

## Team report of Group 14 — CACI WS 20/21 Special Work Performance 1 (19.11.2020)

Luisa Hahn (577736), Vincent Gurgul (573476), Mert Erdinc (615804), Jan Lorch (617032)

### Dataset “City Trips”

In the following we analyzed the given dataset „City Trip Dataset “. The initial overview is instrumental in order to contextualize the hypotheses, findings and assumptions. All calculation and plotting work were done in R/RStudio respectively. Collaboration was done through GitHub.

The given Dataset consisted of 266 participants, each represented as individual rows. Taken datapoints from answered questions is stored in 462 columns total.

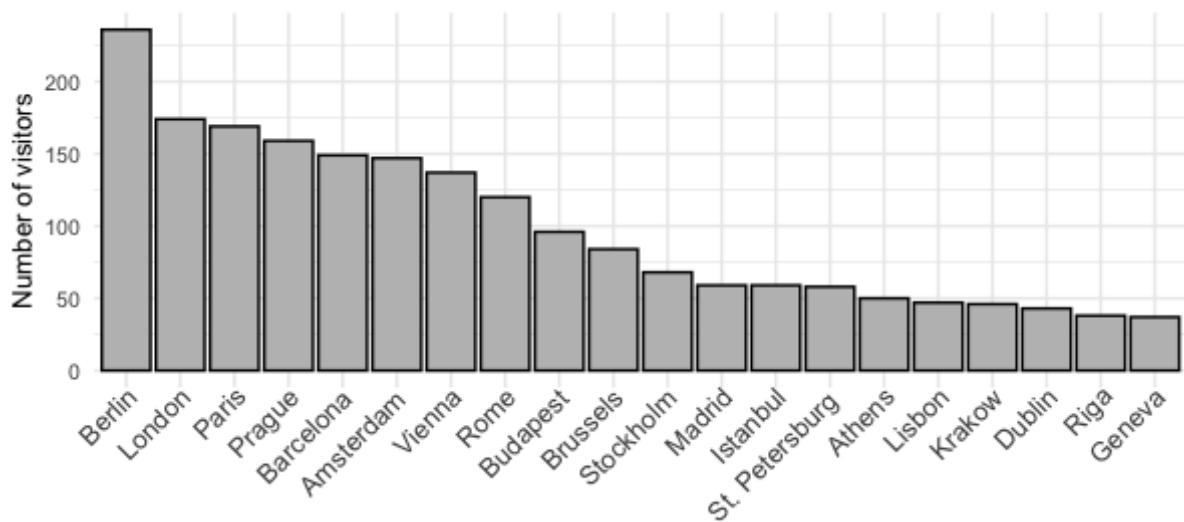
#### Personal Data of the Participants

113 out of 226 participants were male

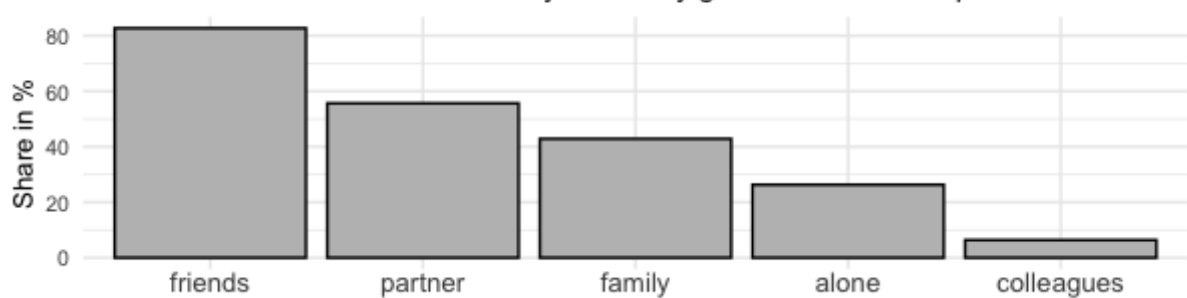
153 out of 226 participants were female

the mean age was 25.5

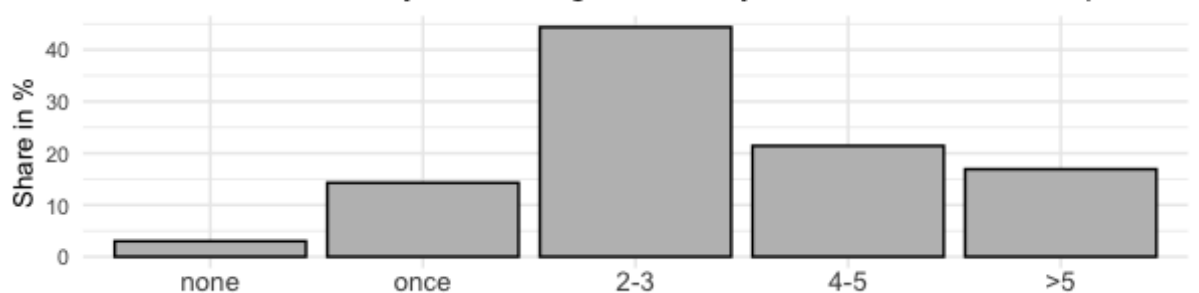
Question 1: Which of the following cities have you visited?



Question 3: With whom do you usually go on a weekend trip?

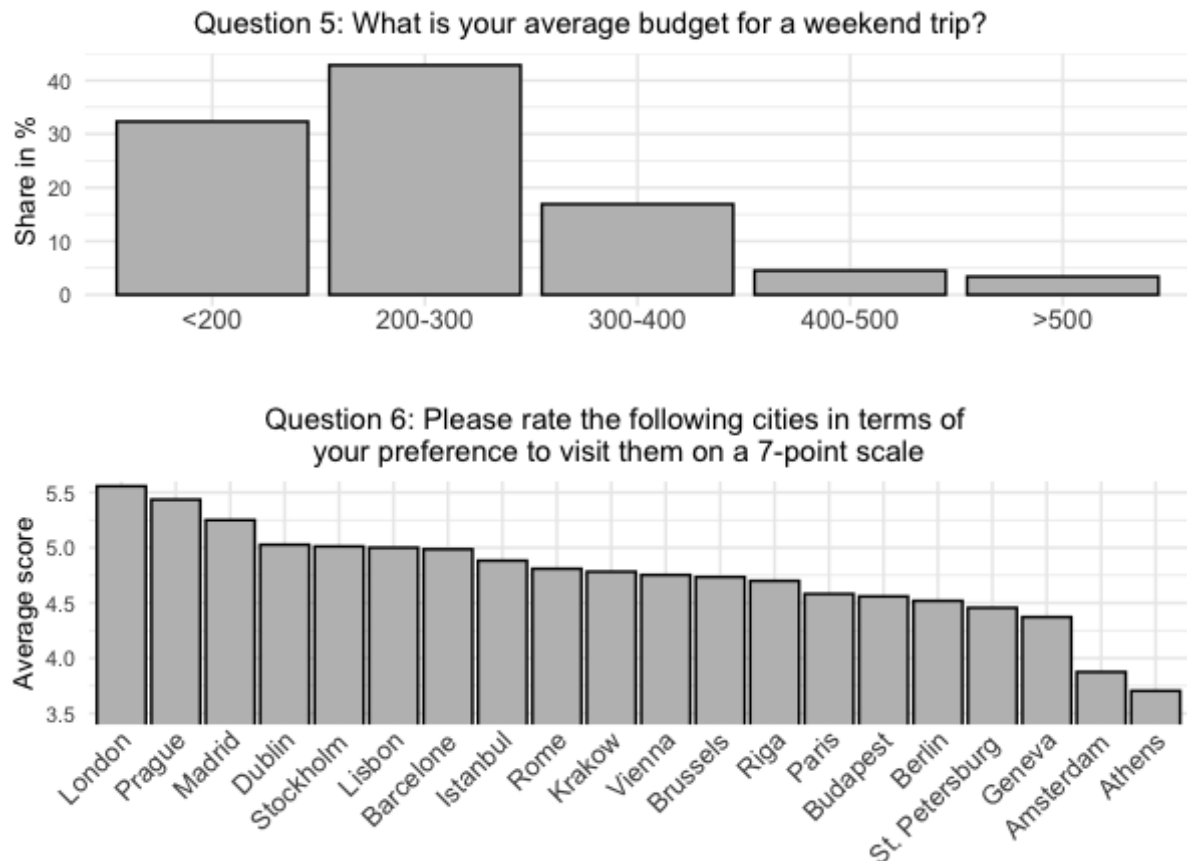


Question 4: How many times during 2015 have you been on a weekend trip?



