Cole Hoffman

BIOS 7719

Data Visualization: Tableau Report

Part 1: Developing Answers to Analytical Research Questions

Chosen dataset: whiskeys.xlsx

Research questions

1. The manager of the local whiskey distillery is looking to showcase a couple new whiskey products to their customers. The manager would like to choose whiskeys that have high approval ratings (>85) but are not too expensive, preferably less than $50. Which brands would you recommend to the manager? And are whiskey ratings and price correlated?

2. The same local distillery wants to showcase whiskeys from around the world at their annual holiday party. The manager would like to offer a diverse selection of whiskeys based on country of origin and category (bourbon, rye, single malt, etc.). The price and customer rating of each whiskey are not of interest since the manager just wants to conduct preliminary research. The manager is wondering whether certain countries are more likely to produce specific categories of whiskeys, or whether the categories are evenly distributed by geographical location.

3. As business continues to grow, the manager finds themselves being too busy to oversee the operations at the local distillery and hires a new assistant manager to take over. The new assistant manager is curious as to whether barrel-aged whiskeys lead to changes in ABV, as well as whether aged whiskeys have higher customer ratings. The assistant manager requires the help of an analyst to investigate this and report back on the results.

Question 1: Response

This is an important question since local restaurants and bars often sell a variety of products, even some that are not their own, to appeal to a broader customer base. In this scenario, the manager of the local distillery may want to maximize profits without having to expand the overhead and production costs for their own business, thus relying on a distributor to provide a larger supply of whiskeys to the distillery. Before answering the

manager's specific question, I generated a scatterplot to investigate whether customer rating and price of each whiskey are positively correlated (Figure 1). This plot contains all whiskeys valued under $500. Three whiskeys were removed when imposing this filter because those whiskeys valued higher were massive outliers (one costed $3000) that distorted the overall relationship. Removing these observations was also reasonable since the objective is to provide the manager with a general overview of the trend between customer rating and price, especially for whiskeys valued at $50 or less. This visualization also lets the manager know that there are plenty of whiskey products worth less than $50 that exceed a customer rating of 85. This resulted in 60 unique brands and 66 distinct whiskeys that satisfy these metrics. For an even simpler visual representation, I decided to keep the $50 maximum price in place, but restrict the customer ratings from 90-100. The results include 31 unique whiskey products across 28 brands (Figure 2).

