Lab 1: Exploratory Data Analysis (2%)

In this exercise, you will be introduced to the basic concepts and notation used in data mining tasks. You will also practice exploratory data analysis (EDA) as a first step in the process of discovering information in a data set.

At the end of the exercise you should be able to:

- load a data set from a CSV file to Pandas dataframe and generate descriptive statistics for its numerical and categorical attributes,
- plot the distribution of numerical and categorical attributes in a data set with a variety of potting techniques,
- utilise EDA for getting an intuitive insight into a data set.

Task 1

Download the following dataset and example Jupyter notebook:

- loans_train.csv
- Lab 1 - Exploratory Data Analysis.ipynb

Task 2

Download the dataset: bank.csv

This dataset is part of the Bank Marketing dataset from the UCI repository. Open the dataset in Excel and make yourself familiar with the meaning of the columns (refer to the description of the dataset at the UCI repository).

Task 3

Open a new Jupyter notebook. In your own notebook, perform EDA with the bank data set by following the examples in the provided notebook “Lab 1 - Exploratory Data Analysis.ipynb”.

- Formulate a particular question that might be interesting to answer by exploring this dataset.
- Calculate statistics for the numerical and categorical attributes.
- Use at least two different plotting techniques to plot the distribution of some numerical attributes. Draw short conclusions (in a markdown cell).
- Generate two plots with the combined distribution of attributes and draw conclusions from them (in a markdown cell).

- Make sure your conclusions are related to the question you have formulated.

Submission

Save your Jupyter notebook and name *lab1.ipynb*.

Submit your notebook in the Course Tools > Assignment section by 23-Feb-2024 23:59. Late submissions will not be accepted.