## Team report of Group 14 — CACI WS 20/21 Special Work Performance 1 (19.11.2020)

Luisa Hahn (577736), Vincent Gurgul (573476), Mert Erdinc (615804), Jan Lorch (617032)



### Analyzing the behavior of “low budget” travelers

In this operation, we intended to examine the given city trip data set for a possible causal relationship between both traveling budget as stated in question 36 and traveling frequency both related to unique city visits as indicated in columns 4-23 and overall city trips within the year 2015 as stated in question 35. Base hypothesis for this operation is that there is a direct causal relationship between a smaller budget and fewer unique, as well as fewer yearly city trips.

In the first step, we extracted a subset of data from the main data dump, only containing data respective to travelers on the lowest budget.

In the next step, for lack of better technical possibilities, we manually assigned the sum of all people from the new „low budget“ subset, who stated they visited a certain city to a new variable.

The new dataframe df is now filled by column with the city names and the respective counter of low budget visitors respectively, overall.

For the unique cities visited dataset, the data points could just be added as the options 0 and 1 directly mathematically translate to a unique visit or no unique visit yet.

For the overall cities visited in 2015, we constructed a weighting scheme.

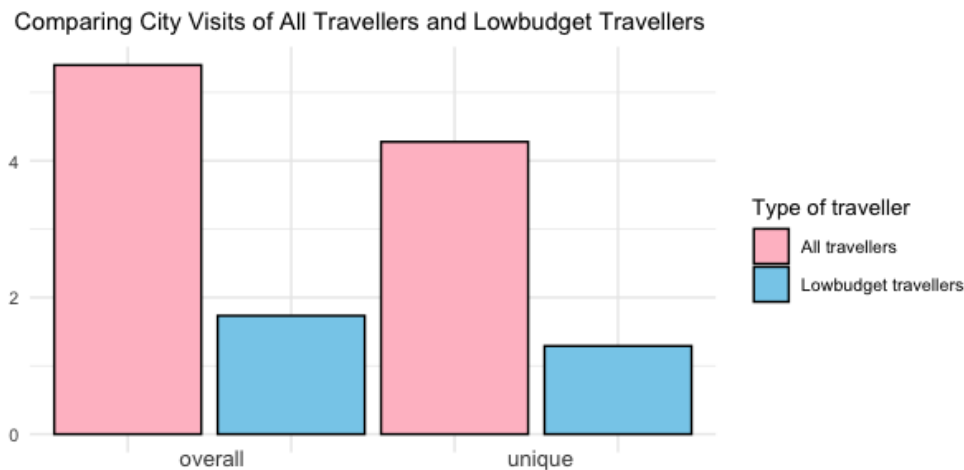
| Answer | Mathematical translation |
|--------|--------------------------|
| 1      | 0                        |
| 2      | 1                        |
| 3      | 2,5                      |
| 4      | 4,5                      |
| 5      | 6                        |

## Team report of Group 14 — CACI WS 20/21 Special Work Performance 1 (19.11.2020)

Luisa Hahn (577736), Vincent Gurgul (573476), Mert Erdinc (615804), Jan Lorch (617032)

Averaging those answers over any given answer, one data point is equal to 2.8 city trips in 2015.

We tried to make use of the given data, even though there are some critical issues, such as unprecise answer options, as in question 36. Assuming the logical interior consistency of our chosen weighting model, and the comparability of the datasets (low budget and average travelers), that it was applied to we can make the following reasonable assumptions.



