U.S. Organization Capstone Project Business Intelligence Solution

## 

Ben Hebbel

Colorado State University Global

MIS-480: Capstone-Business Analytics and Information Systems

Professor: Justin Bateh

10/9/22

U.S. Organization Capstone Project Business Intelligence Solution

The data analysis in a well planned BI solution gives valuable insights for an organization to identify opportunities, stay competitive in the market and increase customer experience. This paper will cover the rationale for choosing Kay Jewelers, the problem or theory, and the data set that will be analyzed. It will also show and explain the data analysis and the results from the analysis, and lastly, it will explain the benefits of a business intelligence solution.

**Kay Jewelers**

Kay Jewelers was founded in 1916 in Reading, Pennsylvania, when customers could buy eyeglasses, razors, and appliances among the jewelry (Kay, 2022). Kay Jewelers believes that everyone deserves to receive and give the gift of love. They are devoted to providing beautiful jewelry for all (Kay, 2022). Kay Jewelers now has more than one thousand stores across the United States and is part of Signet Jewelers.

I chose this company because it works well with the diamond data set. When selecting a data set to work with, I considered what U.S. organization would coincide with the data. I have briefly used the data set before to perform visualizations. The business solution will help Kay Jewelers pick quality diamonds at a lower cost.

**Diamonds Theory**

Kay Jewelers, as the name suggests, is a jewelry company. Their mission is to be able to give everyone the gift of love. But diamonds can be expensive, and the theory is that by analyzing the diamond data set, Kay Jewelers will be able to find quality diamonds for less. There are several ways that will determine the price of the diamond. The cut, color, clarity, and carat are just a few ways to determine the price of the diamond. These factors will be used to save Kay Jewelers money.

The cut of a diamond is used to determine the price of the diamond based on seven aspects: brightness, fire, scintillation, weight ratio, durability, polish, and symmetry (FordgeJeweleryWorks, 2022). The first three aspects deal with the appearance of the diamond. The industry standard grading system used is the GIA international grading system (FordgeJeweleryWorks, 2022). The scale used to grade the cut of a diamond goes from excellent, being the best, to poor, being the worst.

The carat of a diamond is more straightforward than the cut of a diamond. For example, the American dollar is broken down into one hundred pennies; the carat of a diamond is broken down into a point system. For instance, if a diamond is thirty points, then the carat of the diamond is 0.30 carats (FordgeJeweleryWorks, 2022). The color of the diamond uses the GIA grading system. The scale gives a letter grade to the diamond, with D being the best and Z being the worst. The diamond is also compared to a diamond of similar color.

These aspects will be used in the analysis to help Kay Jewelers determine the best quality diamonds for a low price. Being able to find good quality diamonds will help ensure that Kay Jewelers fulfills its mission. The results from the analysis should shed light on the subject and allow Kay Jewelers to pick and choose diamonds that fit their high standards.

**Data Set**

The data set being used in the BI solution is called diamond. This data set contains over fifty thousand data points, covering all the aspects used to determine the price of the diamond. The data set also includes the various measurements of the diamond. These aspects include depth, table, x, y, and z. The x, y, and z values are measured in millimeters. The data set also contains the price of each diamond, which will be used to help display how the cut, clarity, carat, and color determine the price of the diamond. The data will be accessed in real-time via a stored CSV file because it is easier to read into Python or R-studio. The CSV file will be stored in a class folder on OneDrive. OneDrive is Microsoft's personal cloud service that connects you to all your files (Microsoft, 2022). Since the data is stored in the cloud, it provides easy access to the data set while still being secure. Since the file is in a class folder, it allows me to avoid setting up a working directory to read the data set into the BI tool environment.

**Data Analysis**

## Reading in the Data

This block of code shows the data set being read into the R-studio environment. The data set is put into a data frame called diamond. The head() function is used to display the first five values in the diamonds data set. The library ggplot2 is also loaded into the environment to help create the data visualizations.

library(ggplot2)  
diamond <- read.csv('diamonds.csv')  
head(diamond, 5)

## X carat cut color clarity depth table price x y z  
## 1 1 0.23 Ideal E SI2 61.5 55 326 3.95 3.98 2.43  
## 2 2 0.21 Premium E SI1 59.8 61 326 3.89 3.84 2.31  
## 3 3 0.23 Good E VS1 56.9 65 327 4.05 4.07 2.31  
## 4 4 0.29 Premium I VS2 62.4 58 334 4.20 4.23 2.63  
## 5 5 0.31 Good J SI2 63.3 58 335 4.34 4.35 2.75

