Haircut Data Research

Jeppe Dyekjaer, Oliver Dehnfjeld, Jonas Hansen

October 6, 2021

1 Abstract

A hairdresser has been collecting data in regards to tipping amounts every day of the week as well as customer data such as age, marital status and culture. In this paper we use this data to get a better understanding of possible co-relations. Our findings suggest that non married individuals tip higher than married individuals and that Europeans tip lower than other cultures. In regards to tip amount on given weekday no solid suggestions could be made. An important point to this study is the perhaps questionable sample size and reliability of data. Foundations for ideas to study further is perhaps the biggest takeaway.

2 Introduction

We have been given a set of data from a hairdresser, who has noted the amount of tips she is getting from her customers. Furthermore, she has noted the day, the time of the day, the age of the customer, the marital status and from which culture the customer is from. In this paper we wish to better understand this data set and find out, if we can derive some interesting observations.

All footnotes containing In/Out are references to cells in the jupyter notebook provided.

3 Data Processing

At our first glance of the data set we concluded that it contained several inconsistencies, and descriptive notes of the data set itself.

We got rid of the descriptive information about the data set[1] and removed unnecessary rows[2].

Our second step was to begin smoothing out inconsistencies. To do this, we converted all the data in the 'Time' table to 24 hour format[3]. We then removed rows in the 'Culture' column, which contained mistypes, like 'H' and 'V'[4]. The 'Married' column had mistypes like 'Kid', which we changed to 'Not Married', as we expect kids to not be married[5]. In the "Day" column we had inconsistent shortenings of days, which we chose to prefix to the first 3 letters[6]. Finally we grouped the data in the 'Culture' column into continents, as having Phillipino separate to Asian makes for even smaller groups, and it could potentially give a misleading picture of how well different cultures tip[7]. We identified an outlier in the 'Tip amount' column with the value 40. The mean of all tips in the data set is 6.626086957, and is therefore more than 80 percent lower than the outlier, and we decided to remove the outlier from the data set[8].

[1] In [1], In [5]
[2] In [3]
[3] In [9]
[4] In [12]
[5] In [13]
[6] In [13]

4 Methods

We are using Data-Driven Data Science, as we have a set of data, and we want to find new insights from that data, and then derive theses from those findings.

4.1 Types of Data

We are using observational data, as this data set is made by a hairdressers observations of customers in her shop. The data in the 'Married' and 'Culture' columns are non-parametric nominal data, while the data in the 'Age', 'Day', 'Time' and 'Tip amount' columns are parametric interval data. All the data is structured data, but it is a mix of quantitative and categorical data.

5 Analysis

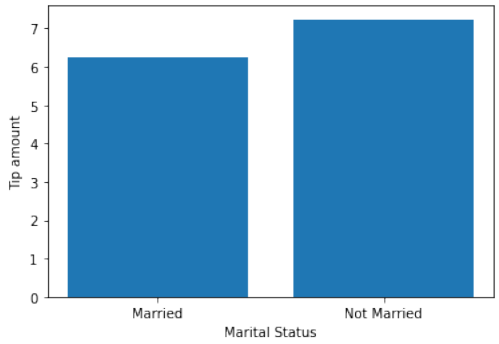
We have decided that it is interesting to examine whether there is a correlation between marital status and the tip amount given by customers[9]. In order not to be misleading, we have removed all data where the age is under 18, we do this because we believe that it is highly unlikely that a person under 18 is married and they are likely not paying for their haircut themselves, but rather their parents are. If we had kept this data we deem that it would possibly skew our results.

6 Findings

We have shown the findings of our research beneath. To see the logic behind the graphs and bars, please see the provided notebook and the given In/Out number in the footnotes.

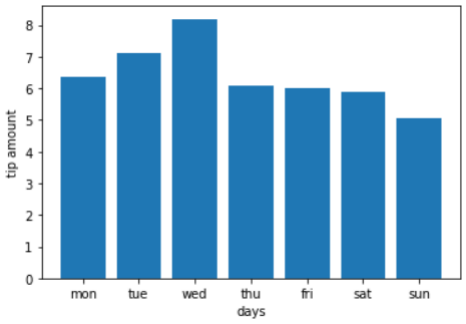
[7] In [12]
[8] In [12]
[9] In [16]

6.1 Tip amount and marital Status



The data suggests, that not married customers (7.235772) tip significantly higher than married (6.240260) customers[10].