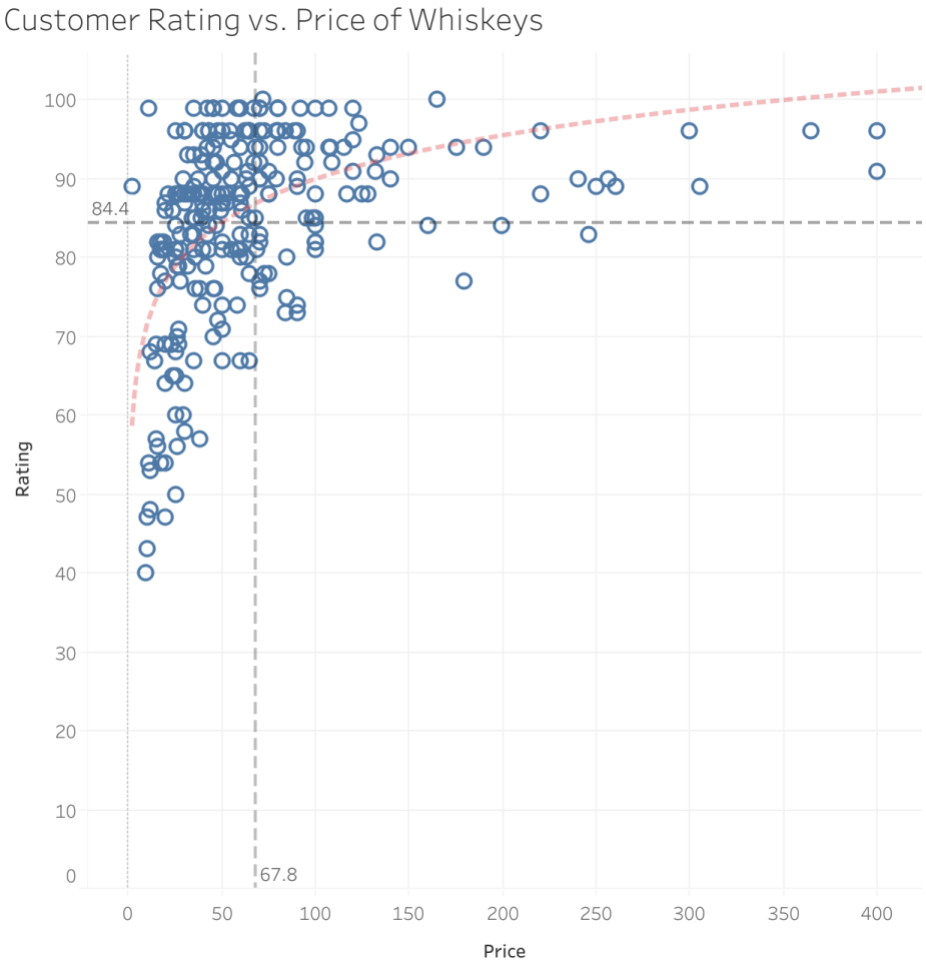

Figure 1: Scatterplot showing customer rating vs. price of whiskeys. Average customer rating of 84.4 and average price of $67.80 displayed with black dashed lines, and the red trend line details the logarithmic association between customer racing and price.