**T1:** Travelers on the lowest budget, below 200€, only travel to about 30.16% as many unique visitors as the average traveler does.

**T2:** Travelers on the lowest budget, below 200€, only travelled about 32.09% as often in 2015 as the average traveler did.

### Potential further analysis:

In order to get a deeper understanding of the „budget“ component in the given dataset and its influence on both perceptions of respective cities and implications for the overall dataset regarding perceived attractiveness and in-practice affordability.

One possible solution structure could be the ordering of all cities by overall perceived attractiveness rating. Their absolute positional ranking should then be indicated by a counting id.

Comparing that counting id of a respective city, to the counting id in a travel frequency ranking among „low budget“ travelers in a new „cost-effectiveness“ ID could lead to some interesting discoveries. In the case of positive or negative causal relationships, certain cities could be identified as being overpriced for how expensive they are and vice versa.

Cities ranking high in both desirability and affordability could then be described as „secret tips“ for city trips.

### Analyzing safety perception by gender

An observation we found interesting to look at in the given data set, was if females and males are feeling any differently regarding the safety in big cities. One could assume that women feel not as safe as men, because women might be physically weaker. Thus, we chose to look at two different cities to see if we were able to confirm this assumption. Looking at the average age of

Luisa Hahn (577736), Vincent Gurgul (573476), Mert Erdinc (615804), Jan Lorch (617032)

the participants in the study, we can assume that most of them drive through the city at night. Therefore, a good infrastructure is important to them to get through the city safely and therefore it might be a good indicator to compare if results differ when looking at cities with very different infrastructures. Since Berlin has in comparison to Rome a really good infrastructure, looking at those two cities should give us a good picture if the gender influences the safety feeling of people. The results are shown in the following visualization.

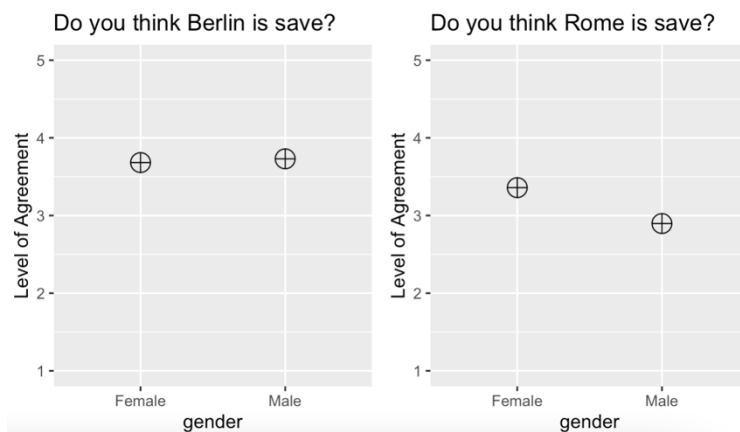


Figure 8

The results derived by the data were not as intuitive as we anticipated. In Berlin, women in the given data set feel nearly as safe as men, with both genders rather agree with the statement “Berlin is safe.” (Figure 8). But as already mentioned, that could be explained by the good infrastructure the city has. However, the results we got for Rome were surprising to us. Women in the given data set feel safer in Rome than men do (Figure 8). A possible explanation for that might be the small statistical population in the data set. Additionally, we do not know in which area in Rome the people in our data set live in. It might be that women live in safer areas than men anyway.

### Comparing Locals’ and Tourists’ perception of a given city being “too touristic”

Another thing we asked ourselves, was if locals consider their city as too touristic compared to tourists. The hypothesis we had, was that locals perceive their city as too touristic whereas tourists will not have such an impression. However, the data did not confirm our hypothesis. In Berlin as well as in Paris and Stockholm, the tourists agreed more with the statement “The city is too touristic.” than locals did. To get more information about this result we looked at the average level of agreement from locals and tourists over all cities. To avoid biased data, we only included cities where we had locals in the data set. However, the result still was not confirming our assumption.

