**Statistical Values of Numerical Variables**

**Sana Mourad, Rand Al-Shoubaki, Leen Daqa, Noor Yazeed**

A picture containing logo

Description automatically generated

**Professor Mohammed Azzeh, Data Visualization**

**Summer Semester 2022**

**Summary**

During the data visualization process of the given data set we found that there were 4000 rows and 30 columns. After cleaning there were still 4000 rows meaning there were no nulls or duplicates found in the data, also we ended up with 27 columns. Some of the attributes in the final dataset were added by us for example : customer\_age, employee\_age, employee\_months\_working, orderDate\_quarter and promotionDate\_quarter. We also found that changing the customer\_City and Branch\_Location columns to North, South, East, and West relative to Amman -the center of Jordan- was more helpful.

Statistical Values of Numerical Variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Mean | Max | Min | Skewness | Kurtosis |
| Quantity | 64 | 180 | 1 | 0.70 | -0.34 |
| Price | 495 | 1000 | 2 | 0.02 | -1.20 |
| Rate | 6 | 10 | 1 | 0.02 | -1.20 |
| Customer Age | 45 | 72 | 10 | 0.04 | -0.75 |
| City Population | 555086 | 2473400 | 89400 | 1.89 | 2.55 |
| City Density | 232 | 723 | 4 | 0.83 | -0.14 |
| Employee Age | 30 | 42 | 18 | -0.11 | -1.25 |
| Employee Months Working | 5 | 11 | 0 | 0.33 | -0.92 |

# Categorical Unique Values

|  |  |
| --- | --- |
| Employee Role Description | Cashier, Captain |
| Employee Type | Fixed-Term, Part-time, Shift-Workers, Full-Time |
| Employee Working Hours | 6h, 12h, 8h |
| Customer Type | Normal, Member |
| Payment Method | Cash, E-Wallet, Credit card |
| Department | Gifts, Designer, Kids, Activewear, Sale, Young Adult, Men, Women, Beauty, Home |
| Holiday | 1 (is Holiday), 0 (not Holiday) |
| Promotion Type | No Promotion, Free Gift with Purchase, Giveaways, Online, Promotion, Sales Promotion, Discount Promotions, Tiered Promotions |

# Branch Location

|  |  |  |  |
| --- | --- | --- | --- |
| North | East | South | West |
| Irbid | Mafraq | Ma’an | Balqa |
| Ajloun | Zarqa | Aqaba | Madaba |
| Jerash | Amman |  | Karak |
|  |  |  | Tafilah |

