Faculty of Computing and Information Technology

BACHELOR DEGREE IN MANAGEMENT MATHEMATICS WITH COMPUTING (HONOURS)

BAMS3043 Mathematical and Statistical Software

Data Exploration on Men's Shoe Prices

Tutorial Group   : RMM3S1G2

Tutor   : Dr. Tan Yan Bin

Submission Date   : 22 August 2021

Source of Dataset   : [Men's Shoe Prices](#)

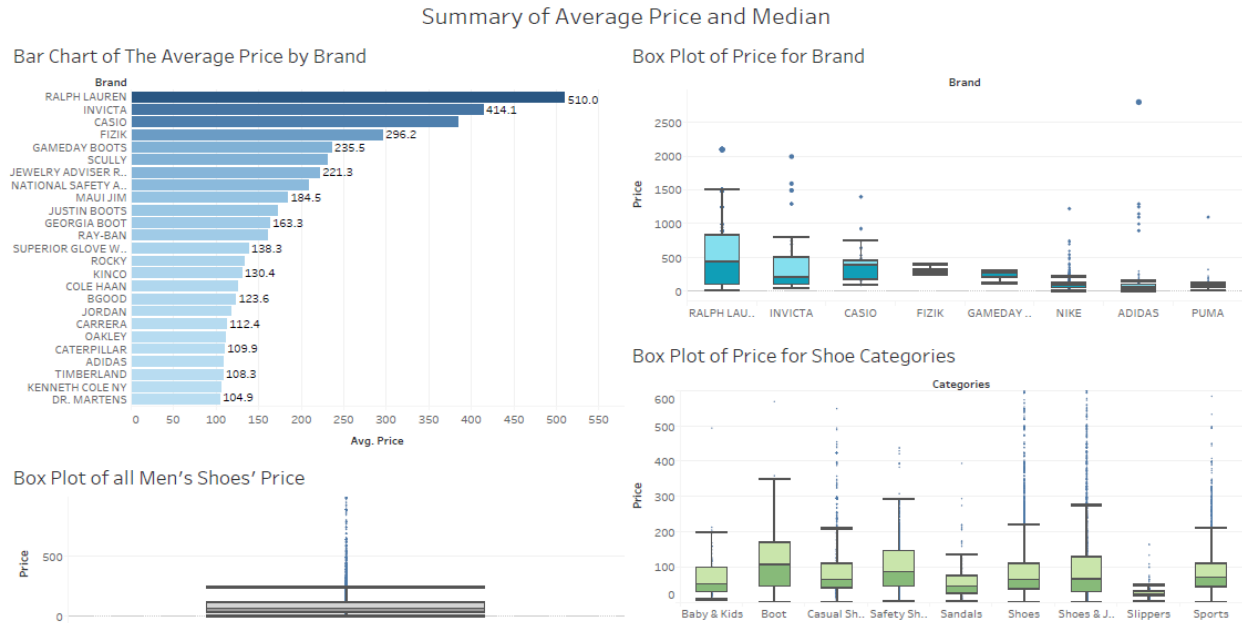| Name	Student ID
Ong YiLiang	20WMR09187
Kang Tien Wey	20WMR09182
Chan Teik Chun	20WMR09179

Table of Contents

Figure 1: Summary of Average Price and Median

Figure 1 has shown the dashboard of the Summary of Average Price and Median which contains the Bar Chart Of The Average Price By Brand, the Box Plot of Price for all men's shoes, the Box Plot of Price for Brand and the Box Plot of Price for Shoe Categories. First of all, the brand is filtered by choosing only the brands that have more than 50 observations and with only the currency of USD for this data analysis.

The bar chart of the average price by brand has shown the average price of every brand that is more than 100 average price. The bar chart is sorted in descending order which is from the highest price to the lowest price. Thus, the bar chart shows that Ralph Lauren has the highest average price with $509.3, followed by Invicta which has the second highest average price with $412.9. Meanwhile, we notice that there is a portion of extreme value in the sample data. So, mean might not be a good statistical measure in this circumstance because the average price will shoot up or drop extremely due to the extreme value. So, median might be another better option for us to measure the data.