6.2 Tip amount and days

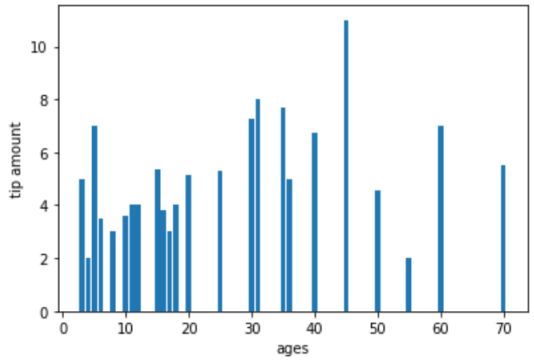



```
sat    53
fri    41
tue    38
mon    34
wed    31
thu    23
sun     9
Name: Day, dtype: int64
Tip amount median count for all days: 32.714285714285715
```

In the bar chart above,[11] we have decided to analyze the average tip amount per day in a week, to figure out when the customers pay the highest tip amount. As seen on the bar chart we can conclude that Wednesday is the day of the week where customers paid the highest tips on average at the hairdresser. But why is Wednesday the highest tip amount day? Based on the tip amount count for Wednesday as seen in the table[12], Wednesday is just below the tip amount median count for all days but still has the highest average tip amount. That concludes that fewer people on Wednesday pay a much higher tip amount since Wednesday has got the number it has. The reasoning for this seem uncertain and it could well be up to pure coincidence.

[10] In [16]
[11] In [15]
[12] In [19]

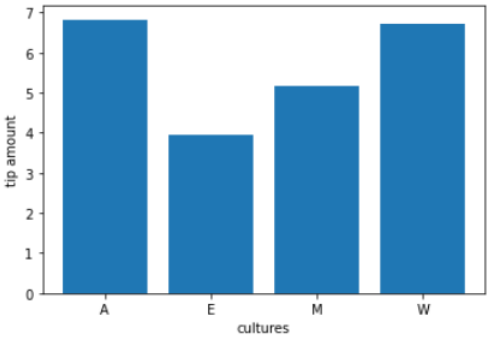
6.3 Tip amount and ages



In this bar chart we compare the average tip amount based on ages for various people. The age 45 has a remarkably higher average tip amount compared to the other ages. That makes sense since people in that age might have finished their educations quite many years ago and have gained a stable income during the time since their finished education. Hence they have more money to give higher haircut tips. The age 5 is also high, because parents of children in that age most likely are paying the haircut bills for them and some extra tips.[13]

6.4 Tip amount and cultures

In this bar chart we compare the average tip amount based on cultures (continents) among people. Asia (A) and America (W) pay the highest tips on average. It makes sense that Europe (E) has the lowest average tip amount since European people are more modest to pay tips compared to other nationalities.[14] The average income of European inhabitants is very high so the lower average tip amount most likely does not relate to economy but might rather be a cultural reason.[15]



7 Conclusion

Through our approach of Data-Driven Data Science[16] and after sanitizing our data[17] we derived the suggestions that married customers tip higher than non mar-

[13] In [17]
[14] Tipping Around the World: A Global Gratuity Guide: https://www.westernunion.com/blog/global-tipping-guide/
[15] In [18]
[16] Section 4
[17] Section 3

ried customers[18]. When examining tip amounts on different weekdays our data suggests that Wednesday is the tip-big day. However, we noted that it is below median count of tips. This tells us that even though the tip amounts were higher, there were fewer tips noted. Hence, we can not with very high certainty say that there is any correlation here.

Upon investigating tip amount and ages we found that the age 45 had a remarkably higher tip amount compared to the rest.[19]. We believe that this could be due to 45 being an age where most are financially stable. The results here are also vague, as when you include ages below 18, chances are, that it is actually the parents who are providing the payment.

Finally, we dug into perhaps what could be seen as most controversial: Tip amount and cultures [20]

The data suggests that Europeans are on the lower end when it comes to tipping. This is not a far fetched suggestion, as culture in Europe have it, that tipping is not as much of a norm.

A common denominator in this study is the shortcoming of data in the data set, as well as question to its reliability. This concludes that suggestions made in this paper in of themselves are not groundbreaking, but should rather be used to choose a focus-point for further studies.

7.1 Ethical Concerns

It might not be ethical to divide the amount of tips by culture in this study. The findings could be used to portray a bad image for people from certain cultures, and the study could be skewed to support scientific racism. Even if that is not the case, we do not know how the hairdresser classified cultures. A person defined in the data set as Asian could very well be of "American culture" and likewise for the rest of the classifications.

However, we decided to keep the findings of culture and tipping, as we found an interesting tendency, which we believe is hard to skew. What we also found out in this analysis is that the tip amount for some cultures may not necessarily relate to how rich or poor the people are but could also relate to gestures, manners or traditions in the cultures.

[18]Section 6.1
[19]Section 6.3
[20]Section 6.4