**Data Science library project:**

This project aims to create custom python functions to facilitate the development of data science projects.

**Feature Engineering:**

- **OK** Function to characterize the dataset, checking for nulls, duplicated values, types of data, number of numerical and categorical variables and describing it with df.describe()

- **OK** Function to check and plot the distributions of each variable

- **OK** Function to remove outliers based on quantiles

- **OK** Function to remove outliers with an IsolationForest algorithm

- Function to remove outliers based on statistical parameters such as the distribution and its standard deviation

- **OK** Function to encode categorical variables with different methods, like OneHotEncoding and others

- **HALT** Function to normalize feature values

**Feature Selection:**

- Function to check for correlated features, plot correlation matrix

- Function to remove correlated features

- Function to remove features based on Lasso or recursive methods

**Pipeline:**

- Custom pipeline using sklearn’s pipeline that returns a DataFrame object, along with its metadata (to be decided)

**Models:**

- Function to save the model/ pipeline as a .joblib or .pickle file

- Function to train different models using the same syntax (take RandomForest and LightGBM as an example: they come from different libraries, so they use different syntaxes, and the goal here is to avoid that)

- Function to make predictions with the models using the same syntax

- Function to evaluate the performance of different models using various metrics, such as F1 score and confusion matrix

- Function to optimize the parameters of the models

- Function to evaluate the importance of each feature used in the model through SHAP values