Thus, the dashboard of Summary of Average Price also shows two box plots of price analysed for different brands and categories. These box plots will tell us about the Interquartile

Range (IQR), median, maximum score, minimum score and outliers for the different brands and the different categories of shoes. For example, we can see that Ralph Lauren has the largest interquartile range of average price and highest maximum score. Furthermore, it obviously shows that the price distribution of Invicta is skewed to the right while Casio is skewed to the left. For the Box plot of average price of shoe categories, Slippers has the smallest interquartile range and has the lowest maximum score compared with other categories.

From the box plot of all men's shoes, 50% of the men's shoes are in between the price of $36 to $118 and the median is $65. With the too specific numerical value here, it is difficult to bring us some meaningful analysis. Yet, by determining the grouped data of price, it gives a better tolerance for decision making. Meanwhile, we will use the information from box plots for further chart plotting. So, there were some price distribution charts plotted in figure 2.
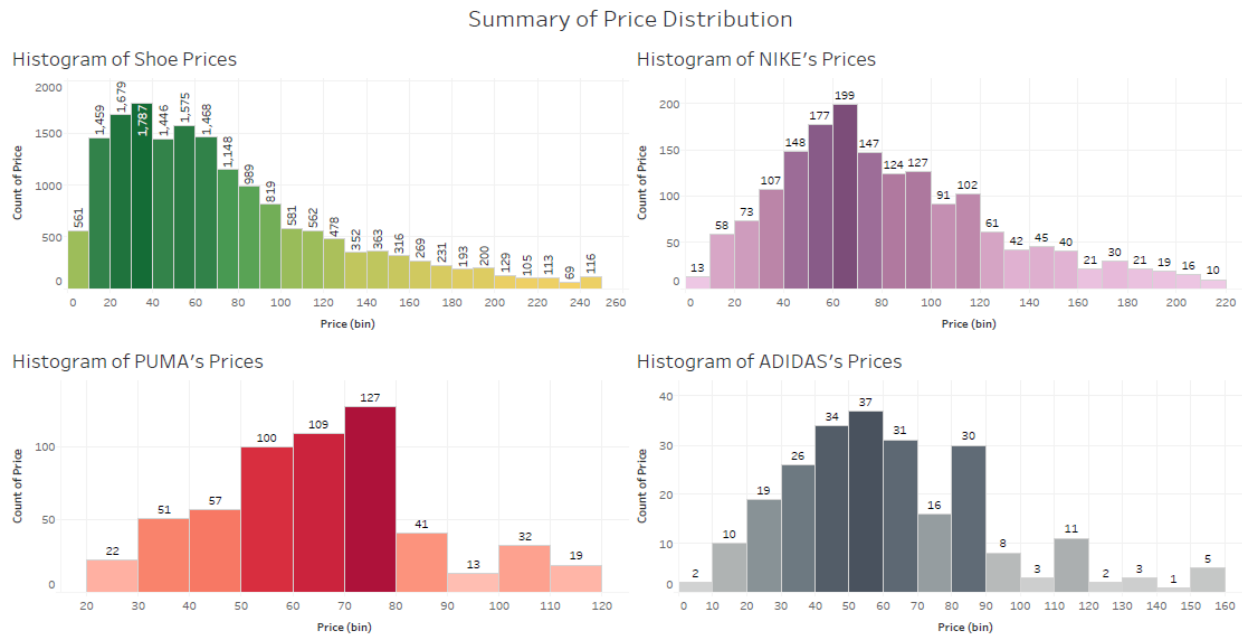

Figure 2: Summary of Price Histogram

From figure 2, the dashboard shows the summary of price distribution of men's shoes in the form of histogram. The reason why Nike, Adidas and Puma are specially picked out by us is because these brands are more popular and hot in our domestic market. So people might be more interested in having the information of these brands.

These histograms are illustrated with the range of price according to the boxplot in figure 1 from the lower whisker to the upper whisker which had included and considered 99.3% (Michael Galarnyk, 2018)[1] of data in it. Whereas the 0.7% of the data on outliers are excluded from the histogram as the outliers are mostly an extreme value and it is meaningless to be considered here. Meanwhile, the interval of the price is set as 10. By defining the skewness of distribution, those histograms could help management to decide the shoe's price to target as many customers as well.

Firstly, the histogram of shoe prices from the range of $0 to $260 is positive skew and the modal interval is in between $30 to $40 with the frequency of 1787. With this information, one of the uses would be encouraging a new merchant or businessman who considers to startup a shoe business to produce a shoe with a selling price between $30 to $40 or slightly a wide modal interval within the interquartile range from figure 1. So, they will be able to sell the shoes at the price within the ability of most people and eventually be able to target more customers by following the current market price distribution.

