

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	8
5.1. Rozkłady wartości atrybutów	8
6. Wnioski	8
6.1. Wnioski dotyczące danych	8
6.2. Korelacje między wartościami atrybutów	8
7. Uwagi na temat jakości danych	9
7.1. Dane brakujące	9
7.2. Dane niespójne	9
7.3. Dane niezrozumiałe	9
7.4. Punkty oddalone	9
8. Podsumowanie	9

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	numer uczestnika, unikalny w całym zbiorze
id	Numeryczny	numer uczestnika, unikalny tylko w obrębie danego spotkania
gender	Nominalny	płeć uczestnika (1 - mężczyzna, 0 - kobieta)
idg	Numeryczny	numer identyfikuje, którą z kolei osobą danej płci jest zadany uczestnik np. 1 - pierwsza kobieta, 2 - druga kobieta, 1 - pierwszy mężczyzna, 2 - drugi mężczyzna
condtn	Numeryczny	warunki przeprowadzenia eksperymentu: 1=limited choice, 2=extensive choice
		cokolwiek by to znaczyło
wave	Numeryczny	numer wydarzenia, w której uczestnik brał udział; dookreśla inne informacje
round	Numeryczny	ilość poznanych osób przeciwnej płci w danym wydarzeniu; tożsame z ilością rund (a.k.a. rozmów, randek), które odbyła dana osoba z osobą innej płci
position	Numeryczny	numer stanowiska (a.k.a stolika) przy którym siedział uczestnik, gdy rozmawiał z osobą zdefiniowaną w kolumnie 'partner'
positin1	Numeryczny	numer stanowiska (a.k.a stolika) przy którym uczestnik siedział na początku wydarzenia; tam odbył pierwszą rozmowę
order	Numeryczny	określa, która z kolei rozmowa (a.k.a. randka) miała miejsce w danym wydarzeniu dla tej osoby; 1 - moja pierwsza rozmowa, 2 - moja druga rozmowa itd.

partner	Numeryczny	numer uczestnika danego spotkania (kolumna id), z którym dana osoba rozmawiała w danej rundzie
pid	Numeryczny	unikalny numer osoby (kolumna iid), z którą rozmawiała zadana osoba
match	Nominalny	1 - jeśli dana osoba wybrała daną osobę jako potencjalnego partnera, 0 - przypadek przeciwny
int_corr	Numeryczny	korelacja między ocenami zainteresowań (np. bardzo lubię przejażdżki rowerem, nienawidzę grać w siatkówkę) uczestnika i partnera w Czasie 1 (tj. przed spotkaniem) wartości z przedziału $[-1, 1]$ gdzie -1 oznacza pełną niezgodność, 1 pełną zgodność, 0 brak korelacji
samerace	Nominalny	1 - jeśli uczestnik i partner są tej samej rasy, 0 - jeśli nie są
age_o	Numeryczny	wiek partnera
race_o	Nominalny	rasa partnera (w zbiorze wartości 1-4) Black/African American=1 European/Caucasian-American=2 Latino/Hispanic American=3 Asian/Pacific Islander/Asian-American=4 Native American=5 Other=6
pf_o_att pf_o_sin pf_o_int pf_o_fun pf_o_amb pf_o_sha	Numeryczny	preferencje osoby mierzone w Czasie 1 (tj. przed spotkaniem), uwzględnia każdą z mierzonych 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/hobby
dec_o	Nominalny	1 - partner jest zainteresowany daną osobą, 0 - partner nie jest zainteresowany daną osobą
attr_o sinc_o intel_o fun_o amb_o shar_o	Numeryczny	preferencje partnera mierzone w Czasie 1 (tj. przed spotkaniem), uwzględnia każdą z mierzonych 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/hobby
prob_o	???	Co to jest??????
met_o	???	Co to jest??????
age	Numeryczny	Wiek uczestnika; z ankiety rejestracyjnej na wydarzenie
field	Nominalne	Nazwa kierunku studiów uczestnika w języku angielskim; z ankiety rejestracyjnej na wydarzenie

