

1. What did you do?

As per the instruction given by Professor Murry in the class, I logged into Teradata website and opened SAS Visual Analytics. I was new to SAS, hence I watched videos available in Assignment 1, Assignment 2 and Assignment 3 of SAS Visual Analytics Course Content available on the Teradata website. Additionally, I watched few YouTube videos to gain more understanding on SAS Visual Analytics. This helped me to understand that SAS Visual Analytics tool is similar to Tableau and Micro Strategy. The last assignments gave a good hands-on experience on both the tools. Hence, I was very excited and confident to learn SAS Visual Analytics.

Designing Business Intelligence Reports

I read the business case of Assignment 3 and wrote 6 business questions for developing business intelligence report for Insight Toy Corporation that would help to build a new strategy for best selling toy product. It was important to select appropriate graphs to solve these business questions to ease the analysis.

1.1 What is the customer feedback different facility continent? Which continent has the highest and lowest customer satisfaction?

There were many measures that reflected customer feedback. Hence, I used 3 measures, such as Customer Satisfaction, Facility Efficiency and Sales Rep Rating. I wanted to compare it for different facility. Since, it was comparison I used bar graph representation. Additionally, graph can be **drilled down to understand the specific of each continent**. This graph would help to focus on high and low customer feedback continent and analyze the continent that needs to be focused upon.

1.2 How does cost of sales, product sales and profit changes in different years?

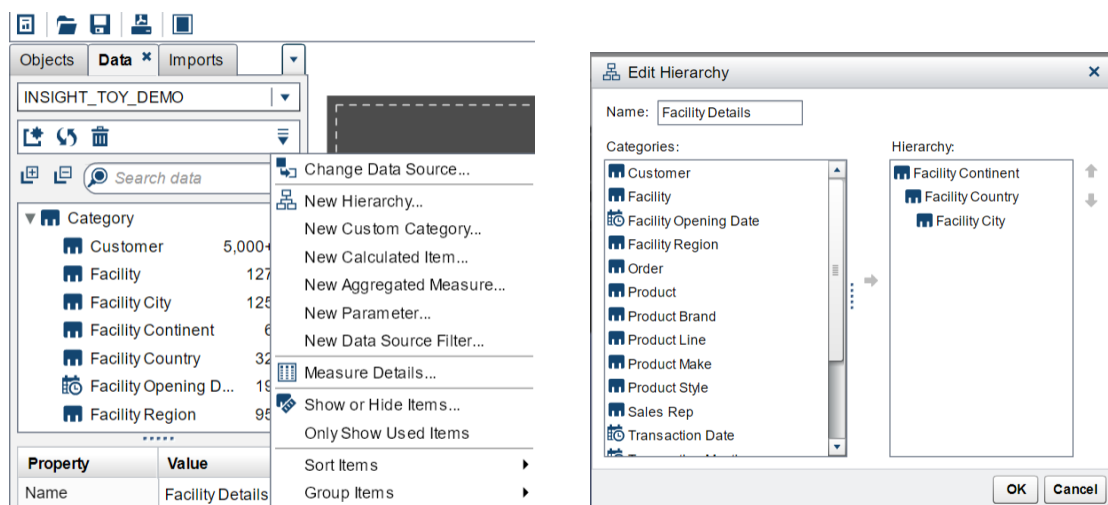
To answer the above question, I added **Transaction year in category and Product cost, product price and product sale in measures**. The trend helped me to understand that Cost of sale and product sales varies in the same way. This was not giving me a clear representation for giving insights. Hence, I decided to create a calculated field for profit. The **calculated field** had the **Product sale -Product cost of sale** and was named Profit. I dragged the profit field to the measures. This helped to gain insights of variation of profit for different years. Aggregation of all the measures was changed to average.

1.3 How does profit vary for different product lines across years?

From the above question, we gain insight that profit was low in 2009. So, I decided to **check the profit for different product line from 2008 to 2012** and understand whether profit is affected by product line across these regions. I dragged product line and transaction year to category and profit to measure. I can drill for different Product line. Additionally, I sliced the transaction year from 2008 to 2012.

1.4 How does product cost and product sales vary across different facilities regions?

To answer the above question, I needed more than one categorical and measures. The facility region consists of Facility Continent, Facility Country and Facility City. I tried to plot dual axis bar chart. The representation was cumbersome. Hence, I checked the other option in Object Menu. I realized that I can use Tree map to show this representation. I watched the video on Tree map again and learned about hierarchy. But the video did not give any information on how to create Hierarchy. Hence, I found one YouTube video to learn the same. I clicked on down arrow below the data source and selected **New Hierarchy**. I dragged Facility Continent, Facility country and Facility state to Hierarchy. The representation was in Hierarchical order. I renamed the Hierarchy name to Facility details.



Finally, I order to view product sales and cost of product in different facilities, I dragged facility details to category, product cost of sale as size measure and product sale to color. This helped me to give a new representation of **Tree map**. I doubled clicked on each region in Tree map to view the variation in product sale and product cost for each region. **This helped me to Drill down from Facility continent → Facility Continent→ Facility City. Additionally, I could roll up from Facility city → Facility Country → Facility Continent.**

1.5 How are units affected for different product lines?

To answer the above question, I needed the comparison of **Units Capacity, Units targeted, Actual Units and Units discarded for different product line**. Hence, bar chart would be the best way of representation. Hence, I selected bar chart, **dragged product line to category. I added Units Capacity, Units targeted, Actual Units and Units discarded to measures**. After looking at the results, I realized that I could check the units affected in each product line for year 2008, 2009 and 2010. Hence, I sliced the data by selecting dropdown option from the object and adding transaction year to represent trends in 2008,

2009 and 2010. Looking at the above results, **I diced the data by adding Facility continent and drilled down to the continent that had highest and lowest profit.**

1.6 How does product quality affect the product line for different years?

To answer the above question, I needed trend of different years. Hence, I choose line graph for this representation. **I dragged Product Quality to measures. Additionally, I dragged product line and transaction year to category.** This helped to view dicing across three field. Hence, I could view the trend of product quality for different product line in different year. To gain more detailed view **I decided to dice and slice the data by using Transaction month as a slider and Facility Continent as a drop down.** This representation helped me to slice, dice, drill down and roll up the detail of quality for each product line in different year for different facility continent. Answering the above question helped me to understand that I need to focus on Africa and North America.

Assignment 4: Using Data Visualization for Problem Solving

I skimmed through assignment 4 of SAS Visual Analytics Course Content and wrote down the hypothesis. Working on Assignment 3 enhanced my confidence to use SAS Visual Analytics tool. Hence, I opened the Heart data set from the SAS data source and skimmed through all the categories and attributes that could be useful to answer the below hypothesis. Reading the word hypothesis, I felt that I need to perform some coding and statistical analysis to answer the question. But when I carefully skimmed through the data, I understood that these hypotheses could be easily proved by data visualization.

- **H1: The weight and cholesterol levels are closely related.**

To verify the H1 hypothesis, I plotted line graph of weight status and cholesterol. The line graph would help me to understand the correlation of weight status and cholesterol. When I dragged Cholesterol to measure and weight status to category. The resulting graph calculated the sum of cholesterol values. Therefore, I changed the aggregation to average to gain more precise results of cholesterol in each weight category.

- **H2: Men are usually more obese than women.**

To prove the above hypothesis, I had to compare weights of men and women. Since the hypothesis demanded comparison, I used bar graph for the visualization. First, I thought to use weight as the measure. But then realized that hypothesis demand comparison of men and women that are overweight. Hence, I used frequency as a measure. I dragged sex and weight status to category. This resulted in frequency of men and women in each weight status. Additionally, we can drill down to check the frequency of men and women by slicing the weight status for overweight men and women.

- **H3: Women usually smoke less than men, but their cholesterol level is higher.**

To prove the hypothesis, I had to compare cholesterol levels as well as smoking of men and women. Hence, I choose bar graph to show comparison. As there were two measures, I chose dual axis bar graph to reflect the comparison. The sex was dragged in the category. Cholesterol and smoking were dragged in the measures. Additionally, I changed the aggregation for both Cholesterol and smoking to average.

- **H4: The blood pressure is higher for people with higher cholesterol level.**

To prove the hypothesis, I used the dual axis line graph to verify the correlation between cholesterol and blood pressure. I was well versed that blood pressure is combination of systolic and diastolic values. The upper reading of Blood pressure represents the systolic values whereas the lower reading is the diastolic value. Since blood pressure was not directly available. I used systolic and diastolic measures and dragged them to the measure section of each axis. I changed the aggregation for both measure to average. I added Cholesterol status to category. Additionally, to reverify the hypothesis, I plotted a line graph of blood pressure status and cholesterol values.

Additionally, the business case in assignment 4 suggested to find some insight about the people who suffered from Coronary heart disease. To gain these insights, my approach was first to understand the frequency of cause of death caused by Coronary heart disease.

Which cause results in highest deaths?