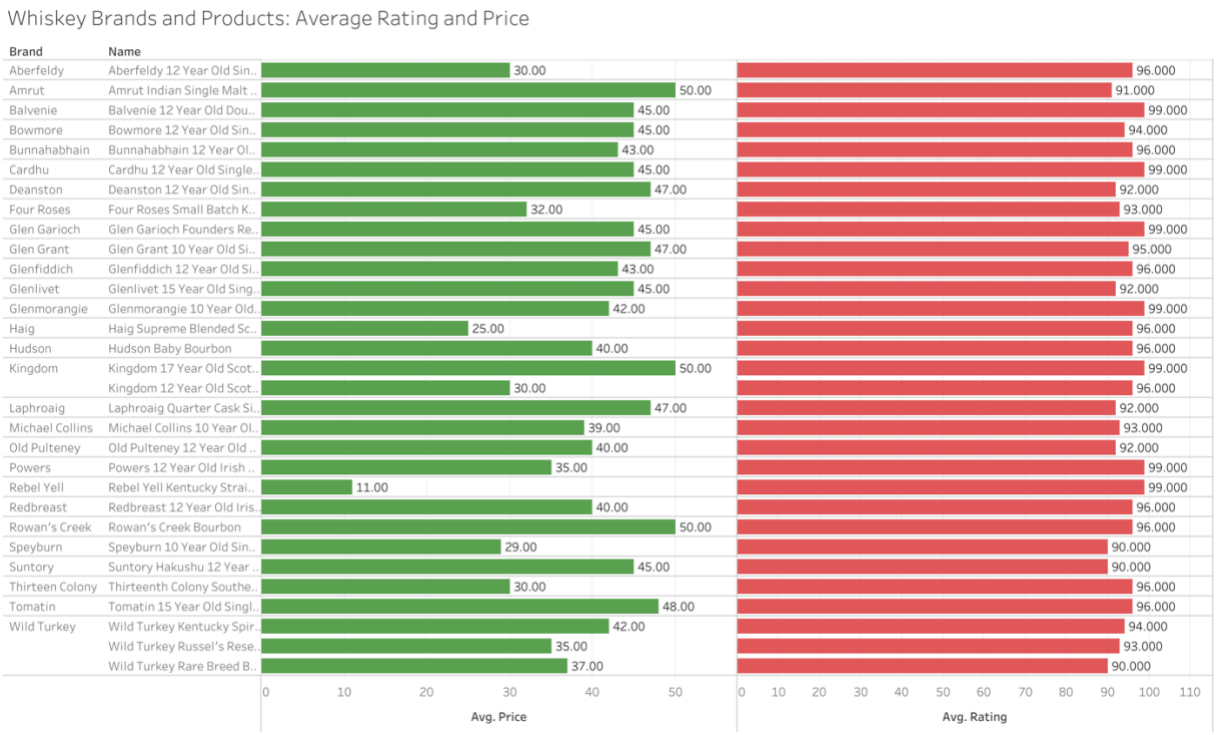

Figure 2: Grouped bar chart that displays the average rating and price of whiskey names and brands for all whiskey products valued $50 or less with a customer rating of 90-100.

I would immediately recommend that the manager consider selling the Rebel Yell whiskey since it has a customer rating of 99 and only costs $11. The Haig Supreme Blended Scotch is also a viable option with a 96 customer rating and $25 cost. The Whyte & Mackay branded whiskey could be an excellent option with a customer rating of 89 and only $2; of note, this product is not pictured because it falls just below the lower threshold of the 90-100 customer-rated whiskeys. Perhaps the manager would be interested in buying several products from the same whiskey brand; if so, the Wild Turkey brand sells three whiskeys with customer ratings of 90-94 ranging from $35-$42 each. The Kingdom brand sells two whiskeys with customer ratings of 96 and 99, costing $30 and $50, respectively. Although these products cost a little more, the brand may be willing to sell larger quantities of their product to the manager if they decide to buy in bulk. This decision may also forge a mutually beneficial business partnership between the local distillery and the brand while increasing profits for the local distillery.

Question 2: Response

This question is exploratory in nature where the number of unique whiskey products by type across countries can be analyzed via the current whiskey dataset. Since country of origin and type are categorical variables, Tableau intuitively constructs a table that provides the number of whiskey products for each type broken down by country. I considered using a Dorling cartogram which would provide a spatial representation of the number of whiskeys made in each country; however, only a few nations are major whiskey producers, and most of them are located far apart on the map, leading to a sparse spatial data visualization. Furthermore, Ireland and Scotland are both major whiskey producers, and yet are small enough countries to where the data points obscure the entire nation on the map. This makes it impossible for the user to visually comprehend the data without changing the map's scale. Consequently, used a color-coed summary table to display the number of whiskey products broken down by type and country of origin (Figure 3). Lighter shades of blue indicate fewer whiskeys, whereas darker shades suggest a higher number.

Whiskey Counts by Country and Type

Country	Type	Count of Whiskey
Canada	Blended	25
	Rye	3
	Single Malt	2
England	Single Malt	1
France	Blended	1
	Single Malt	1
India	Single Malt	5
Ireland	Blended	23
	Flavored	1
	Grain	1
	Pure Pot Still	2
	Single Malt	14
Japan	Blended	1
	Single Malt	3
Scotland	Blended	21
	Campbeltown	2
	Highlands	27
	Islands	9
	Islay	17
	Lowlands	1
	Speyside	32
Taiwan	Flavored	1
	Single Malt	7
USA	Blended	2
	Bourbon	51
	Corn	8
	Flavored	3
	Grain	1
	Rye	15
	Single Malt	2
	Unaged	2

Count of Whiskey: 1 – 51

Figure 3: Table of color-coded whiskey counts by country and type.

Based on the results in Figure 3, the number of whiskey products, as well as type, are not equally distributed across the countries. For instance, the United States is the only nation that produces Corn (8 products) and Bourbon (51) whiskeys, and Scotland is the sole producer of Campbeltown (2), Highlands (27), Islands (9), Islay (17), Lowlands (1), and Speyside (32). These two nations produce the most whiskey, followed by Ireland. There appears to be an association between type and country of origin; for example, the North American nations (United States and Canada) are producers of blended, rye, and single

malt whiskeys; Asian nations (India, Taiwan, and Japan) all produce single malts; and Scotland and Ireland produce many blended whiskeys.