field_cd	Nominalne	<p>Numeryczna reprezentacja atrybutu 'field', gdzie:</p> <p>1= Law 2= Math 3= Social Science, Psychologist 4= Medical Science, Pharmaceuticals, and Bio Tech 5= Engineering 6= English/Creative Writing/ Journalism 7= History/Religion/Philosophy 8= Business/Econ/Finance 9= Education, Academia 10= Biological Sciences/Chemistry/Physics 11= Social Work 12= Undergrad/undecided 13=Political Science/International Affairs 14=Film 15=Fine Arts/Arts Administration 16=Languages 17=Architecture 18=Other</p>
undergrd	Nominalne	szkoła, w której uzyskano tytuł licencjata; z ankiety rejestracyjnej na wydarzenie
mn_sat	Numeryczne	średni wynik z egzaminu SAT ze szkoły, którą ukończyła dana osoba; z ankiety rejestracyjnej na wydarzenie; proxy na określenie inteligencji
tuition	Numeryczne	czesne za szkołę, którą ukończyła dana osoba; z ankiety rejestracyjnej na wydarzenie; tylko dla licencjatów
race	Nominalne	<p>Rasa uczestnika, gdzie:</p> <p>Black/African American=1 European/Caucasian-American=2 Latino/Hispanic American=3 Asian/Pacific Islander/Asian-American=4 Native American=5 Other=6</p>
imprace	Numeryczne	Jak ważne jest dla Ciebie (w skali od 1 do 10), aby osoba, z którą się umawiasz, była tego samego pochodzenia pochodzenia rasowego/etnicznego?
imprelig	Numeryczne	Jak ważne jest dla Ciebie (w skali od 1 do 10), aby osoba, z którą się umawiasz, była tego samego pochodzenia religijnego?
from	Nominalne	pochodzenie uczestnika
zipcode	Nominalne	kod pocztowy miejsca pochodzenia uczestnika
income	Numeryczne	średni dochód gospodarstwa domowego w miejscu pochodzenia uczestnika; z Census Bureau; brak danych = pochodzenie zagraniczne lub niepodany kod pocztowy
goal	Nominalne	<p>główny cel uczestnika wzięcia udziału w wydarzeniu, gdzie:</p> <p>Seemed like a fun night out=1 To meet new people=2 To get a date=3 Looking for a serious relationship=4 To say I did it=5 Other=6</p>