To gain insight, I selected pie chart and added cause of death to category. I found few missing values. I sliced the data by removing the missing values. The remaining data was used for our analysis. I added labels for category name and percent of value from pie chart properties.

What is the average male and female age of diagnosis and death with Coronary heart Disease?

To answer the above question, I plotted a dual axis bar graph by dragging sex to category. I sliced and diced by adding age of death, age of diagnosis for Cause of death as Coronary heart disease. Additionally, I aggregated the measures to reflect the average values.

What is effect of cholesterol, smoking and weight on coronary heart disease?

I was curious to understand how can I add multiple charts in one section. Hence I dragged 3 pie chart object in one section. Each pie chart included the weight status, smoking status and cholesterol status as category whereas, weight, smoking and cholesterol as measure respectively.

What is the average cholesterol level and smoking rate for different weight status?

To answer the above question, I plotted a dual axis bar graph by dragging weight status to category. I sliced and diced by adding smoking and cholesterol for Cause of death as

Coronary heart disease. Additionally, I aggregated the measures to reflect the average values.

What is range of weight and height for people that died due to Coronary heart disease?

To answer the above question, I plotted a scatter plot of weight and height. I aggregated to the average value. Finally, I added drop down to reflect the height and weight for coronary heart disease.

How does cholesterol level and smoking level changes for different blood pressure status?

To answer the above question, I plotted a dual axis bar graph by dragging weight status to category. I sliced and diced by adding smoking and cholesterol for Cause of death as Coronary heart disease. Additionally, I aggregated the measures to reflect the average values.

2. What were the Results?

Assignment 3: Designing Business Intelligence Reports

2.1 What is the customer satisfaction, facility efficiency and Sales Rep Rating for different facility continent? Which continent has the highest and lowest customer satisfaction?

The below visualization helps to understand that **Africa has highest Sales Rep Rating** where as **Oceania has the lowest Sales Rep Rating**. Oceania has highest Facility Efficiency and Asia has lowest Facility Efficiency. **Customer Satisfaction is highest in Africa, but lowest Oceania**. Hence, the reports suggest that we need to focus on Oceania to improve Customer Satisfaction and Sales Rep Rating. Additionally, Asia has low values for all 3 measures. Therefore, it is important to focus on Asia to improve customer feedback.



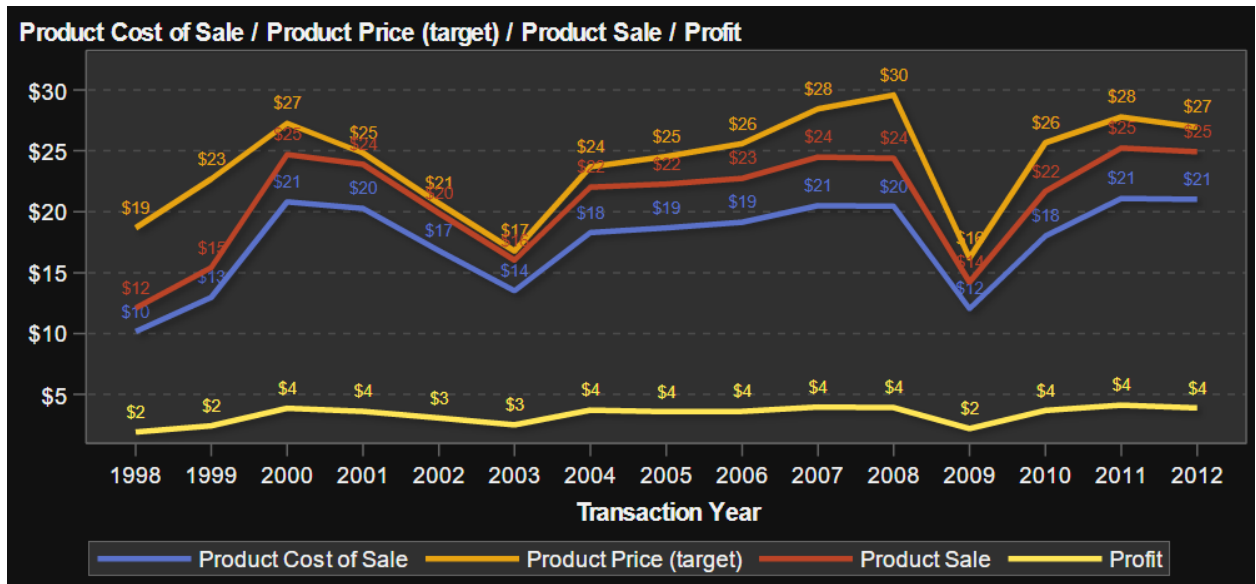
2.2. How does cost of sales, product sales and profit changes in different years?

The approximate average profit for all years is \$4. We can also see that the profit decreases in 2003 and 2009.

In 2003, Product cost of sale and product sale decreases drastically from 2001 but the profit is affected by \$1 whereas it increased by \$1 in 2004. Whereas, **in 2009 there is a drastic change in product cost of sale and product sale. But the profit varies by \$2.**

Additionally, even if the Product sale is high, the cost of good is also high, therefore the profit remains constant at \$4.

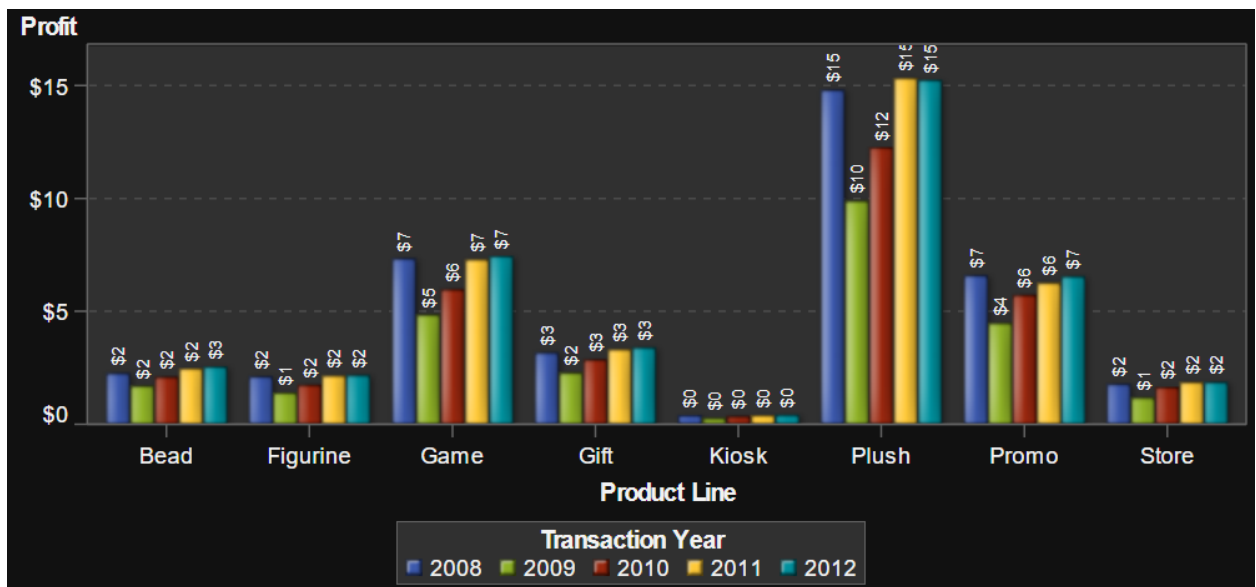
In 2002 and 2003, the product sale is approximately close to targeted product price. Whereas, in other years the product sale is less that target price. **These insights would help to drill down and understand the insights for 2009.**



2.3. How does profit vary for different product lines across years?

The profit for Kiosk is \$0. The company needs to find reasons for no profit for Kiosk and needs to focus on ways to increase profit.

