

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	3
5. Wyniki eksploracyjnej analizy danych	11
5.1. Rozkłady wartości atrybutów (histogramy)	11
5.2. Punkty oddalone i rozkłady wartości atrybutów (wykresy pudełkowe)	46
6. Wnioski	73
6.1. Wnioski dotyczące danych	73
6.2. Korelacje między wartościami atrybutów	74
7. Uwagi na temat jakości danych	79
7.1. Dane brakujące	79
7.2. Dane niespójne	79
7.3. Dane niezrozumiałe	79
7.4. Punkty oddalone	79
8. Podsumowanie i wnioski ogólne:	79

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 21 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

Celem eksploracji danych ze zbioru „Speed Dating Experiment” jest znalezienie odpowiedzi na pytania:

- Czy ludzie potrafią dokładnie przewidzieć swoją postrzeganą wartość na rynku randkowym?
- Sprawdzenie, jaki atrybut najmocniej wpływa na dobór partnera przeciwnej płci.

Kryteria sukcesu, które zostaną przyjęte w celu oceny skuteczności eksploracji danych, obejmują:

- wysoka korelacja ($\geq 0,6$) między przewidywaną a rzeczywistą wartością uczestników na rynku randkowym
- zidentyfikowanie cech, które mają największy wpływ na postrzeganą wartość uczestników
- przeprowadzenie analizy istotności atrybutów ze wskazaniem najistotniejszego

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, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/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, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/hobby Metoda oceny (domniemana po przeglądzie danych): skala 1-10, gdzie 1 - najmniej istotne, 10 - najbardziej istotne
prob_o	???	Niemożliwe do określenia, niewskazane w opisie zbioru. Prawdopodobnie atrybut oceniany w skali 1-10, gdzie 1 - najmniej istotne, 10 - najbardziej istotne
met_o	???	Niemożliwe 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	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	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 zainteresowania/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, pozostałewydarzenia: 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, pozostałewydarzenia: 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 poszukują 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 zainteresowania/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, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/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 szczery! 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, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/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, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/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.

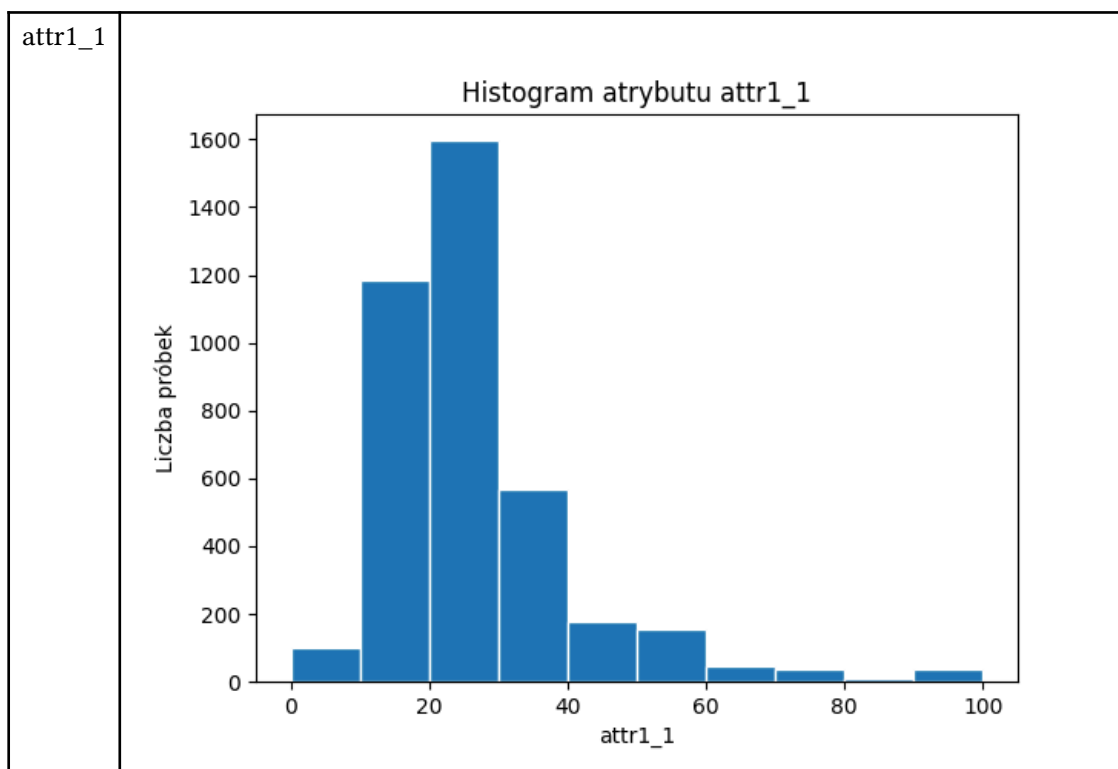
attr4_2 sinc4_2 intel4_2 fun4_2 amb4_2 shar4_2	Numeryczne	<p>Ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerść, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/hobby obrazująca to, co jest najbardziej pożądaną cechą u partnera przeciwnej płci (własna opinia nt. 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, szczerść, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/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 szczerzy!</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>

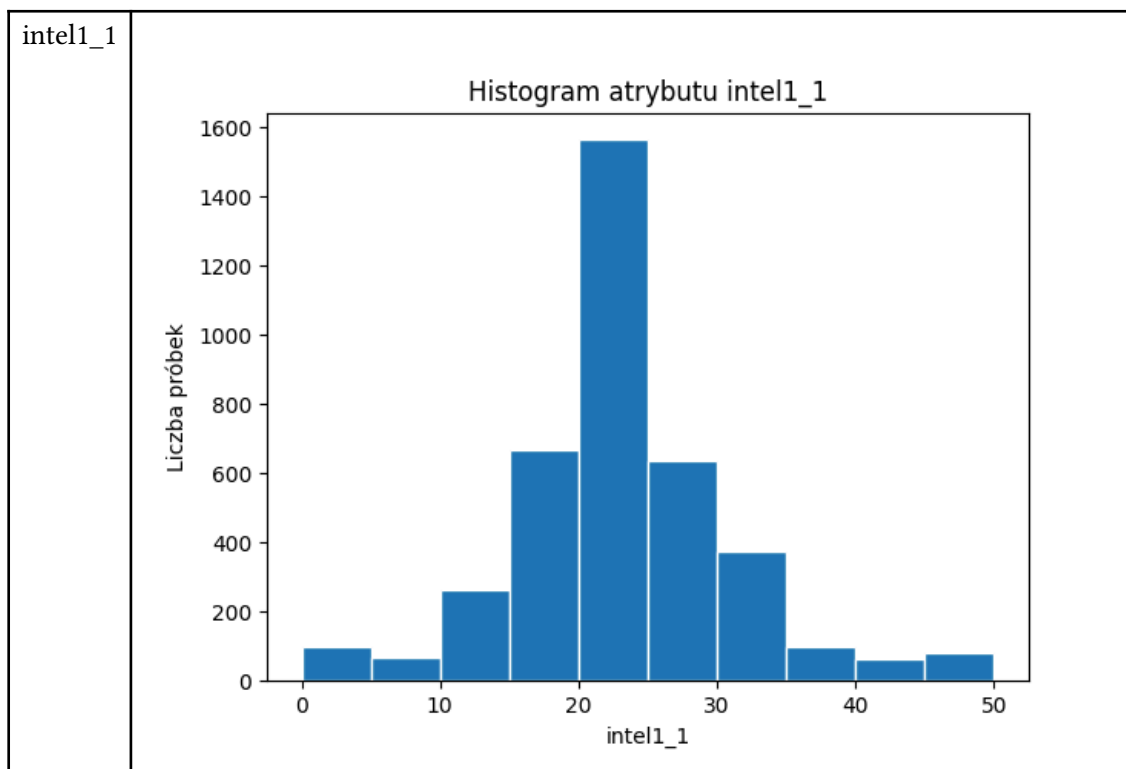
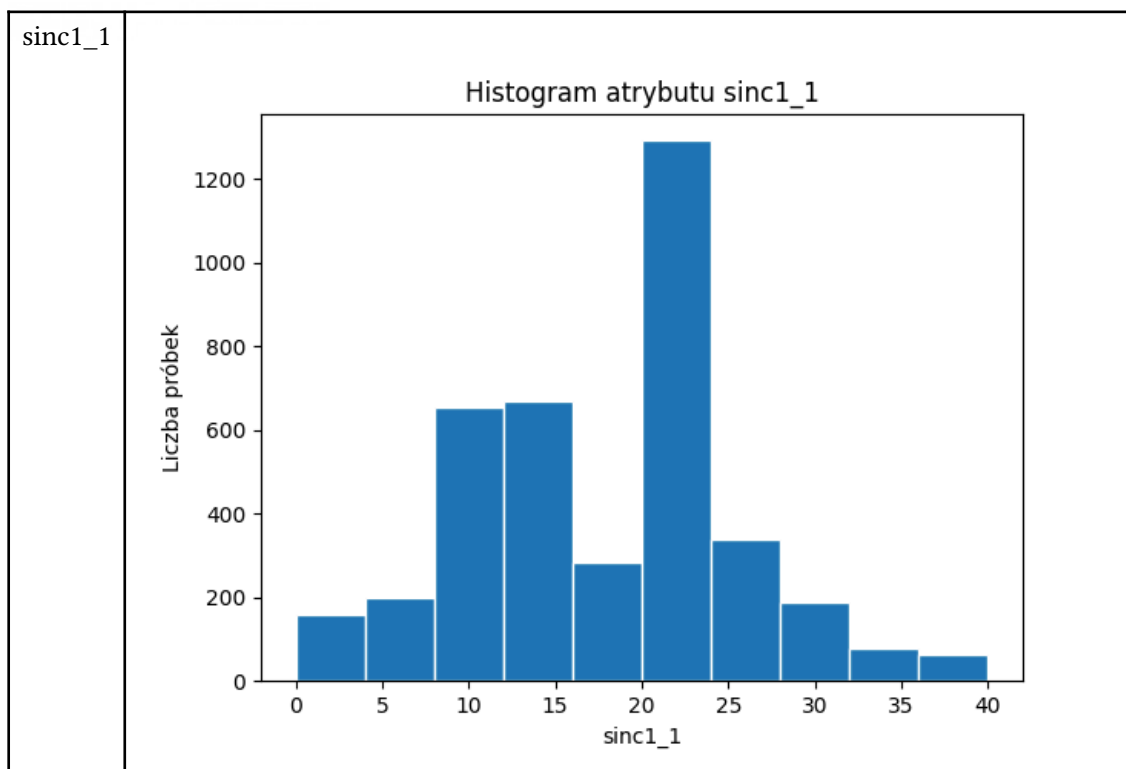
num_in_3	Numeryczne	<p>{JEŚLI BYŁEŚ NA CO NAJMNIEJ JEDNEJ RANDCE}</p> <p>Jeśli tak, to z iloma?</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p> <p><i>Definicja atrybutu niezrozumiała (również po angielsku).</i></p>
attr1_3 sinc1_3 intel1_3 fun1_3 amb1_3 shar1_3	Numeryczne	<p>Ocena istotności każdej z 6. cech: atrakcyjność fizyczna, szczerość, inteligencja, poczucie humoru (fun), ambicja, wspólne zainteresowania/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 zainteresowania/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 zainteresowania/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 zainteresowania/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. Bądź szczery!</p> <p>Metoda oceny: skala 1-10, gdzie 1 = okropny, 10 = świetny</p> <p>Odpowiedź uzyskana 3-4 tygodnie po wydarzeniu.</p>

