Eksploracja danych

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