

PROJEKT Z PRZEDMIOTU EKSPLOACJA 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	11
5.1. Rozkłady wartości atrybutów	11
6. Wnioski	11
6.1. Wnioski dotyczące danych	11
6.2. Korelacje między wartościami atrybutów	11
7. Uwagi na temat jakości danych	12
7.1. Dane brakujące	12
7.2. Dane niespójne	12
7.3. Dane niezrozumiałe	12
7.4. Punkty oddalone	12
8. Podsumowanie	12

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 badanych w różnych momentach oraz wynikach randek.

Celem uproszczenia opisu atrybutów, niektóre z nich mają w opisie pytania. Atrybut taki odpowiada na to pytanie.

W raporcie używa się słów i zwrotów:

- wydarzenie - jedno z 24 spotkań speed datingu
- runda / randka / rozmowa - spotkanie dwóch osób i rozmowa między nimi, tzw. randka
- uczestnik - osoba biorąca udział w wydarzeniu, uczestnik rozmowy; kobieta lub mężczyzna
- 6. cech - sześć cech, które były badane w eksperymencie: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/hobby
- stanowisko / stolik - rozróżnialne między sobą miejsca, w którym znajdował się uczestnik podczas rozmowy z osobą przeciwnej płci
- Czas 1 - moment przed wydarzeniem, w którym uczestnicy wypełniali ankiety dotyczące ich preferencji i oczekiwań

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

Opis atrybutów przedstawia poniższa tabela. Warto nadmienić, że przez zbieranie danych o preferencjach nawet do 3-4 tygodni po zakończeniu wydarzenia i opublikowaniu rezultatów szybkich randek nie dla każdego zestawu parametrów jest dostępna pełna informacja.

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
wave	Numeryczny	numer wydarzenia (które z kolei), w którym 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), gdzie: 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, szczerść, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby. Metoda oceny (domniemana po przeglądzie danych): podział 100 punktów, więcej = ważniejsze
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, szczerść, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby Metoda oceny (domniemana po przeglądzie danych): skala 1-10, gdzie 1 - najmniej istotne, 10 - najbardziej istotne

prob_o	???	Niemżliwe do określenia, niewskazane w opisie zbioru. Prawdopodobnie atrybut oceniany w skali 1-10, gdzie 1 - najmniej istotne, 10 - najbardziej istotne
met_o	???	Niemżliwe do określenia, niewskazane w opisie zbioru.
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	Numeryczna reprezentacja atrybutu 'field', gdzie: 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
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	Rasa uczestnika, gdzie: Black/African American=1 European/Caucasian-American=2 Latino/Hispanic American=3 Asian/Pacific Islander/Asian-American=4 Native American=5 Other=6
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	pohodzenie 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	główny cel uczestnika wzięcia udziału w wydarzeniu, gdzie: 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
date	Nominalne	Jak często chodzisz na randki? Gdzie: Several times a week=1 Twice a week=2 Once a week=3 Twice a month=4 Once a month=5 Several times a year=6 Almost never=7
go out	Nominalne	Jak często wychodzisz na imprezy? Gdzie: Several times a week=1 Twice a week=2 Once a week=3 Twice a month=4 Once a month=5 Several times a year=6 Almost never=7
career	Nominalne	Jaką ścieżkę kariery chciałbyś obrać?