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

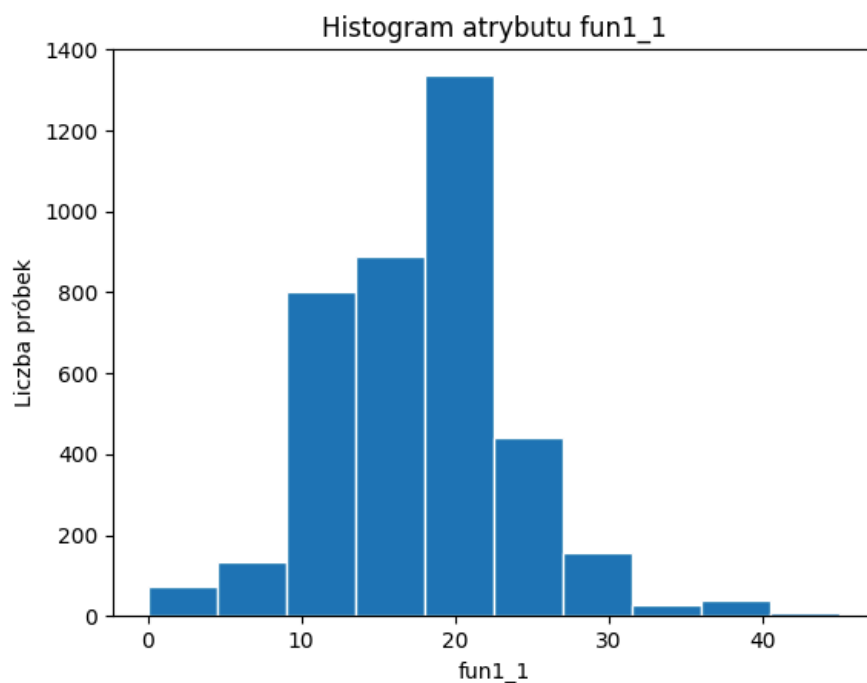
5. Wyniki eksploracyjnej analizy danych

5.1. Rozkłady wartości atrybutów (histogramy)

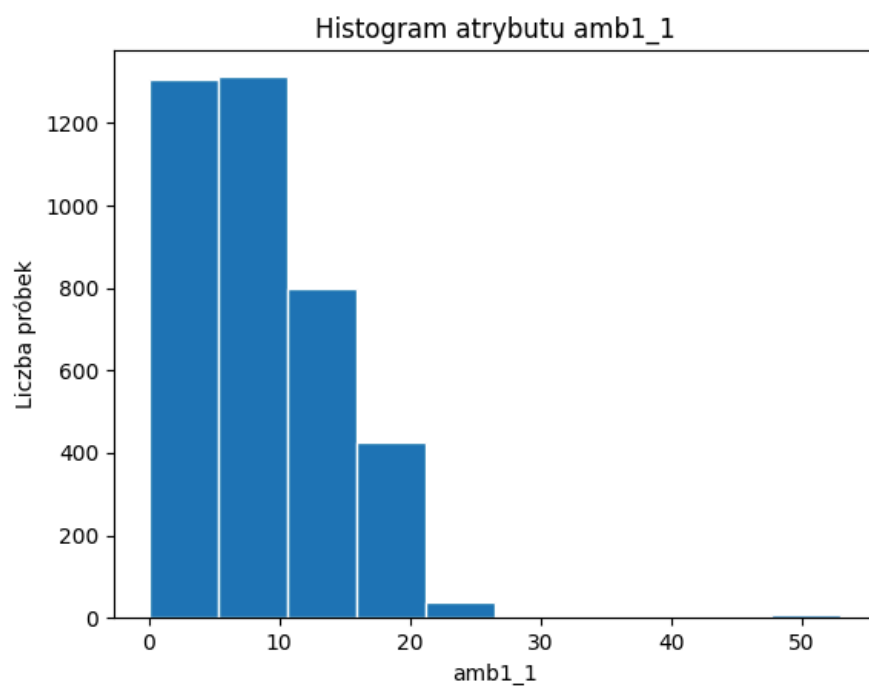




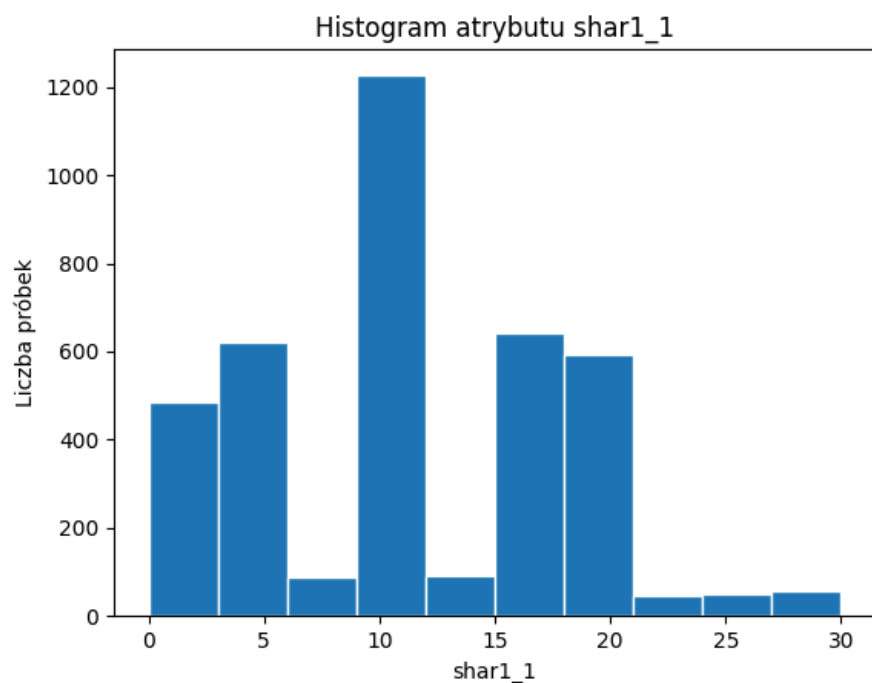
fun1_1



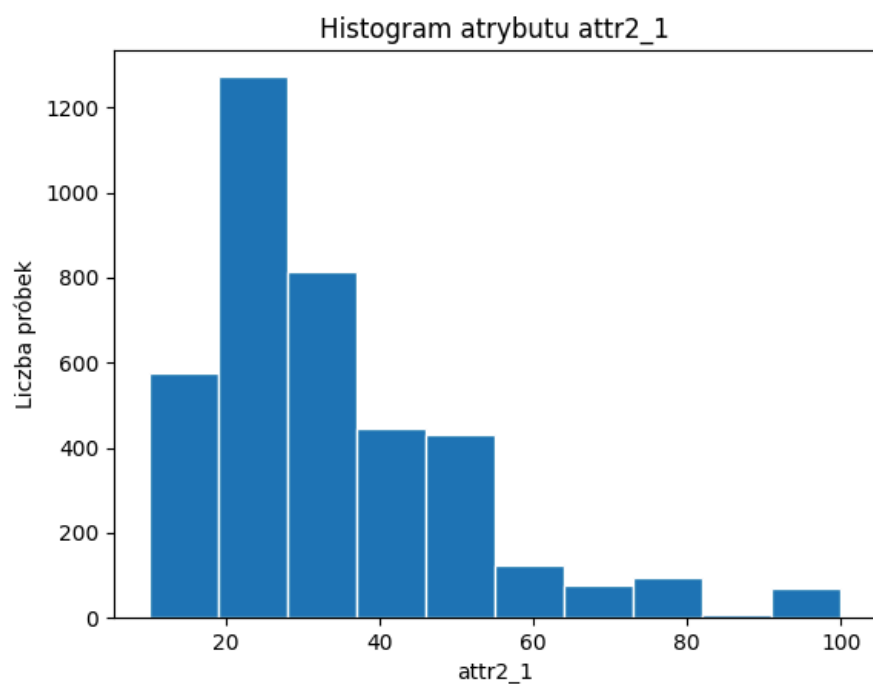
amb1_1



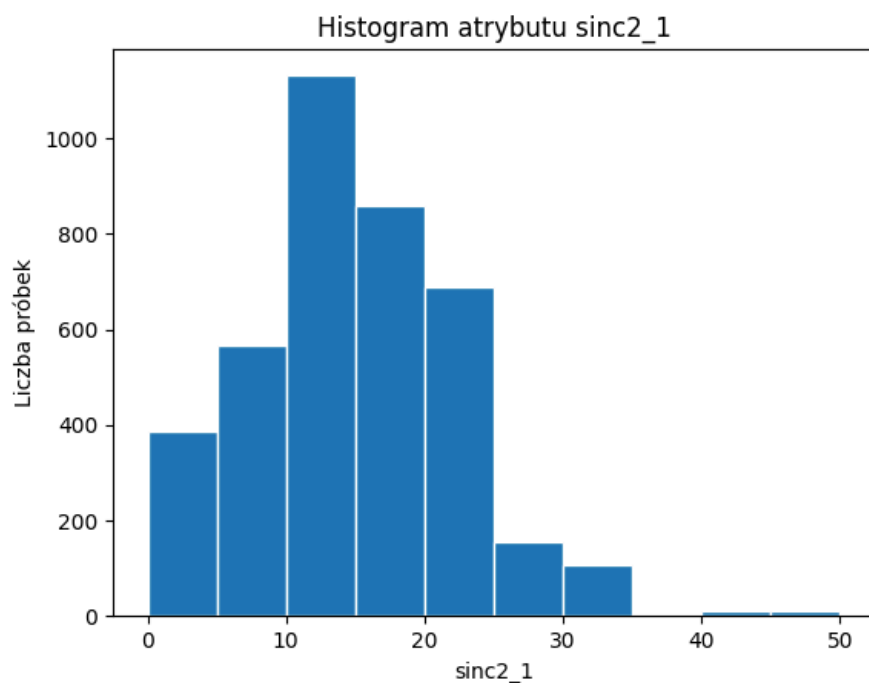
shar1_1



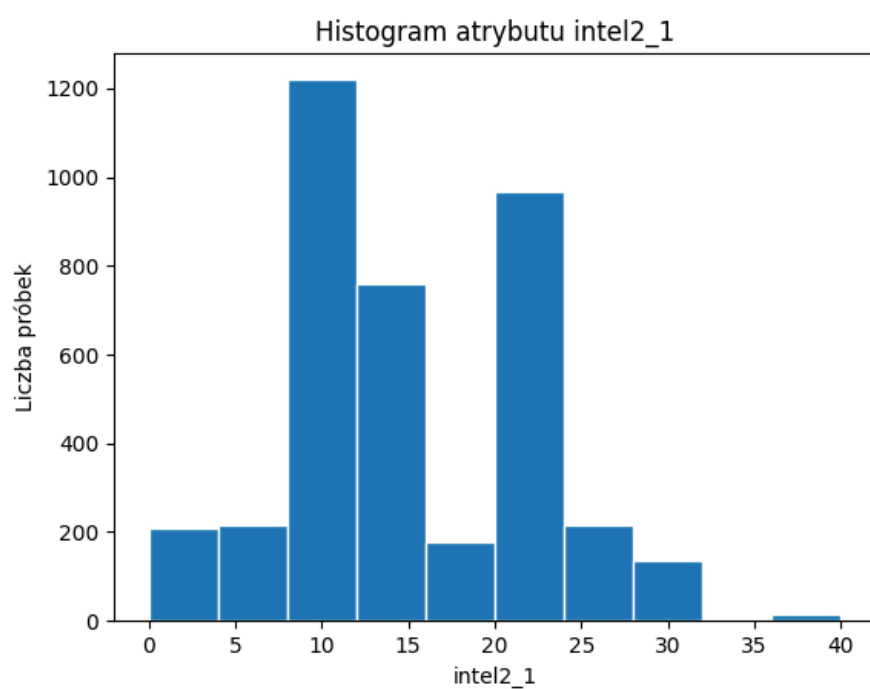
attr2_1



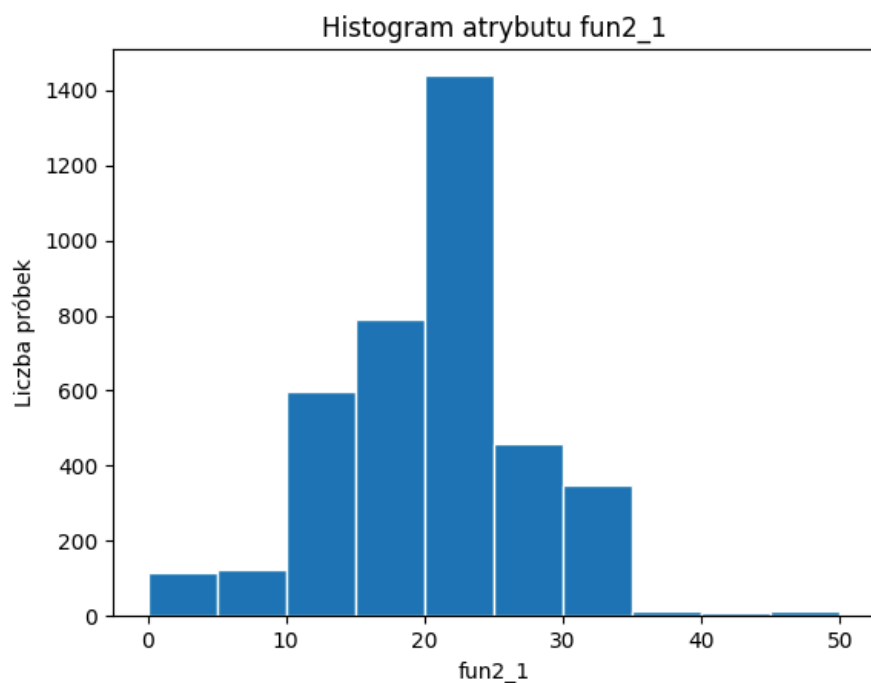
sinc2_1



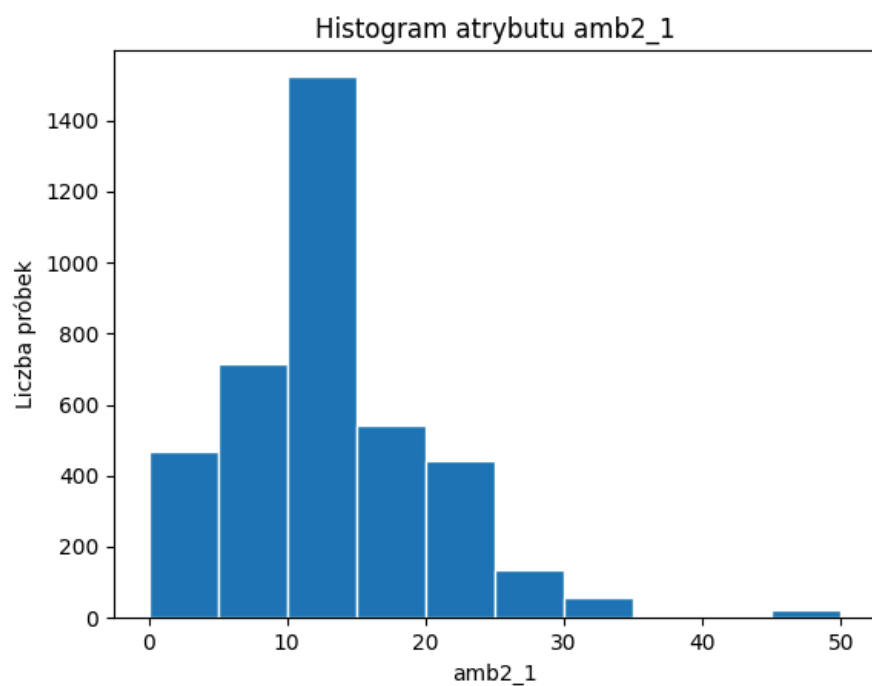
intel2_1



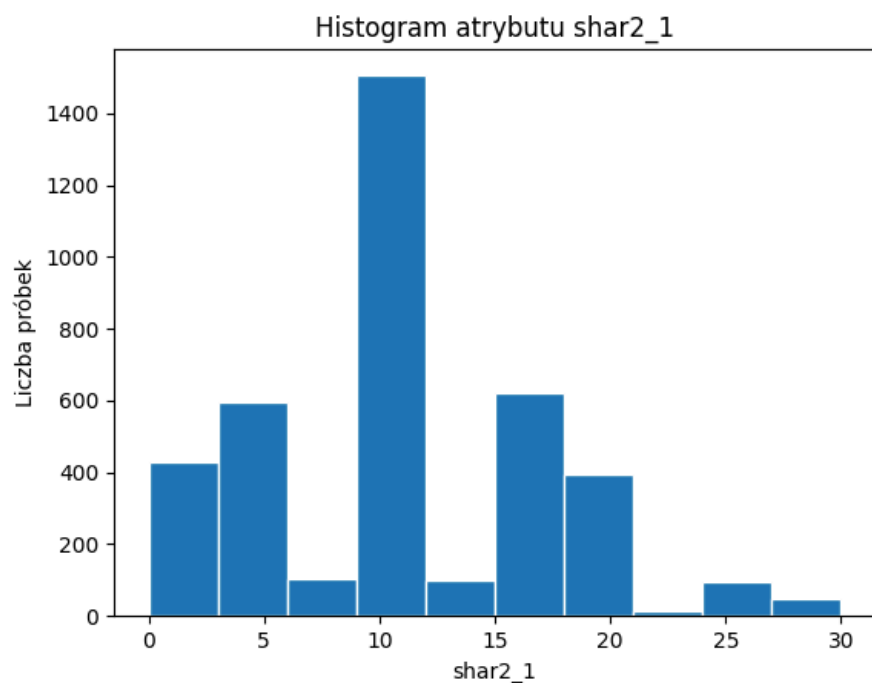
fun2_1



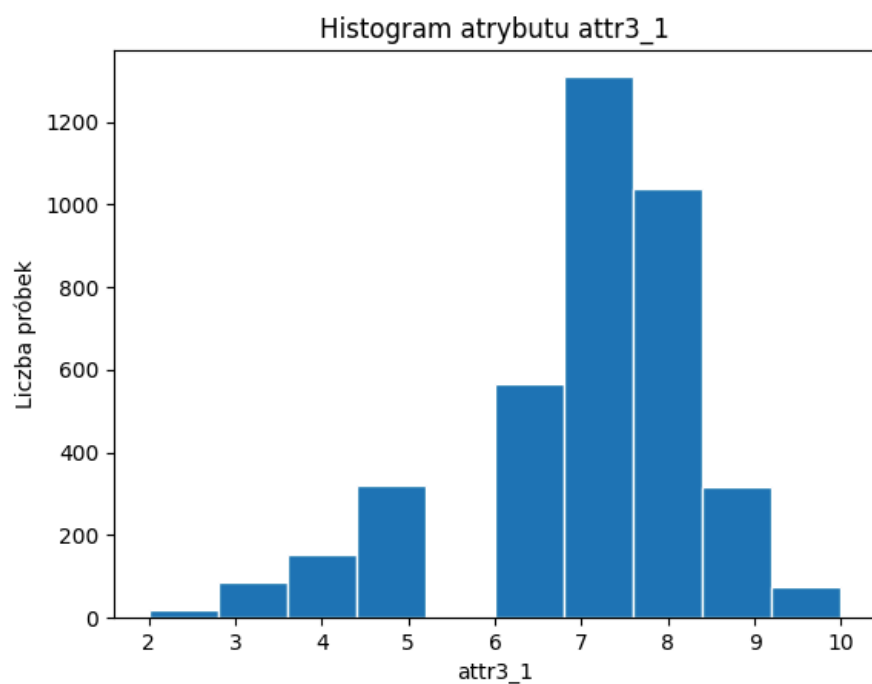
amb2_1