Next, the dashboard shows that the price distribution of Nike and Adidas is slightly skew to the right. Coincidentally, the modal intervals for Nike and Adidas are quite similar, which is from $60 to $70 and $50 to $60 respectively. It shows that most of the people are still willing to spend more for a branded shoe as compared with the modal interval of overall men's shoe. While, the normal distribution of Puma is almost symmetric and the price of modal interval is higher than Nike and Adidas which is from $70 to $80 with the frequency of 127.
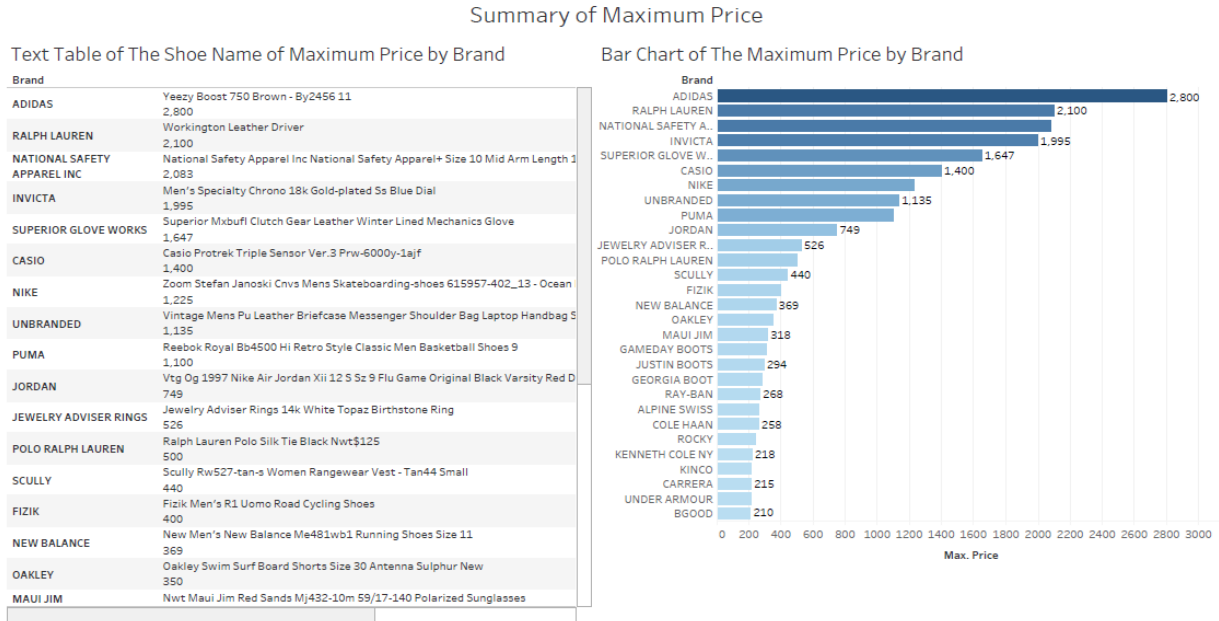
---

[1] Understanding Boxplots

Figure 3: Summary of Maximum Price

Figure 3 has shown the dashboard of Summary of Maximum Price which contains the Text Table of the shoe name of Maximum Price by brand and the Bar Chart of the Maximum Price by brand. Similarly, in this dashboard we have excluded the brand that has less than 50 observations and only consider the currency of USD. Meanwhile, the bar chart is tidy up by only including the maximum price that is at least 200.