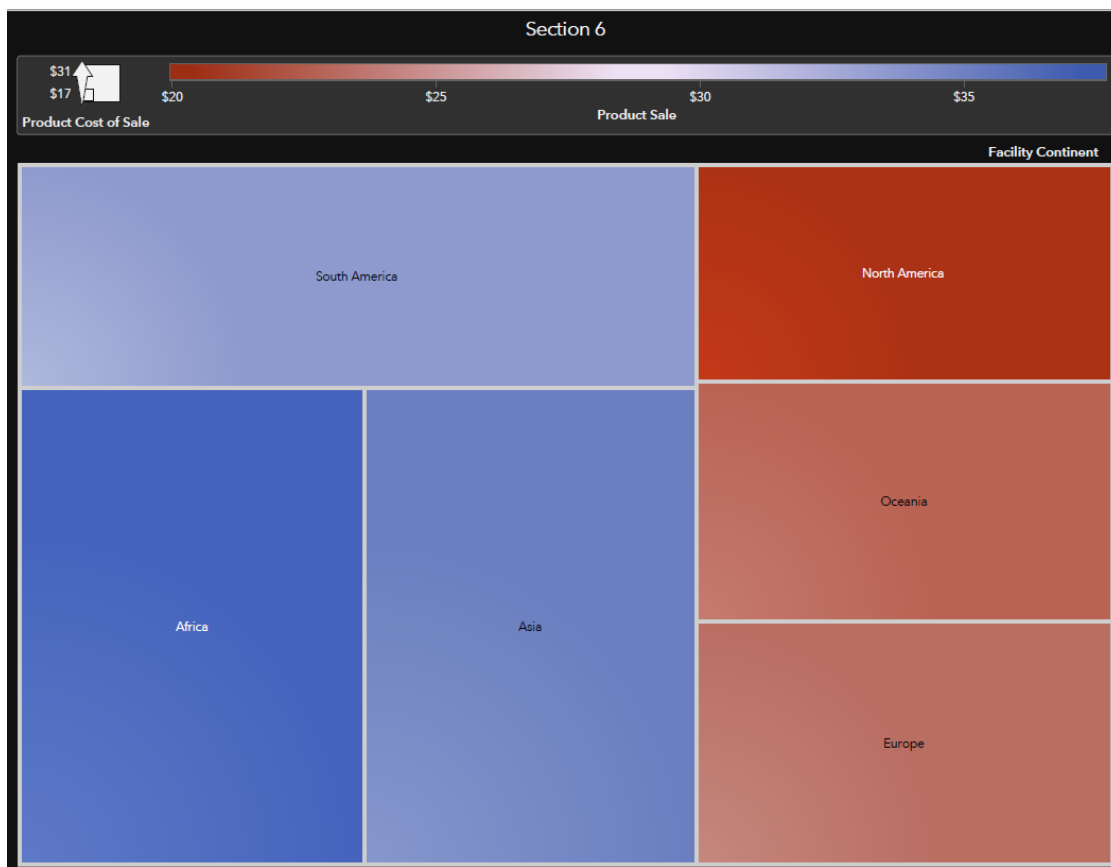
The Profit for Plush product line is very high. As observed the profit trend of Plush was very high in 2008 and drastically decreases in 2009. Then, eventually increases in 2010, 2011 and 2012. **This insight would help to understand that the profit in 2009 is highly affected by Plush product line.** Additionally, we can view that the profit ranges from \$2 to \$7 for other product lines. Hence, company can focus on ways to increase the profit for other product lines.



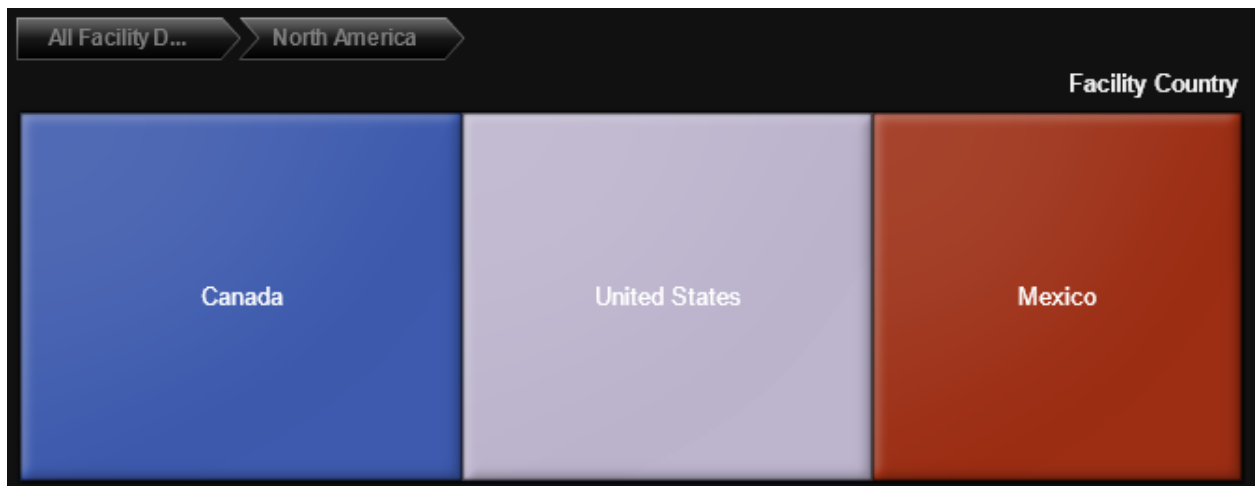
2.4 How does product cost and product sales vary across different regions?

Tree map helped to give a good insight for Facility Regions that had high and low product cost and product sales. This would help company to focus on region that has low sales and help to increase the sales. Additionally, the company can increase the production in region that gives high sales.

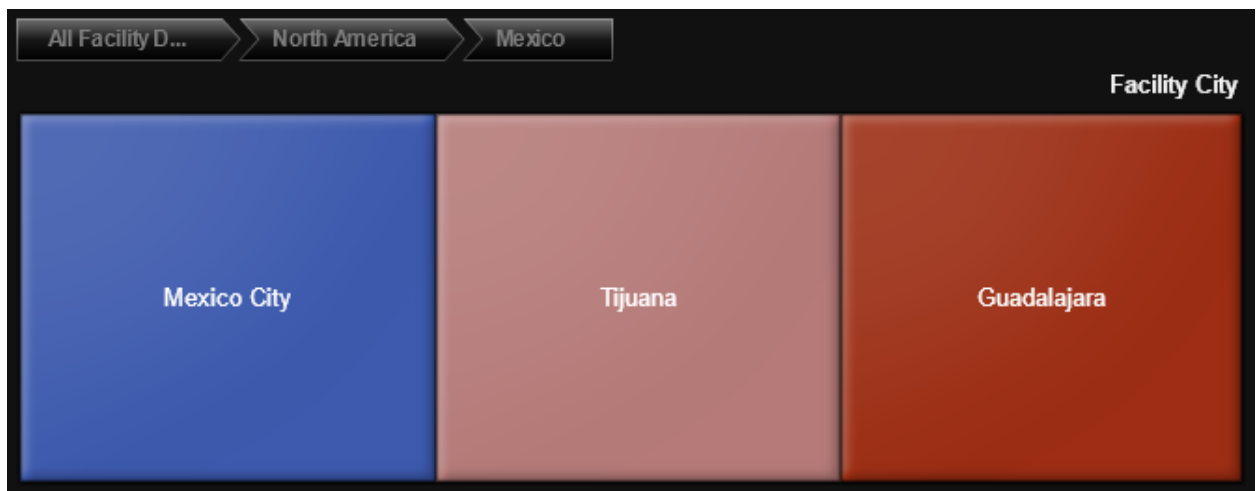
As you can view in the below tree map, **Africa (Product sale \$38 and Product cost \$31)** has the highest Product sale and Product cost, whereas **North America (Product sale \$20 and Product cost \$17)** has the lowest product cost and sales.



I drilled down the region that has lowest product and cost by double clicking on it. Now I could compare the tree map for Facility countries. I found that Canada has the highest Product cost (\$21) and Product Sales (\$18) whereas, Mexico has lowest (Product sale \$18 and Product cost \$15).



I drilled down Mexico again to view the representation based on Facility city. To my notice, Mexico City has the highest Product cost and sales whereas Guadalajara has the lowest. If you hover over each region, we can find the Product cost and Product sales for each region.



The drill down and roll up in tree map representation gives a good insight on product cost and sales of all facility regions. The minimum cost of goods sold was \$18 and product sale (\$15) for Guadalajara and highest cost is of Mexico City (product sale \$19 and product cost \$15)

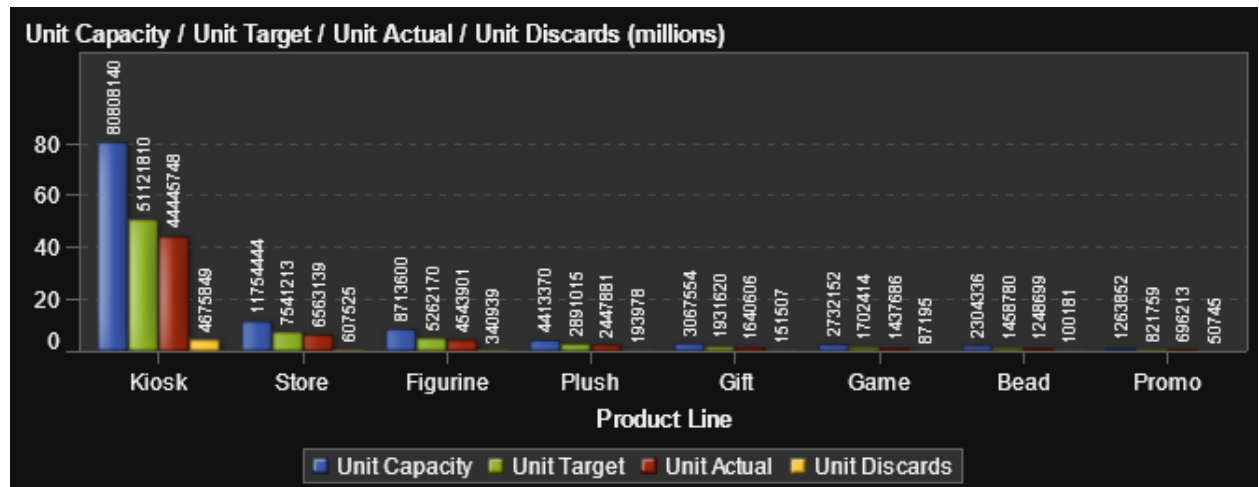
Hence, we can say that product sale is not uniform for different states in a country whereas the Product cost is same.

