|  |  |
| --- | --- |
| **Title of your Project** | Uni4U |
| Want to study in Germany? We've got your back! | |

|  |  |
| --- | --- |
| **Short summary (4-5 sentences)** |  |
| After graduation, students in Germany can be overwhelmed by the variety of universities to choose from for their future. However, it gets even more difficult as the living conditions vary from city to city and need to be factored in when deciding where to study. Therefore, we decided to simplify this decision by coding and creating an interactive dataset, that shows the best university according to specific preferences. | |

**- The first two parts will also appear on your Digital Shaper certificate!**

|  |  |
| --- | --- |
| **Main section** |  |
| * **Introduction**: What are the ideas behind your project? for which project was a solution sought?   After a fun kickoff and getting to know our group, we started looking for problems in our daily lives as well as in the world. We quickly settled on the issue of finding the right university since we either have struggled or are struggling with it.   * **Methodology:** Which methods/tools from your learning tracks did you use to solve the problem? How exactly did you go about it?     **Feel free to include matching images!**  Then the search for datasets related to the issue started. On Kaggle we quickly found a dataset containing a ranking of the best universities in the world. After deciding which categories, we would like to keep and which to drop, we got to work and started altering the dataset by dropping all irrelevant rows and columns. For instance, we kept categories such as quality of education and faculty and ignored others like citations and publications. At this point we decided we only wanted to focus on German universities but ran into a little trouble when we tried filtering those out since not all of the universities had the city in the name or contain mutated vowels.  After a little help from our mentor Johnny, we were able to figure it out and continued working with the data. While doing so, we also looked for additional data to further extend our dataset by adding relevant information about the living conditions in the different cities.  We found two other datasets about rental apartments and population per city which we used to factor in some of the living conditions in the different cities. When we started working on this dataset, we had to figure out how we could create a column with the city names for each city without manually typing in each one. Eventually we managed that as well with little help.  We now had to figure out how to merge the data sets. After some trial and error, we again reached out to our mentor Johnny who paved the way for us. Afterwards, we summed up the information by calculating the mean, median or true and false percentages and merged the university data set with the living conditions added those to our university data set.  To help the prospective students make better decisions we created a notebook where they can evaluate our data by themselves. For example, we developed a boxplot diagram to look at rent prices for a selected city and a scatterplot to get a glimpse of correlations between two selected criteria of interest, amongst others. Additionally, we created a Google Data Studio report which allows the user to interact with our data and use our dataset without having to understand python.  Here are some examples what is possible with our notebook:   * We could answer questions like: Which university has the best quality of education while at the same time is located in a city where the total rent median is not more than 700€? Answer: University of Göttingen * We were able to create a barchart with the median rent prices (you could of course choose another factor) per city to visually compare them to each other.      * There is also the possibility to get a comparable set of boxplots with the total rent, base rent and service charge for a wished city. Here you can see the results for Hamburg:      * We also have detected a medium strong correlation between the national rank of a university and the respective median of rent prices in the respective city. So this means that the better the rank of the university the higher the median of the renting prices in the respective city.      * **Your learning journey:** How did the cooperation in the group work, were you able to apply the skills you learned, what was fun?   We had a fantastic team spirit, as more expert Python users took it upon themselves to upskill the rest of the team, writing code in a very pedagogic manner. We applied many of the useful skills taught through TechLabs such as interactions with GitHub repositories and advanced data analysis.  We feel like we emerge from this course with not only a wealth of useful skills, but also a good feeling of camaraderie   * **Results of the project:** What exactly is the solution to your problem? What are the key results?   Thanks to our project, the user is now able to choose a university based on categories such as quality of education and faculty, national rank and other important categories whilst also having information about the living conditions in a particular city such as rent, living space available Internet speed and so on.  Potential students can now make a well-rounded data driven decision before embarking on their educational journey! | |

|  |  |
| --- | --- |
| **Further details** |  |
| Team:  All Data Science Track   * Marek Lorenz * Janine Hoffmann * Katharina Hess * Fabian Seeger * Lara Jahncke * Reema Kundar   Mentor:   * Johnny Kessler * Optional: * Google Data Studio: <https://datastudio.google.com/reporting/72741ea9-79e5-4303-b4ff-b8638db1d1c5>, * GitHub: [GitHub - reemakk/group-3](https://github.com/reemakk/group-3) | |

.