In the bar chart of the maximum price by brand, the brand has been listed in descending order. By looking at the bar chart, we know that Adidas is the brand that has the most expensive shoe which cost $2800. By referring to the text table, the table told us the most expensive shoe in Adidas is Yeezy Boost 750 Brown - By2456 11.
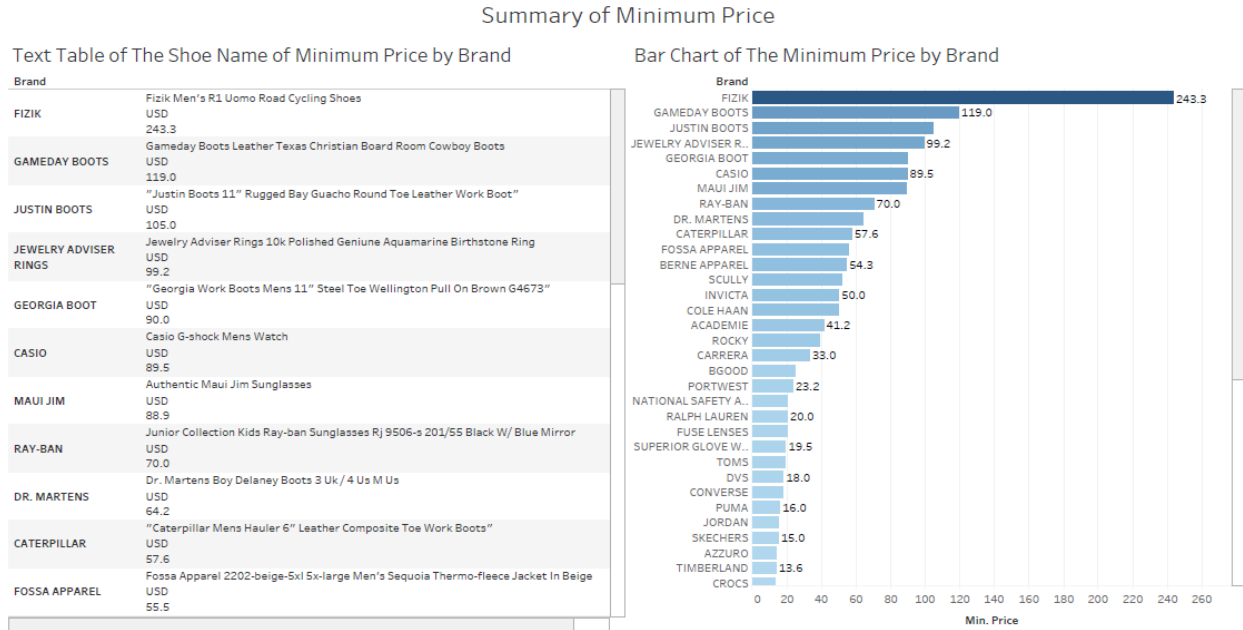
Figure 4: Summary of Minimum Price

Figure 4 has shown the dashboard of Summary of Minimum Price which contains the Text Table Of The Shoe Name Of Minimum Price By Brand and the Bar Chart Of The Minimum Price By Brand. In this dashboard, we have filtered it by choosing only the brands that have more than 50 observations and with the currency of USD for the data analysis too.

In the bar chart of the minimum price by brand, the brand has been listed in a descending order too. Through the bar chart, we can see that Fizik has the highest minimum price among all of these brands. It may represent that Fizik is a brand not that budget-friendly to the people as compared to the others. If people are likely to buy Fizik shoes, their spending will be at least $243.3. By referring to the table, we will get the shoe name that has the minimum price in each brand according to the bar chart respectively too.
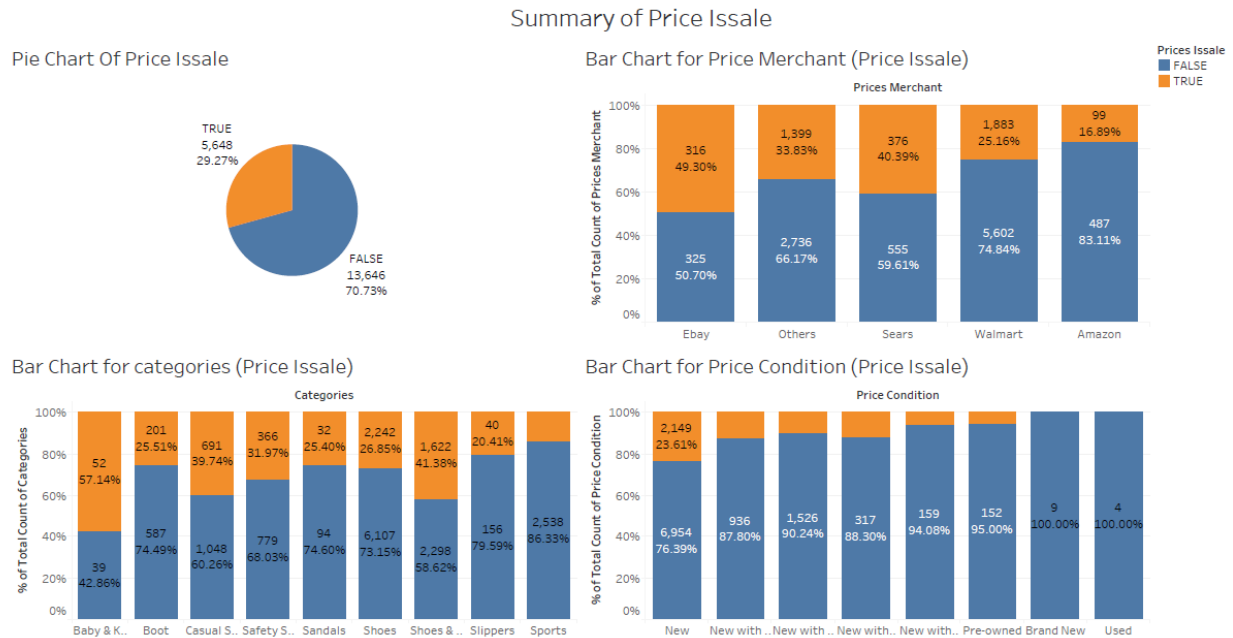
Figure 5: Summary of Price Issale