Finally, with the help of Tree map, we can drill down and roll up and find product cost and sale for every region. But for our analysis we can now focus on region Africa and North America.

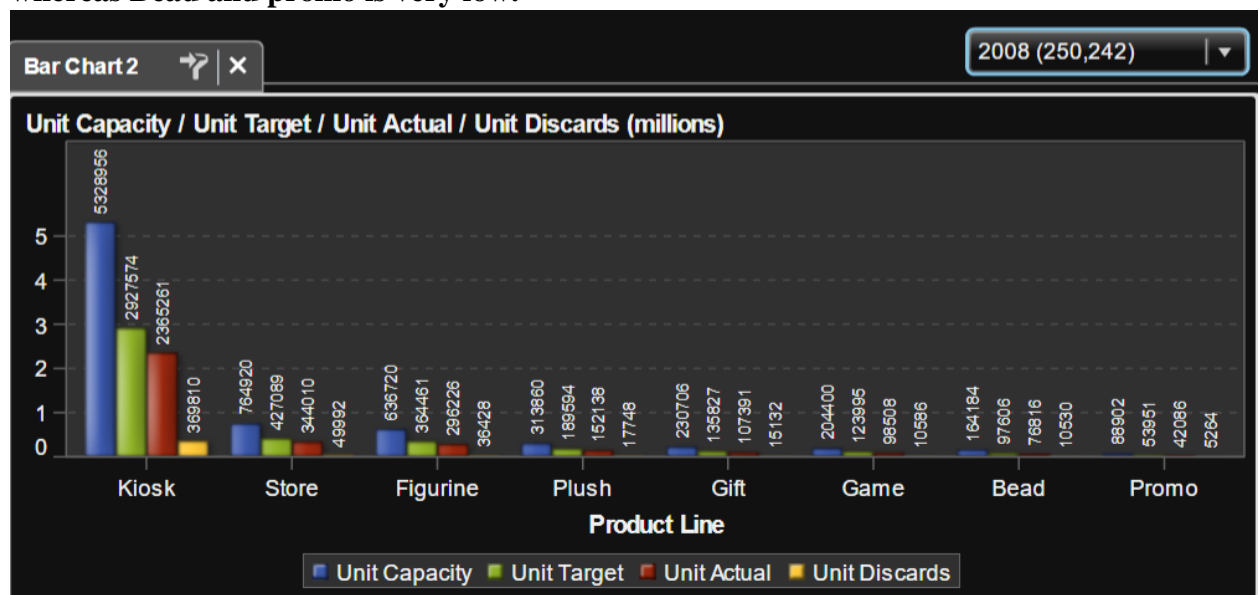
2.5 How are units affected for different product lines?

The below graph represents comparison of Unit capacity, Units target, Units Actual and Units discarded for different Product Line.

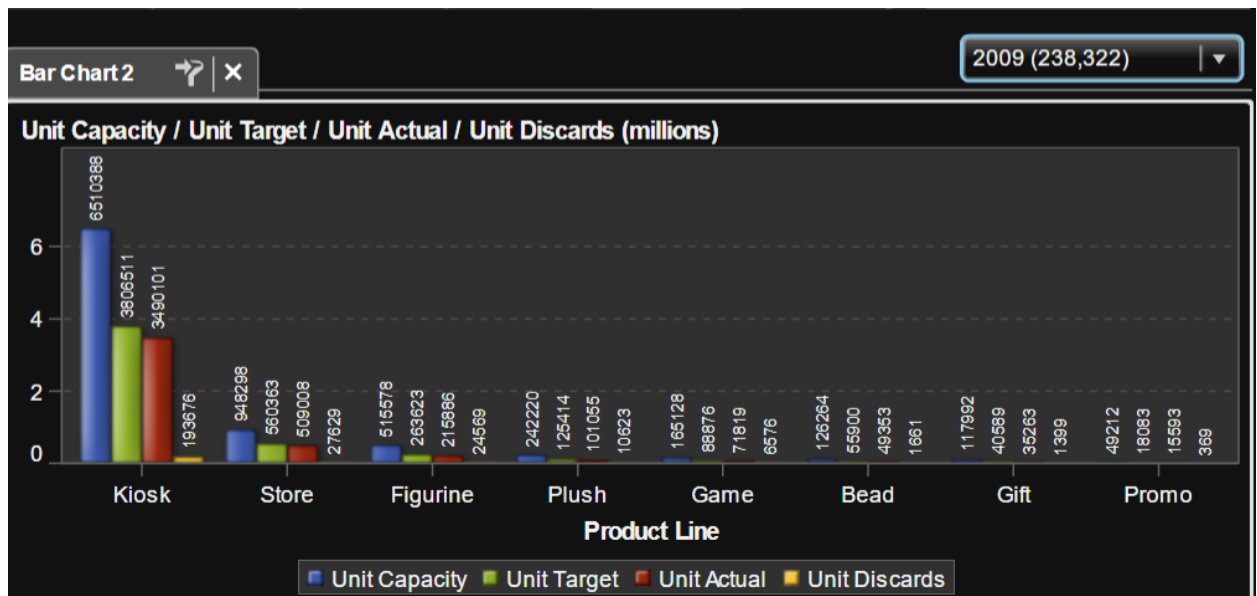
We can see that **Kiosk** has the highest Unit capacity, Units target and Units Actual. Additionally, 4675849 units were discarded for Kiosk. Games, Bead and Promo has very less units.



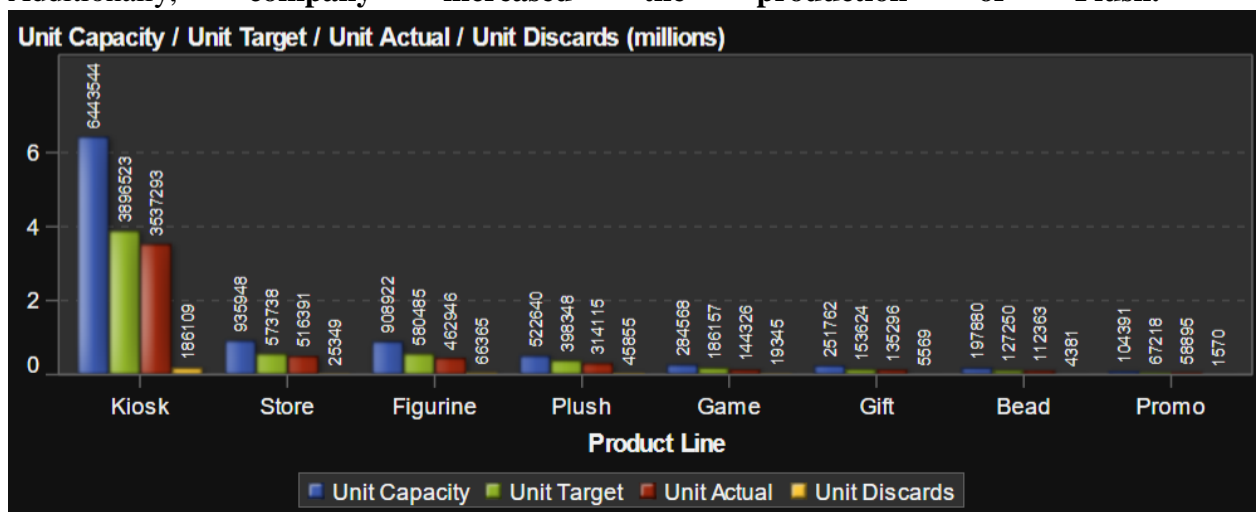
When we slice and drill down for 2008 we can see that units for Kiosk is very high whereas Bead and promo is very low.



For year 2009, capacity for Kiosk declined considerably but units targeted, and actual units were high in comparison to other product line. We can also say that **less capacity utilization is done for Kiosk**. Whereas, the utilization (i.e units target and actual in comparison to unit capacity) for other product lines is better. Even if the capacity was less utilized by company for Kiosk, **they still had discarded units**. As you can see in the graph for profits, Kiosk had \$0 profit. This could be reason one of the reasons for less profit in 2009.