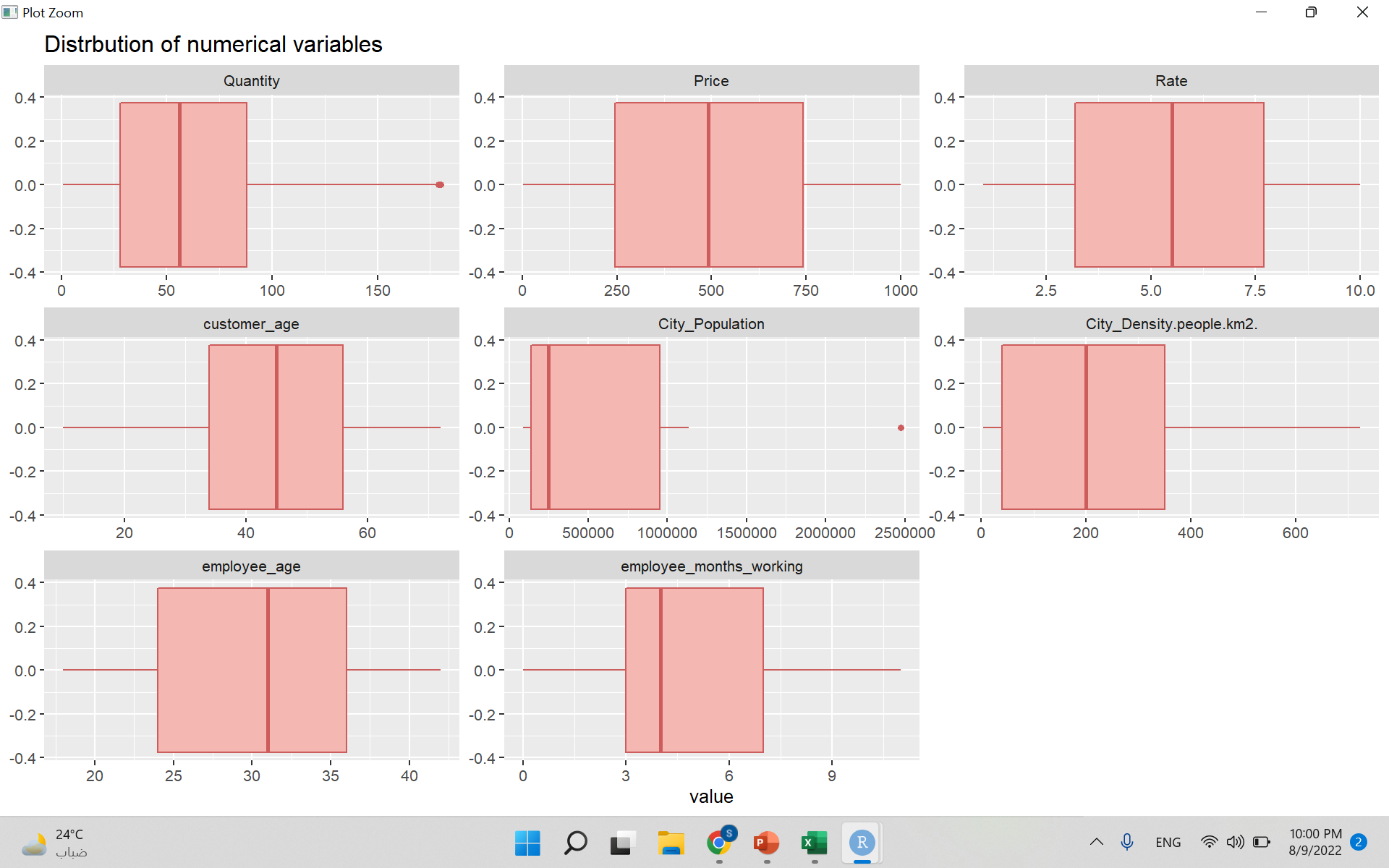
# Quarter Distribution

|  |  |
| --- | --- |
| Quarter 1 | January-March |
| Quarter 2 | April-June |
| Quarter 3 | July-September |
| Quarter 4 | October-December |

Figure 1 Title: Distribution of numerical variables (using Histogram)

Figure 1. This Histogram shows the distribution of numerical variable however we found that it does not show the distribution clearly, so we decided to be more accurate by using skewness function. The methodology used was creating long dataSet of only numerical variable and using facet\_wrap() function on variable which is the name of each numerical variable.

Figure 2 Title: Distribution of numerical variables (using Box Plot)

Figure 2: This Boxplot shows the distribution of numerical variable according to median and quartiles (IQR: interquartile range), there were only outliers in quantity and city population.

The methodology used was creating long dataSet of only numerical variable and using facet\_wrap() function on variable which is the name of each numerical variable.

Chart, bar chart

Description automatically generatedFigure 3: Relationship between Price, Department and Holiday

Figure 3: This histogram shows the relationship between price and department according to Holiday, The Sale and gifts departments sold more during holidays while beauty, designer, women and kids' department sold less in holidays.

Figure 4 Title: Relationship between Quantity and price

Scatter chart

Description automatically generated with low confidence

Figure 4: This figure shows the correlation between Quantity and price and there was no correlation since all the data were distributed throughout the plot.

Chart, bar chart

Description automatically generatedFigure5 Title: Relationship between Ordered day and promotion Type

Figure 5: This plot shows the relationship between when an order was placed and the type of promation at that time of the year the most Common promotion was giveaways, and there was no significant difference between number of promotion each quarter.

Figure 6 Title: How Price is affected by the location.Chart, bar chart

Description automatically generated Figure 6: This plot shows the difference of prices for each customer city, products were the most expensive in the West region , while the products were cheapest in the South

Figure 7: Relationship between PriceChart, bar chart

Description automatically generated and Promotions.

Figure 7: This plot shows the difference of prices for each promotion quarter the price is Not affected by the quarter when there was a promotion at that time of the year.

Figure 8: Relationship between price according to Customer type.Chart, histogram

Description automatically generated Figure 8: This plot shows the price increasing/decreasing according to customer Type the figure shows that there was sligty increase in price for members

Figure 9: Rate according to department type(ErrorBar). Chart, bar chart

Description automatically generated

Figure 9: This figure Displays Error Bar for Tax Rate according to each each Department, as we know the error bar length correlates to accuracy and certainty of the variable which means that the accuracy of the Tax rate of each deparetment is pretty similar to the rest and we noticed that the Kids and Home have the highest tax Rate. Beauty and women have the highest accuracy.

Conclusion

In conclusion, we saw that the online branch is the main option for customers and the most selling departments are Kids, Women and Beauty, most of the customers are between 20 and 60 years old. Prices are the highest in the West, and lowest in the South. The retailer should do promotions to increase the number of people using credit cards as the payment method as it is safer and easier. We also recommend the store improve their rewards program to encourage members to spend more at the store.

More features could be added, for ex: online sales should have their own dataset and study, since there was a huge number of sales on the online branch. We recommend that the store has more than one rewards program ex: Students, Elderly, Orphans, Military, and Medical Staff. There should also be rewards given to employees based on how much they sell to incentivize them to be more loyal to their job. It will also probably be better if there are more than two employee roles. They should reduce the number of individual products in order to easily alter stock according to trends and what is needed at the time.