

Visualization final assignment: Mall Customers

Antoine LESNES (Student ID: 243XI08X)

1. Introduction

The aim of this study is to understand how mall customers spend depending on their age, gender and income. We will for this use a dataset giving for 200 mall customers their spending score (on a scale from 1 to 100), their annual income in dollars, their age and their gender to see if there is a correlation between the different attributes of the customers and their spending scores.

2. Method

We will for that try to visualize the data to shed light on the importance of one or the other characteristic of the customer on its spending. To show if there is a correlation between income and spending, a simple scatterplot seems to be the best solution. Indeed, we will see a line if there is a correlation and a more homogeneous scatterplot if there isn't. We will then separate the genders by color (blue for males and orange for girls) to see if one gender spends more than the other. To see if there is a correlation with age, we will represent our points in different shade of blue and orange, going from lighter colors for young people to darker colors to older people. There are also some sliders to help with choosing a certain age range and to set a spending score to facilitate the analysis of the data. To facilitate the visualization of the impact of gender, I used a pie chart to represent the number of males and females that are selected.

3. Result

First, by using the min Spending score slider, we see that the proportion between males and females doesn't vary depending on spending scores, so neither of the genders seem to spend more than the other. We see a difference in the population depending on age however since there is slightly more females than males overall in the mall, but we see that with older people, there is a lot more males in the shop than females. We then see on the scatterplot a big mass of points in the center of the graph that tells us that people with average spending power seem to have a spending score in adequation with their income. However, we see that it doesn't seem to be a correlation between spending score and income as both the poorer and richer people seem to either spend a lot or not that much. We also see by selecting only young or old people that older people seem to spend less than young people.

4. Discussion

It is interesting to see that there is no correlation between gender and spending score as some stereotypes says that women are biggest spenders than men. Also, we see that younger people tend to spend more than older people, maybe because they have less experience with money or because young people want more things than older people. It is also interesting to see that middle class people have

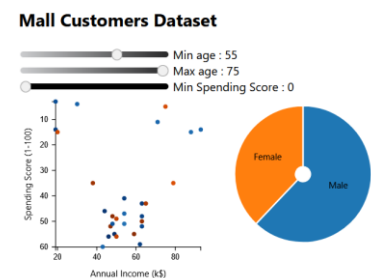


Figure 1: Proportion of males and females with older people

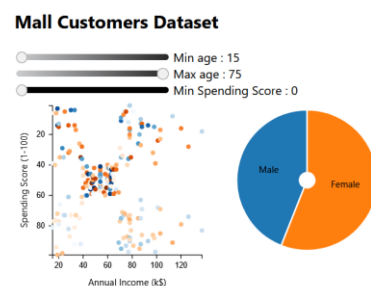


Figure 2: All the population

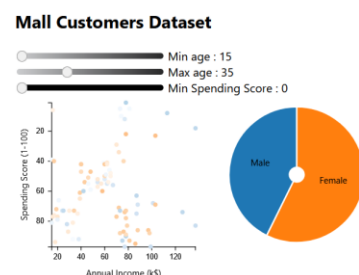


Figure 3: Only the young, that are spending more

more coherence with their spending habits when poorer and richer people either spend a lot or not that much.

5. Conclusion

Gender doesn't seem to be an important factor but age matters a lot in spending habits. However, contrarily to what we may think, poorer people don't tend to spend little and richer people doesn't tend to spend that much more than others. It is thus important not to holds conclusions too fast because the reality is often far from what we think is the truth.

6. Reference

Data from: https://github.com/MainakRepositor/Datasets/blob/master/Mall_Customers.csv