Data Engineering - Slot 3: Analyze

As a result of Slot 2 you will have structured data available to you about job postings in Switzerland.

The last slot is focused on performing analysis on this data.

Analysis Step:

For the step we have to achieve two things:

1. **Define 3 Analysis Questions**: Based on the structured data available you define three questions that will form the basis for your data-backed analysis. The questions can range from focusing on the distribution of elements (e.g. job-titles, to the availability of attributes (pensum), to the distribution of attributes.

2. **Implement the 3 Defined Questions:** Use your data in order to attempt to answer the three questions you have defined. **Make sure that these three questions are answered using three different algorithms.**

   Feel free to use any of the following or similar tooling to illustrate your results:

   - Plotting: Matplotlib, seaborn, Pandas
   - Analytics and ML Rapid App Dev: https://www.gradio.app, https://streamlit.io

   The focus should be to explore the functionality for **visualisation** in order to answer your originally defined questions. **Make sure to choose suitable diagrams of visualization.**

3. **Validity:** After you have implemented the visualisation it is time to consider validity. Validity can be defined **as the extent to which a test measures what it purports to measure** (as defined by Guilford New Standards for Test Evaluation).

   What does this mean? When we run an analysis such as trying to answer the question “How many Data Engineering jobs are available in Switzerland?”, we

will go out and collect data, transform it, and visualize or quantify it. In almost all cases we have to approximate as we: - Do not have all data available - The transformation steps are not 100%.
Make sure your analysis shows validated data. Explain, how you validated it.

Deliverables:

The deliverables for this step consist of your code for the analysis and the discussion of your findings and their validity.

- Write your code in a Jupyter Notebook and explain your steps using Markdown cells in the Jupyter Notebook. The notebook shows the results of your last execution.
- Discuss your findings and their validity as a 2-3 page PDF or in Markdown cells in the notebook.

If you have questions or require support please use the following link to book appointments:

https://outlook.office365.com/book/BFHBookingsbsc_data_engineering@bfh.ch/?ae=true