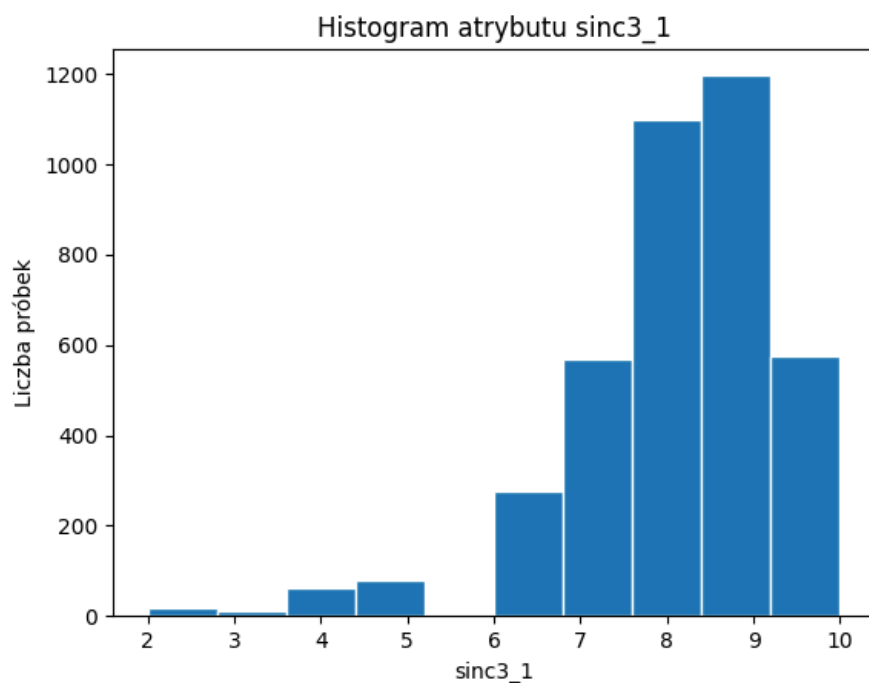
shar2_1



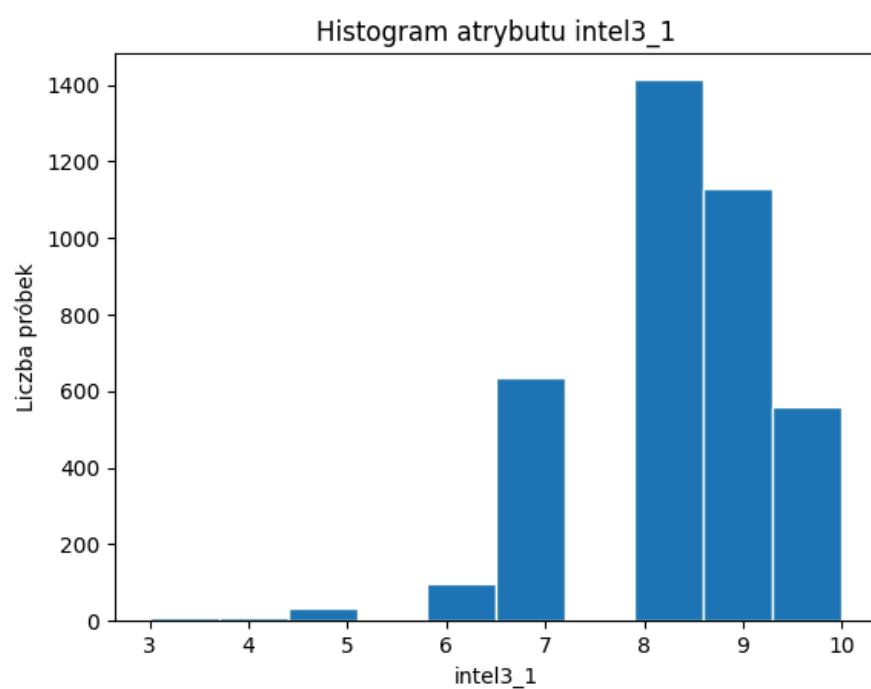
attr3_1



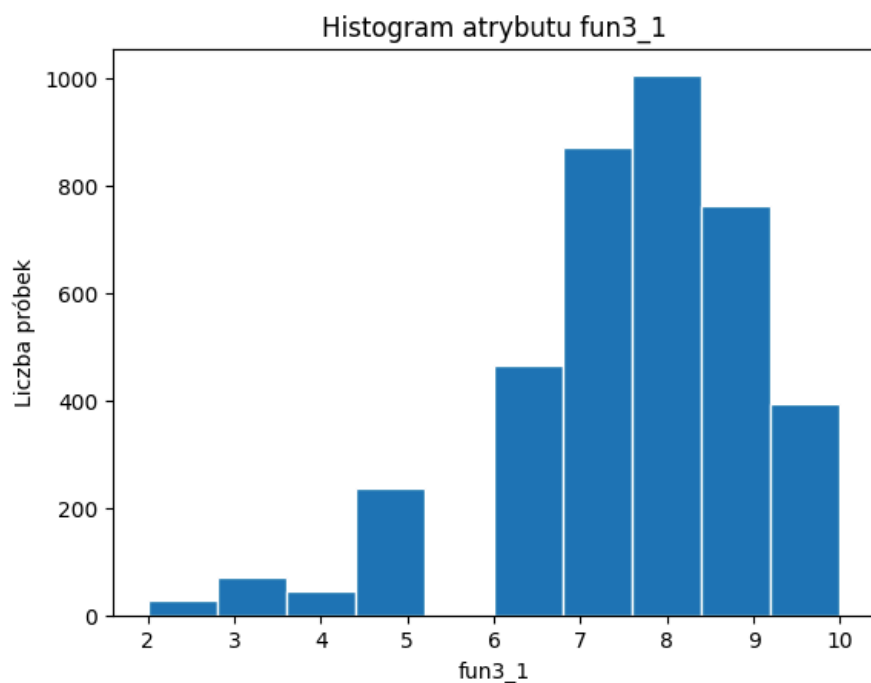
sinc3_1



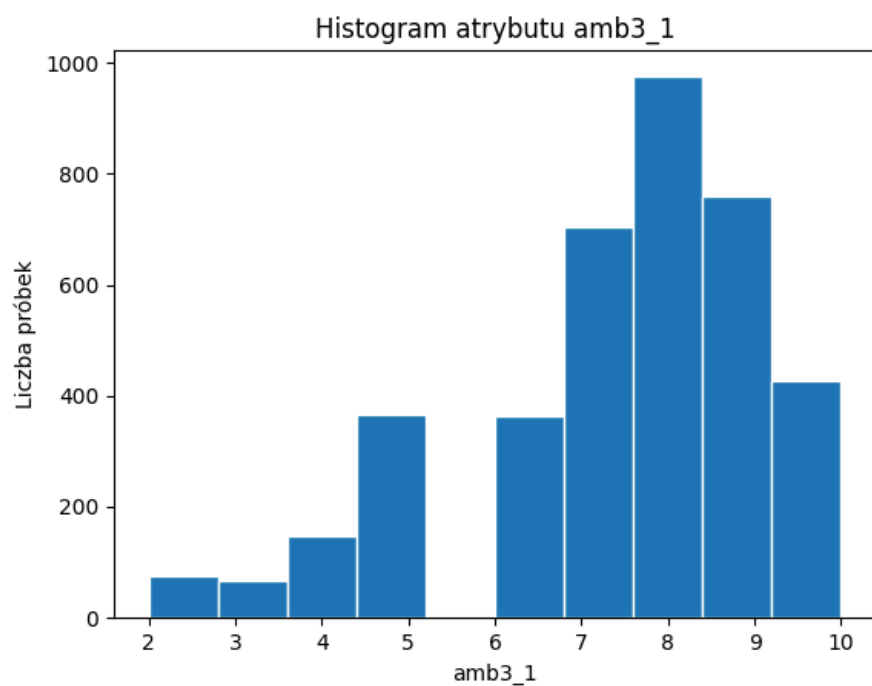
intel3_1



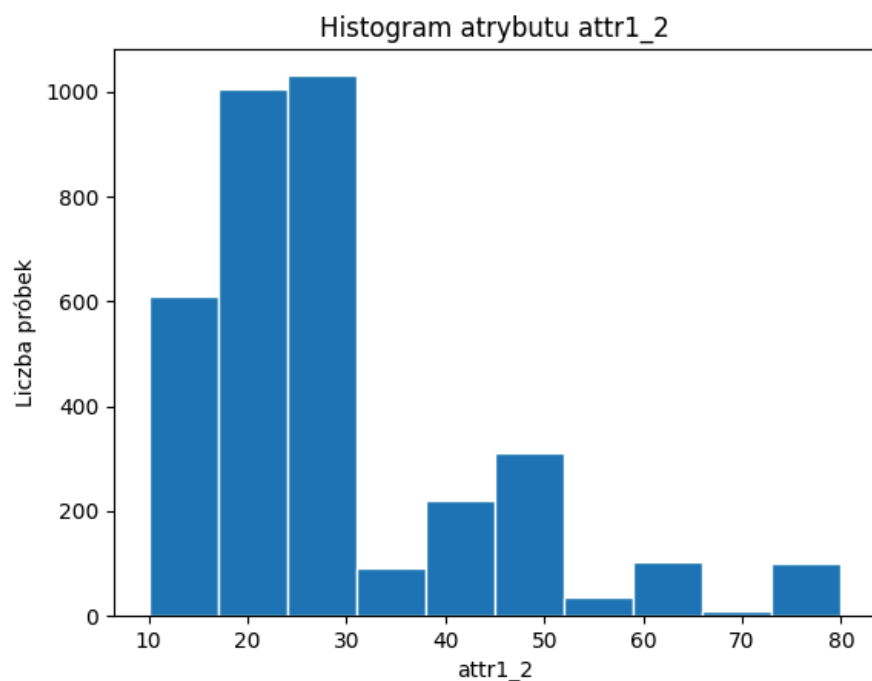
fun3_1



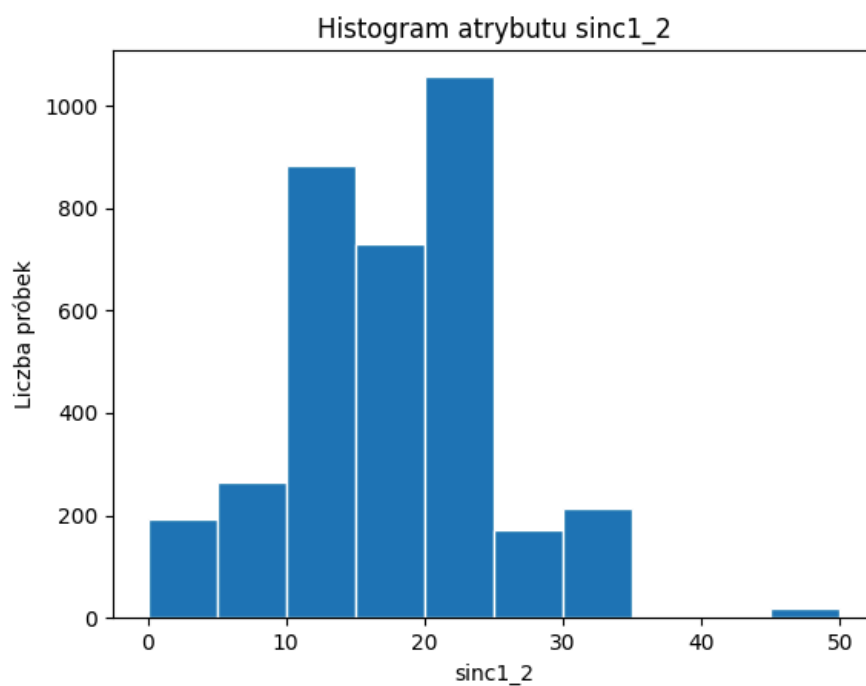
amb3_1



attr1_2

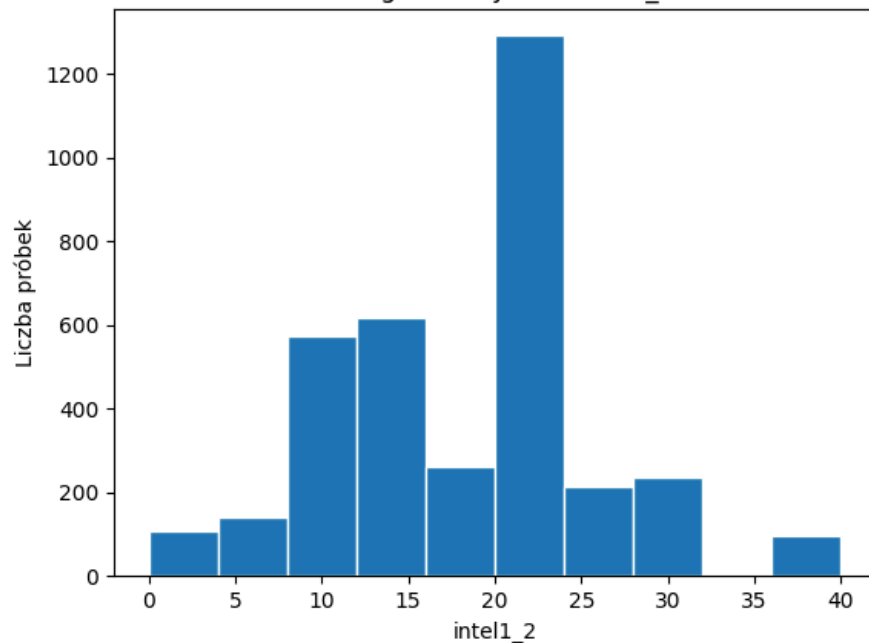


sinc1_2



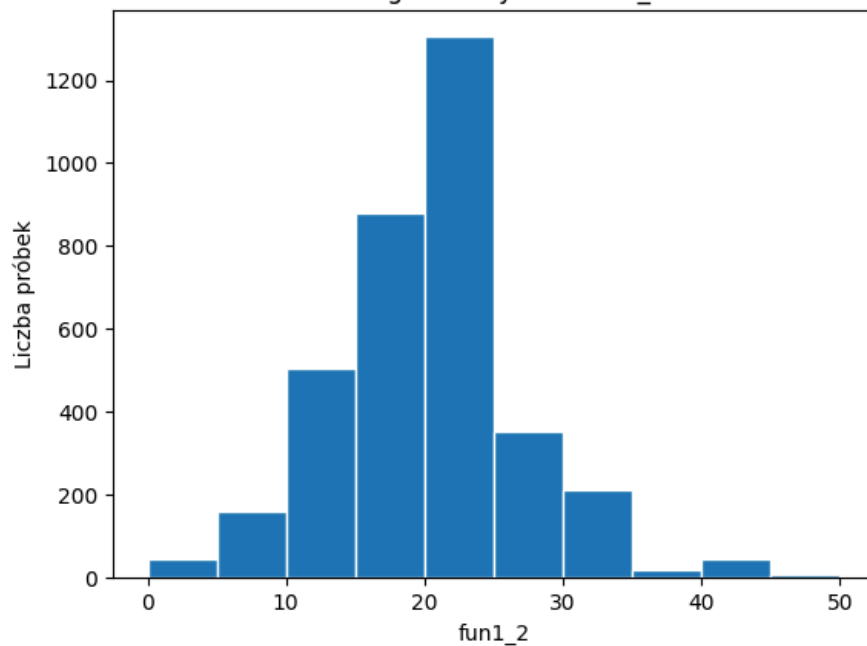
intel1_2

Histogram atrybutu intel1_2

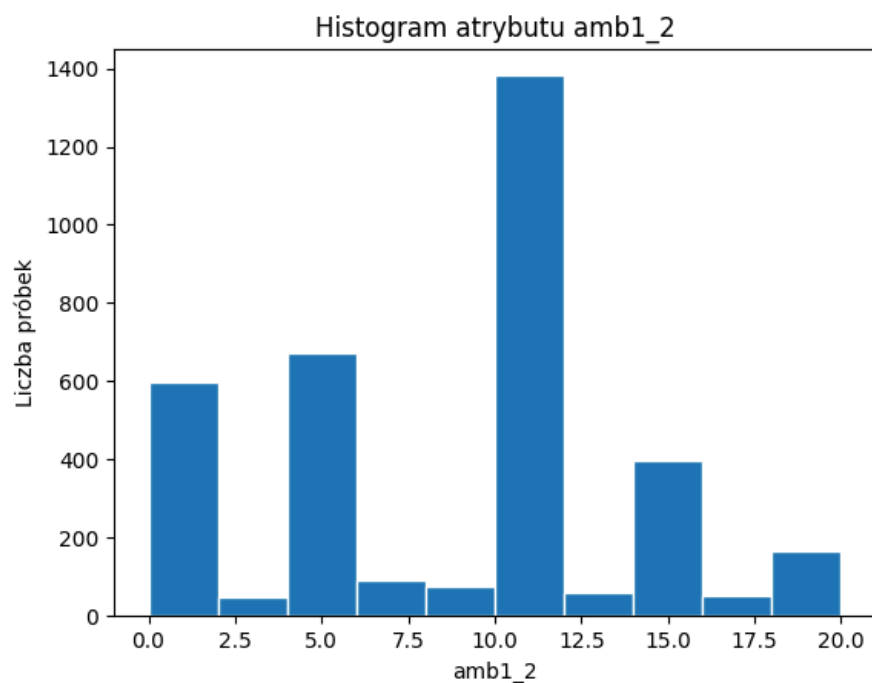


fun1_2

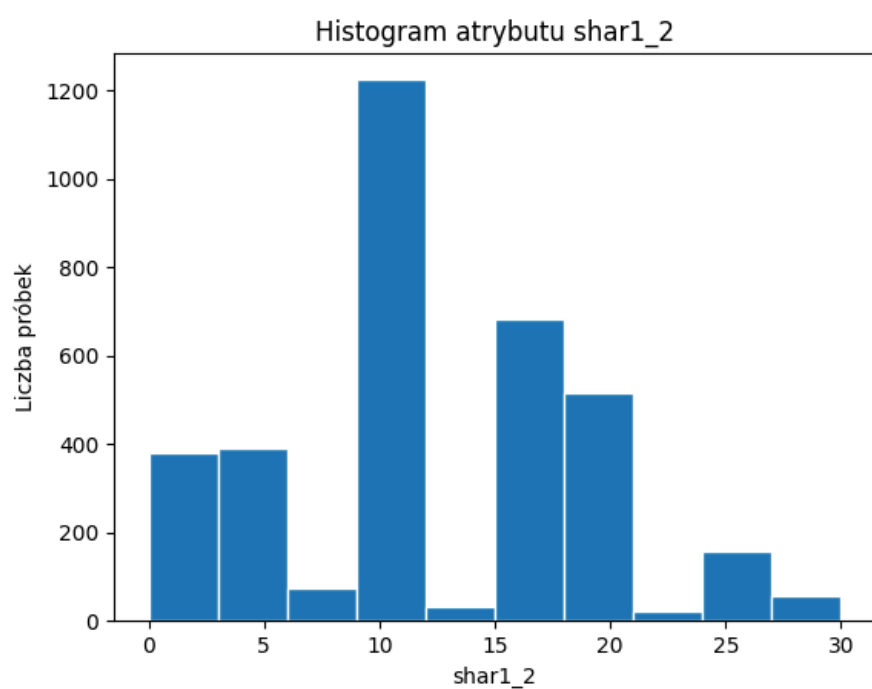
Histogram atrybutu fun1_2

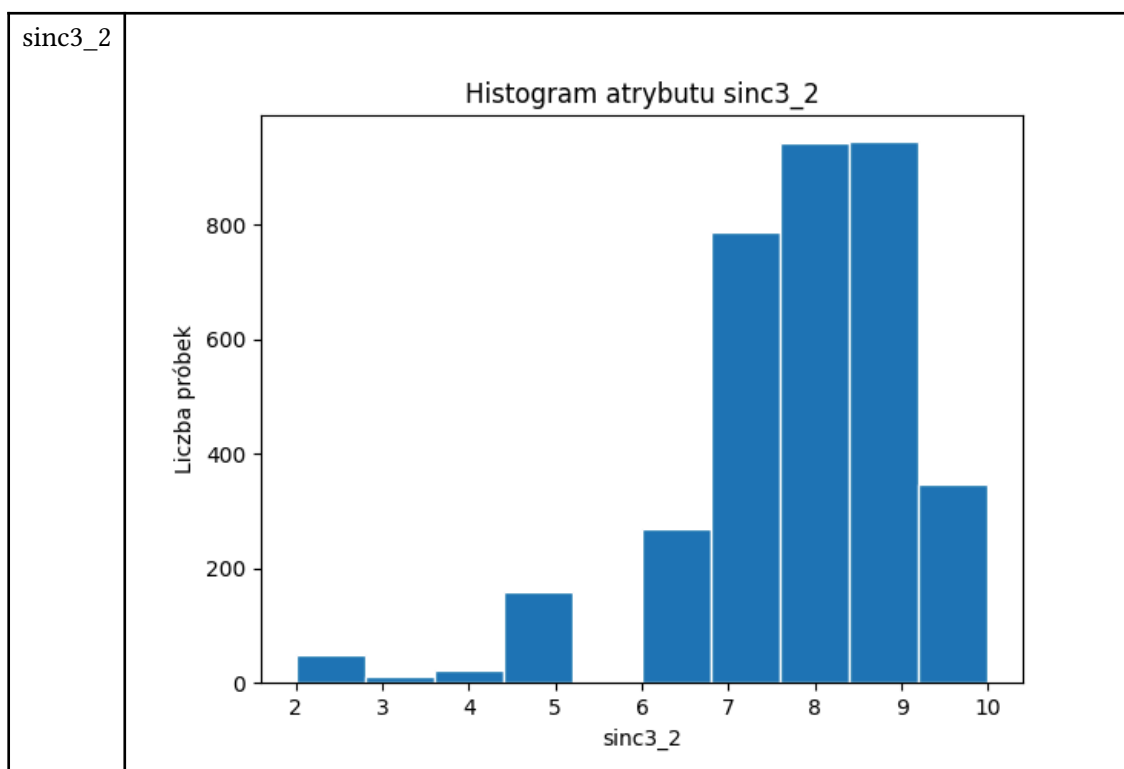
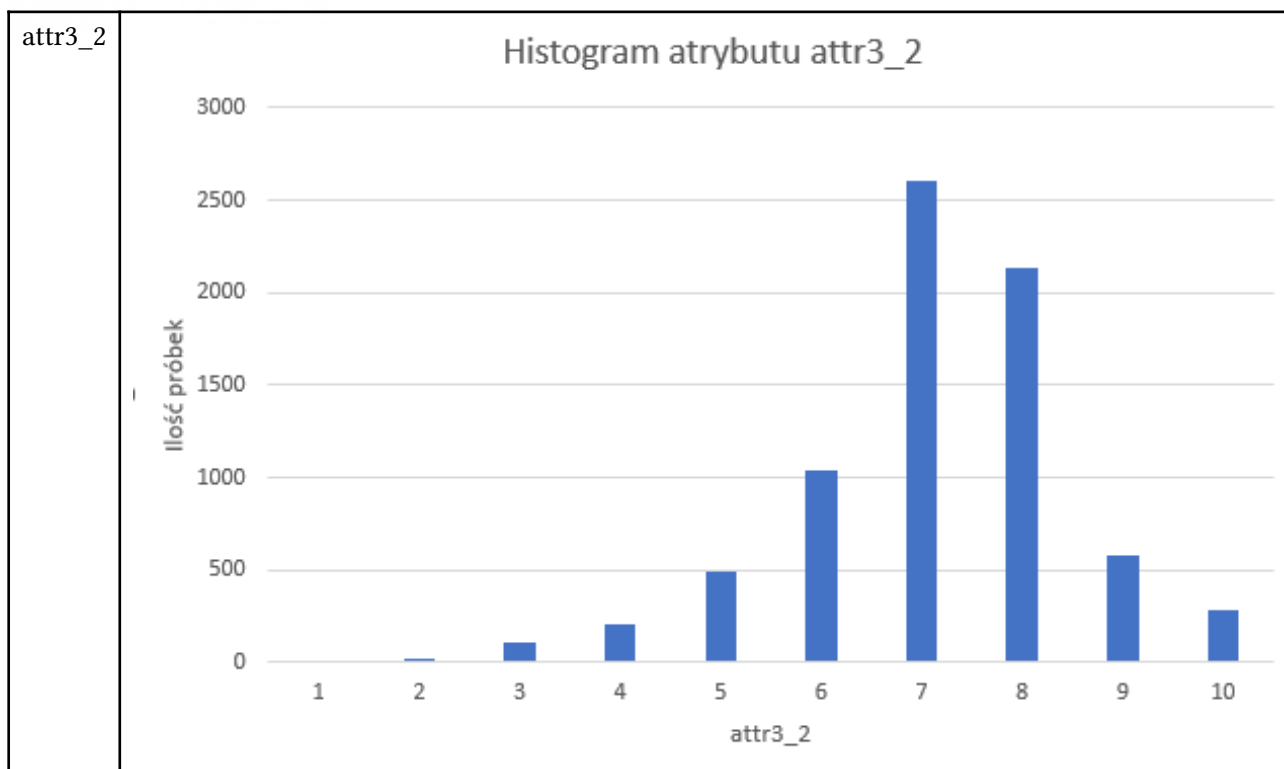