## Summary of Data

This block of code calculates the summary statistics of the data set. The summary() function is used to complete the calculation. This function is used to get a better understanding of the data set. It gives a five number summary showing the mean, median, and maximum values (PluralSight, 2019). It calculates the five number summary for numerical values while it gives the class and mode of non-numerical values.

summary(diamond)

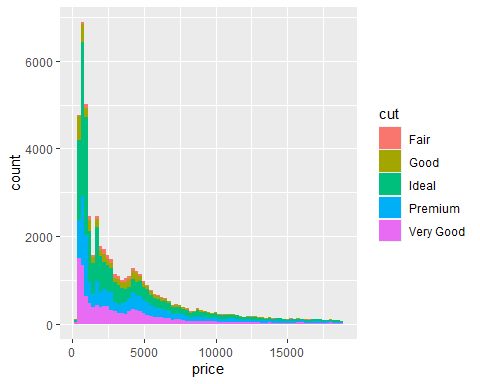
## X carat cut color   
## Min. : 1 Min. :0.2000 Length:53940 Length:53940   
## 1st Qu.:13486 1st Qu.:0.4000 Class :character Class :character   
## Median :26971 Median :0.7000 Mode :character Mode :character   
## Mean :26971 Mean :0.7979   
## 3rd Qu.:40455 3rd Qu.:1.0400   
## Max. :53940 Max. :5.0100   
## clarity depth table price   
## Length:53940 Min. :43.00 Min. :43.00 Min. : 326   
## Class :character 1st Qu.:61.00 1st Qu.:56.00 1st Qu.: 950   
## Mode :character Median :61.80 Median :57.00 Median : 2401   
## Mean :61.75 Mean :57.46 Mean : 3933   
## 3rd Qu.:62.50 3rd Qu.:59.00 3rd Qu.: 5324   
## Max. :79.00 Max. :95.00 Max. :18823   
## x y z   
## Min. : 0.000 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 4.710 1st Qu.: 4.720 1st Qu.: 2.910   
## Median : 5.700 Median : 5.710 Median : 3.530   
## Mean : 5.731 Mean : 5.735 Mean : 3.539   
## 3rd Qu.: 6.540 3rd Qu.: 6.540 3rd Qu.: 4.040   
## Max. :10.740 Max. :58.900 Max. :31.800

## 

## Price Distribution

The code block below shows the distribution of the variable price. It also shows the cut of the diamond. As you can see, most of the diamonds are distributed on the left-hand side of the histogram. The price of a diamond can range from under $1,000 to approximately $20,000. It does show that there are diamonds that do have a "very good" cut rating under $5,000. This graph shows that Kay Jewelers can get diamonds that have a good cut rating for a low price.

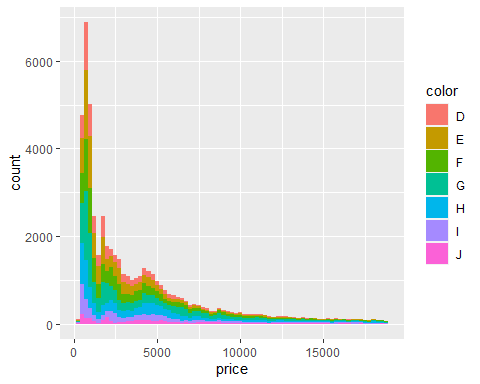
ggplot(diamond, aes(x = price, fill = cut))+geom\_histogram(binwidth = 250)



## Color Distribution

This block of code shows a histogram comparing the price to the color of the diamonds in the data set. The color map shows that only a few diamonds have a color grade of J. It also shows that most of the diamonds have a color grade of D, which is the best, and are under $5,000. This indicates that Kay Jewelers can purchase diamonds for a low price that have the best color grade.

