

# PROJEKT Z PRZEDMIOTU EKSPLORACJA DANYCH

## PIERWSZY ETAP: ZROZUMIENIE PROBLEMU + ZROZUMIENIE DANYCH

### SPEED DATING EXPERIMENT

30.05.2025 r.

## Spis treści

1. Ogólny opis zbioru .....	2
2. Określenie celu eksploracji i kryteriów sukcesu .....	2
3. Charakterystyka zbioru danych .....	2
3.1. Pochodzenie .....	2
3.2. Format .....	2
3.3. Liczba przykładów .....	2
3.4. Ilość zbiorów danych .....	2
4. Opis atrybutów .....	2
5. Wyniki eksploracyjnej analizy danych .....	3
5.1. Rozkłady wartości atrybutów .....	3
6. Wnioski .....	3
6.1. Wnioski dotyczące danych .....	3
6.2. Korelacje między wartościami atrybutów .....	3
7. Uwagi na temat jakości danych .....	4
7.1. Dane brakujące .....	4
7.2. Dane niespójne .....	4
7.3. Dane niezrozumiałe .....	4
7.4. Punkty oddalone .....	4
8. Podsumowanie .....	4

## 1. Ogólny opis zbioru

Zbiór danych zawiera informacje z eksperymentu „Speed Dating”, który miał na celu zbadanie, jakie czynniki wpływają na decyzje dotyczące wyboru partnera. Eksperyment polegał na organizacji serii szybkich randek, podczas których uczestnicy mieli okazję poznać się nawzajem w krótkim czasie. Zbiór danych zawiera informacje o uczestnikach, ich preferencjach oraz wynikach randek.

???

## 2. Określenie celu eksploracji i kryteriów sukcesu

????

## 3. Charakterystyka zbioru danych

### 3.1. Pochodzenie

<https://www.kaggle.com/datasets/annavictoria/speed-dating-experiment>

### 3.2. Format

.csv

### 3.3. Liczba przykładów

8378 rekordów

### 3.4. Ilość zbiorów danych

1

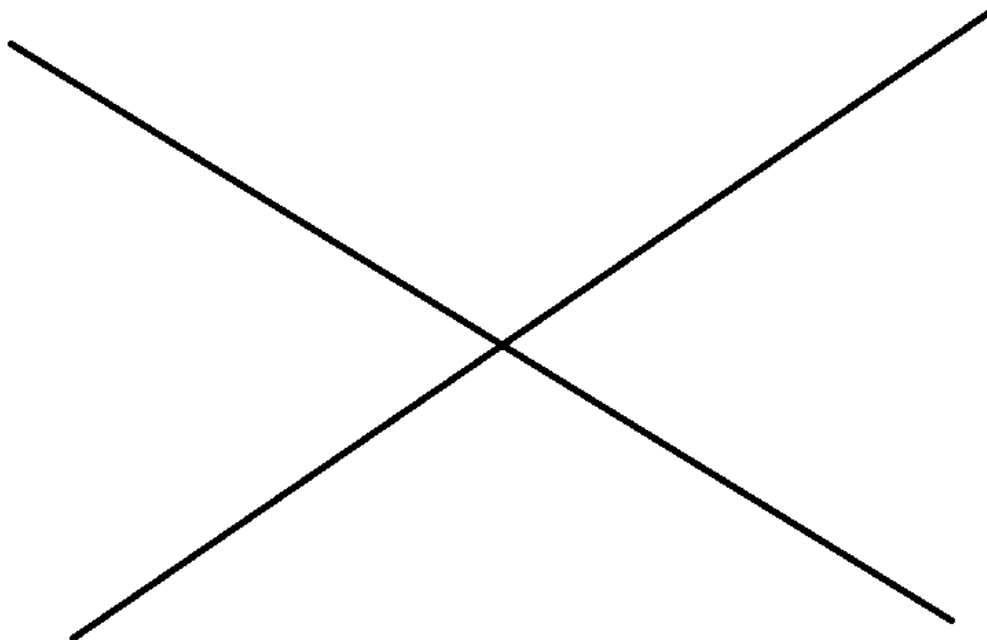
## 4. Opis atrybutów

abcd

Nazwa	Typ	Znaczenie
iid	Numeryczny	unikalny numer uczestnika
id	Numeryczny	numer uczestnika w ramach fali
gender	Numeryczny	płeć uczestnika (1 - mężczyzna, 0 - kobieta)
idg	Numeryczny	numer badanego w obrębie płci, grupy
condtn	Numeryczny	1=limited choice, 2=extensive choice
wave	Numeryczny	numer fali, w której uczestnik brał udział; dookreśla inne informacje
round	Numeryczny	numer rundy, w której uczestnik brał udział
position	Numeryczny	numer stacji, na której partnerzy poznali się
positin1	Numeryczny	numer stacji, na której zachęto zabawę
order	Numeryczny	ilość randek zadanej nocy

## 5. Wyniki eksploracyjnej analizy danych

### 5.1. Rozkłady wartości atrybutów



Rysunek 1: Histogram dla atrybutu <ATRYBUT>.

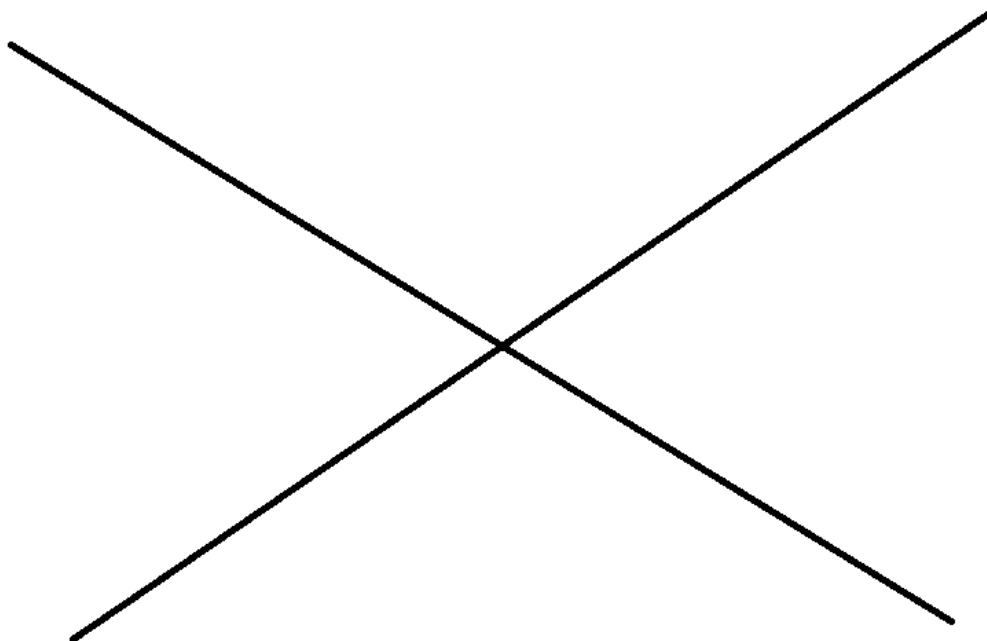
## 6. Wnioski

### 6.1. Wnioski dotyczące danych

abcd

### 6.2. Korelacje między wartościami atrybutów

abcd



Rysunek 2: Macierz korelacji.

## 7. Uwagi na temat jakości danych

abcd

### 7.1. Dane brakujące

xdxdxd

### 7.2. Dane niespójne

abcdabcd

### 7.3. Dane niezrozumiałe

abcd abcd

### 7.4. Punkty oddalone

abcd

## 8. Podsumowanie

xdxdxd