Now by viewing year 2010, we could see the measures taken by the company to increase the profit from **2009 to 2010**. We can see that there was comparatively more utilization of **the capacity for Kiosk but they still had few discarded**. Additionally, more units of Game and bead were produced and less units of gifts were produced in comparison to 2010. Additionally, **company increased the production of Plush.**

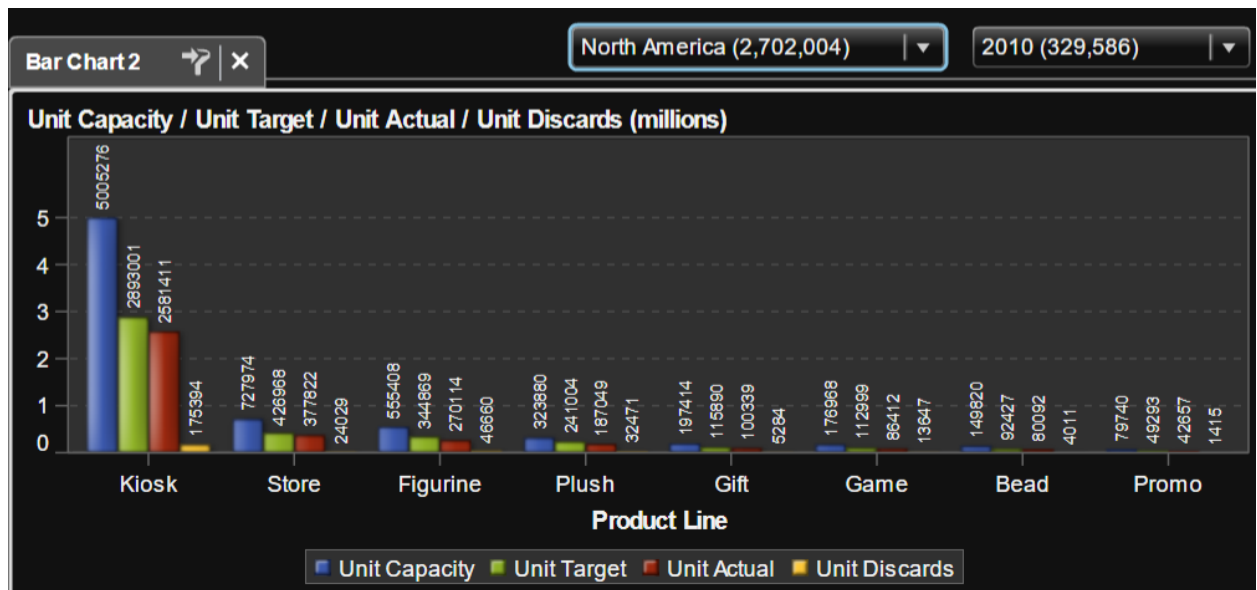


Hence, with the help of the above representation we can conclude that, **Kiosk makes low profit hence company needs to decrease the capacity and product of Kiosk.** Additionally, **Plush gives highest profit. Promo and game also give the high profit.** The company needs to increase the capacity and production of these three product lines.

After adding Facility continent drop down, I **selected Africa for year 2010**. The below results helped me to understand that **Africa focuses on only three Product line (i.e. Plush, Figurine and Game)**. These are the product line that **gives high profit**.



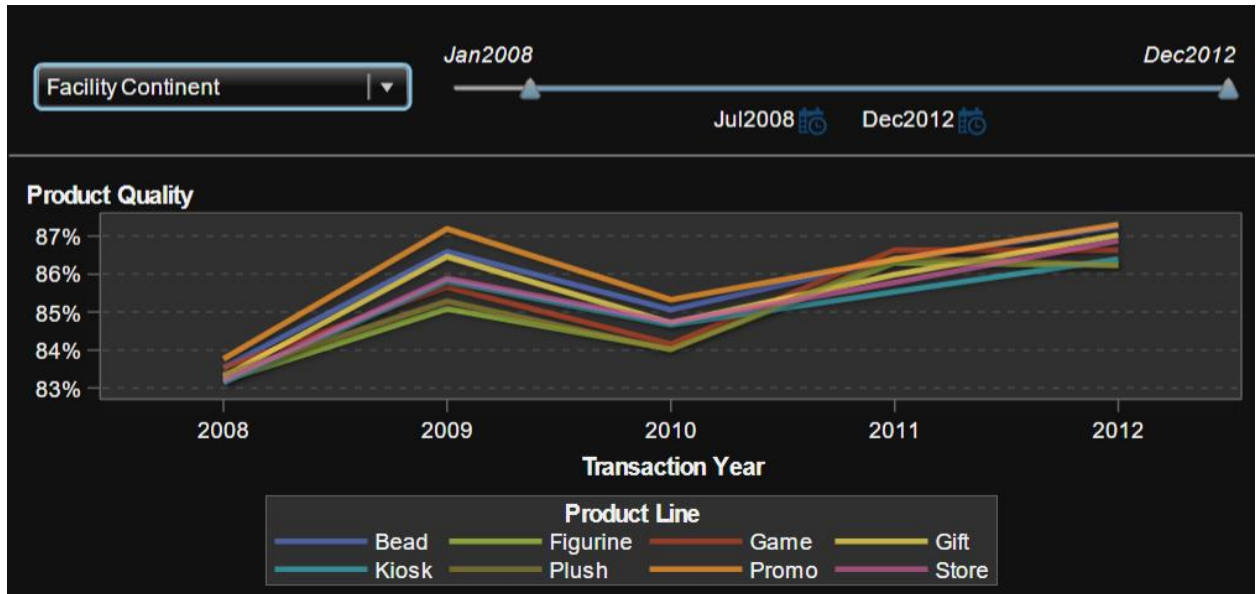
Similarly, I verified the details for **North America**, this helped me to understand that **company focuses on all product lines in North America**. The major focus is given to **Kiosk** and it generates less Profit.



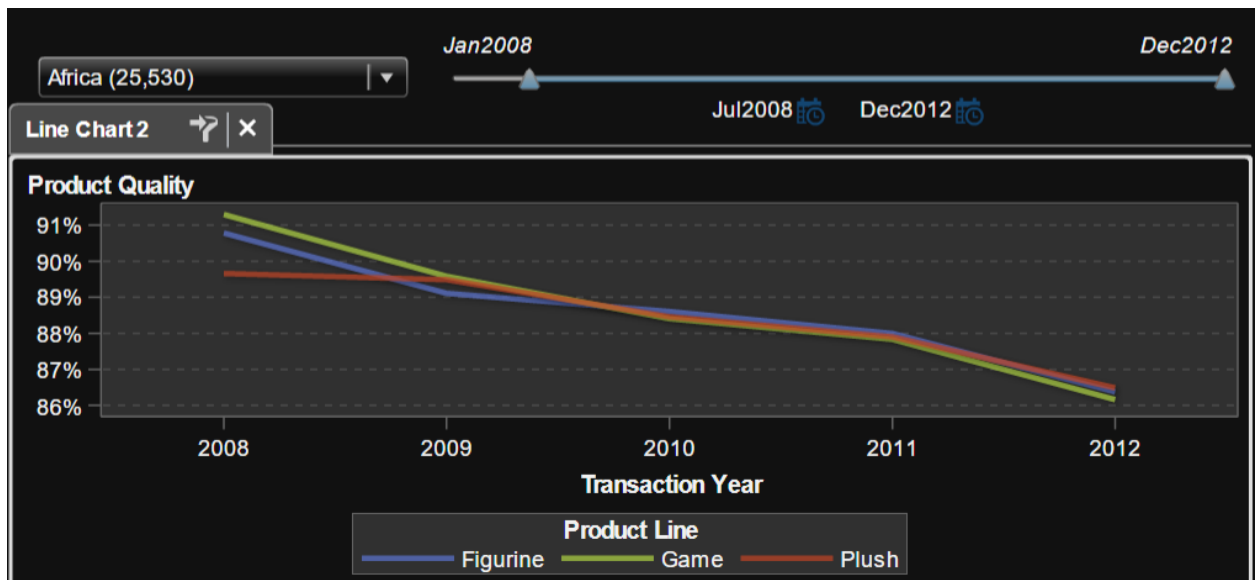
Hence, from the above analysis we can say that North America should focus on product lines that generates more profit.

