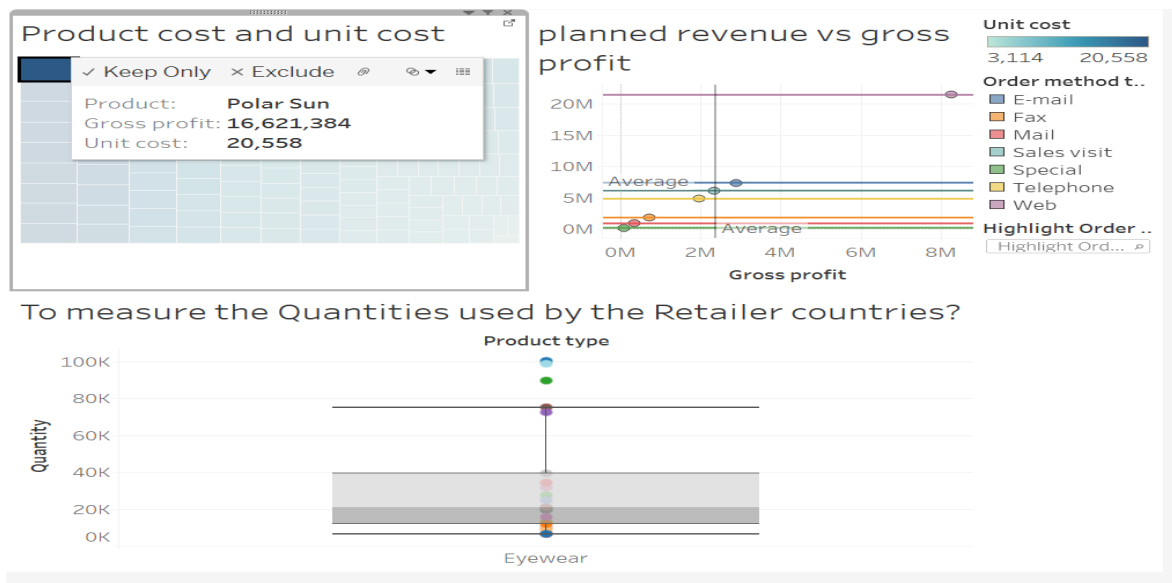


Fig 12 Dashboard 2

Dashboard 2 is created by using case studies such as 3,4 and 5. When product polar sun is used as filter, then its visualizing the gross profit and product type of that product.

Color Blindness Simulator:

Color blindness simulator is a case where provides respective colors to people who suffers color deficiency problems.

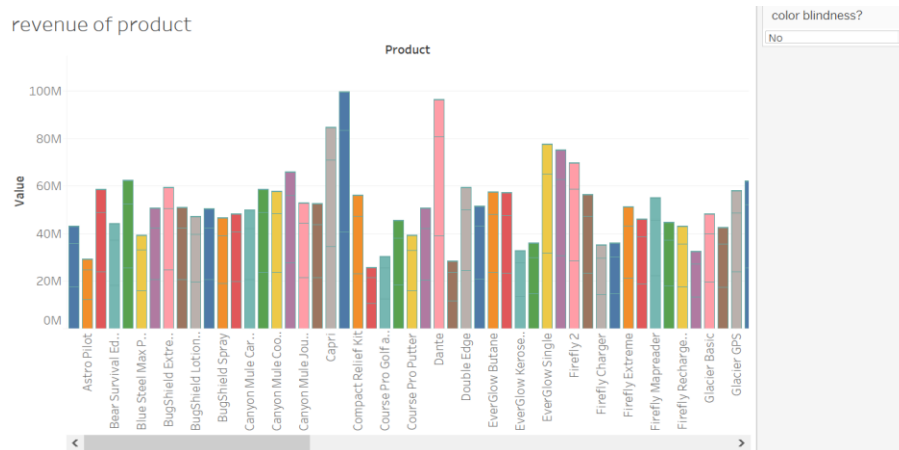


Fig 13 visualization to normal people

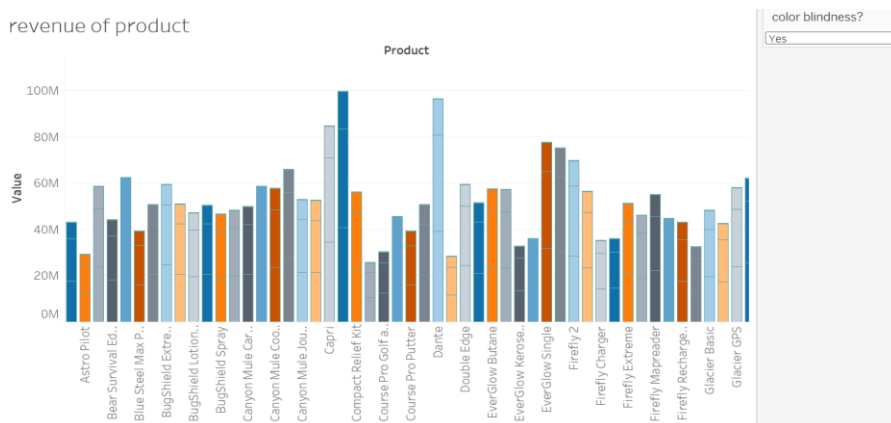


Fig 14 Visualization to colorblind

From above both figures, we can see that when a person is colorblind, the colors are changed from their colors to colorblind colors.

Conclusion:

Visualization helps in understanding the dataset in a clear perspective even in cases of complex datasets. From the report, we conclude that United states, Canada and Australia has high revenues and Italy, Germany, ted states and Canada has more gross profits. People likes to order through web more. Products such as firefly, polar sun, Trendi, Maximus, Kodiak have high gross margin which are calculated from gross profits and investments.

References:

[Keim, 2002]: D. A. Keim, Information Visualization and Visual Data Mining, *IEEE Transactions on Visualization and computer graphics*, 2002.

[Voigt, 2002]: Robert Voigt, An Extended Scatterplot Matrix and Case Studies Information Visualization, Master's thesis, Hochschule Magdeburg-Stendal, Linking and Brushing. 2002

Frane, A. (2015). A Call for considering Color Vision Deficiency When creating Graphics for Psychology Reports, *Journal of General Psychology*, 142(3), 194-211.

Heer, J., Mackinlay, J., Stolte, C., & Agrawala, M. (2008). Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE transactions on visualization and computer graphics*, 14(6).

<http://www.color-blindness.com/2015/05/11/color-blind-check-released/>