From the figure 5 above shown the Summary of Price Issale in a dashboard. Price issale is to determine whether the price is a sale or discounted price. In the pie chart, 70.73% of the shoes sold are not in sales price, while the shoes sold at a discounted price were 29.27%.

Besides, the bar chart for price merchants shows that the seller on Amazon is less likely to give discounted prices, having only 16.89%. While around half of the shoes on Ebay are sold at the discounted price. 40.39% of shoes having discounted prices sold on Sears contrast to 59.61% of items not at sales. Thus, people who aim to get a discounted price should focus and shop more on Ebay, Sear and Walmart.

Moreover, the bar chart for categories shows that the category of Baby & Kids has the highest percentage of items sold at discounted price which is up to 57.14%. The second highest category is Shoes & Jewelry and Casual Shoe which consists around 40%. While the following categories are Safety Shoes, Boots, Sandals and Slipper which are in between 20% to 40% respectively. Lastly, the Sports category has the least count of shoes sold at discounted price which is only 13.67%. By looking at this bar chart, sales and marketing managers may be able to

make a hypothesis for the trending category shoes in current circumstances for further research and analysis.

At last, it is the bar chart for price conditions. The price conditions that have sold at discounted price are New, New with tags, New with Box, New without tags, New without box and Pre-owned. Among these price conditions with sales, New condition of shoes has the highest percentage which is 23.61% while the lowest frequency of condition that have sales is Pre-owned which only consists of 5%. Lastly, it clearly shows that none of the price conditions of Brand New and Used have a discounted price. In the nutshell, this bar chart has shown to confirm the fact that our common sense is true with the data analyzed.
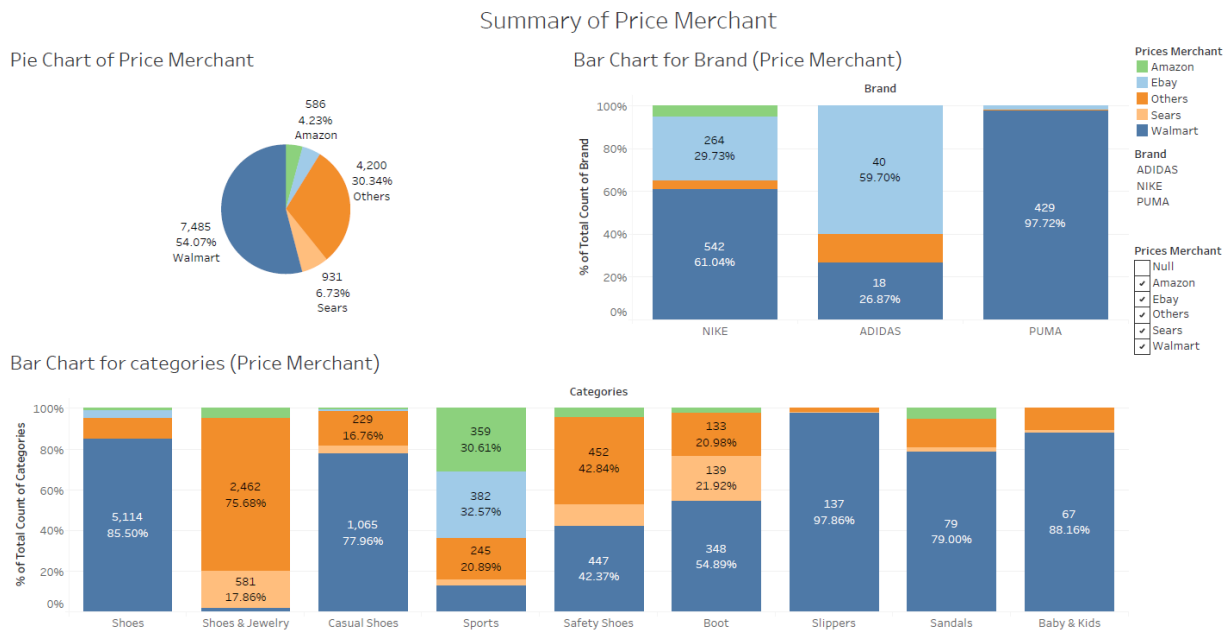