2.6. How does product quality affect the product line for different years?

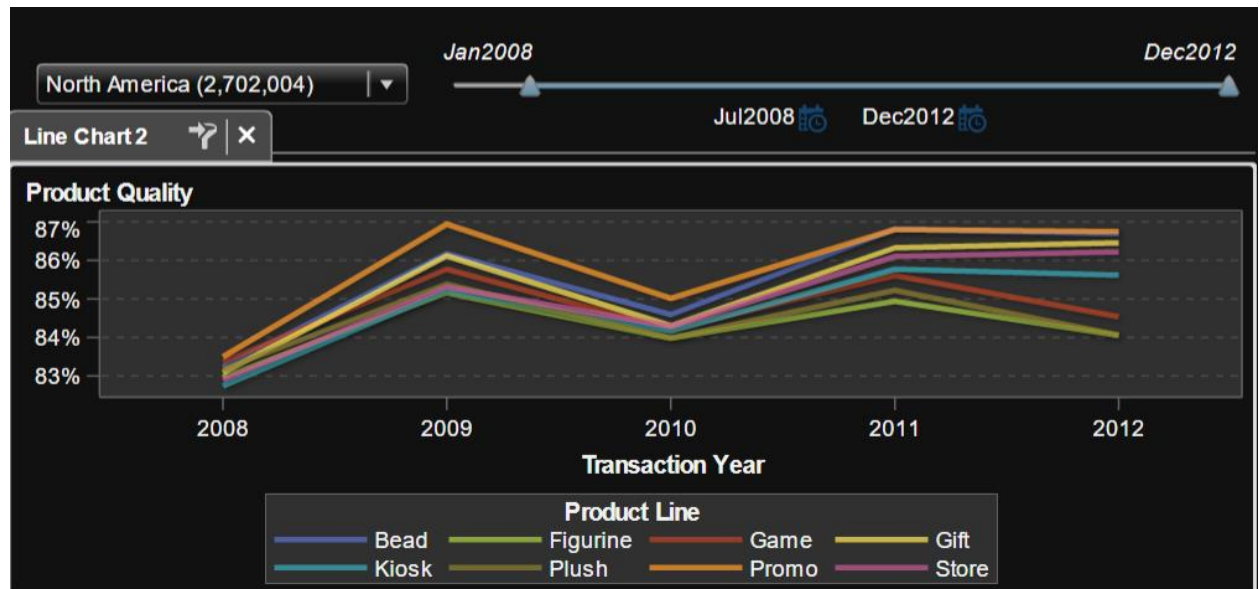
Below representation helps to understand the **trend of product quality for different products in different facility continent from 2008 to 2012**. As observed the quality of **Promo** is high in comparison to other product line whereas, the product quality for high profit generating product line (i.e. **Plush, Game and Figurine**) is low.



When I drill down, for continent **Africa**, I could see that the **Product quality over the years 2008 to 2012 is decreasing**.



Similarly, when we drill down for North America continent, we could see that **Quality for promo is high** where as **quality for high profit generating Product line is low**. Additionally, we can see that Profit was less in 2009 but the quality for 2009 was highest.



We can recommend that the quality needs to be improved for high profit products in Africa and North America.

Recommendation:

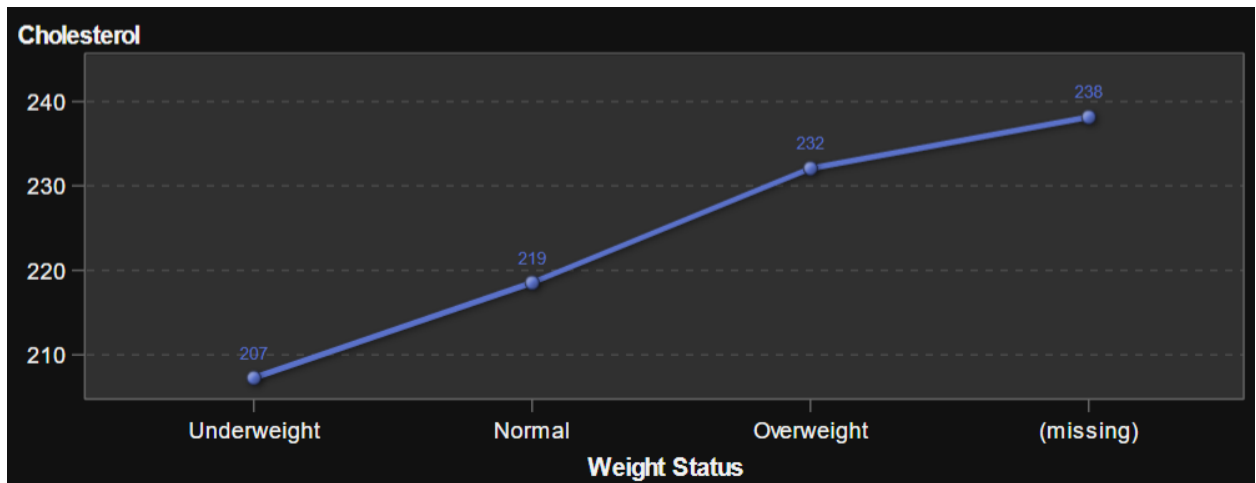
Finally, from the above analysis we can recommend the following:

1. As observed, **North America has the lowest cost of product and cost of sales whereas Africa has the highest**. The cost of product is not uniform for all region in North America. Additionally, customer satisfaction is highest in Africa and comparatively low in North America and Oceania. **Company needs to focus on needs to improve customer satisfaction in these regions and increase more customers satisfaction in all regions that would help to acquire and retain more customers.**
2. **The average profit is \$4**, whereas profit is less in 2009. The company tried to implement measures to increase their profit, but the measures resulted in average profit. The company can focus on the below reason to increase their profit.
 - a. Company should focus on **high profit generating products such as Plush, Figurine, Game and Promo**. Company should **increase the capacity and utilization for the units for high profit generating products and decrease the capacity and production for low profit generating profit (eg. Kiosk).**
 - b. **The quality is decreasing in North America and Africa for high profit generating product. The company should focus on increasing the quality of high profit generating products.**

Using Data Visualization for Problem Solving

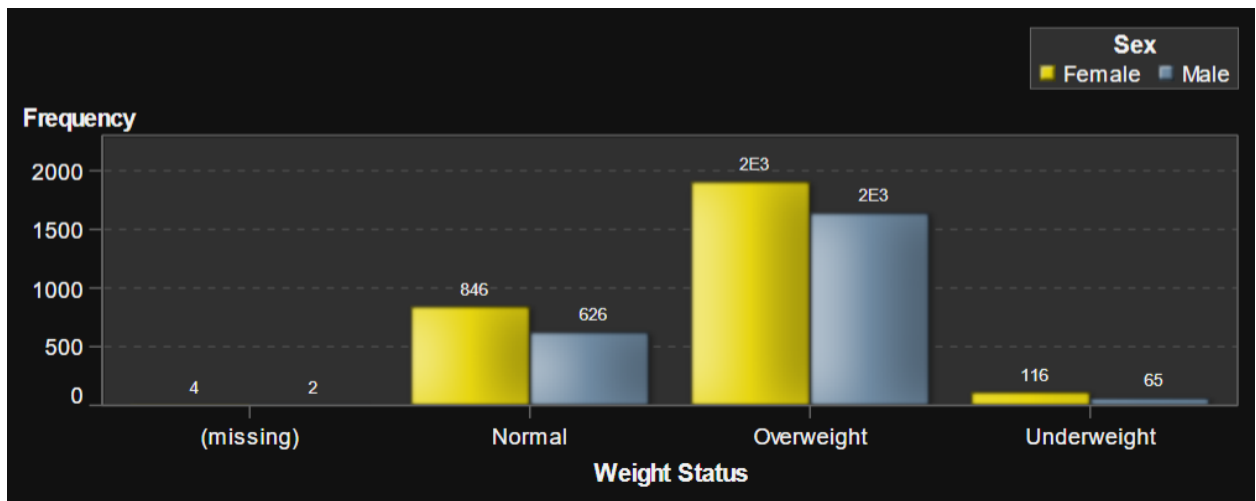
- **H1: The weight and cholesterol levels are closely related.**

The below line graph reflects that the cholesterol increases with increase in weight. There is positive correlation between cholesterol and weight status. Hence, we can say that the above hypothesis is true.



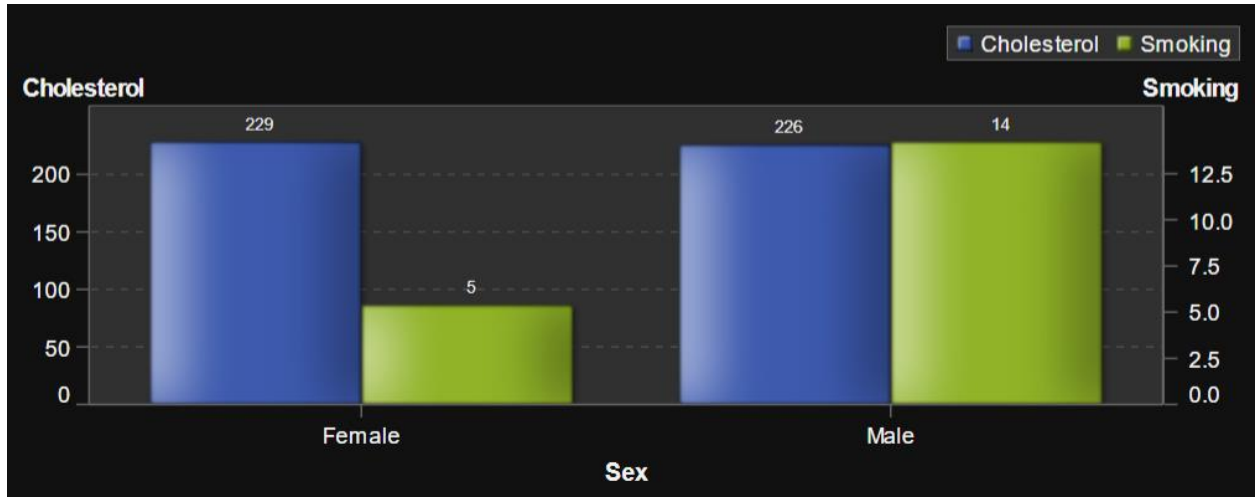
- **H2: Men are usually more obese than women.**

The below graph reflects the comparison between weights men and women. As observed in the below graph, number of women is more in each of the weight status than men. If we carefully observe the overweight section, the number of women is more than men. With the help of below graph, we can say that women are more obese than men. Hence the proposed hypothesis, men are more women than men is false.