amb1_2

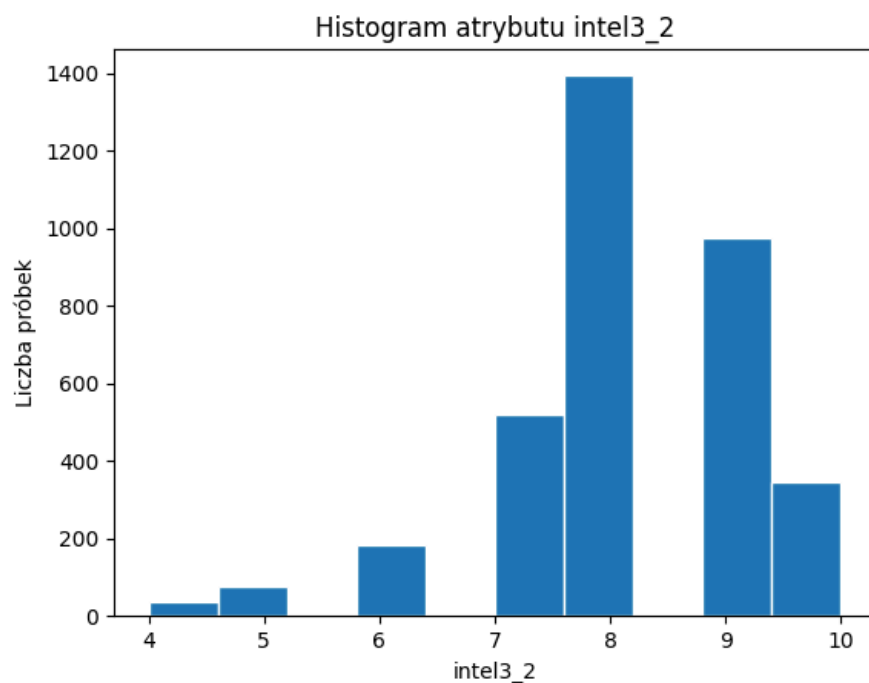


shar1_2

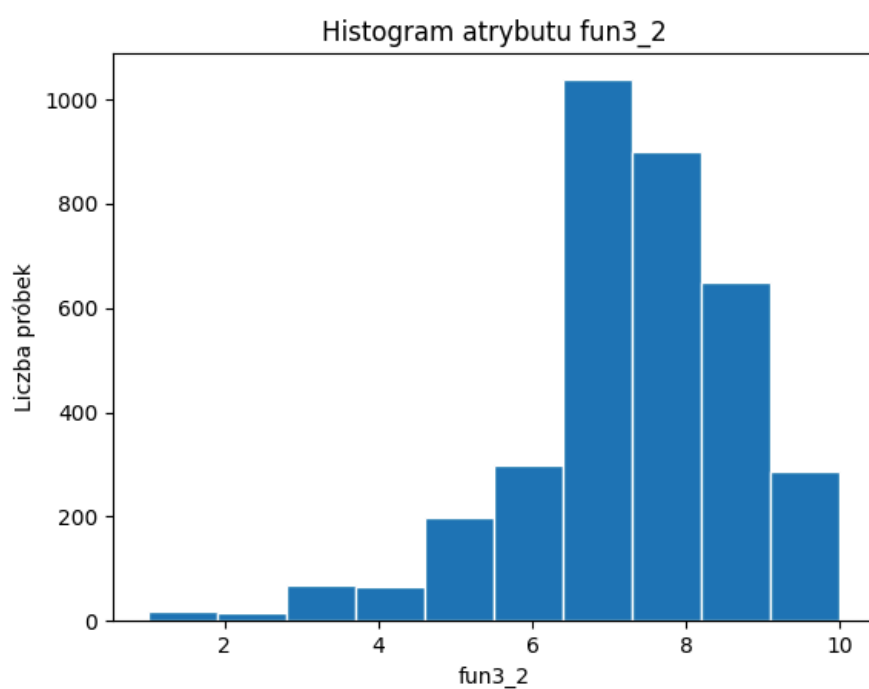




intel3_2

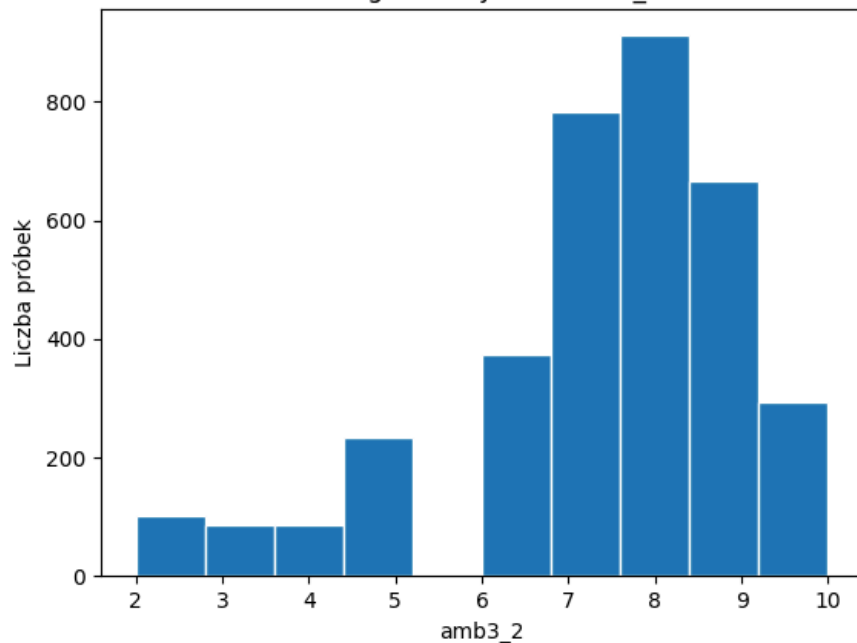


fun3_2



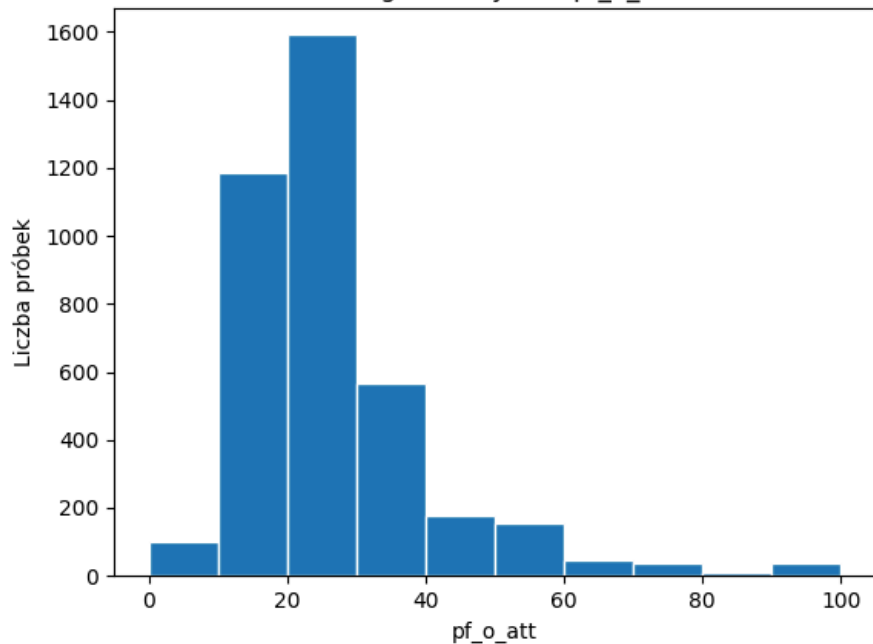
amb3_2

Histogram atrybutu amb3_2

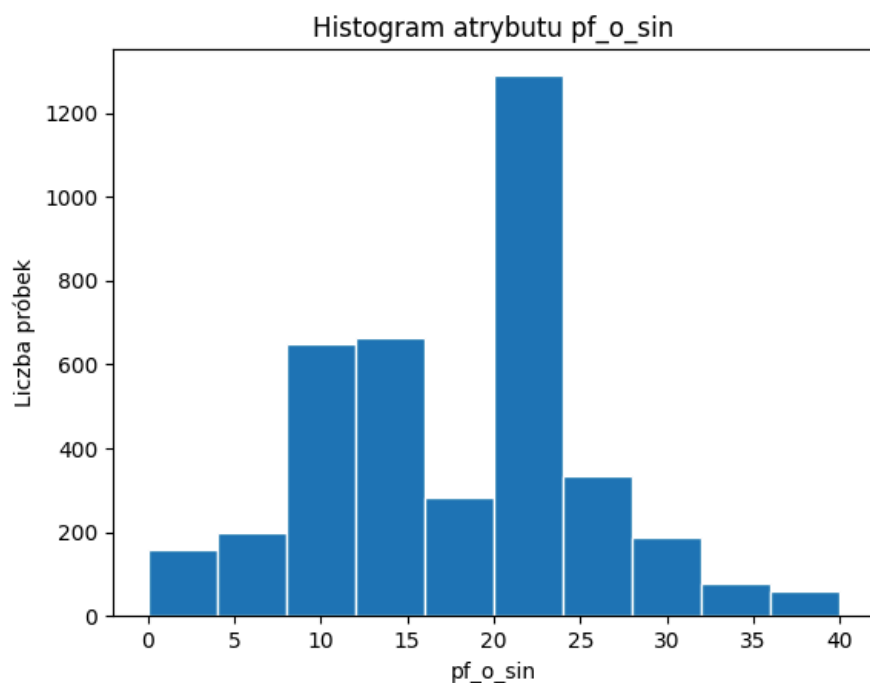


pf_o_att

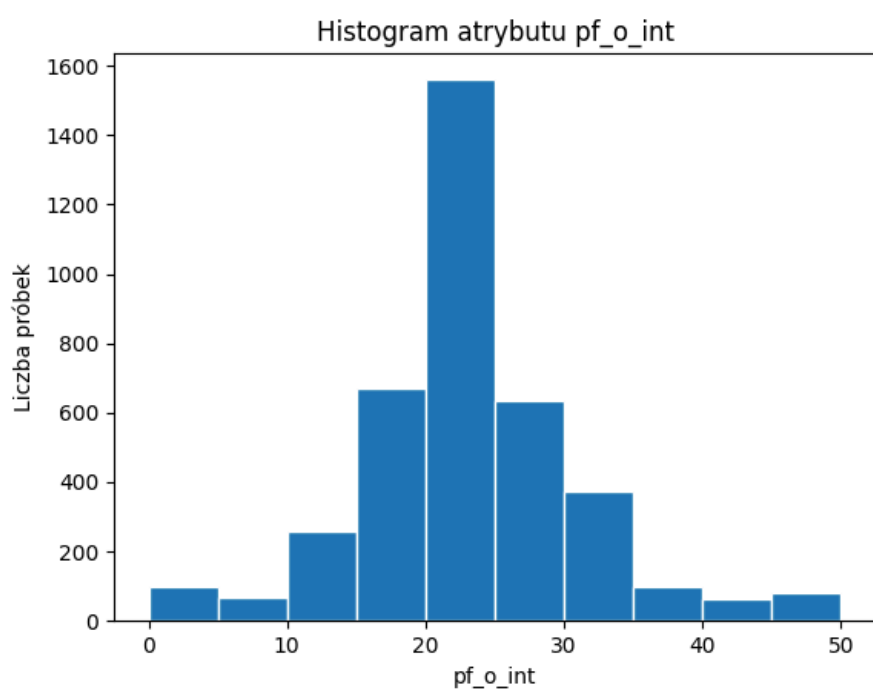
Histogram atrybutu pf_o_att



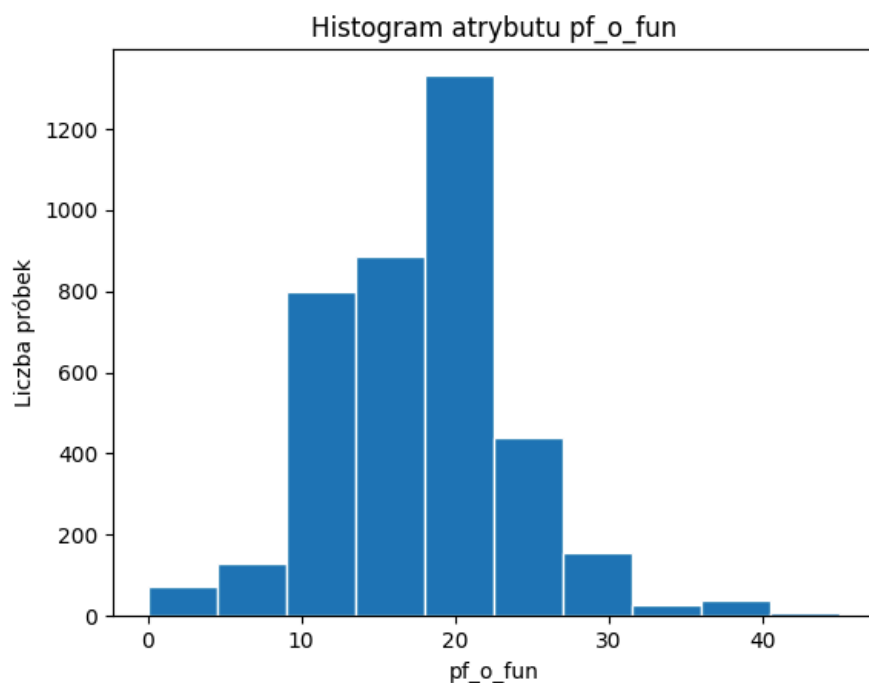
pf_o_sin



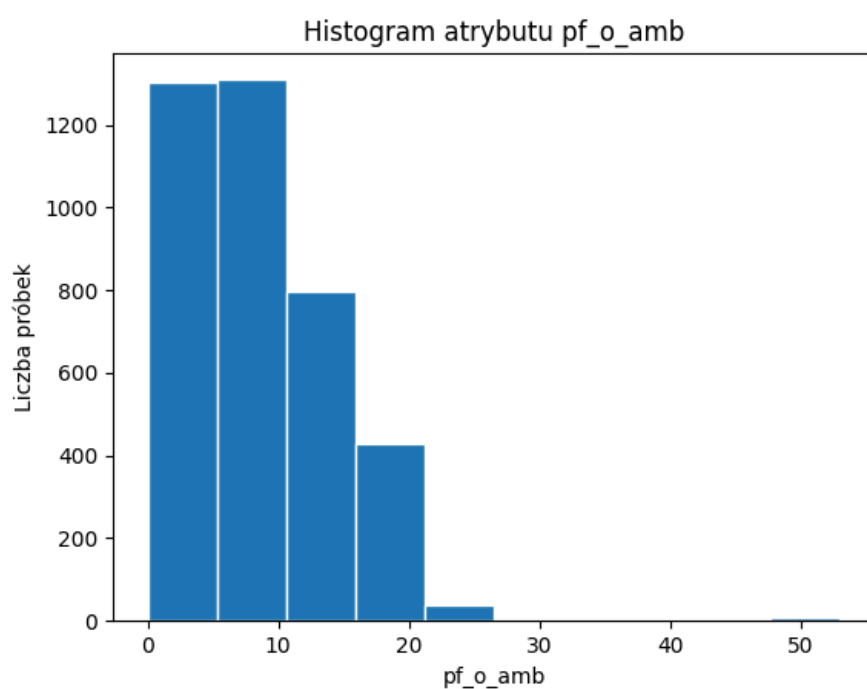
pf_o_int



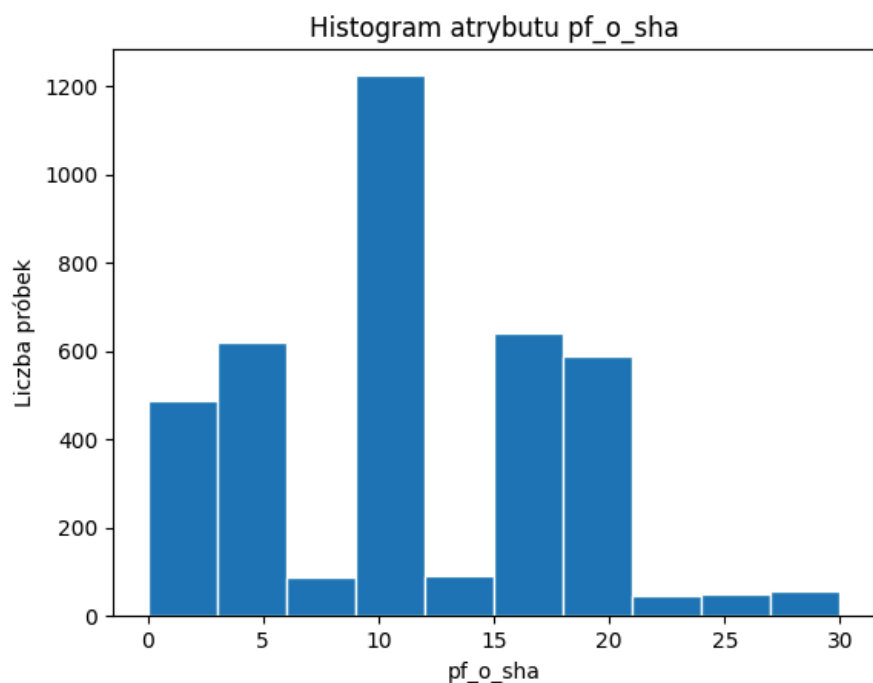
pf_o_fun



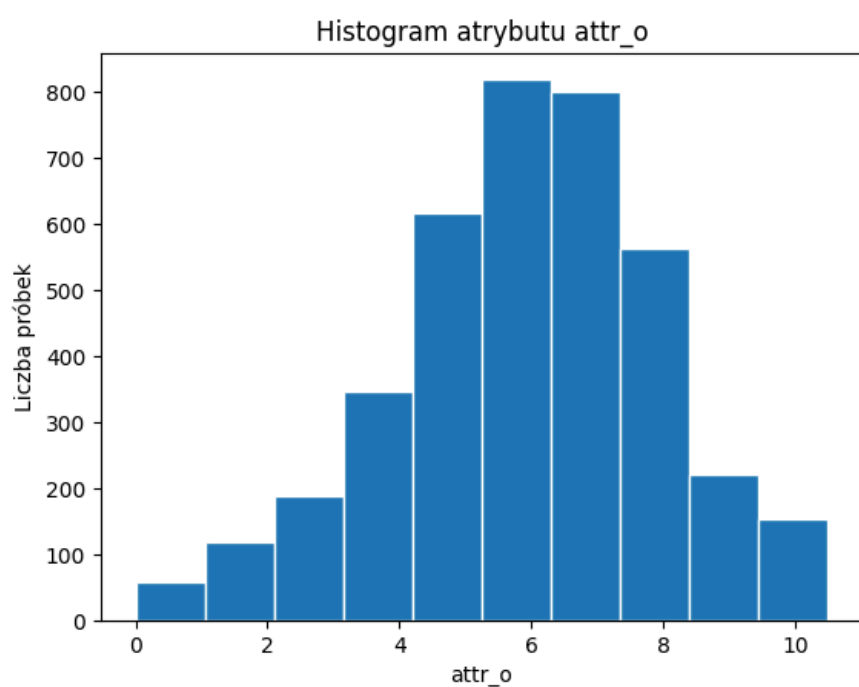
pf_o_amb

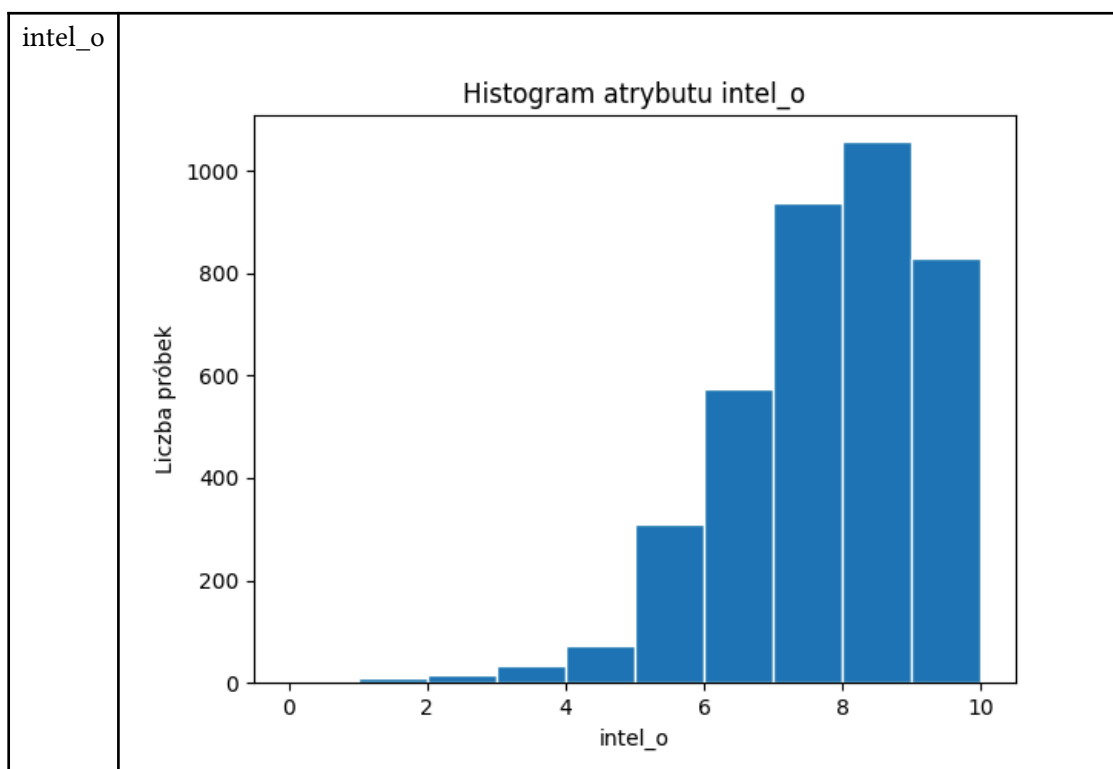
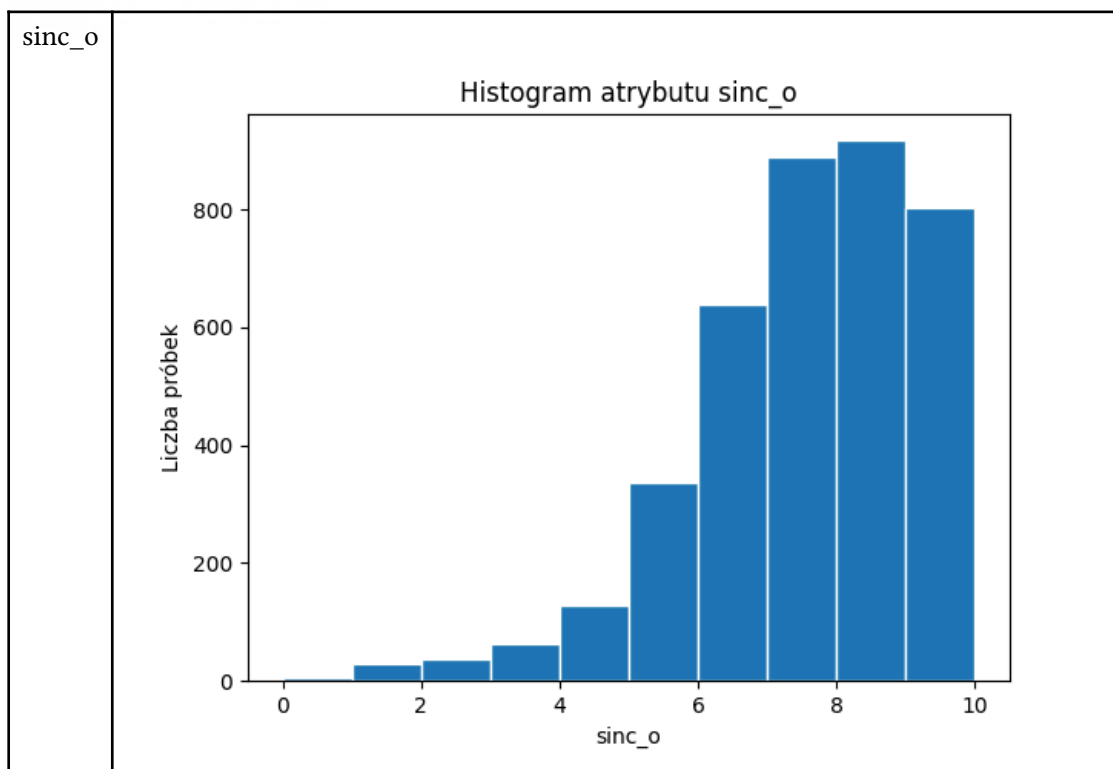


pf_o_sha

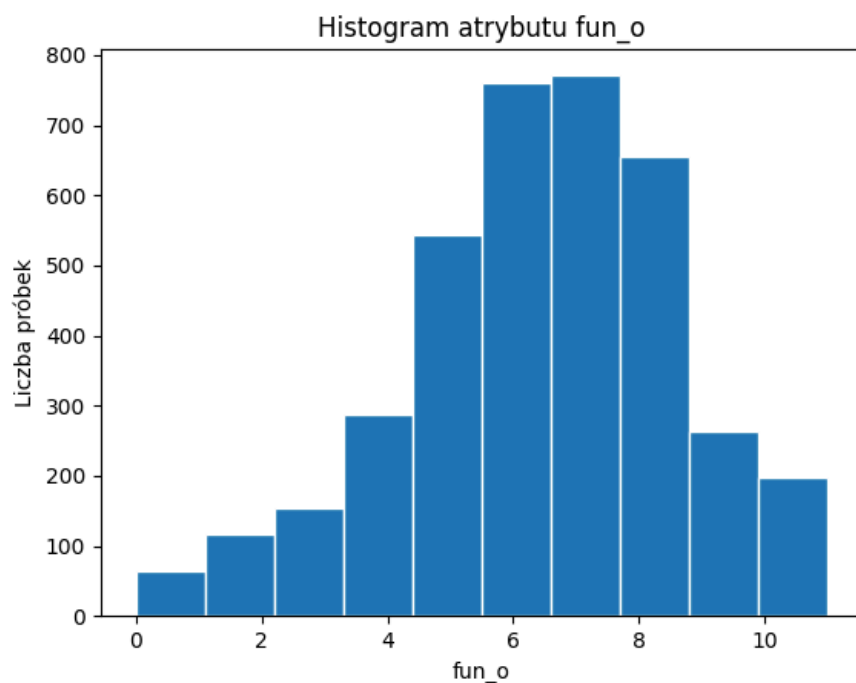


attr_o

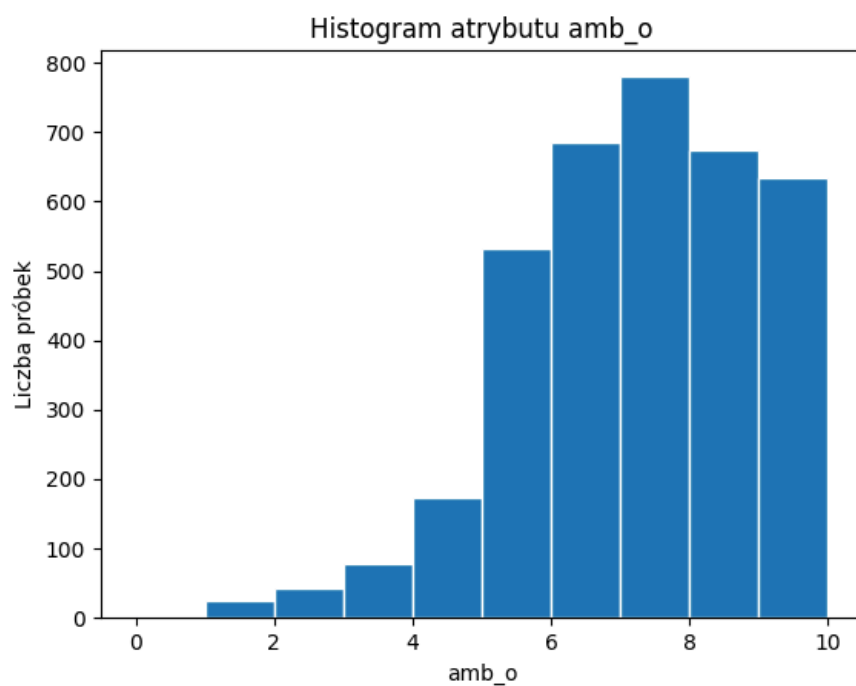


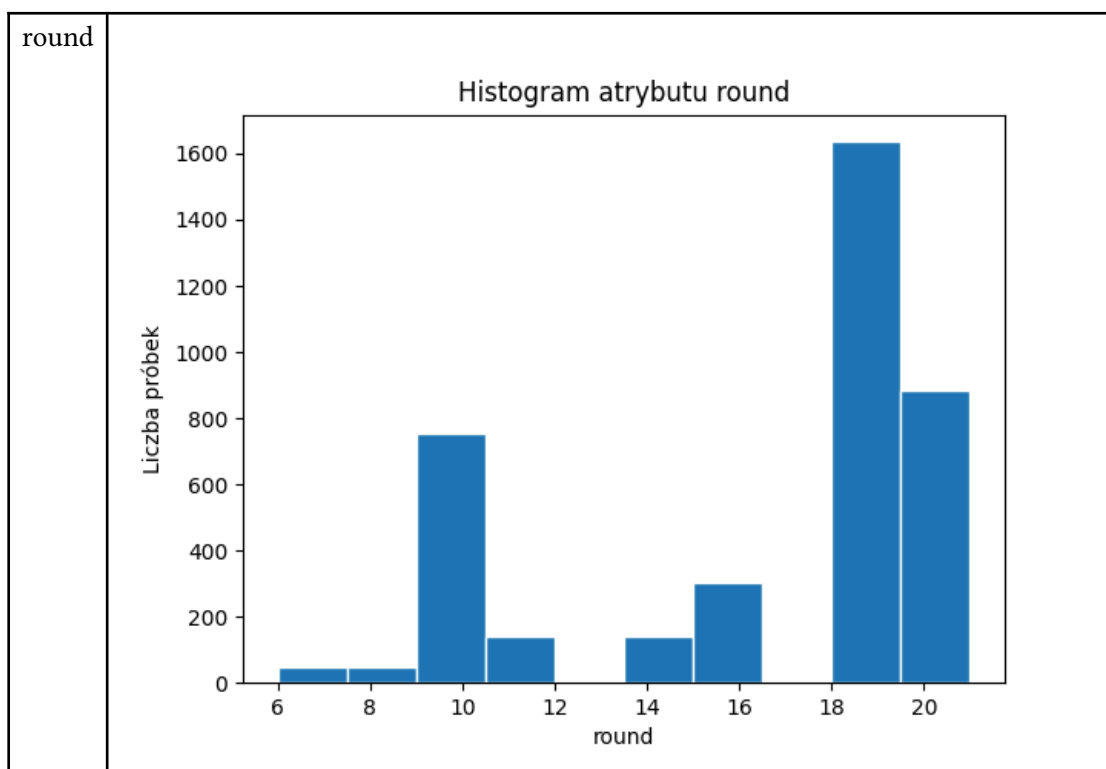
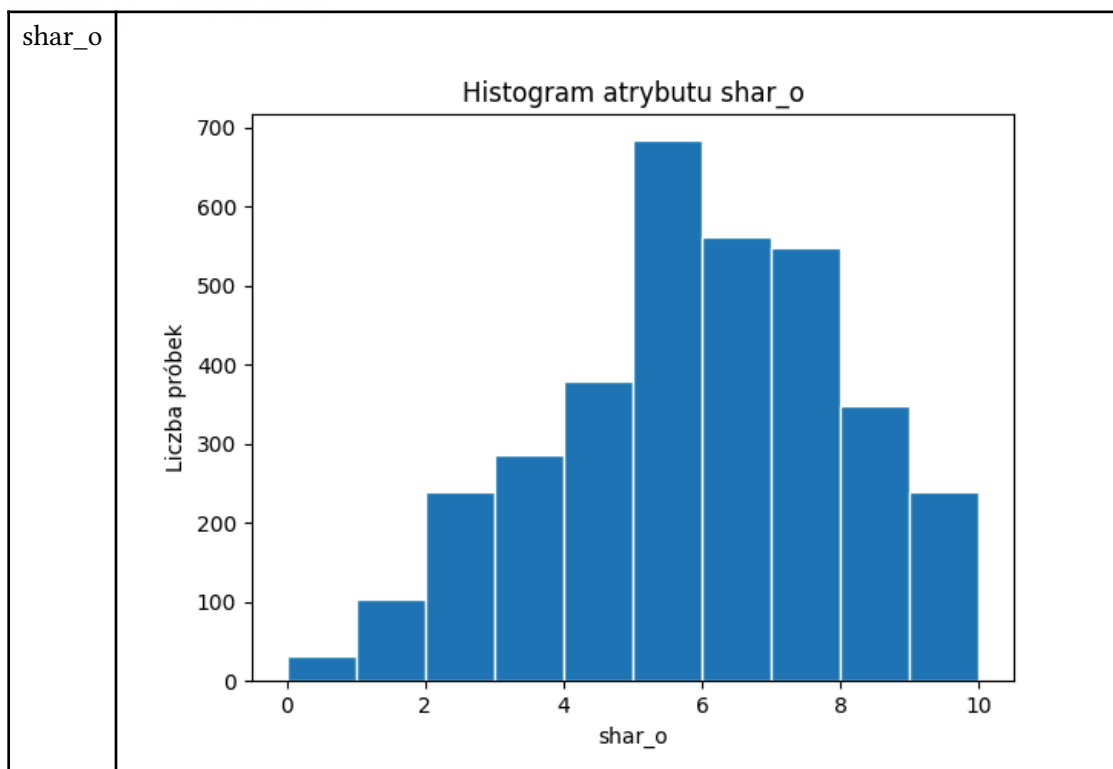