career_c	Nominalne	<p>Zakodowana odpowiedź na pytanie o ścieżkę kariery, gdzie:</p> <p>1= Lawyer 2= Academic/Research 3= Psychologist 4= Doctor/Medicine 5=Engineer 6= Creative Arts/Entertainment 7= Banking/Consulting/Finance/Marketing/Business/CEO/Entrepreneur/Admin 8= Real Estate 9= International/Humanitarian Affairs 10= Undecided 11=Social Work 12=Speech Pathology 13=Politics 14=Pro sports/Athletics 15=Other 16=Journalism 17=Architecture</p>
sports tvsports exercise dining museums art hiking gaming clubbing reading tv theater movies concerts music shopping yoga	Numeryczne	<p>Odpowiedź na pytanie: Jak bardzo jesteś zainteresowany {nazwa aktywności korespondująca z nazwą atrybutu}, w skali od 1 do 10?</p> <p>Mniej oczywiste atrybuty: tvsports: oglądanie sportów yoga: joga lub medytacja</p>
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 numer 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 numer 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 szczerzy!) Ocena tylko w skali 1-10</p> <p>4: tego poszukuję inne osoby (własna opinia nt. preferencji innych) wydarzenia numer 6-9: ocena w skali 1 - 10, gdzie 1 = bardzo nieistotne, 10 = bardzo istotne, wydarzenia numer 10-21: 100 punktów do rozdysponowania, więcej = bardziej istotne UWAGA! Nie mierzono tych parametrów dla wydarzeń numer 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. Dotyczy osoby, z którą właśnie zakończono rozmowę.
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”)?

attr1_s sinc1_s intel1_s fun1_s amb1_s shar1_s	Numeryczne	Określ ważność każdej z 6. cech: atrakcyjność fizyczna, szczerść, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby w skali od 1 do 10, gdzie 1 - wcale nieważne, 10 - bardzo ważne. Ankieta przeprowadzona w połowie wydarzenia.
attr3_s sinc3_s intel3_s fun3_s amb3_s	Numeryczne	Oceń swoją opinię na temat własnych atrybutów w skali od 1 do 10 (1 = okropne, 10 = świetne). Bądź szczerzy! Ankieta przeprowadzona w połowie wydarzenia.
satis_2	Numeryczne	Ogólnie rzecz biorąc, jak bardzo jesteś zadowolony z ludzi, których spotkałeś? Gdzie 1=zupełnie niezadowolony, 10=bardzo zadowolony Odpowiedź uzyskana dzień po wydarzeniu.
length	Nominalne	Ocena czasu trwania rundy (4 min), gdzie: Too little=1 Too much=2 Just Right=3 Odpowiedź uzyskana dzień po wydarzeniu.
numdat_2	Nominalne	Oceń liczbę „szybkich randek”, które odbyłeś, gdzie: Too little=1 Too much=2 Just Right=3 Odpowiedź uzyskana dzień po wydarzeniu.
attr7_2 sinc7_2 intel7_2 fun7_2 amb7_2 shar7_2	Numeryczne	Ponowna ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerść, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby. Metoda oceny: podział 100 punktów, więcej = ważniejsze Odpowiedź uzyskana dzień po wydarzeniu.
attr1_2 sinc1_2 intel1_2 fun1_2 amb1_2 shar1_2	Numeryczne	Ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerść, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby obrazująca to, czego dany uczestnik poszukiwał u partnera przeciwnej płci. Metoda oceny: Wydarzenia 6-9: ocena w skali 1 - 10, gdzie 1 = bardzo nieistotne, 10 = bardzo istotne, Pozostałe wydarzenia: 100 punktów do rozdysponowania, więcej = bardziej istotne Odpowiedź uzyskana dzień po wydarzeniu.

