

# Alireza Taghizadeh

Customer segmentation challenge for Analysts

# Required python packages

- Several common python packages are used here, e.g. pandas, scikit-learn, matplotlib, jupyterlab.
- I have used poetry to make a list of required packages.
- Jupyter notebook is used to have both code and documentations.

# Data cleaning

- Firstly, I have investigate the input data using pandas. For cleaning data, I decided to remove ID columns ("customer\_id", "credit\_account\_id").
- For missing values, I have decided to remove them (it is also possible to fill in with).
- For the column "initial\_fee\_level", there were several outliers (by looking at the data), and I have discarded them.
- For categorical columns such as gender and branch, I have used a LabelEncoder due to simplicity (I could also use a OneHotEncoder).

To have a better feeling, I have also plotted the distributions of data.

# What are the most important factors for predicting whether a customer has converted or not?

- I have considered two analysis to identify the most important factor: Principal component analysis (PCA) and k best feature selection.
- Principal Component Analysis (PCA): this is the most common approach with a strong mathematical backbone. In this approach, data in a high-dimension space is projected it into a lower-dimensional sub-space. Using PCA, we can see that almost 97% of data can be captured using the principal component 1. Looking at the direction of this component, we can see that it is almost aligned with the "initial\_fee\_level". So, this is the most important feature, which affects the converted.
- SelectKBest: It is using scikit-learn tools to select the most important feature. This method also confirms that initial\_fee\_level is the most important feature.

# Final conclusion

- The most important feature affects the customer conversion is initial\_fee\_level.
- If initial\_fee\_level is larger than 230, the costumer will be converted in more than 77% of cases.
- If initial\_fee\_level is smaller than 15, the costumer will not be converted in more than 85% of cases.