

# Analiza EDA przed modelowaniem.

Matematyka i Analiza Danych, II rok

EDA, ang. *exploratory data analysis*



Dane

# Dane

a) jakie rozszerzenie, jak wczytać

# Dane

- a) jakie rozszerzenie, jak wczytać
- b) jakie zmienne, jaki wymiar danych

# Dane

- a) jakie rozszerzenie, jak wczytać
- b) jakie zmienne, jaki wymiar danych
- c) zrozumienie co zawierają dane, jaką informację niosą poszczególne kolumny

# Dane

- a) jakie rozszerzenie, jak wczytać
- b) jakie zmienne, jaki wymiar danych
- c) zrozumienie co zawierają dane, jaką informację niosą poszczególne kolumny
- d) czy dane są pełne, ewentualne braki danych



# Problemy?

Co w przypadku “dużej” tabeli?

Analiza

# Analiza

a) ogólne statystyki

# Analiza

a) ogólne statystyki

b) analiza rozkładu jednej zmiennej (zmienne ilościowe i jakościowe)

# Analiza

- a) ogólne statystyki
- b) analiza rozkładu jednej zmiennej (zmienne ilościowe i jakościowe)
- c) korelacje zmiennych

# Analiza

- a) ogólne statystyki
- b) analiza rozkładu jednej zmiennej (zmienne ilościowe i jakościowe)
- c) korelacje zmiennych
- d) analiza dwóch lub więcej zmiennych (na podstawie korelacji zmiennych)

# Analiza

- a) ogólne statystyki
- b) analiza rozkładu jednej zmiennej (zmienne ilościowe i jakościowe)
- c) korelacje zmiennych
- d) analiza dwóch lub więcej zmiennych (na podstawie korelacji zmiennych)
- e) testy statystyczne

Model



# Model

a) podział zbioru względem targetu (stratyfikowany czy nie)

# Model

- a) podział zbioru względem targetu (stratyfikowany czy nie)
- b) jaki typ modelu wybrać

Zadanie