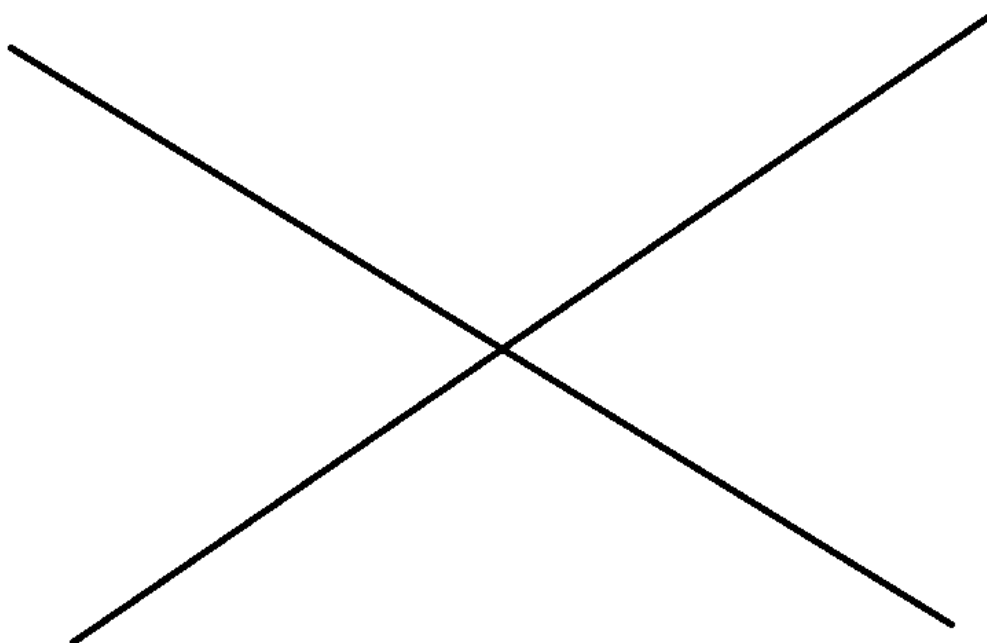
attr4_2 sinc4_2 intel4_2 fun4_2 amb4_2 shar4_2	Numeryczne	<p>Ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby obrazująca to, co jest najbardziej pożądaną cechą u partnera przeciwnej płci (własna opinia nt. preferencji innych).</p> <p>Metoda oceny: 100 punktów do rozdysponowania, więcej = bardziej istotne</p> <p>Odpowiedź uzyskana dzień po wydarzeniu.</p>
attr2_2 sinc2_2 intel2_2 fun2_2 amb2_2 shar2_2	Numeryczne	<p>Jak myślisz, czego szuka płć przeciwna na randce? (własna opinia nt. preferencji innych osób, ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby).</p> <p>Metoda oceny: 100 punktów do rozdysponowania, więcej = bardziej istotne</p> <p>Odpowiedź uzyskana dzień po wydarzeniu.</p>
attr3_2 sinc3_2 fun3_2 intel3_2 amb3_2	Numeryczne	<p>Jak myślisz, jak się oceniasz? Oceń swoją opinię na temat własnych atrybutów w skali od 1 do 10, gdzie 1 = okropny i 10 = świetny. Bądź szczery!</p> <p>Odpowiedź uzyskana dzień po wydarzeniu.</p>
attr5_2 sinc5_2 intel5_2 fun5_2 amb5_2	Numeryczne	<p>Jak myślisz, jak postrzegają Cię inni? Oceń siebie tak, jak myślisz, że inni oceniliby Cię w każdym z poniższych atrybutów, w skali od 1 do 10 (1 = okropny, 10 = świetny)</p> <p>Odpowiedź uzyskana dzień po wydarzeniu.</p>
you_call	Numeryczne	<p>Z iloma osobami, z którymi miałeś dopasowanie, skontaktowałeś się, aby umówić się na randkę?</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>
them_cal	Numeryczne	<p>Ile osób, z którymi miałeś dopasowanie, skontaktowało się z Tobą?</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>
date_3	Nominalne	<p>Czy byłeś na randce z kimś, z kim miałeś dopasowanie? Gdzie: 1=tak, 2=nie</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>
numdat_3	Numeryczne	<p>{JEŚLI BYŁEŚ NA CO NAJMNIJ JEDNEJ RANDCE}</p> <p>Z iloma partnerami byłeś do tej pory na randce?</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>