There was barely any difference in the data set about how people felt about the city being too touristic. The mean for locals was 3.41 and the one for tourists was 3.49, so both rather agreed with the statement. A possible explanation for the data in Berlin, Paris and Stockholm might be that the tourists are sightseeing in the cities they visit and thus, are spending more time at touristic places. The locals in our data set might avoid such places and therefore get a whole different feeling for the city as tourist do.

## Team report of Group 14 — CACI WS 20/21 Special Work Performance 1 (19.11.2020)

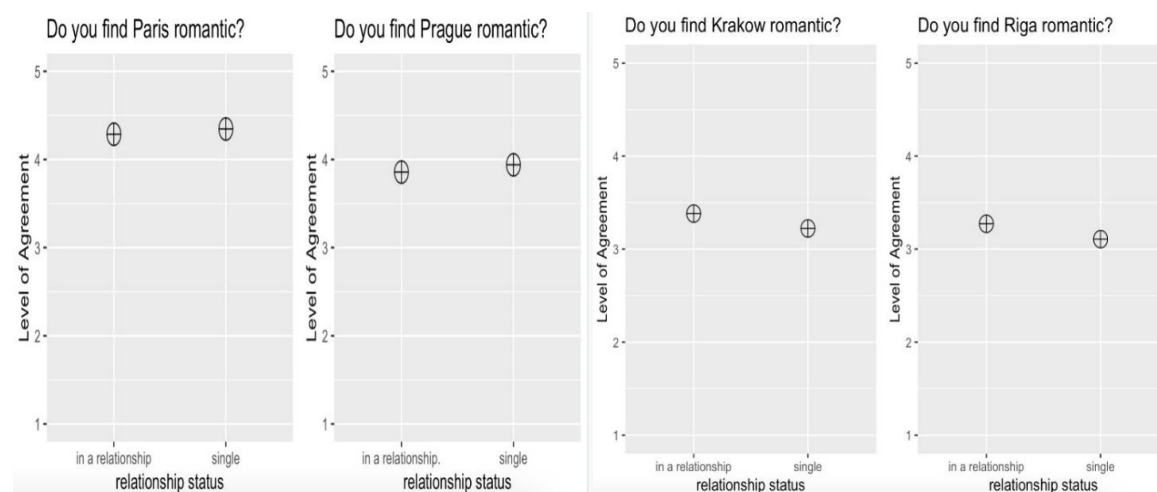
Luisa Hahn (577736), Vincent Gurgul (573476), Mert Erdinc (615804), Jan Lorch (617032)

### Perceived romantic quality comparing Prague and Paris

The underlying assumption behind the two graphs above is that couples might have found a particular city more romantic as they are spending time in the city together and might engage in more “romantic” activities in the city compared to single people, possibly causing them to perceive it as more romantic.

Our hypothesis was that both Prague and Paris are considered above average in romantic quality thus allowing for differentiation in larger orders of magnitude. While single people found Paris and Prague more romantic than the couples, there is no significant difference in their preferences as the difference is less than 5%.

Additionally, to reduce the possibility that the conclusions above resulted from the nature of the cities on the left side, on the right side, we have examined Krakow and Riga as well. Contrary to the graph on left, these cities were perceived as slightly more romantic by the couples compared to singles. With more data, the difference between how the singles and couples perceive how romantic a city is could potentially further change. The data set that we had seems to indicate that there is no significant correlation between the partnership status and perceived romantic quality of a city.



### Perceived friendliness of Istanbul, compared to English-speaker friendliness

With the data we have got from 77 people about Istanbul, we have observed that the mean perceived friendliness of the city is 3.416 with a median of 4 whereas the mean perceived English-speaker friendliness is 3.091 with a median of 4. As expected, this indicates that Turkish people living in Istanbul are friendly, but they don't speak in English with foreigners for various unknown reasons.

```
> summary(Istanbul_friendly)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
1.000  3.000  4.000  3.416  4.000  5.000   189

> summary(Istanbul_language)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
1.000  2.000  3.000  3.091  4.000  5.000   189
```

Takeaway: Istanbul is regarded as more friendly overall than it is English-speaker friendly, opening new questions about English language proficiency in Turkey's capital.