Figure 6: Summary of Price Merchant

According to figure 6, the price merchant refers to a list of third-party merchants found on e-commerce websites selling this product at this price. Through the pie chart, most of the merchants or sellers are more likely to sell the shoes on Walmart which occupy up to 54.07%.

Next, the brands that are chosen in the Bar Chart for Brand are Nike, Adidas and Puma. Among these brands, mostly the Puma men's shoes are sold on Walmart e-commerce which consist of as high as 97.72%. Adidas products could be more commonly found in Ebay and Walmart which is 59.70% and 26.87% respectively. In addition, from Figure 5, we know that almost 50% of men's shoes are having sales on Ebay. This could be good news and the benefit for Adidas fans to start to shop at Ebay.

Moreover the third bar chart shows the percent of total count of categories. Sport Shoes are sold almost evenly at different e-commerce sites excluding Sear which is on Amazon, Ebay, and Walmart with the percentage of 30.61%, 32.57% and 13.3% respectively. And yet, this bar chart also fulfilled the analysis for the bar chart of Nike, Adidas and Puma, which showed that these sports brands are not selling any men's shoes at Sear. On the contrary, people who shop at Sear would be more likely to get the men's shoe in the categories of Shoe & Jewelry, Safety Shoes and Boots.
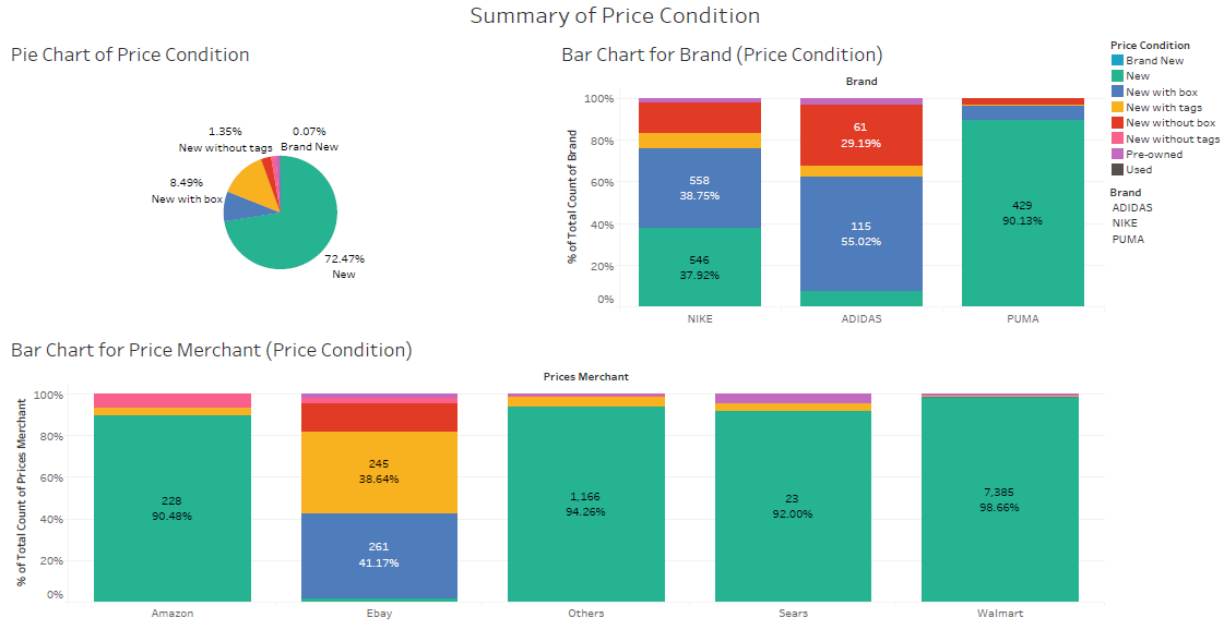
Figure 7: Summary of Price Condition