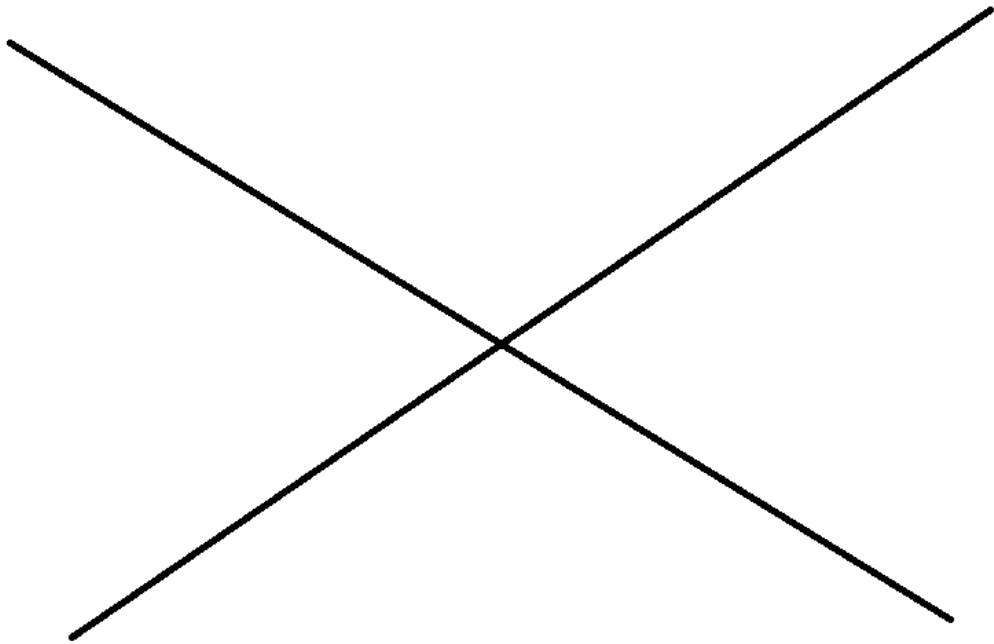
date	Nominalne	<p>Jak często chodzisz na randki? Gdzie:</p> <p>Several times a week=1</p> <p>Twice a week=2</p> <p>Once a week=3</p> <p>Twice a month=4</p> <p>Once a month=5</p> <p>Several times a year=6</p> <p>Almost never=7</p>
go out	Nominalne	<p>Jak często wychodzisz na imprezy? Gdzie:</p> <p>Several times a week=1</p> <p>Twice a week=2</p> <p>Once a week=3</p> <p>Twice a month=4</p> <p>Once a month=5</p> <p>Several times a year=6</p> <p>Almost never=7</p>
career	Nominalne	Jaką ścieżkę kariery chciałbyś obrać?
career_c	Nominalne	<p>Zakodowana odpowiedź na pytanie o ścieżkę kariery, gdzie:</p> <p>1= Lawyer</p> <p>2= Academic/Research</p> <p>3= Psychologist</p> <p>4= Doctor/Medicine</p> <p>5=Engineer</p> <p>6= Creative Arts/Entertainment</p> <p>7= Banking/Consulting/Finance/Marketing/Business/CEO/Entrepreneur/Admin</p> <p>8= Real Estate</p> <p>9= International/Humanitarian Affairs</p> <p>10= Undecided</p> <p>11=Social Work</p> <p>12=Speech Pathology</p> <p>13=Politics</p> <p>14=Pro sports/Athletics</p> <p>15=Other</p> <p>16=Journalism</p> <p>17=Architecture</p>

sports tvsports exercise dining museums art hiking gaming clubbing reading tv theater movies concerts music shopping yoga	Numeryczne	Odpowiedź na pytanie: Jak bardzo jesteś zainteresowany {nazwa aktywności korespondująca z nazwą atrybutu}, w skali od 1 do 10? Mniej oczywiste atrybuty: tvsports: oglądanie sportów yoga: joga lub medytacja
exphappy	Numeryczne	Ogólnie rzecz biorąc, w skali od 1 do 10, jak szczęśliwy spodziewasz się być z ludźmi poznanymi podczas speed-datingu?
expnum	Numeryczne	Jak myślisz, ile z 20 osób, które spotkasz, będzie zainteresowanych randką z Tobą?