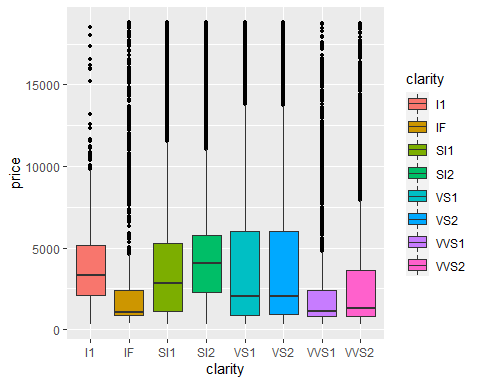
ggplot(diamond, aes(x = price, fill = color))+geom\_histogram(binwidth = 250)



## Box Plot

This block of code shows a box plot comparing the clarity of a diamond to the price. The display shows that there are many outliers in the data. The box plot shows the median, upper and lower quartile, and whiskers; anything outside the box is considered extreme (DataCamp, 2020). The outliers in the plot show that each clarity category has expensive diamonds, but all the boxes have a median below $5,000.

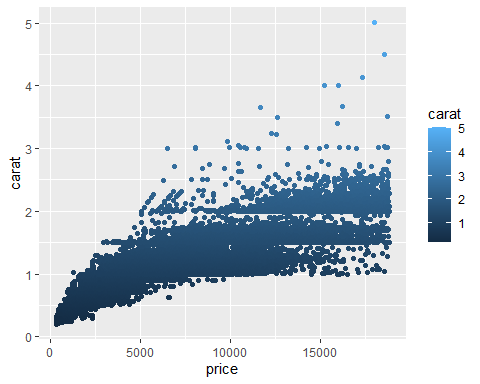
ggplot(diamond, aes(x=clarity, y=price, fill = clarity)) + geom\_boxplot(outlier.colour = 'black', outlier.shape = 16, outlier.size = 1)



## Scatter Plot

The scatter plot compares the diamonds' price and carat in the data set. This graph shows that the carat highly affects the price of a diamond. The majority of diamonds in the data set are at or below three carats. The graph shows that diamonds that are one carat and above have a price of around $3,000 to approximately $20,000. Diamonds that are above three carats are going to be more expensive.

ggplot(diamond) + geom\_point(mapping = aes(x = price, y = carat, color = carat))



**GitHub Screenshot**

**Benefits of a BI Solution**

Business intelligence solutions can and will provide insights into the organization. It gives leaders a holistic view of their operations and benchmark results over the larger organization (Tableau, 2022). With a holistic perspective, leaders can identify opportunities. The next benefit that business intelligence provides is data-driven decisions. Accurate data and faster reporting capabilities provide an organization with better business decisions (Tableau, 2022). Lastly, it can give an organization a competitive advantage. Businesses can be more competitive when they understand the market and how they perform in it (Tableau, 2022). Knowing where the market is and how you perform enables companies to adjust accordingly, increasing their performance as a business and in the market.

**Conclusion**

A well planned business intelligence solution can give an organization valuable insight to identify areas of opportunity, stay competitive in the market, and increase customer experience and satisfaction. The theory that was analyzed for Kay Jewelers was, can they purchase quality diamonds for less? The analysis answered that theory, and yes, Kay Jewelers can purchase quality diamonds for less if the diamond has a carat of three or less. The carat of a diamond highly affects the price of the diamond. Kay Jewelers can purchase quality diamonds for less and still accomplish its mission to give everyone the gift of love. This shows how important it is to utilize BI tools to help businesses grow in their industry. Using the diamond data set, Kay Jewelers could lower their costs, giving them a competitive advantage in the market.

**References**

Kay Jewelers. (2022). *Why Kay?*.

<https://www.kay.com/why-kay?icid=FOOTER:WHYKAY>

FordgeJeweleryWorks. (2022). *How Do I Determine A Diamonds Value?*.

<https://www.forgejewelryworks.com/determine-diamonds-value/>

Data Camp. (2020). *Box Plot in R Tutorial*.

<https://www.datacamp.com/tutorial/boxplot-in-r>

Codecademy. (2021). *What is R Used For?*.

<https://www.codecademy.com/resources/blog/what-is-r-used-for/>

GeeksforGeeks. (2022). Introduction to R Markdown.

<https://www.geeksforgeeks.org/introduction-to-r-markdown/>

PluralSight. (2019). *Interpreting Data Using Descriptive Statistics With R*.

<https://www.pluralsight.com/guides/interpreting-data-using-descriptive-statistics-r>

Tableau. (2022). The top 7 benefits of business intelligence (BI).

<https://www.tableau.com/learn/articles/business-intelligence/enterprise-business-intelligence/benefits>