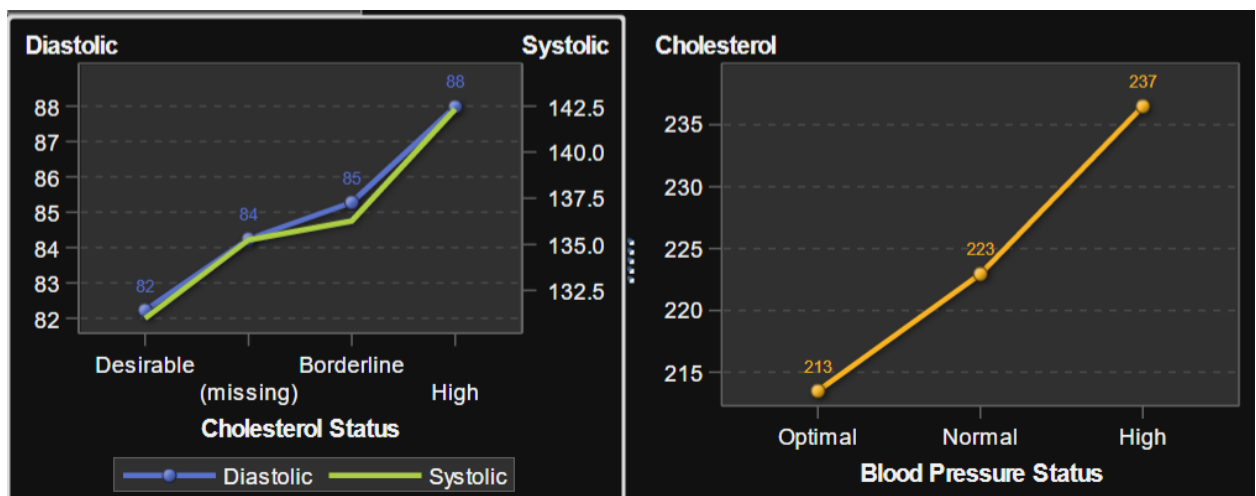
- **H3: Women usually smoke less than men, but their cholesterol level is higher.**

The below result reflects that men smoke more than women, whereas women have high cholesterol than men. Hence, we can say that hypothesis that ‘Women usually smoke less than men, but their cholesterol level is higher’ is true.



- **H4: The blood pressure is higher for people with higher cholesterol level.**

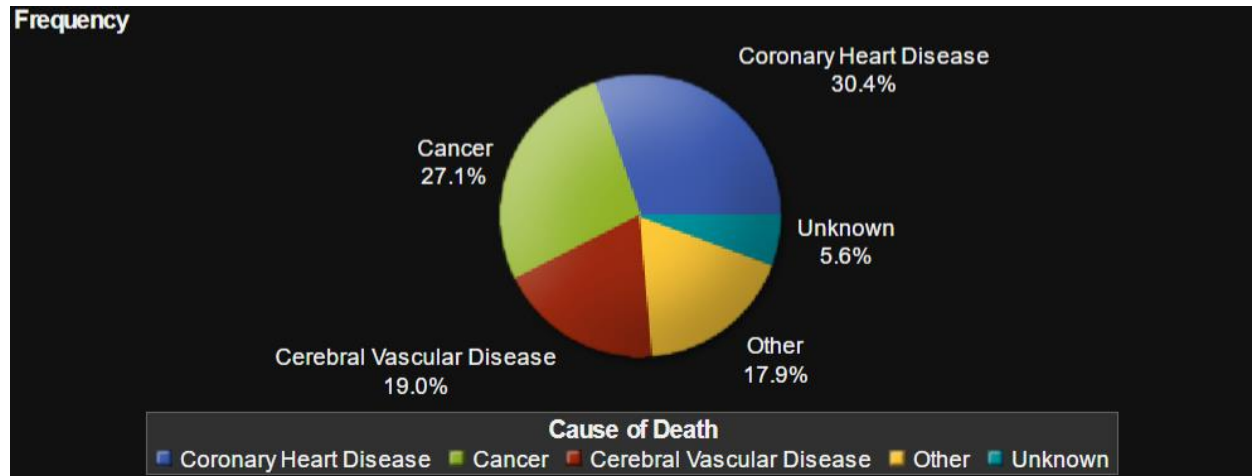
The below graph reflects association that both systolic and diastolic readings are high for patient with high cholesterol level. This reflects a positive correlation between cholesterol status and systolic reading as well as cholesterol status and diastolic readings. Additionally, I reverified using the blood pressure status and cholesterol measure. Same results were reflected. i.e. high blood pressure status and high cholesterol value. There was a positive correlation between two attributes. Hence, we can say that the hypothesis ‘Blood pressure is higher for people with higher cholesterol level’ is true.



Observation of Coronary Heart Disease

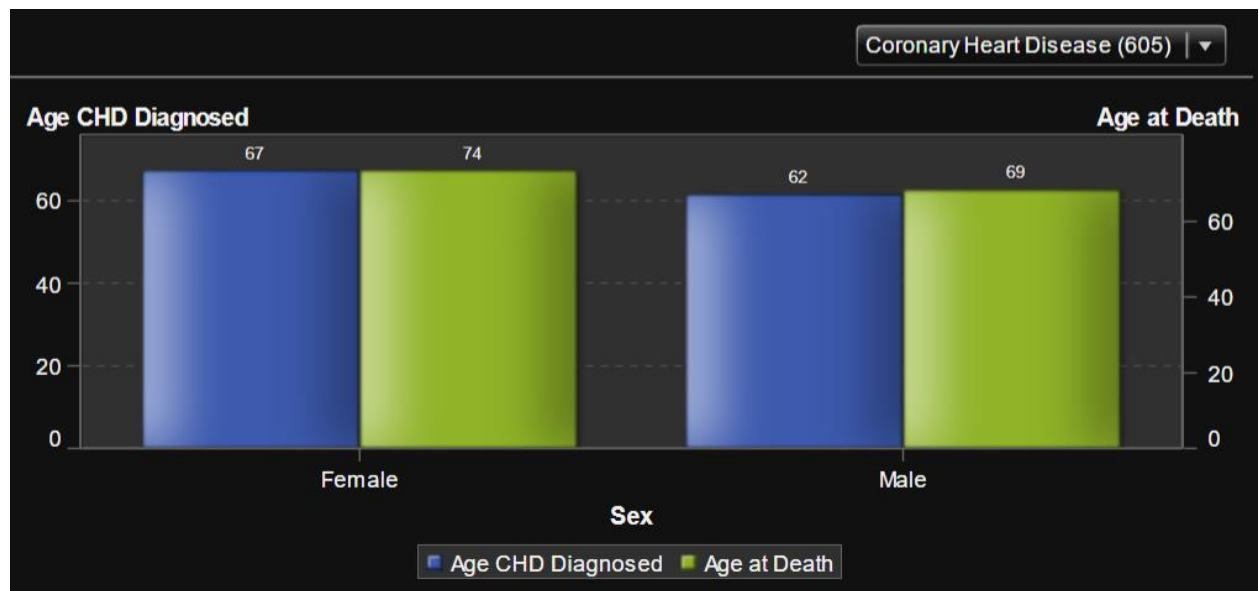
2.1 Which cause results in highest deaths?

The below result suggests that death due to Coronary heart disease is maximum. Approximately, 30% of people died due to Coronary Heart Disease



2.2. What is the average age of diagnosis and death with Coronary heart Disease?

The below result suggests that Age at which Coronary heart disease is diagnosed and age of death is higher in Female than Male. We can say that Men are diagnoses at a early stage then female. Additionally, men that has coronary heath disease die at early age than female.



The below pie chart represents that people with coronary heart disease are more with high cholesterol, have heavy smoking rate and are overweight. Coronary disease is less in people who are under weight or having light smoking rate.