In short, I would recommend that the manager consider how many whiskeys they can afford to offer without going over their budget. Once that is determined, it would be worthwhile to consider pricing of each whiskey since prices may vary based on type and country of origin. Then, the manager can determine the number of whiskeys to offer based on geographical location and type. I encourage the manager to choose single malts from England and India since those are the only whiskeys produced in those nations. I also recommend the blended whiskeys from France and Japan, not only because those are the only two blended whiskeys, but because France and Japan do not produce much else. There is only one flavored whiskey hailing from both Taiwan and Ireland, so it would be worth including those in the party. Since Scotland and the United States produce the most whiskey, perhaps offering more whiskeys from these two countries would allow for equal geographical representation.

Question 3: Response

This question is also exploratory and can be largely answered by using the given dataset. Since this visualization compares three different variables, I wanted to create a heatmap since I felt this would be an effective visualization for the assistant manager to gain further insight into how the aging process might affect the customer rating and alcohol content of whiskey. For the purposes of creating the heatmap, I converted ABV from a continuous measure to a dimension. I decided to bin the age values into groups which helped simplify the visualization without contributing to any significant loss of information. Customer rating was kept as a continuous measure, and the average ratings were reported for each age group and ABV. Customer ratings are coded in a red-blue divergence color scheme so that the viewer can easily distinguish between high and low ratings.

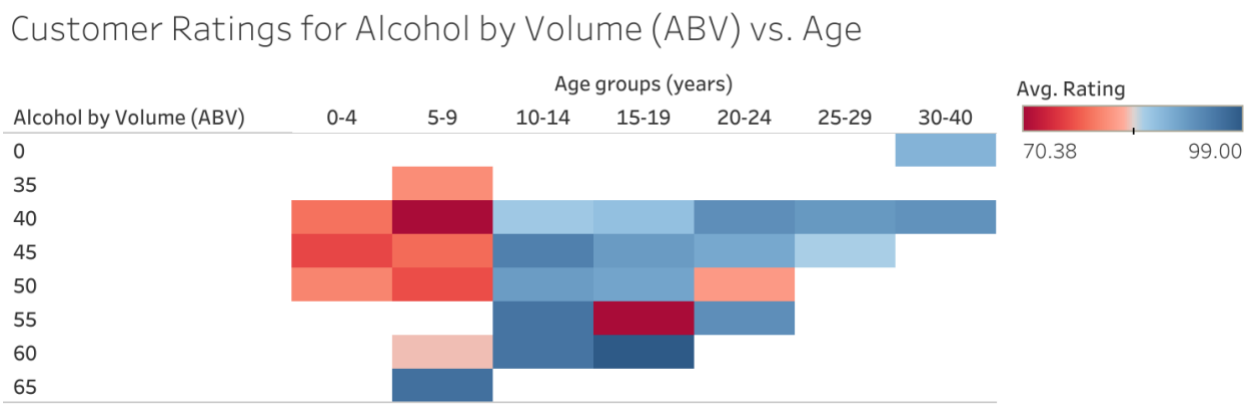

Figure 4: Heatmap of customer ratings of whiskey products sorted by alcohol by volume (ABV) and age. Customer ratings are color-coded based on a diverging red-blue color scheme; age is grouped by 5 years except for the 30-40 year age group; alcohol by volume (ABV) is reported in intervals of 5 ABV units, except for one whiskey with a reported ABV of 0.

Overall, it appears that the aging process has little to no influence on ABV. This is apparent for whiskeys with reported ABV of about 40 since this ABV appears in all age groups (Figure 4). However, the aging process seems to be directly associated with customer rating, where the more a whiskey ages, the higher the customer rating tends to be for that whiskey product. While this trend does not uphold for every single grouping, this does appear to be the general pattern outlined in Figure 4.