attr1_1 sinc1_1 intel1_1 fun1_1 amb1_1 shar1_1 attr4_1 sinc4_1 intel4_1 fun4_1 amb4_1 shar4_1 attr2_1 sinc2_1 intel2_1 fun2_1 amb2_1 shar2_1 attr3_1 sinc3_1 fun3_1 intel3_1 amb3_1 attr5_1 sinc5_1 intel5_1 fun5_1 amb5_1	Numeryczne	<p>Atrybuty opisujące każdą z mierzonych 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby w różnych kontekstach. Kontekst definiuje pierwsza cyfra.</p> <p>Dla pierwszej cyfry równej:</p> <p>1: tego poszukuję u osoby przeciwnej płci wydarzenia 6-9: ocena w skali 1 - 10, gdzie 1 = bardzo nieistotne, 10 = bardzo istotne, pozostałe wydarzenia: 100 punktów do rozdysponowania, więcej = bardziej istotne</p> <p>2: Jak myślisz, czego płęć przeciwna szuka na randce? wydarzenia 6-9: ocena w skali 1 - 10, gdzie 1 = bardzo nieistotne, 10 = bardzo istotne, pozostałe wydarzenia: 100 punktów do rozdysponowania, więcej = bardziej istotne</p> <p>3: Jak myślisz, jak się mierzysz? Oceń swoją opinię na temat własnych atrybutów w skali od 1 do 10 (bądź szczery!) Ocena tylko w skali 1-10</p> <p>4: tego poszukuję inne osoby (własna opinia nt. preferencji innych) wydarzenia 6-9: ocena w skali 1 - 10, gdzie 1 = bardzo nieistotne, 10 = bardzo istotne, wydarzenia 10-21: 100 punktów do rozdysponowania, więcej = bardziej istotne UWAGA! Nie mierzono tych parametrów dla wydarzeń 1-5!</p> <p>5: Jak myślisz, jak postrzegają cię inni? Ocena tylko w skali 1-10, gdzie 1 - okropnie, 10 - świetnie</p>
dec	Nominalne	Ocena chęci ponownego spotkania się uczestnika badania z osobą, z którą właśnie zakończył rozmowę. Gdzie: 1 - tak, 0 - nie
attr sinc intel fun amb shar	Numeryczne	Ocena każdej z 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby w skali od 1 do 10, gdzie 1 - okropnie, 10 - świetnie
like	Numeryczne	Ogólnie rzecz biorąc, jak bardzo lubisz tę osobę? Gdzie: 1=wcale nie lubię, 10=bardzo ją lubię
prob	Numeryczne	Jak bardzo prawdopodobne jest, że ta osoba powie ci „tak”? Gdzie: 1=nieprawdopodobne, 10=niezwykle prawdopodobne
met	Nominalne	Czy znałeś wcześniej tę osobę? Gdzie: 1=tak, 2=nie
match_es	Numeryczne	Ile dopasowań szacujesz, że uzyskasz (dopasowanie ma miejsce, gdy ty i twój partner zdecydujecie „Tak”)

5. Wyniki eksploracyjnej analizy danych

5.1. Rozkłady wartości atrybutów



Rysunek 1: Historgram dla atrybutu <ATRYBUT>.

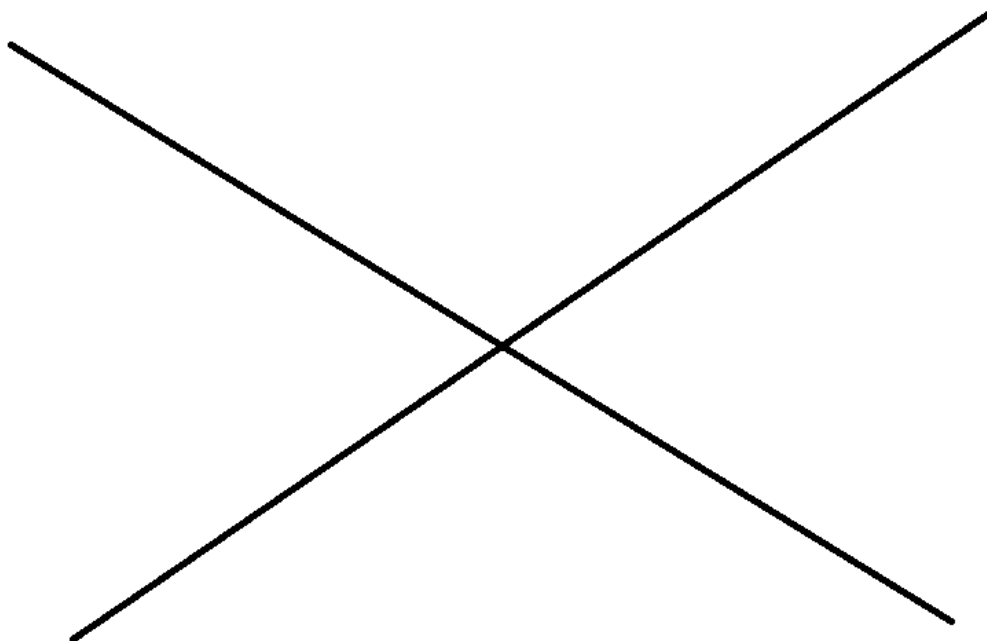
6. Wnioski

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