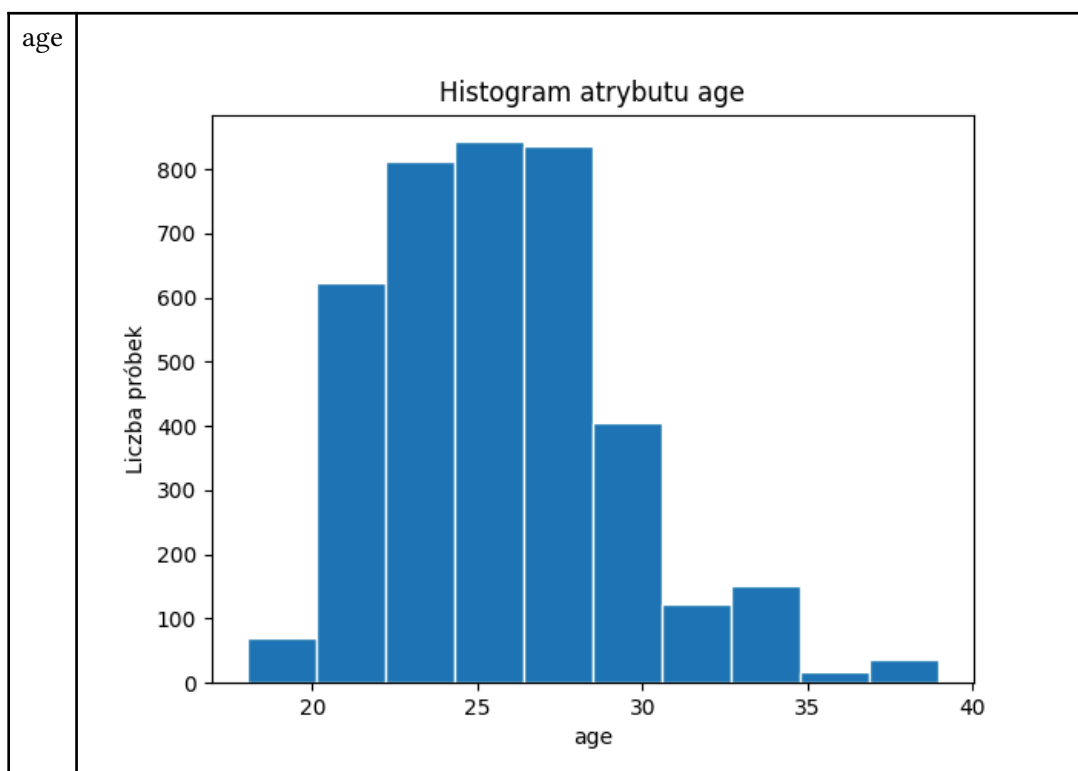
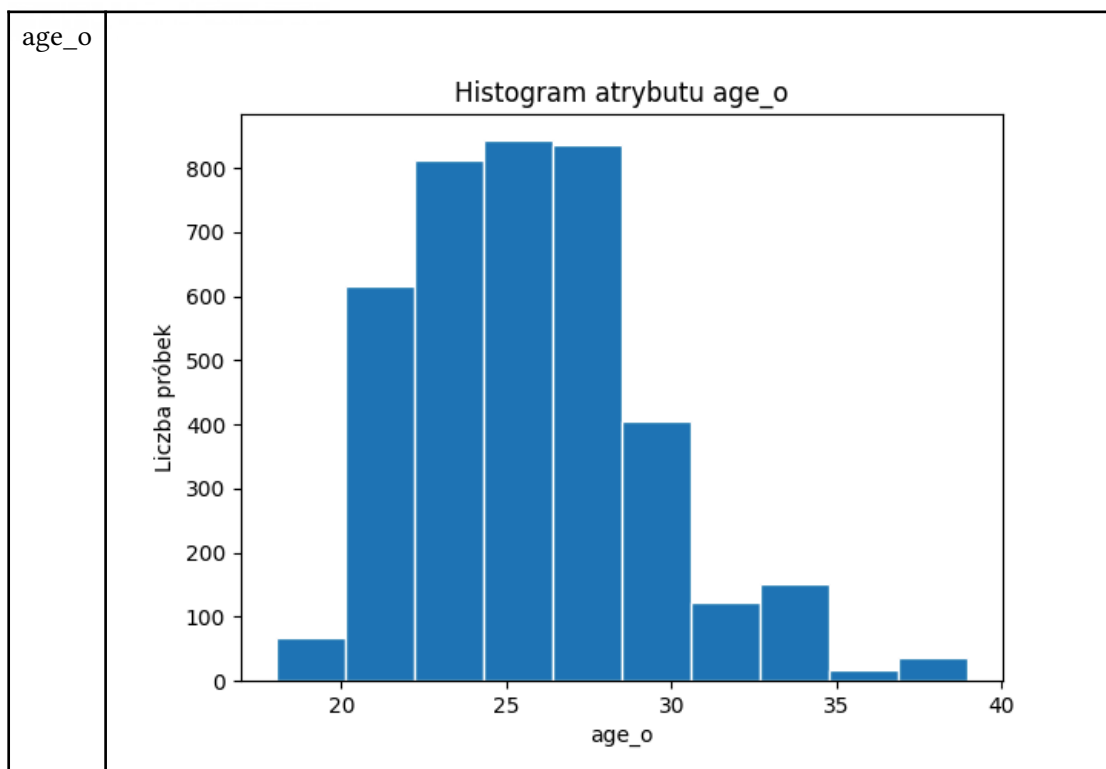
fun_o



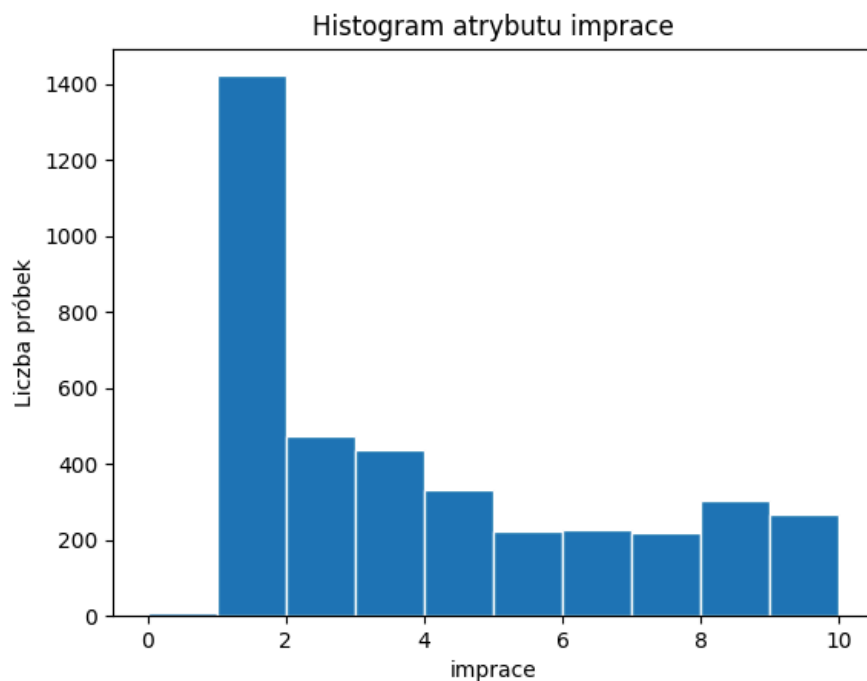
amb_o





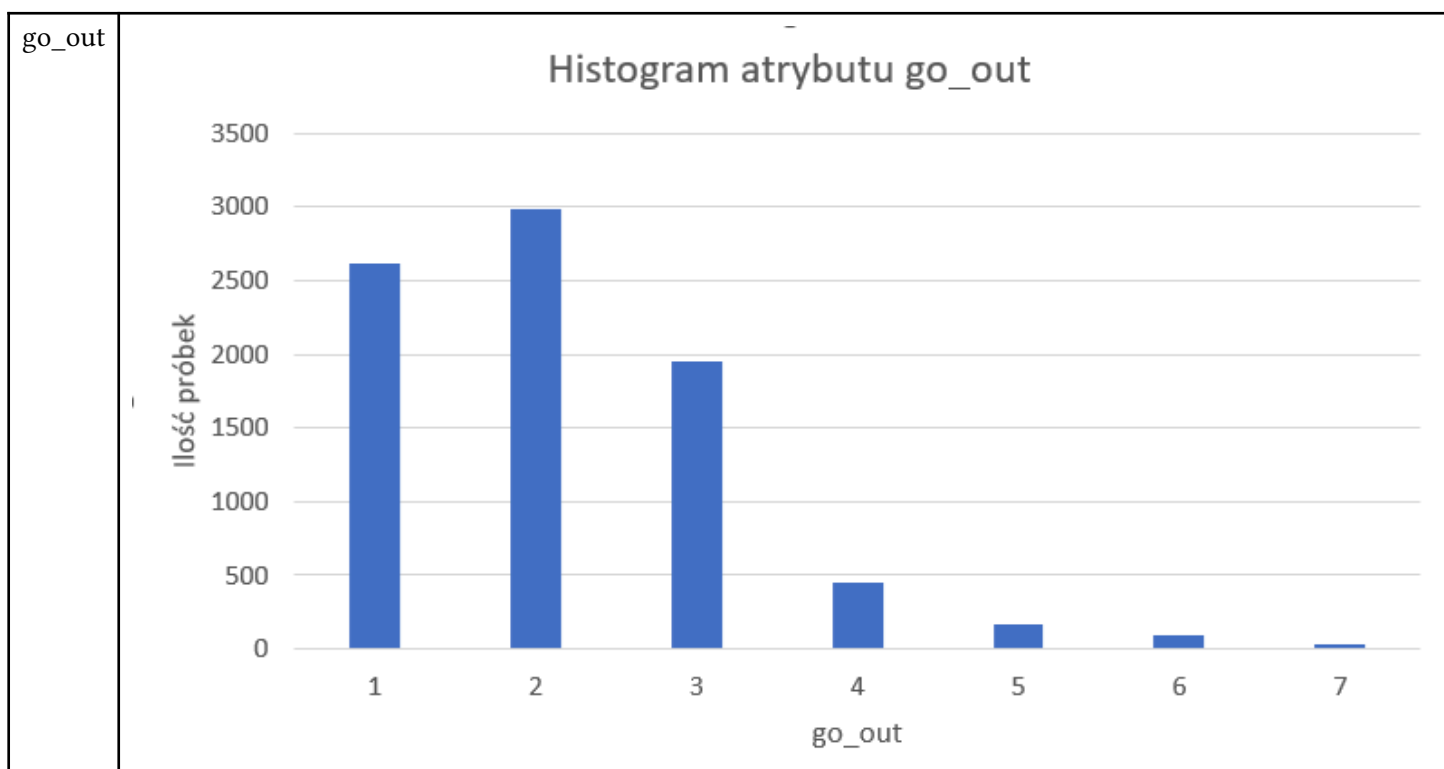
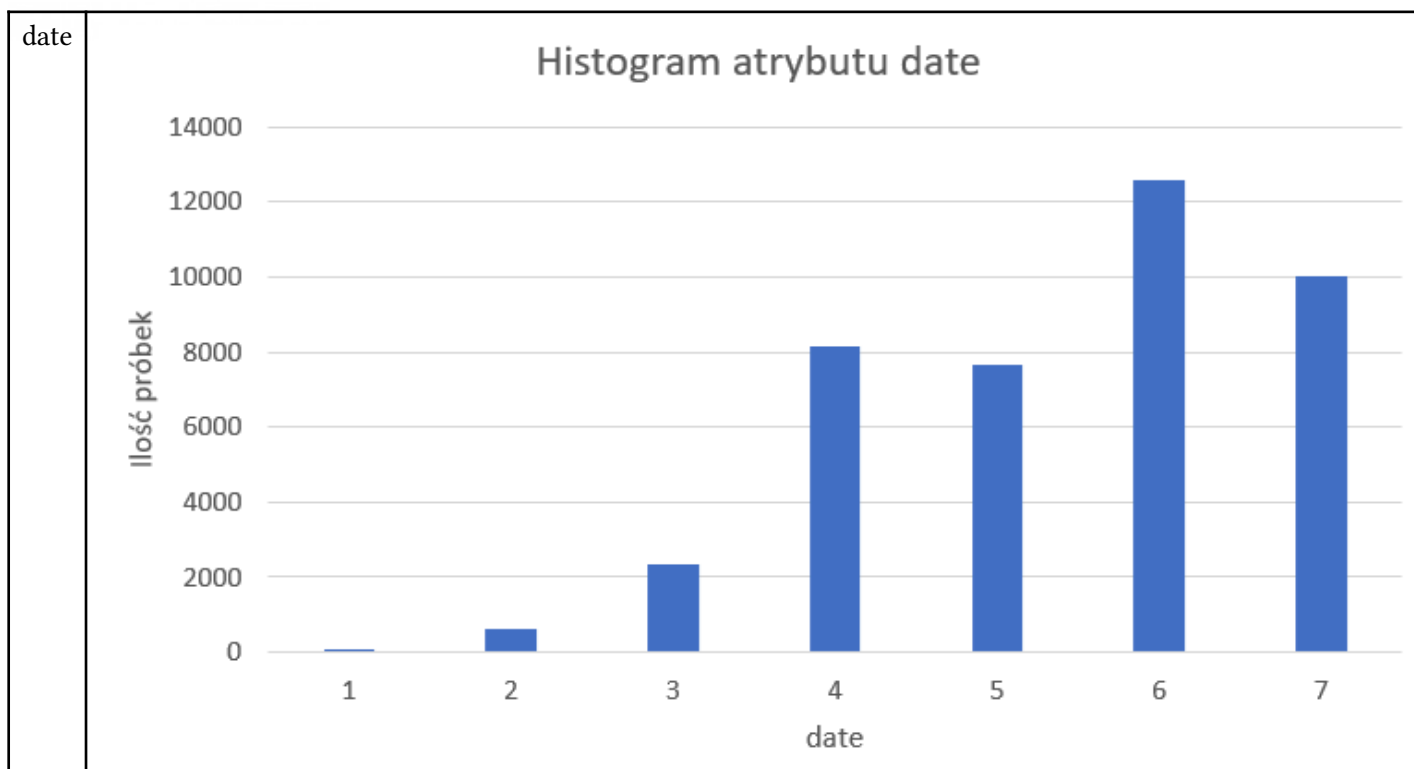


