**Intro**

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 was 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.

In this operation we intended to examine the given city trip data set for a possible causal relationship between both travelling budget as stated in question 36 and travelling 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 being 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 travellers 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 |

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 travellers), that it was applied to we can make a the following reasonable assumptions.

A1: Travellers on the lowest budget, below 200€, only travel to about 30.16% as many unique visitors as the average traveller does.

A2: : Travellers on the lowest budget, below 200€, only travelled about 32.09% as often in 2015 as the average traveller 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 fort he 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“ travellers 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.