Part 2: Unexpected Findings

One unexpected insight that came from exploring the whiskey dataset was the geographic distribution of whiskey production. Initially, I tried to visualize this with a Dorling cartogram that uses dots to represent the number of whiskey of the map of the world with dots signifying the number of whiskey products manufactured in each country. However, this approach did not reveal any key patterns since only a few countries produce whiskey; therefore, the dots were obscured by the sheer size of the globe. When I switched to a color-coded summary table, it became evident that the United States and Scotland were the largest producers of whiskey. It was also when I learned that Bourbon is produced exclusively in the United States. This makes sense since Bourbon is primarily made from corn, and the United States is the world's largest producer of corn. I was also surprised by the sheer variety of whiskey types produced in Scotland, but this makes sense given Scotland's renowned history of whiskey-making. Based on these findings, I wanted to research into why these two countries are the leaders in whiskey production. Apparently, whiskey production thrives in regions that experience significant temperature fluctuations, as this aids in the barrel aging process. Abrupt seasonal changes are common in many

areas of the United States, which might explain why many whiskeys are produced in the country.

One thing I learned from this Tableau assignment is how certain visualization techniques yield different insights. While the world map visualization obscured certain patterns, the summary table provided a clearer understanding of whiskey production across countries and types.

Part 3: A Critique of Tableau

In using Tableau, I learned that it seems to excel at spatially arranging data, as well as maximizing the available design space as more dimensions and measures are added without sacrificing image resolution. However, the whiskey dataset was not very large (284 items and 8 attributes total), and my visualizations were not inherently complex (i.e., large network trees), so perhaps with an excessive number of dimensions, this efficiency may diminish. One of Tableau's strengths is its intuitive handling of data attributes, where it seems to correctly order ordinal variables and appropriately categorizes data that are categorical.

A key advantage of Tableau is its alignment with the what-why-how framework of data visualization that we learned about in class. Tableau suggests suitable visualizations based on selected dimensions and measures, efficiently processes data queries, and summarizes data in a user-friendly manner. Tableau also possesses built-in alignment, ordering, and separation techniques which can be used to enhance data clarity for both the user and viewer. While Tableau does not seem to assist with many front-end processes, such as defining the domain or formulating research questions, it excels at selecting and implementing appropriate visual idioms. The platform provides extensive visual channels for encoding marks, including but not limited to color, shape, size, and volume. Although users may occasionally need to adjust these elements manually to achieve their desired visualization, this flexibility can be expected in any visualization software.

Tableau demonstrates a strong ability to handle and visualize tabular data. Tableau seamlessly applies separation, ordering, and alignment techniques, handling layout density efficiently and ensuring that the visualization design space is properly utilized. These features contribute to Tableau's reputation as a powerful and user-friendly data visualization tool.

I did have some personal challenges while working with Tableau on this assignment, especially since this was my first time using the software. However, my challenges were more rooted in data manipulation tasks, which are not unique to Tableau. For example,

when analyzing customer ratings and price across 284 unique whiskeys in Question 1, some records had missing values (11 did not have customer ratings and 4 were missing price), making it difficult to draw firm conclusions without further data collection. Tableau seemed to allow for interpolation of missing values, but I was hesitant to use that feature without developing a stronger understanding of the underlying missing data mechanism. Additionally, calculating correlations, such as between customer rating and price for Figure 1, required creating a new calculated field using unfamiliar code syntax.

Another challenge involved determining total number of whiskey products by country for Question 2. While Tableau's default settings easily calculated product counts by whiskey type within each country, there seemed to be a learning curve when it came to customizing visualizations beyond the default options. Similarly, converting whiskey age from a continuous measure to a categorical variable required binning values, which once understood, made creating a heatmap for Question 3 straightforward. Lastly, there were 109 whiskey products that were missing a reported age value; however, this was an issue surrounding the dataset and not the Tableau software.

Overall, Tableau automates many processes, making it highly user-friendly. However, this level of automation may be limiting to statisticians who are accustomed to more manual control through coding. While tableau enables users to rapidly create visualizations, its intuitive interface can sometimes make precise data manipulation more challenging. Despite these trade-offs, Tableau appears to be an effective tool for quickly generating complex visualization compared to other statistical software tools like R.