imprace

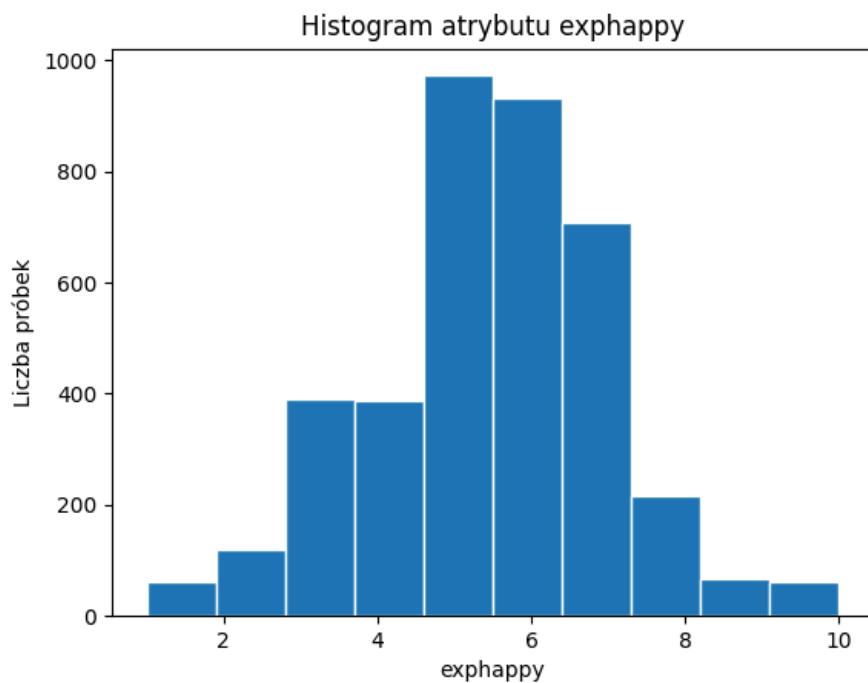


goal

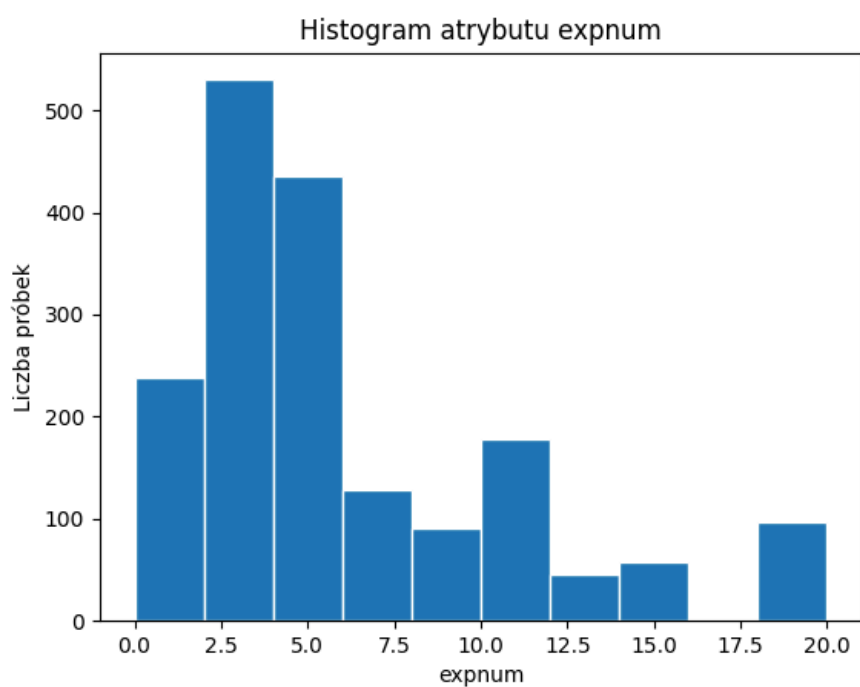




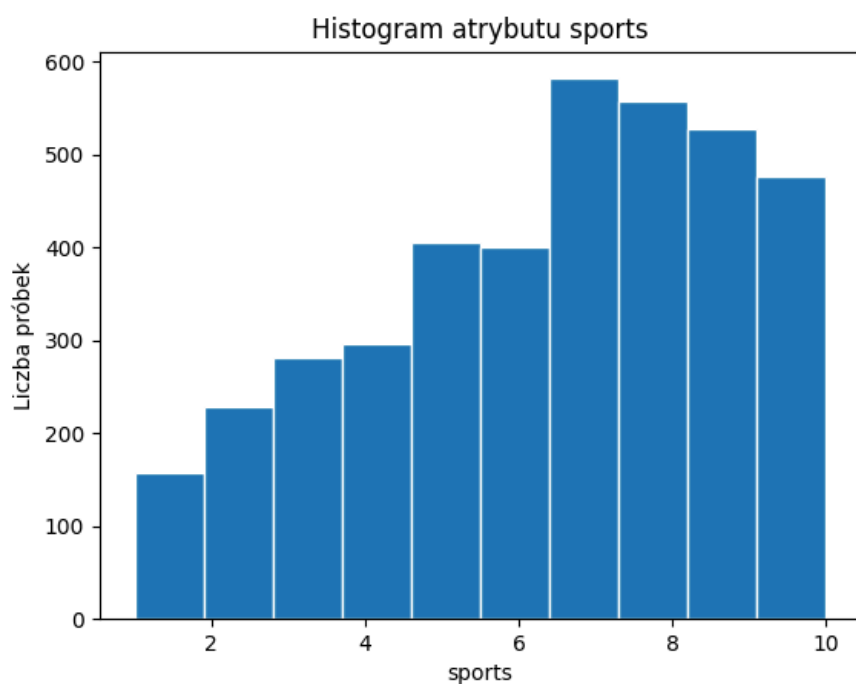
exphappy



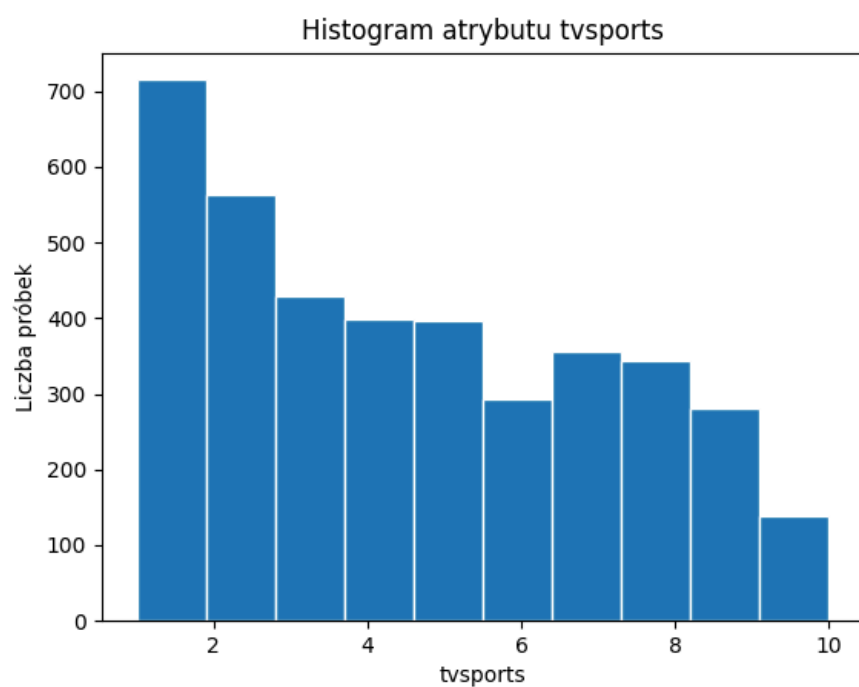
expnum



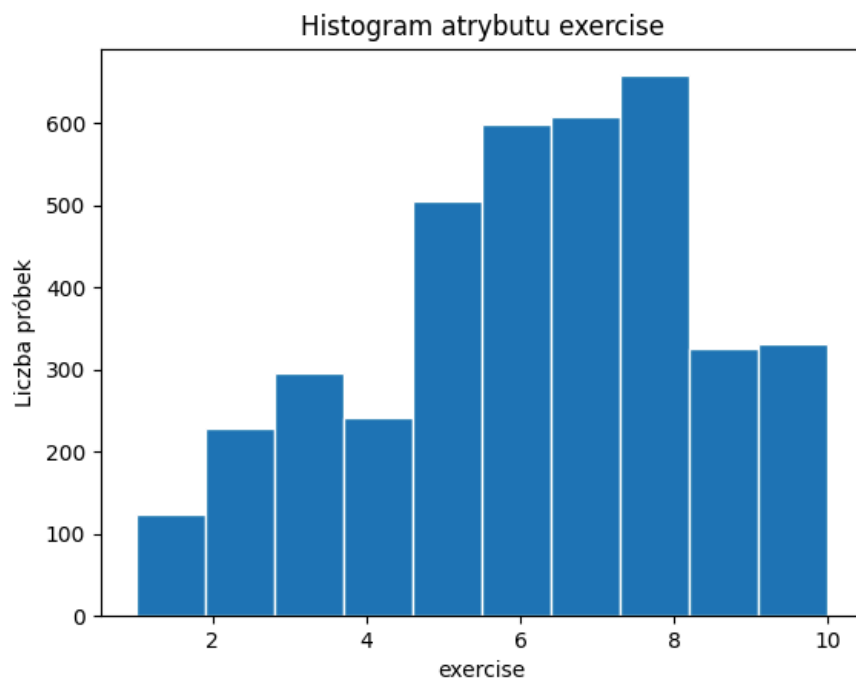
sports



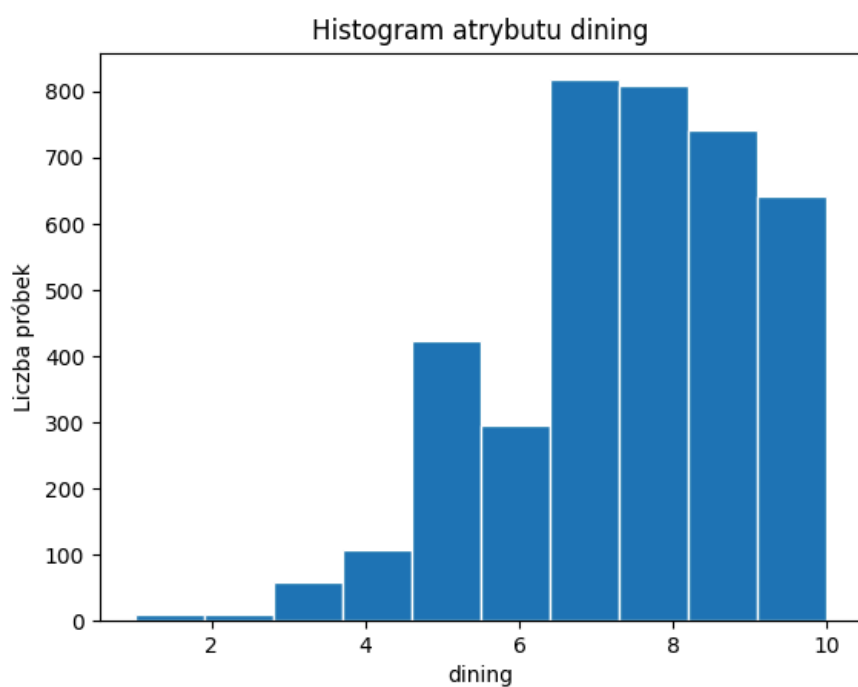
tvsports

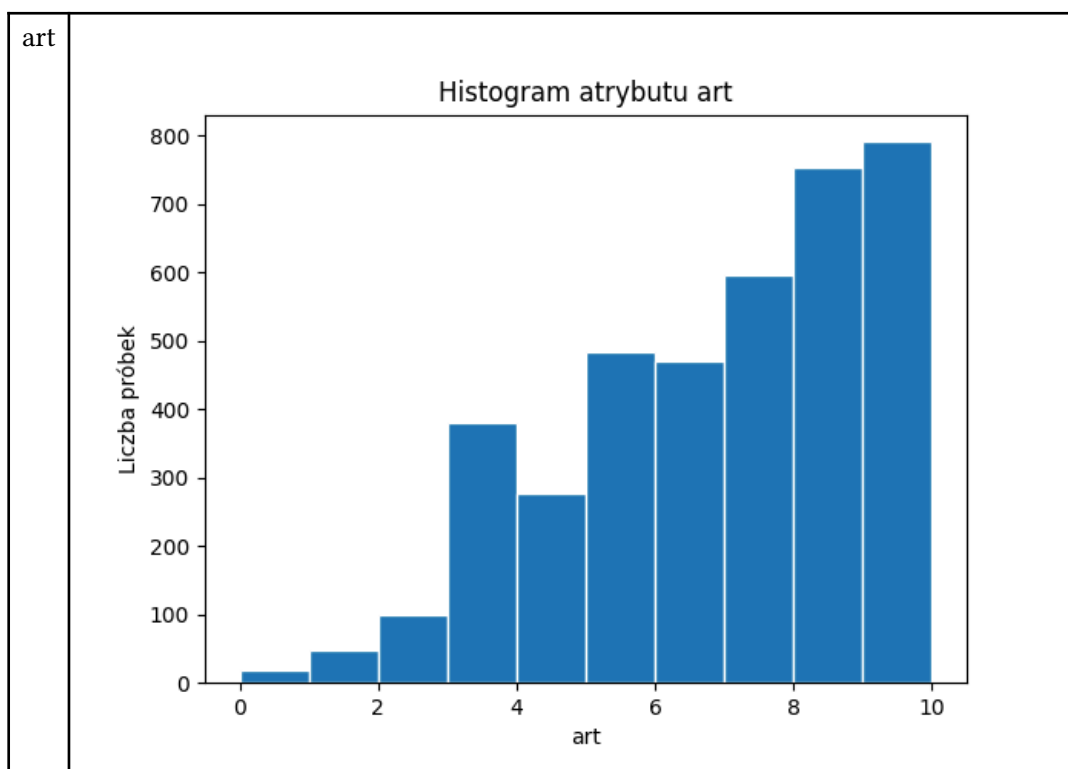
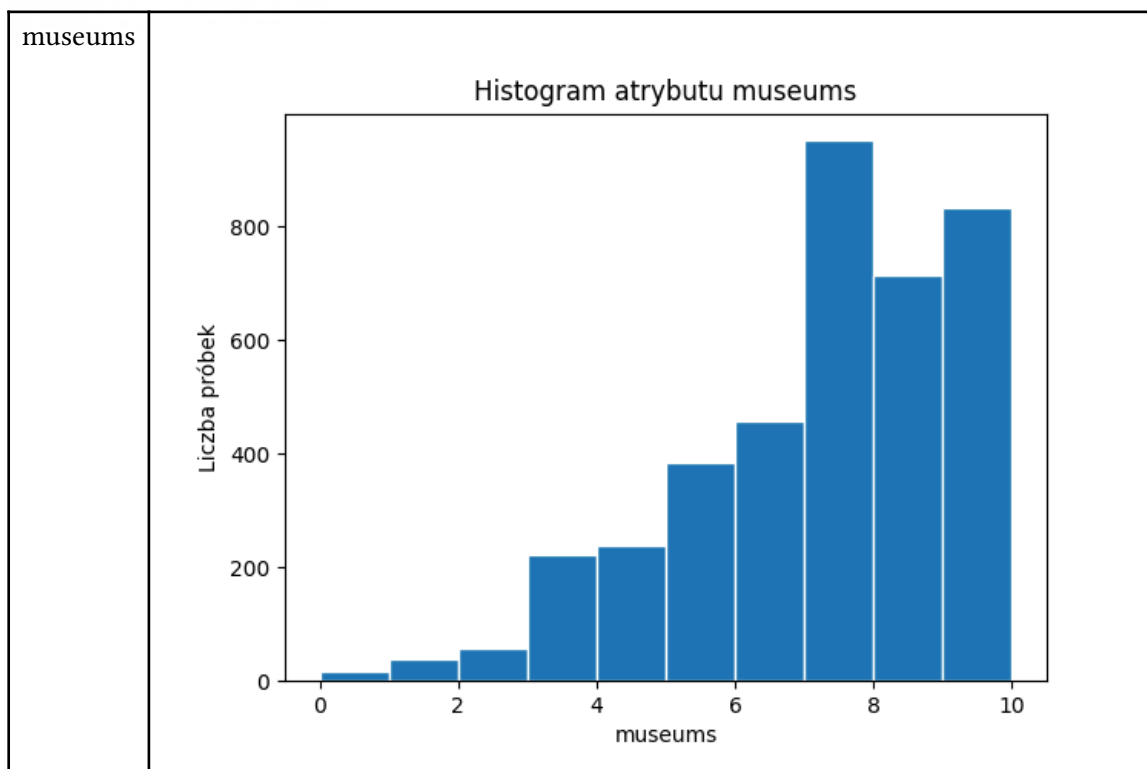


exercise

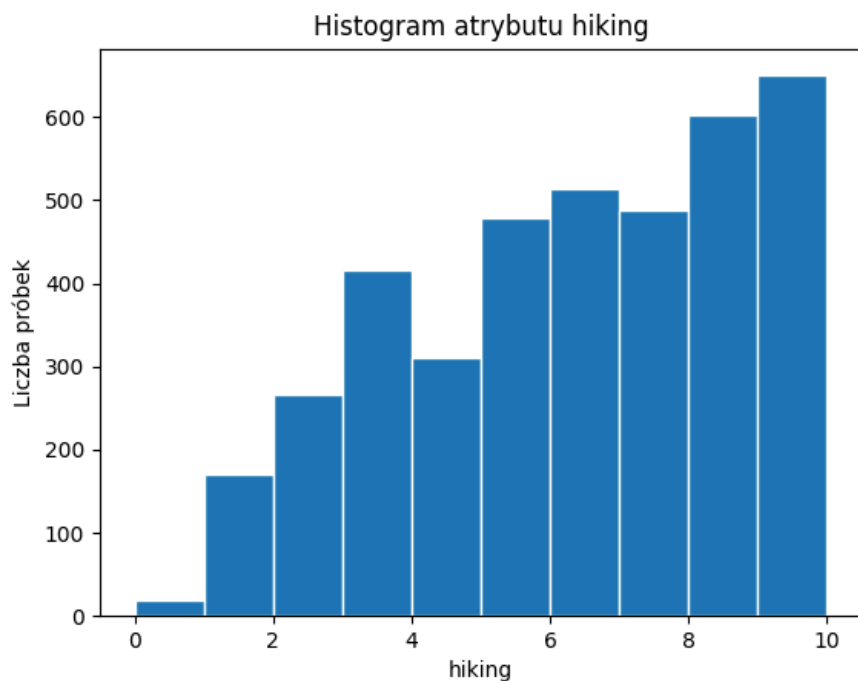


dining

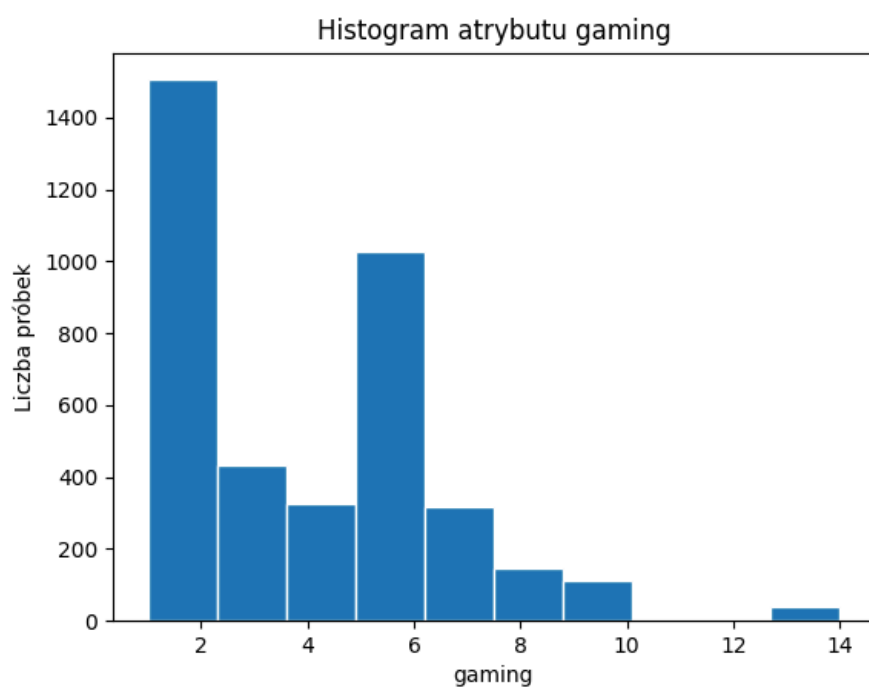




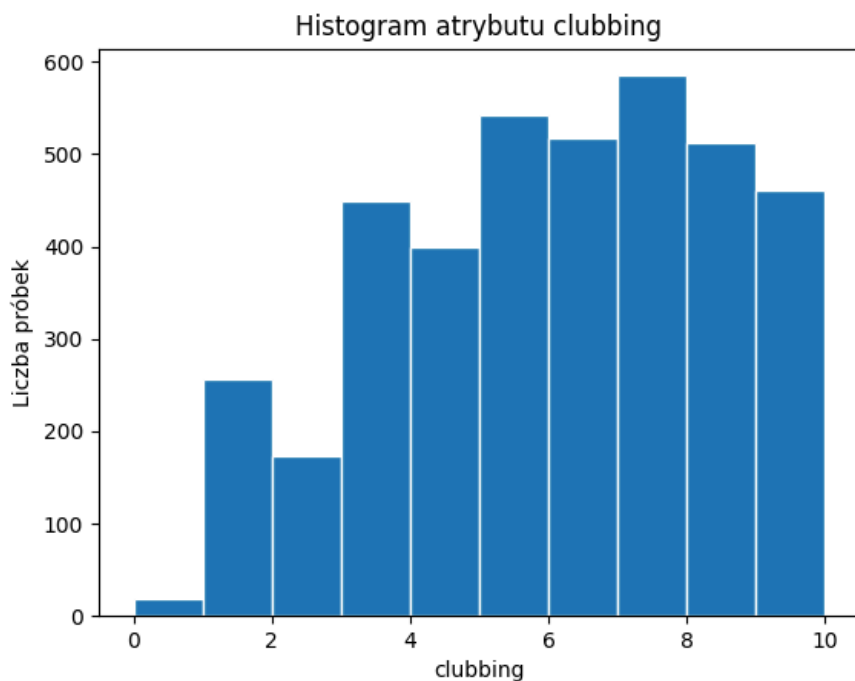
hiking



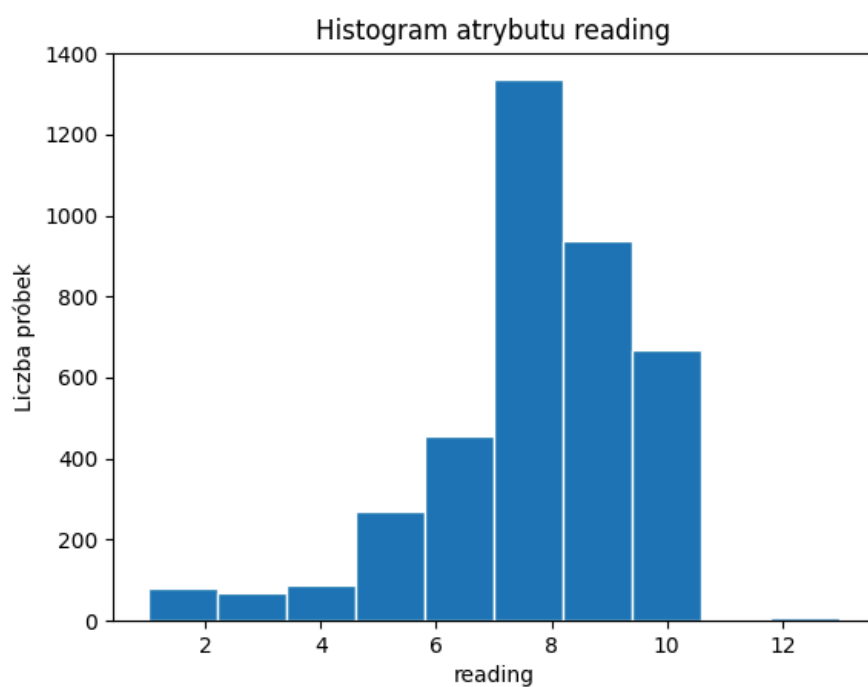
gaming

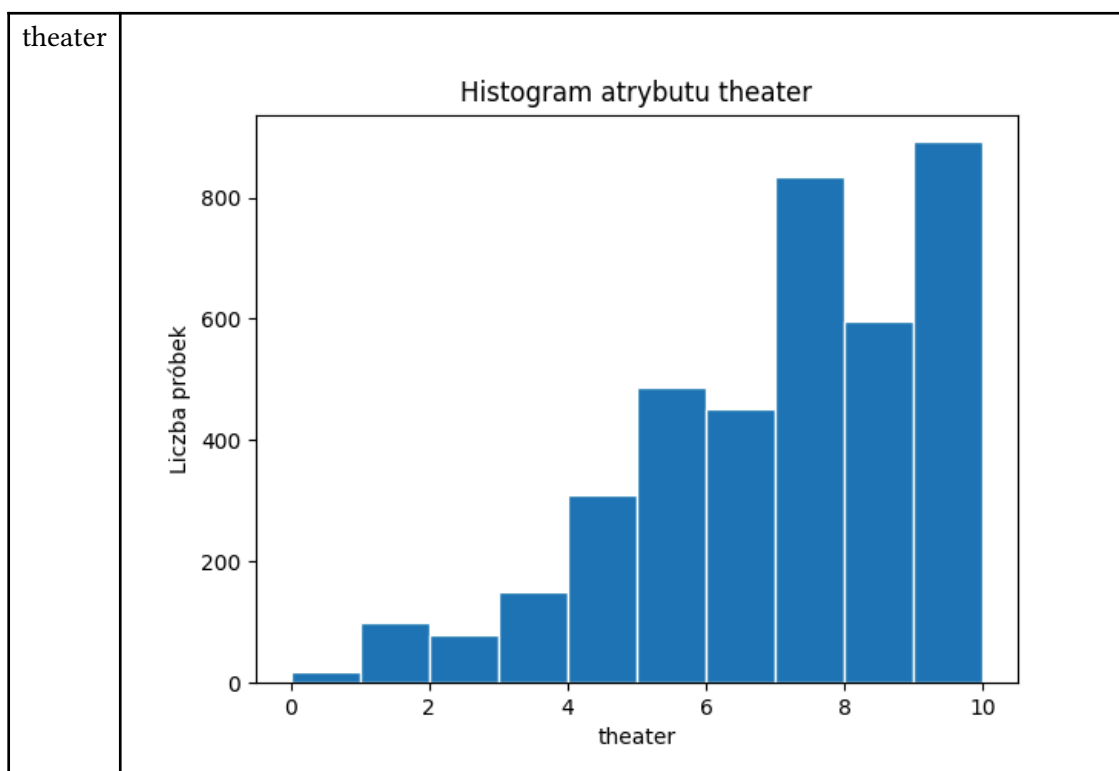
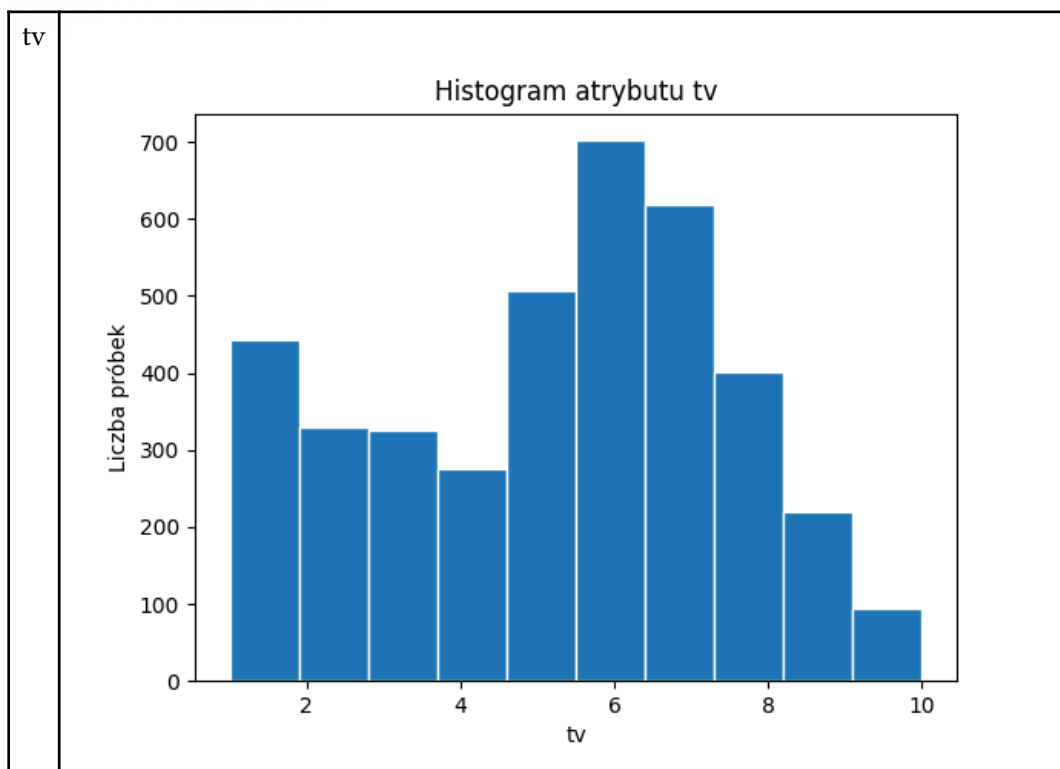


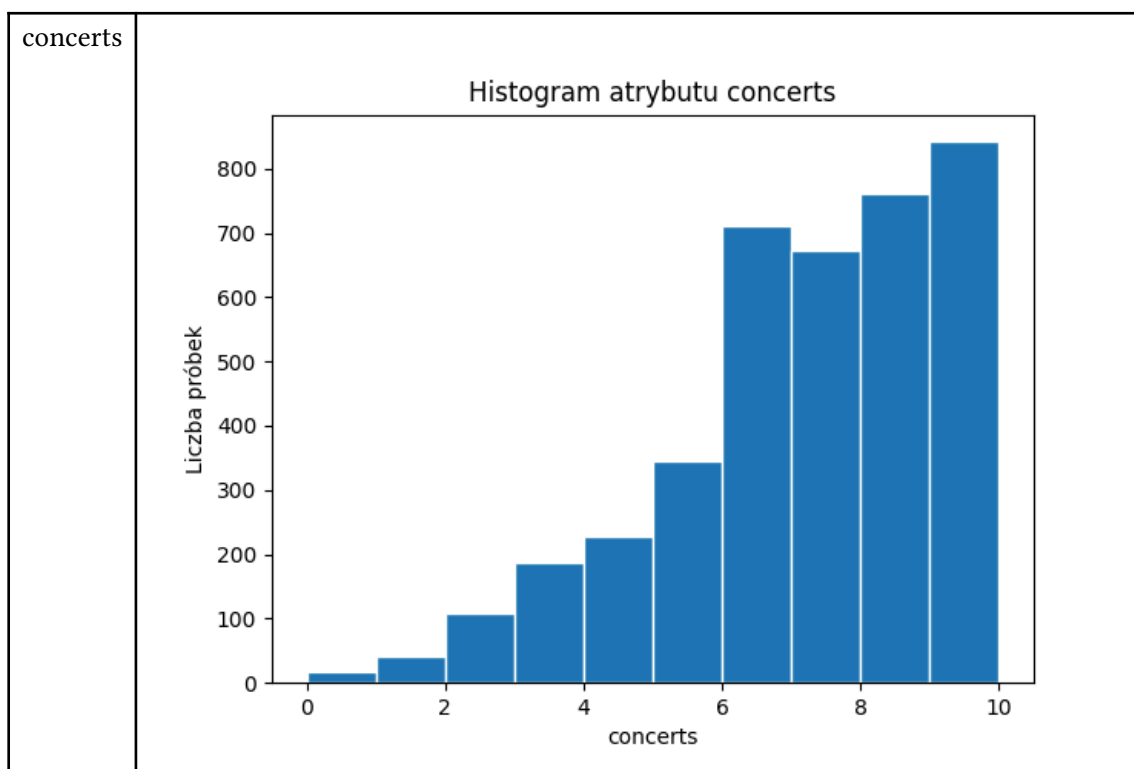
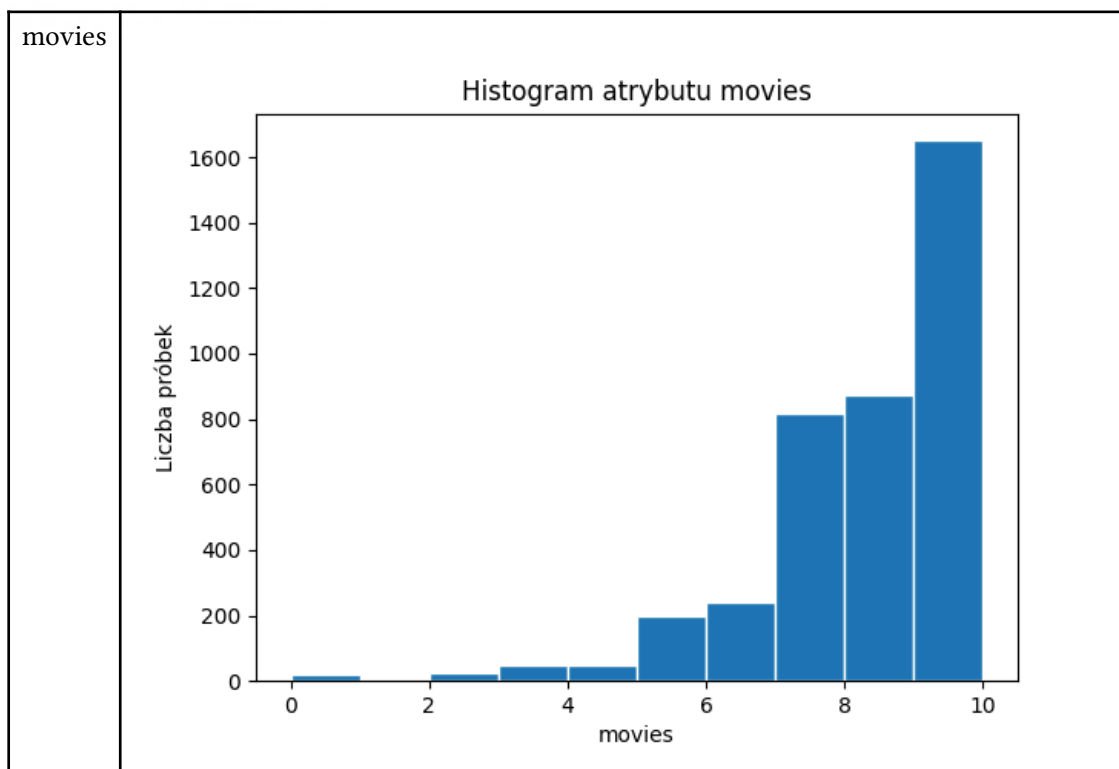
clubbing