Figure 7 has shown the Summary of Price Condition in a dashboard. Firstly, the pie chart shows that most of the condition of shoes sold on e-commerce are the new one that takes up almost 98% which consist of New, New with tags, New with box, New without tags, New without box and Brand New. While the remaining 2% are Pre-owned and Used conditions.

Through the bar chart for the brand, we see that around 30% of Adidas shoes are new without a box, Nike consists of around 15% while Puma is only around 3%. Although there are some men's shoes that are new without a box, it is clearly seen that these popular and hot brands' shoes all have the tags. So, these brands actually did well to prevent their customers from getting fake shoes by ensuring all their shoes have the proper tags.

The Bar Chart for Price Merchant has shown that Amazon is having 6.35% of men's shoes that are new without tags and Ebay consists of almost 3%. It is still good to remind people to be more aware of buying men's shoes on Amazon and Ebay from getting fake products although it is just a small portion of it. In addition, Ebay may have a relatively poor packaging on delivery due to almost 14% of men's shoes are in the condition of new without a box.
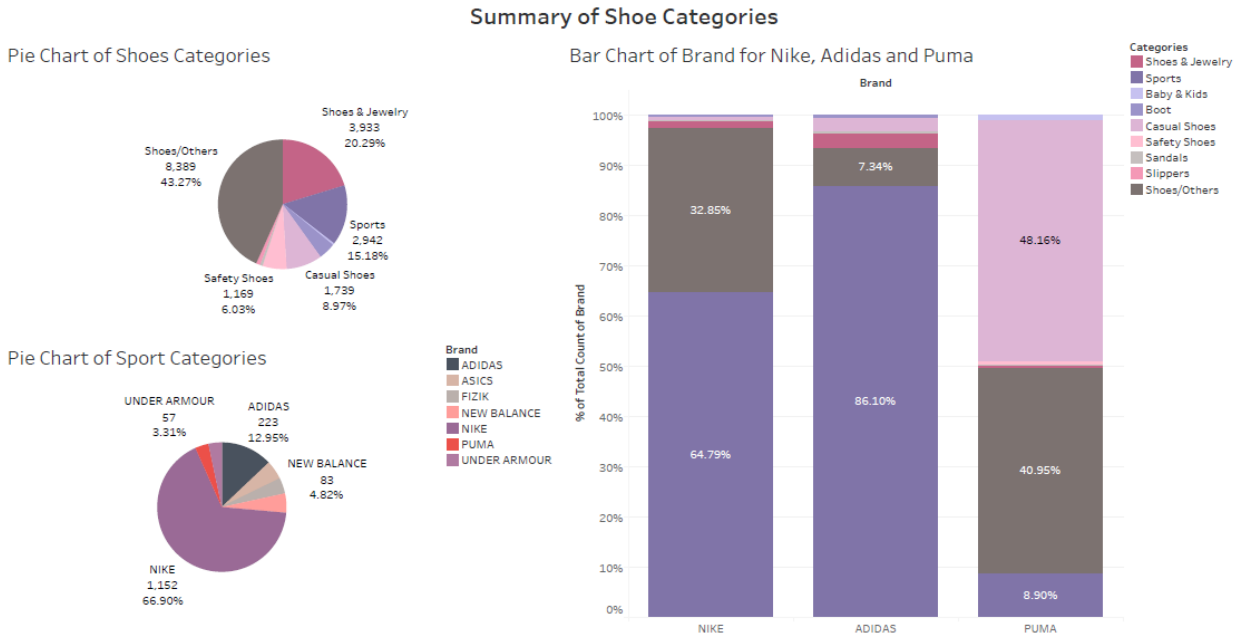
Figure 8: Summary of Shoe Categories.

Figure 8 has shown the Summary of Shoe Categories in a dashboard. From the pie chart of shoe categories 43.27% of shoes are not really able to be classified in a specific category, around 15% of shoes are under the category of sports. While there are almost 9% of casual shoes and safety shoes take up 6%.