num_in_3	Numeryczne	<p>{JEŚLI BYŁEŚ NA CO NAJMNIEJ JEDNEJ RANDCE}</p> <p>Jeśli tak, to z iloma?</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p> <p><i>Definicja atrybutu niezrozumiała (również po angielsku).</i></p>
attr1_3 sinc1_3 intel1_3 fun1_3 amb1_3 shar1_3	Numeryczne	<p>Ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby obrazująca to, czego dany uczestnik poszukiwał u partnera przeciwnej płci.</p> <p>Metoda oceny: 100 punktów do rozdysponowania, więcej = bardziej istotne</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>
attr7_3 sinc7_3 intel7_3 fun7_3 amb7_3 shar7_3	Numeryczne	<p>Ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby obrazująca to, co ma w rzeczywistości największe znaczenie dla uczestnika.</p> <p>Metoda oceny: podział 100 punktów, więcej = ważniejsze</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>
attr4_3 sinc4_3 intel4_3 fun4_3 amb4_3 shar4_3	Numeryczne	<p>Ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby obrazująca to, co jest najbardziej pożądaną cechą u partnera przeciwnej płci (własna opinia nt. preferencji innych).</p> <p>Metoda oceny: skala 1-10, gdzie 1 - bardzo nieistotne, 10 - bardzo istotne</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>
attr2_3 sinc2_3 intel2_3 fun2_3 amb2_3 shar2_3	Numeryczne	<p>Jak myślisz, czego szuka płeć przeciwna na randce? (własna opinia nt. preferencji innych osób, ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainresowania/hobby).</p> <p>Metoda oceny: skala 1-10, gdzie 1 - bardzo nieistotne, 10 - bardzo istotne</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>
attr3_3 sinc3_3 intel3_3 fun3_3 amb3_3	Numeryczne	<p>Jak myślisz, jak się oceniasz? Oceń swoją opinię na temat własnych atrybutów w skali od 1 do 10, gdzie 1 = okropny i 10 = świetny. Bądź szczery!</p> <p>Metoda oceny: skala 1-10</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>

attr5_3 sinc5_3 intel5_3 fun5_3 amb5_3	Numeryczne	Jak myślisz, jak postrzegają Cię inni? Oceń siebie tak, jak myślisz, że inni oceniliby Cię w każdym z poniższych atrybutów. Metoda oceny: skala 1-10, gdzie 1 - okropnie, 10 - świetnie. Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.
--	------------	--

5. Wyniki eksploracyjnej analizy danych

5.1. Rozkłady wartości atrybutów



Rysunek 1: Histogram dla atrybutu <ATRYBUT>.

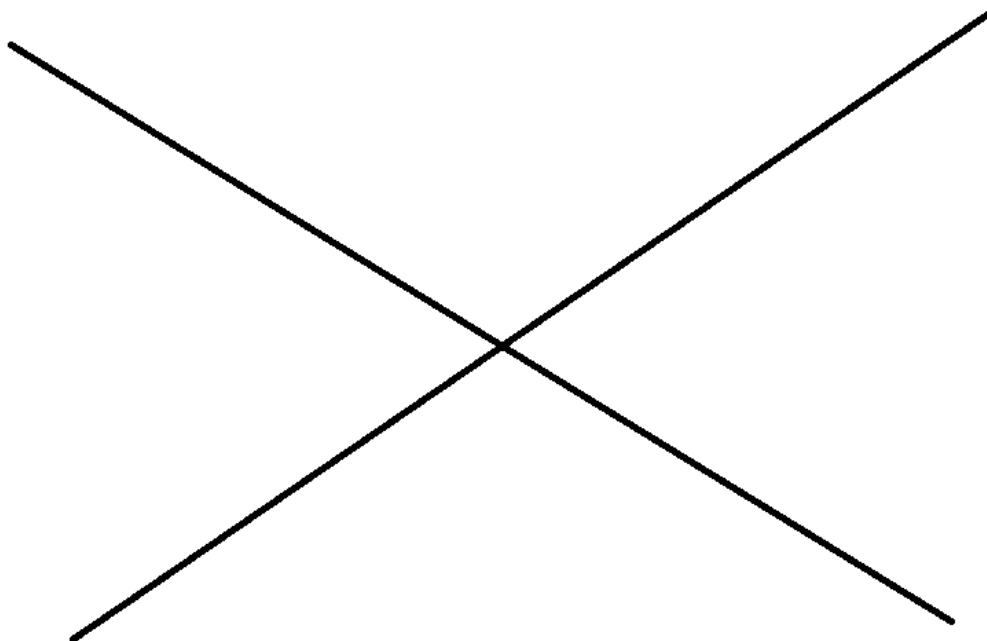
6. Wnioski

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