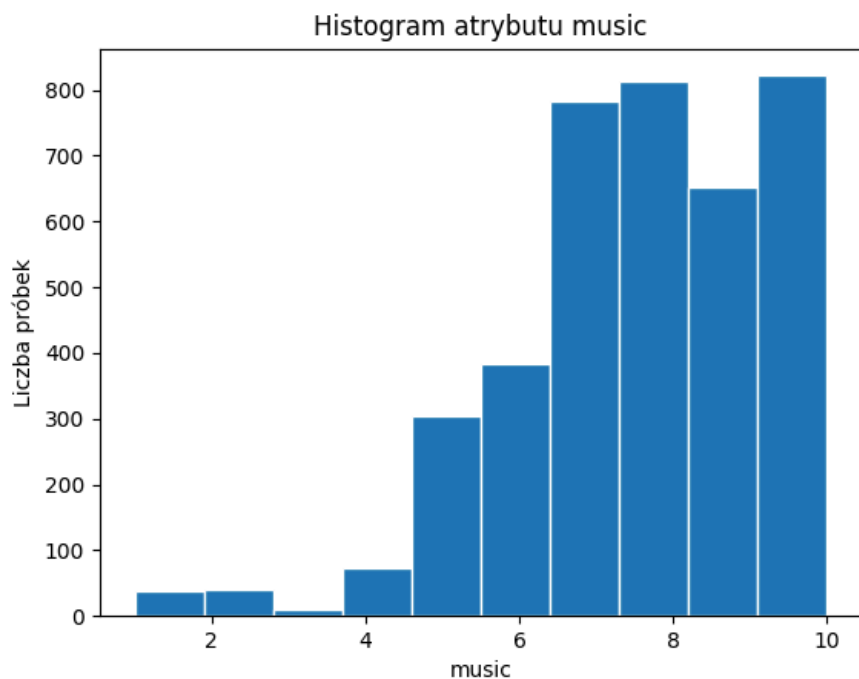
reading



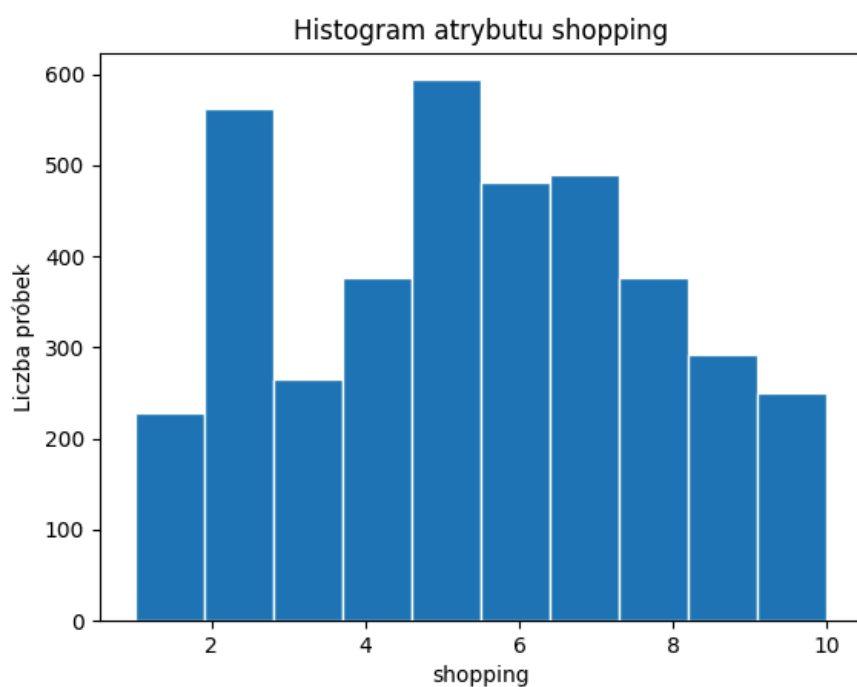




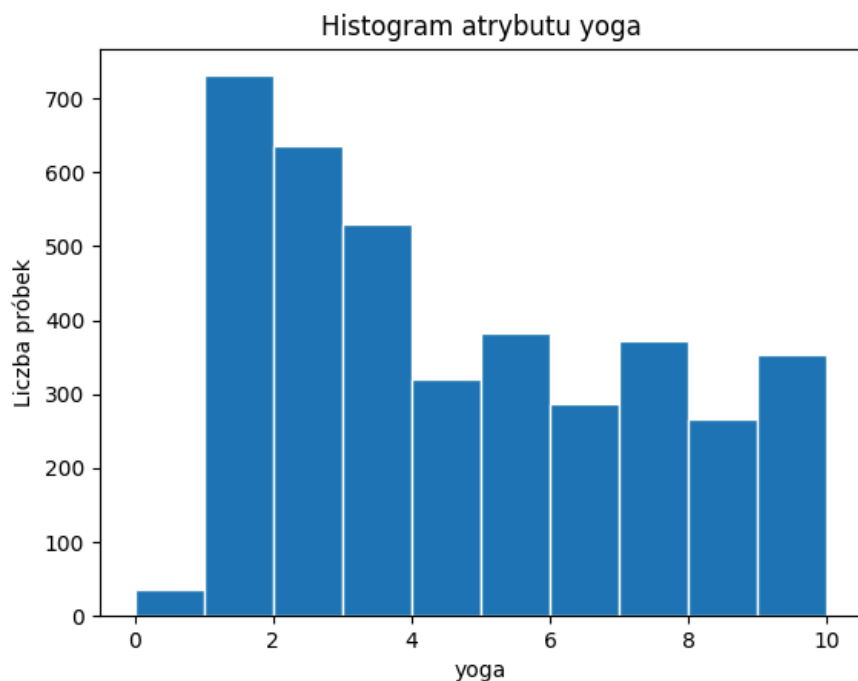
music



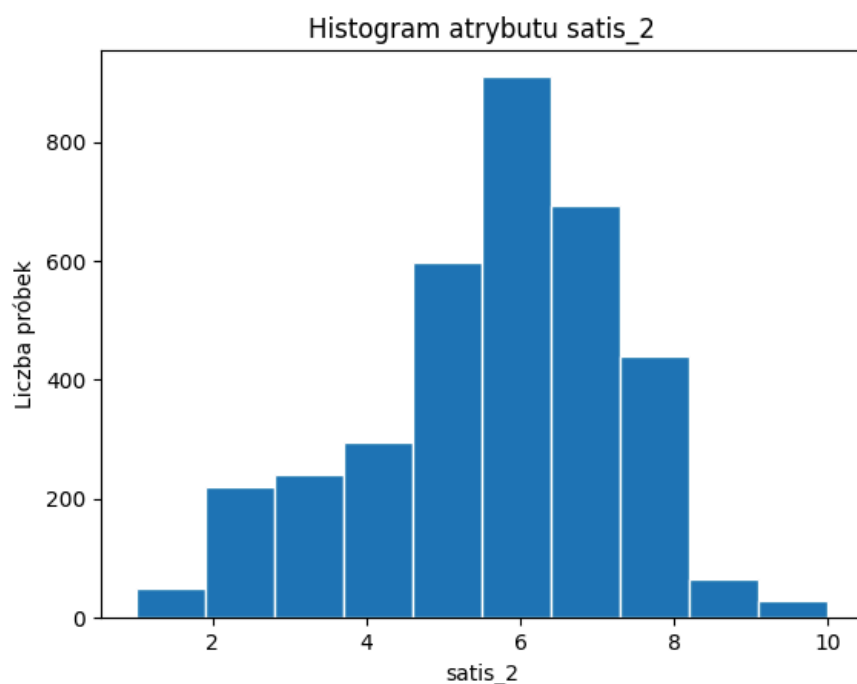
shopping



yoga

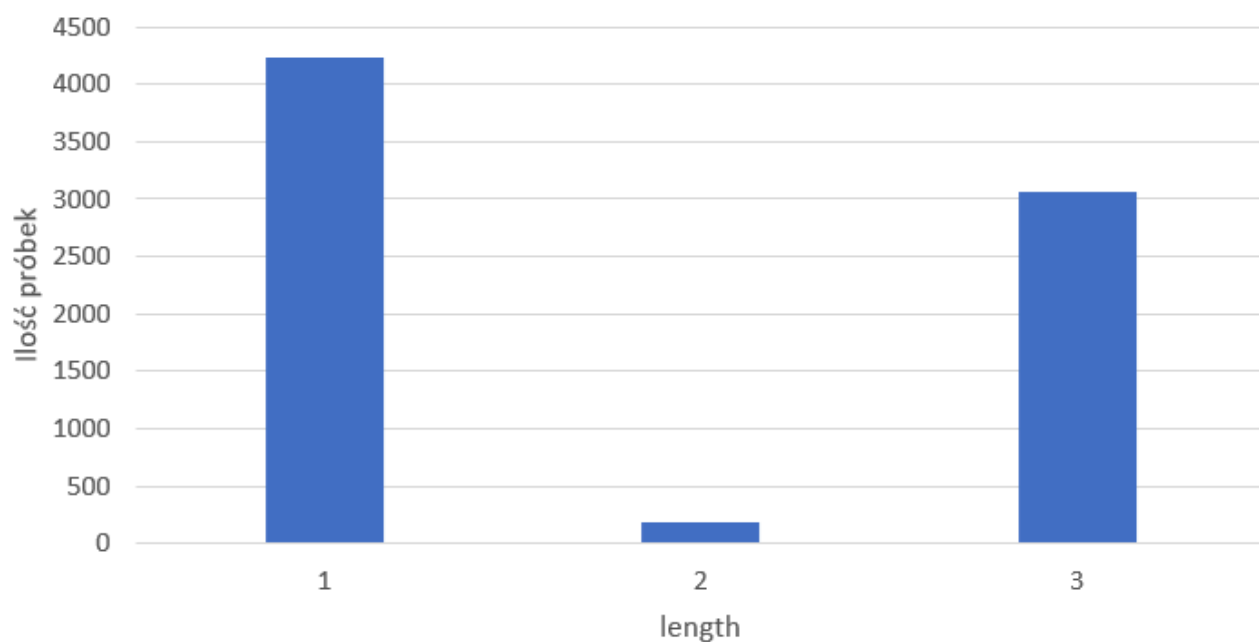


satis_2



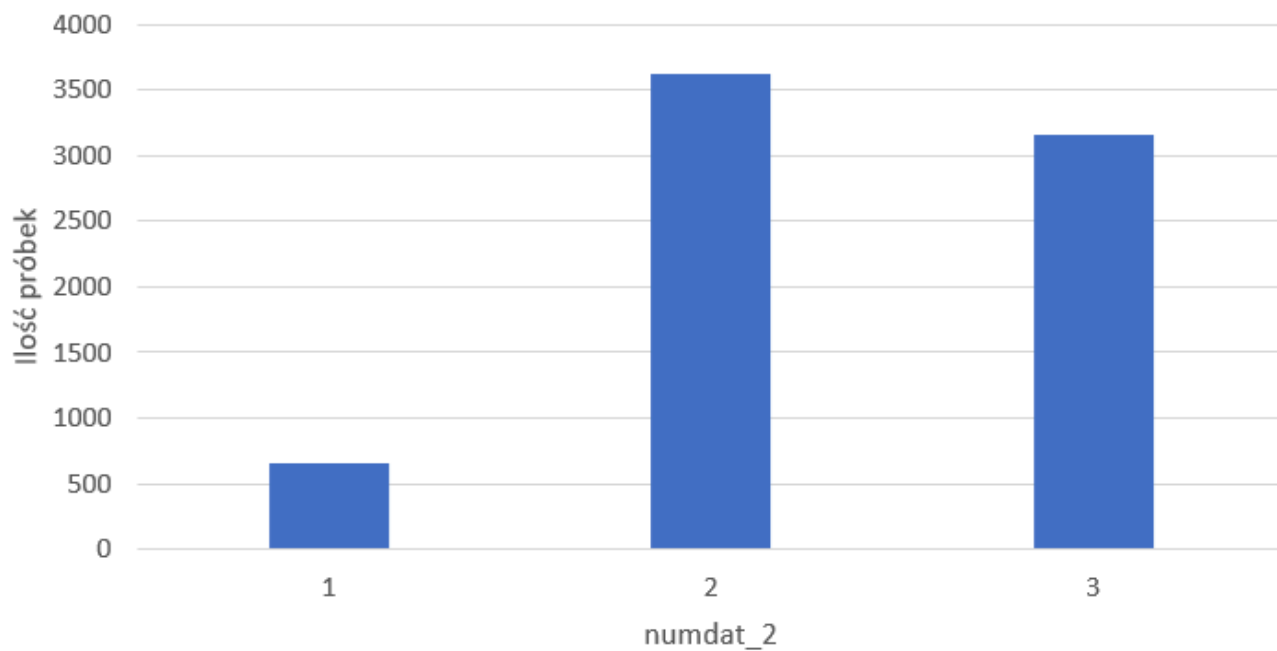
length

Histogram atrybutu length

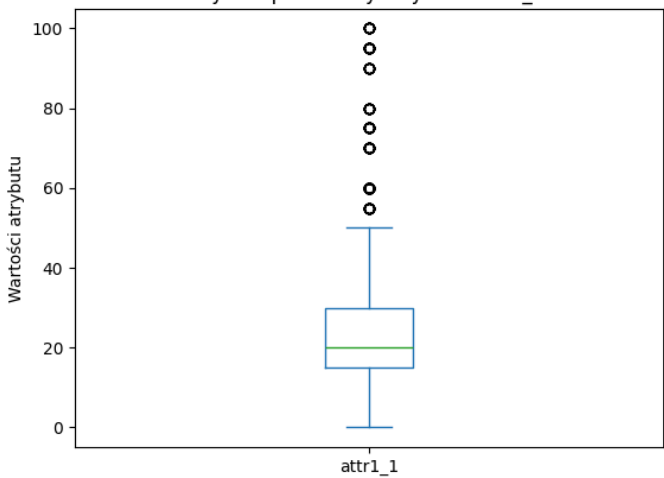
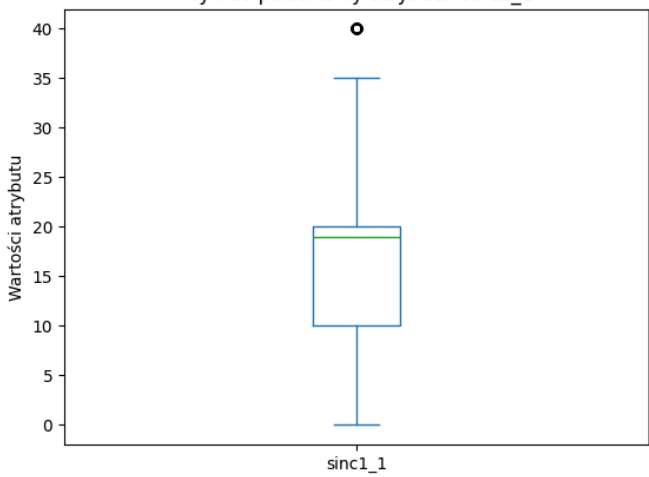


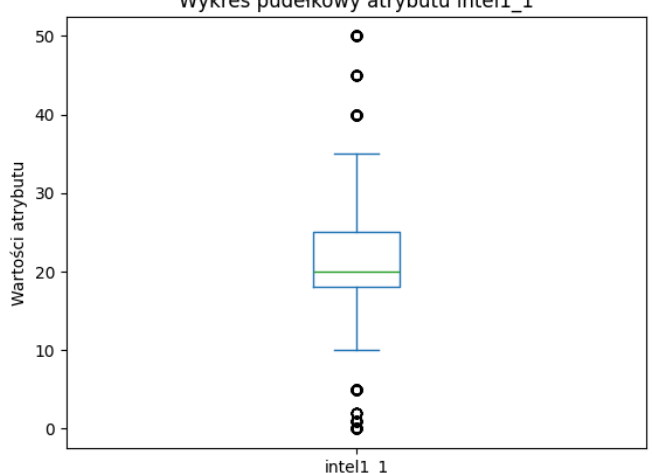
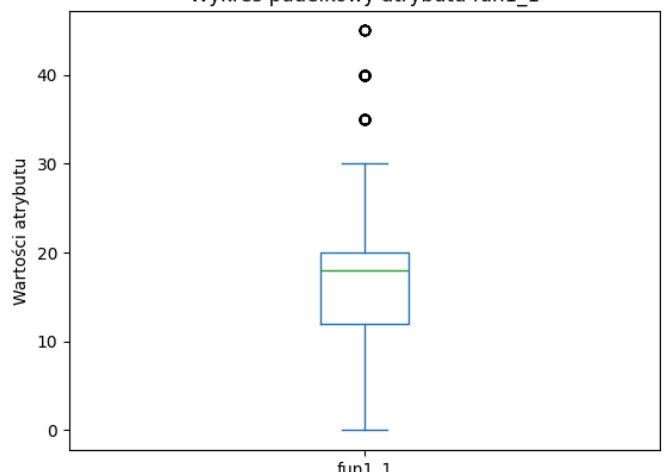
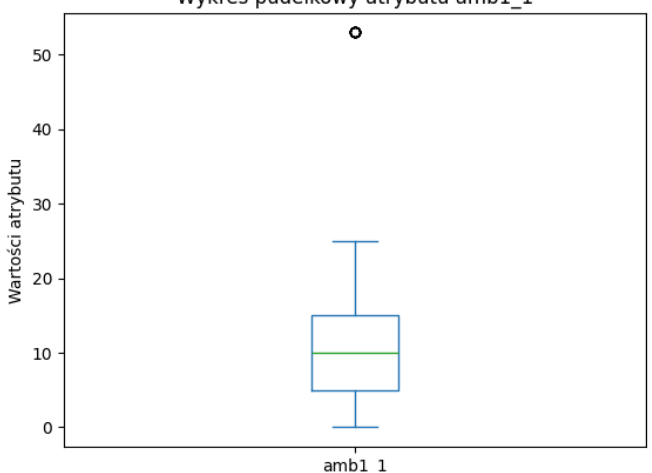
numdat_2

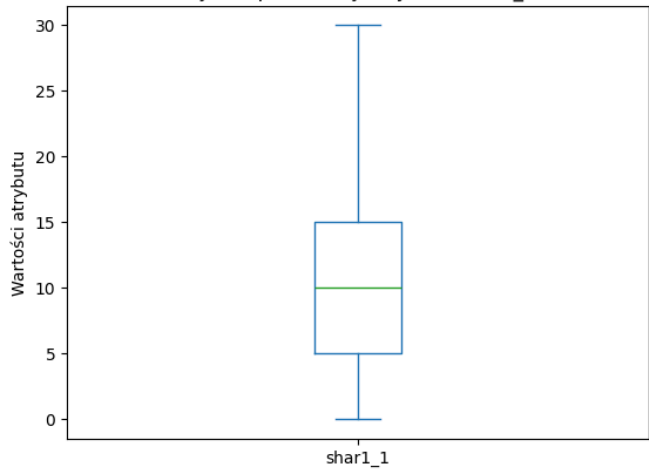
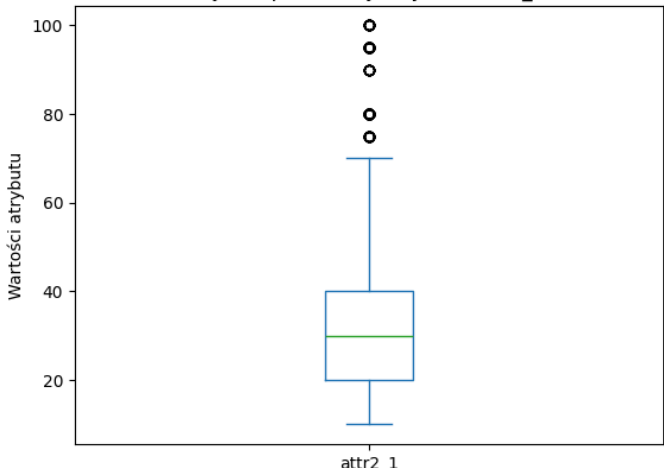
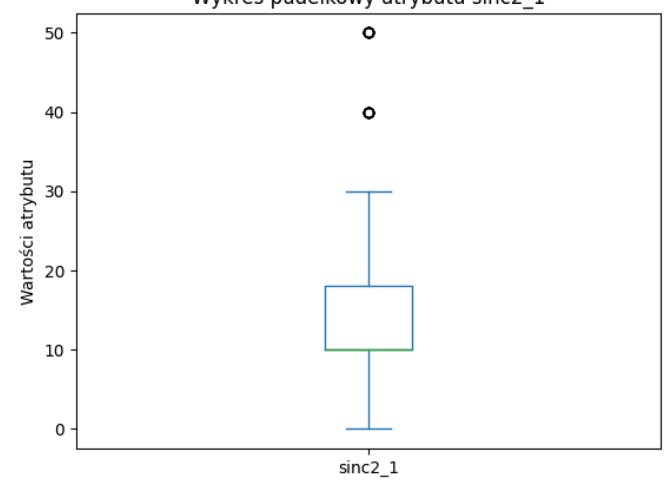
Histogram atrybutu numdat_2