By enlarging the information in the pie of sport category, the pie chart of sport category shows that Nike takes up almost 70% in this category and followed by Adidas, 12.95%. While the rest of the brands take up uniformly in the pie chart which is around 3%. This pie chart indirectly describes that Nike is the magnate in the sport category which is the world's largest supplier and manufacturer of athletic shoes, apparel and other sports equipment.[2]

Although Nike is the magnate in the category of sport, the shoe of sport in Nike just takes almost 65% which is lower than Adidas that takes up 86.1%. Through the bar chart, we can also see that almost 50% of shoes in Puma are casual shoes. Thus, Puma brand is targeted on designs and manufactures casual and athletic footwear.[3]

---

[2] Nike - Statistics & Facts
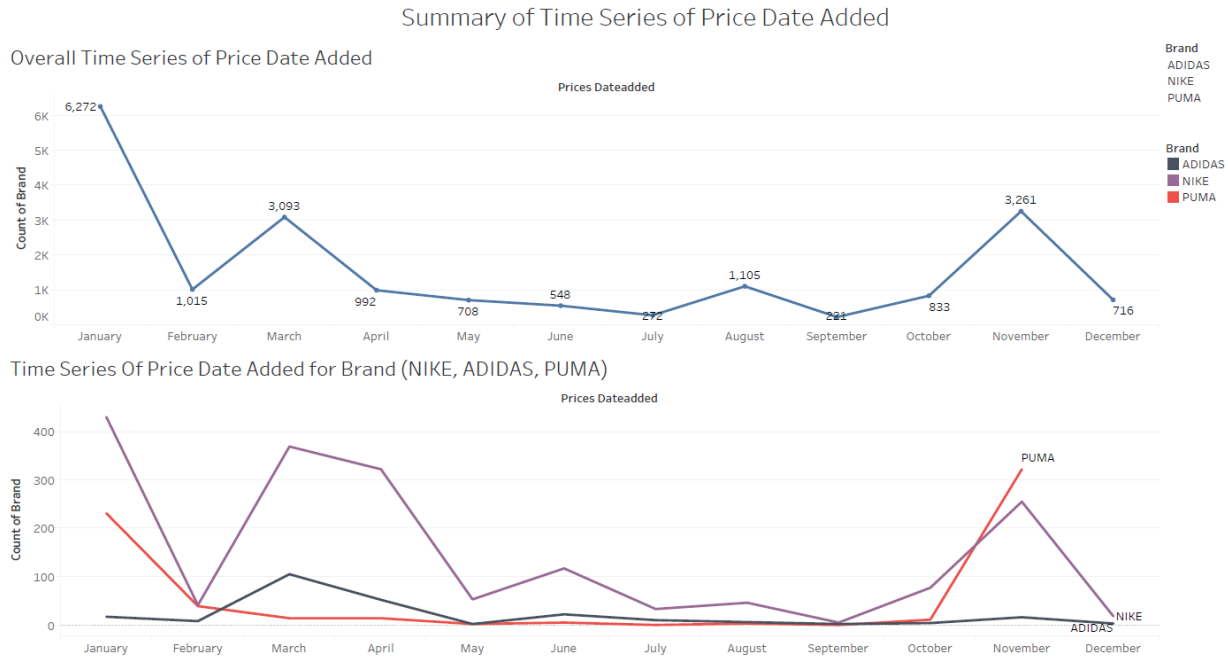[3] Puma - Statistics & Facts

Figure 9: Summary of Time Series of Price Date Added

Figure 9 has shown the dashboard of Summary of Time Series of Price Date Added which include the Overall Time Series of Price Date Added and Time Series of Price Date Added for Brand Nike, Adidas and Puma., which are three major brands we focus on. Price Date Added is the date of this product was first added to the product database.

The chart of the Overall Time Series of Price Date Added has shown that most of the shoes were added in the beginning of the year, more than 6000 shoes were added in January. Based on the overall time series, these brands can focus their sales on January like promotions and advertisements to improve their sales because there will be a lot of different brands and kinds of shoes will be added in January which will lead to a fierce competition.

In the chart of Time Series of Price Date Added for Brand Nike, Adidas and Puma, we will be able to see that their new shoes are added in different months of the year that are distinguished by brands. For example, the brand Nike has added most of their shoes in January, the brand Adidas has added most of their shoes in March, and the brand Puma has added most of their shoes in November.