

Ćwiczenia nr 1

Plik *bank\_marketing\_training* zawiera dane pochodzące z kampanii telemarketingu bezpośredniego przeprowadzonej przez jeden z portugalskich banków. Bank był zainteresowany czy klienci, z którymi się kontaktowano, zdecydowali się na założenie lokaty terminowej w banku. Plik zawiera 26 874 rekordy. Będą nas interesować następujące predyktory:

- *age* – wiek,
- *educations* – wykształcenie,
- *previous\_outcome* – poprzednia odpowiedź,
- *days\_since\_previous* – czas jaki upłynął od poprzedniego kontaktu,

oraz zmienna celu:

- *response* – odpowiedź.

1. Pobierz plik *bank\_marketing\_training*.

2. Wykonaj poniższe polecenia w R:

- Wczytaj plik.
- Dodaj zmienną ID indeksującą obserwacje.
- Zamień wartość 999 na kod braku danych NaN (ang. *not a number*).
- Przypisz kategoriom zmiennej *educations* wartości numeryczne zgodnie ze schematem:
  - illiterate* – 0,
  - basic.4y* – 4,
  - basic.6y* – 6,
  - basic.9y* – 9,
  - high.school* – 12,
  - professional.course* – 12,
  - university.degree* – 16,
  - unknown* – *Missing*.
- Wykonaj standaryzację zmiennej *age*, odejmując od niej jej średnią i dzieląc tę różnicę przez odchylenie standardowe.
- Zidentyfikuj obserwacje odstające ze względu na wartości zmiennej *age*.

3. Powtórz analizę w Pythonie.