2.2 What is effect of cholesterol, smoking and weight on coronary heart disease?



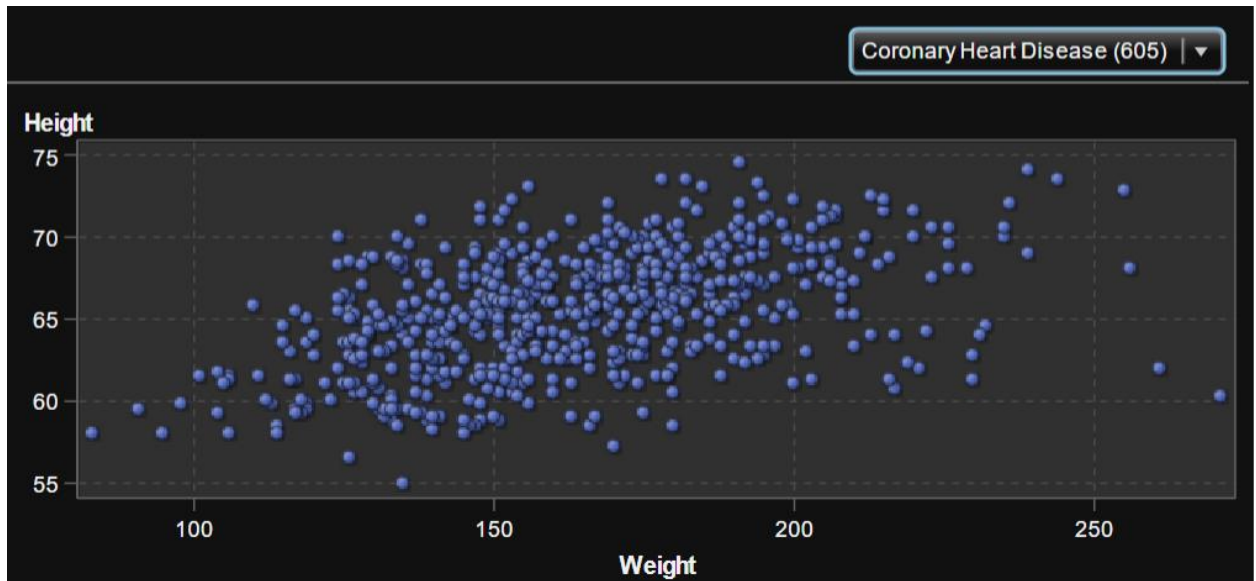
2.4 What is the average cholesterol level and smoking rate for different weight status?

The below graph suggest that overweight people has less smoking habit but a high cholesterol whereas, underweight people has high smoking habit. Additionally, the increase in smoking habit in underweight could lead to death due to coronary heart disease.



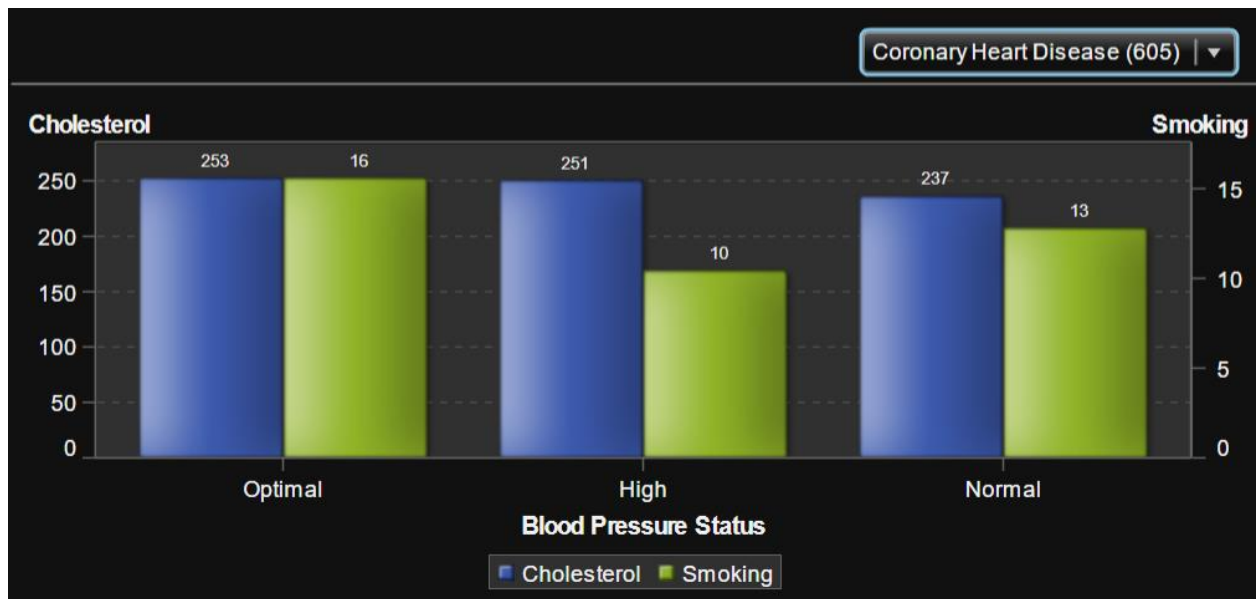
2.5. What is range of weight and height for people that died due to Coronary heart disease?

The below graph suggests that patient that died due to coronary heart disease have height between 55 to 75 and weight between 130 to 230 pounds.



2.6 How does cholesterol level and smoking level varies for different blood pressure status?

The below graph suggest that cholesterol level is very high for people with Optimal and high blood pressure whereas the smoking rate is high for Optimal blood pressure status. Hence, we can say that people with optimal blood pressure are more prone to deaths with Coronary heart disease.



Assertion regarding the potential underlying causes of illness are:

1. Overweight, high smoking and high cholesterol are potential factor that results in coronary heart disease death. Hence, it is important to control weight and cholesterol level.
2. The increase in smoking habit in underweight would make them prone to Coronary Heart disease.
3. People with height between 55 to 75 and weight between 130 to 230 pounds are more prone to Coronary heart disease.
4. People with Optimal blood pressure has high smoking and cholesterol. Hence, they are more prone to Coronary Heart Disease.
5. Men are more prone to Coronary heart disease than women. The average age when the Coronary heart disease is diagnosed is 62 years and age of death is 69 years.

3. What did you learn?

Assignment 3 and 4 gave me a good exposure on SAS Visual Analytics tool. I learned the different ways to create reports, sections and charts.

SAS visual analytics has some good business-related features and data visualization in comparison to Tableau and MicroStrategy. The option such as dual axis bar, dual axis line or dual axis bar line are directly available. Reports are systematically arranged in Properties drop down. We can select the Tiled or Precision layout. Same as MicroStrategy, the reports can be exported in pdf format. Navigation is easy in SAS visual analytics compared to Tableau and MicroStrategy. Additionally, you can change the style of Data skin to Sheen, Matte, etc. This gives a good visual representation.

Unlike Tableau or MicroStrategy, there is no back button or undo button directly available in menu options. Ctrl Z option does not work in SAS Visual Analytics. The report generation of SAS is slow in comparison to Tableau/ MicroStrategy. Whenever you make any changes you need to manually click on refresh button whereas in Tableau or MicroStrategy, reports are auto refreshed and updated.

Tableau provides with dashboard and storyline facility. This is not available in SAS Analytics. Instead we can add charts to Sections. Additionally, there is more variation in colors, shapes, data labels in Tableau. SAS visual analytics has limited variation. I really liked the hierarchy option available in SAS Visual Analytics, but Tableau provides a good variety of geographical, animation and time plots.

Filter option is provided in the user interface of Tableau and micro strategy whereas SAS visual analytics we need to drag controls over to sheets.

As SAS is on a cloud platform. Unlike tableau, I was not able to export the SAS file. The email option in SAS Visual Analytics, emails the link to the report rather than adding pdf report to the email.

In Assignment 3, I learned good insights can be represent in simplest manner by using line graph or pie chart. The business case was an open end. So, it gave me a good exposure to make my own business question to gain insights. I realized it is very important to understand the parameters we would need to gain those insights. Slice, dice, drill down and roll up can be easily represented in SAS Visual Analytics with the help of dual axis graphs, drop down, hierarchy in tree map etc. By perform these operations, I understood to view data in different dimensions and gain relevant insights from them. This insight helped to focus on strength and weakness of the company. Additionally, I could give good recommendation to Insight Toy Corporation for enhancing their profits.

In Assignment 4, the business case had some hypothesis. I learned that we can reject or accept hypothesis by visualization. We do not need statistical analysis such as calculation z-score or t value. I learned to understand correlation with other data. Additionally, I write business question and draw conclusions based on business questions and hypothesis. For example, I could say that despite higher cholesterol level in women, the number of male Colorany Heart Disease victims is more.

References:

http://www.sas.com/en_us/software/business-intelligence/visual-analytics/interactive-demo/full-access.html

<https://www.teradatauniversitynetwork.com/getmedia/430f613d-1f0f-4503-8d58-86c1ac386784/SAS-Visual-Analytics-User-Guide>

