

# Task

For this task we would like you to investigate bank transaction classification. We do not expect you to spend more than a morning or an afternoon on this task.

Last year we enhanced the banking features in the FreeAgent application with a machine learning model to [automatically categorise bank transactions into accounting categories](#). For example our model might suggest an explanation of “Accommodation and Meals” if it detects a Starbucks transaction.

We have provided you with two datasets. The first, ‘bank\_transaction\_features.csv’, represents bank transactions and contains a unique transaction identifier, the transaction description, the transaction amount and the transaction type. The second, ‘bank\_transaction\_labels.csv’, contains the transaction identifier, accounting category and a flag indicating whether the transaction should be used for training or validation.

## Questions

### 1. Exploratory data analysis

- Is there any difference between the number of transactions and their types, amounts and descriptions in each category?
- How does the validation data compare with the training data?

Use appropriate tables or visualisations to support your findings when exploring the data.

### 2. Transaction classification model

- Build a model that can classify bank transactions into accounting categories by using the transaction descriptions. You may optionally include other features.
- Train your model and any preprocessing steps on the training transactions provided.
- Evaluate the performance of your model on the validation transactions provided.
- Are there any patterns in transactions that are misclassified by your model?
- Can you suggest any improvements to the dataset or your model?

Use appropriate metrics to present your findings on model performance.

Please send us any code you have written to answer the questions and a summary of your findings in the format of slides that clearly outline your approach to the task and your results. We don't expect a totally polished answer, but you should describe what you did and why.

Good luck!