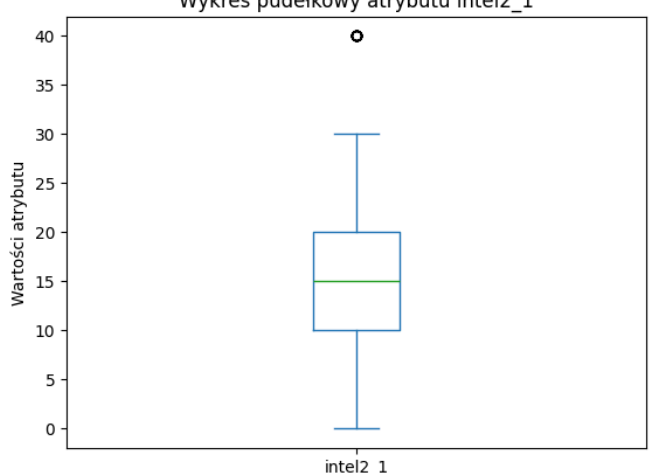
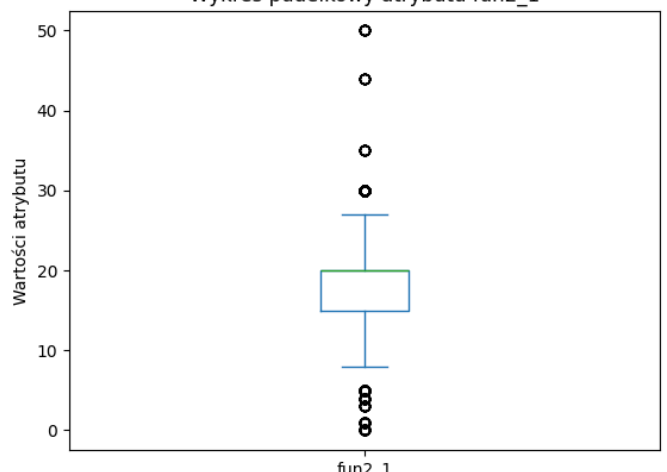
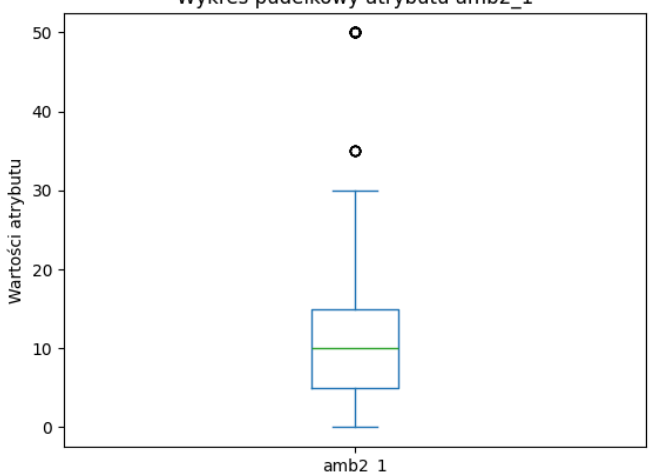


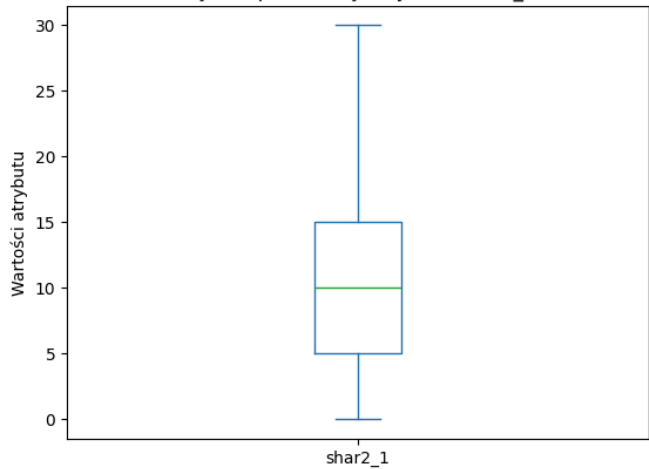
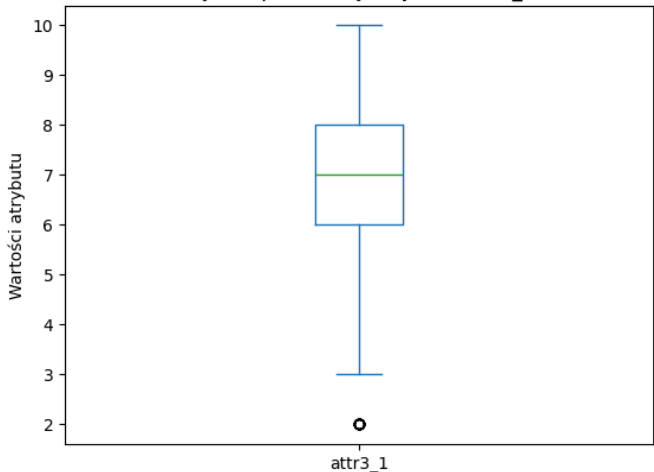
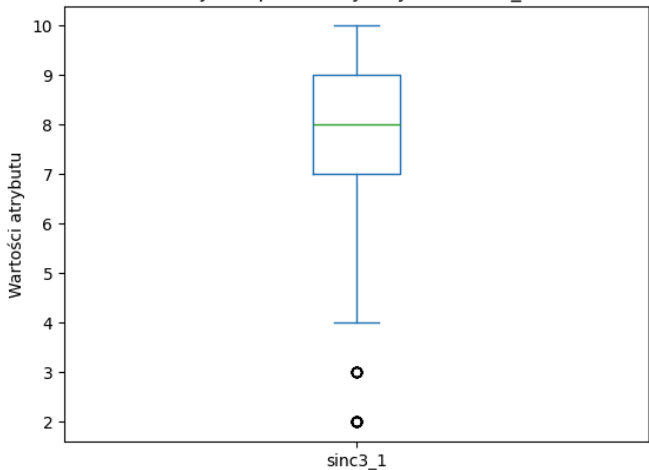
5.2. Punkty oddalone i rozkłady wartości atrybutów (wykresy pudełkowe)

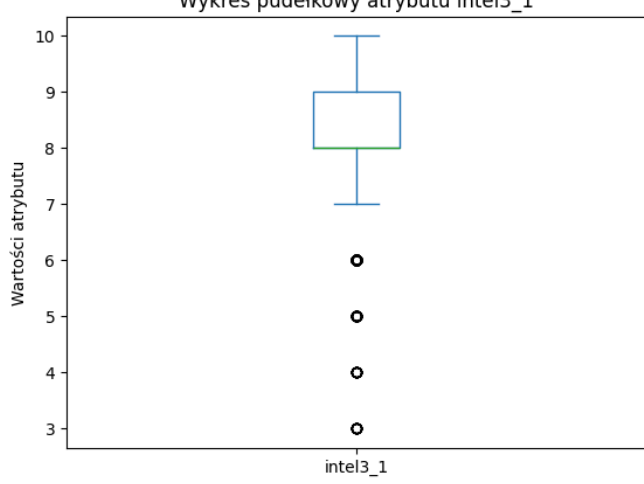
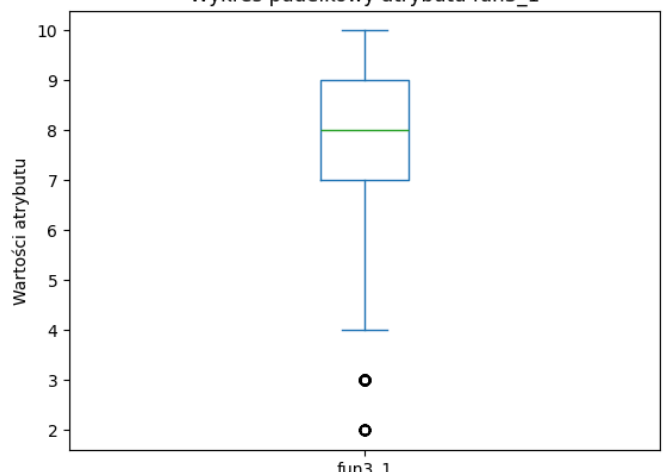
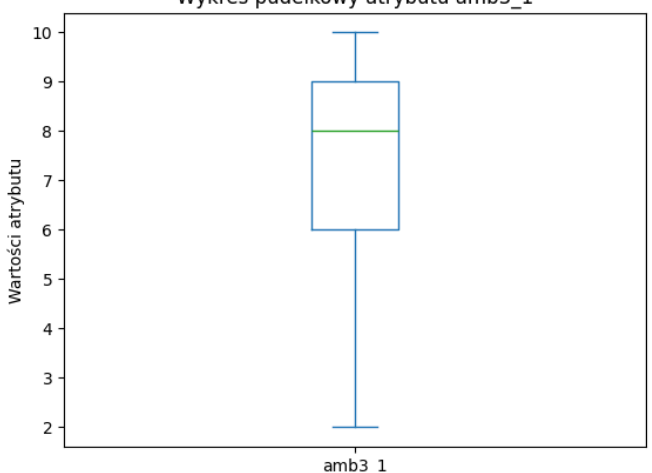
attr1_1	<p>Wykres pudełkowy atrybutu attr1_1</p> 	<p>Mediana: 20.0</p> <p>Przedział wartości występujących najczęściej: [15.0; 30.0]</p> <p>Liczba punktów oddalonych: 1265</p>
sinc1_1	<p>Wykres pudełkowy atrybutu sinc1_1</p> 	<p>Mediana: 19.0</p> <p>Przedział wartości występujących najczęściej: [10.0; 20.0]</p> <p>Liczba punktów oddalonych: 1131</p>

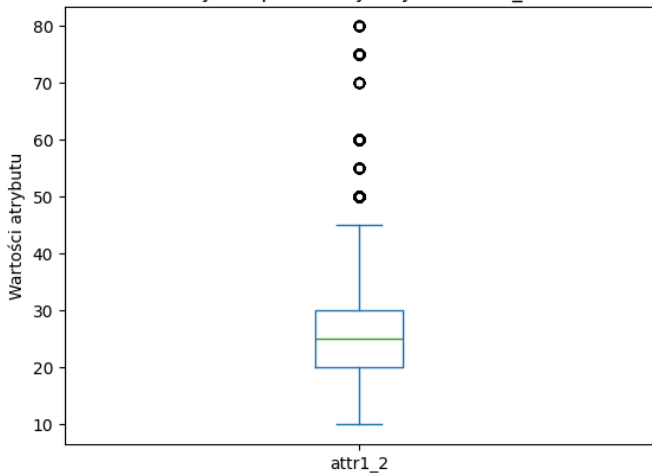
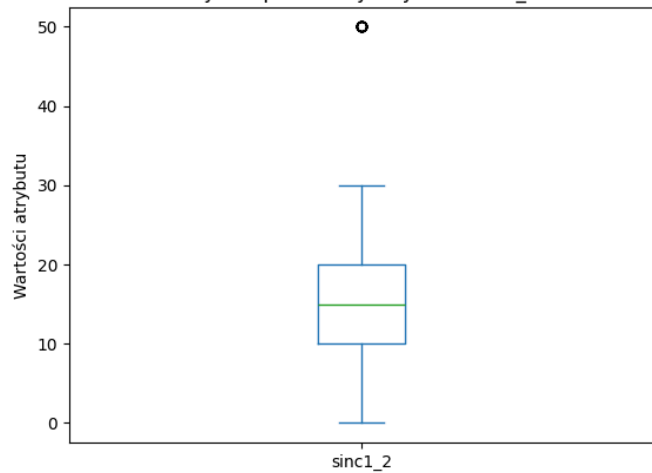
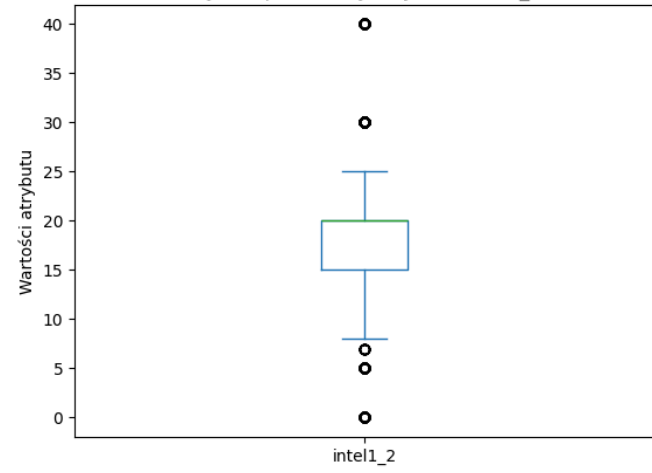
intel1_1	<p>Wykres pudełkowy atrybutu intel1_1</p> 	<p>Mediana: 20.0</p> <p>Przedział wartości występujących najczęściej: [18.0; 25.0]</p> <p>Liczba punktów oddalonych: 1514</p>
fun1_1	<p>Wykres pudełkowy atrybutu fun1_1</p> 	<p>Mediana: 18.0</p> <p>Przedział wartości występujących najczęściej: [12.0; 20.0]</p> <p>Liczba punktów oddalonych: 1694</p>
amb1_1	<p>Wykres pudełkowy atrybutu amb1_1</p> 	<p>Mediana: 10.0</p> <p>Przedział wartości występujących najczęściej: [5.0; 15.0]</p> <p>Liczba punktów oddalonych: 1125</p>

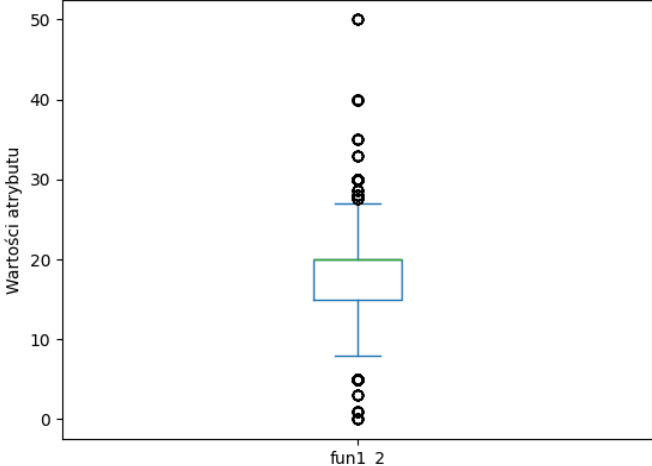
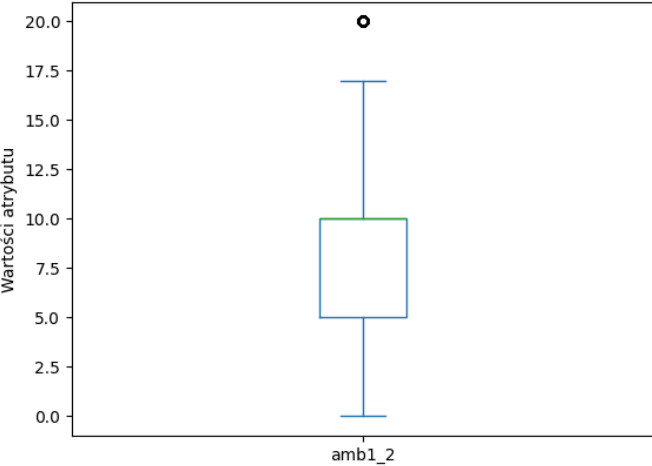
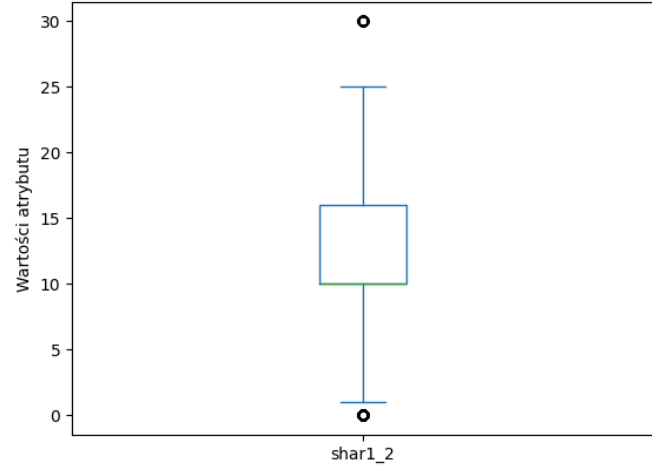
shar1_1	<p>Wykres pudełkowy atrybutu shar1_1</p> 	<p>Mediana: 10.0</p> <p>Przedział wartości występujących najczęściej: [5.0; 15.0]</p> <p>Liczba punktów oddalonych: 1311</p>
attr2_1	<p>Wykres pudełkowy atrybutu attr2_1</p> 	<p>Mediana: 30.0</p> <p>Przedział wartości występujących najczęściej: [20.0; 40.0]</p> <p>Liczba punktów oddalonych: 1464</p>
sinc2_1	<p>Wykres pudełkowy atrybutu sinc2_1</p> 	<p>Mediana: 10.0</p> <p>Przedział wartości występujących najczęściej: [10.0; 18.0]</p> <p>Liczba punktów oddalonych: 1924</p>

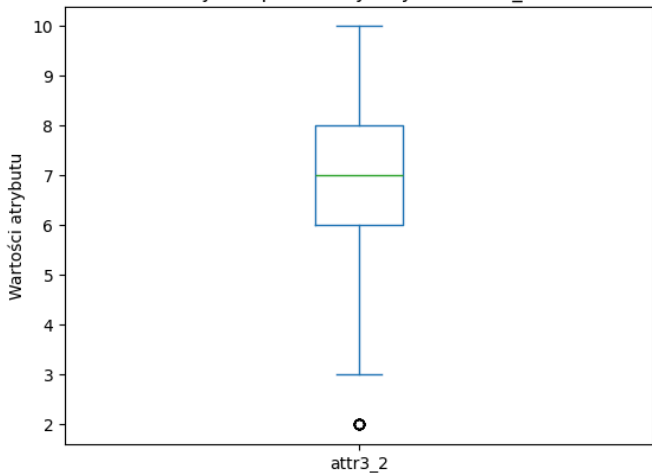
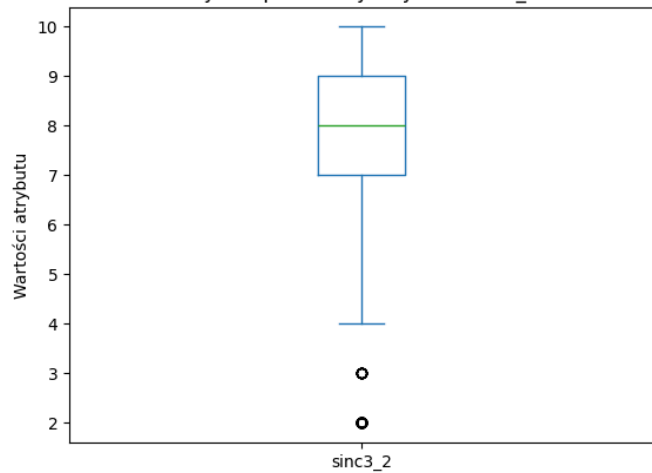
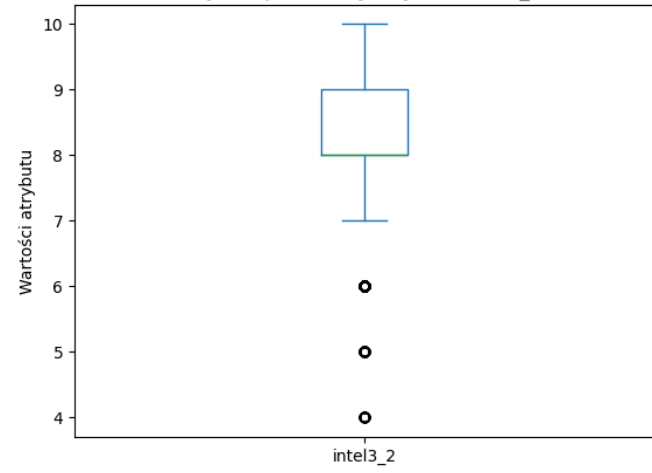
intel2_1	<p>Wykres pudełkowy atrybutu intel2_1</p> 	<p>Mediana: 15.0</p> <p>Przedział wartości występujących najczęściej: [10.0; 20.0]</p> <p>Liczba punktów oddalonych: 872</p>
fun2_1	<p>Wykres pudełkowy atrybutu fun2_1</p> 	<p>Mediana: 20.0</p> <p>Przedział wartości występujących najczęściej: [15.0; 20.0]</p> <p>Liczba punktów oddalonych: 1766</p>
amb2_1	<p>Wykres pudełkowy atrybutu amb2_1</p> 	<p>Mediana: 10.0</p> <p>Przedział wartości występujących najczęściej: [5.0; 15.0]</p> <p>Liczba punktów oddalonych: 1168</p>

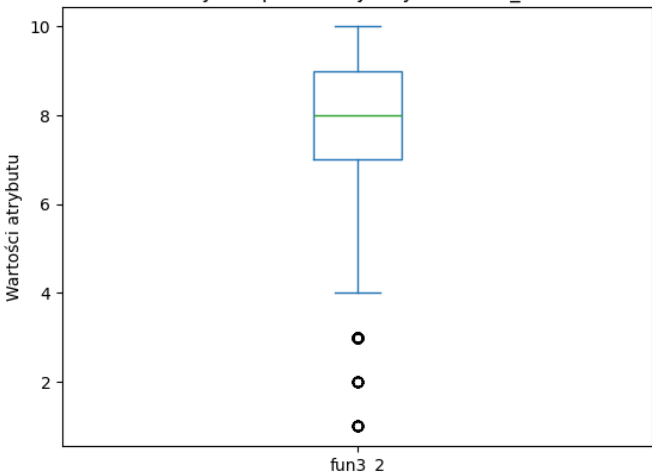
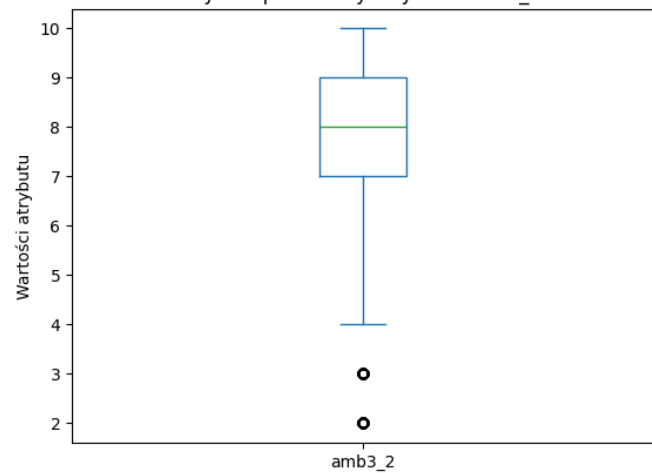
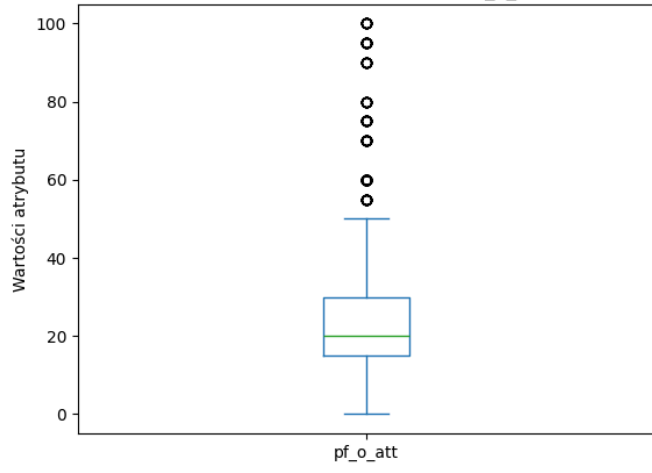
shar2_1	<p>Wykres pudełkowy atrybutu shar2_1</p> 	<p>Mediana: 10.0</p> <p>Przedział wartości występujących najczęściej: [5.0; 15.0]</p> <p>Liczba punktów oddalonych: 1057</p>
attr3_1	<p>Wykres pudełkowy atrybutu attr3_1</p> 	<p>Mediana: 7.0</p> <p>Przedział wartości występujących najczęściej: [6.0; 8.0]</p> <p>Liczba punktów oddalonych: 973</p>
sinc3_1	<p>Wykres pudełkowy atrybutu sinc3_1</p> 	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [7.0; 9.0]</p> <p>Liczba punktów oddalonych: 1021</p>

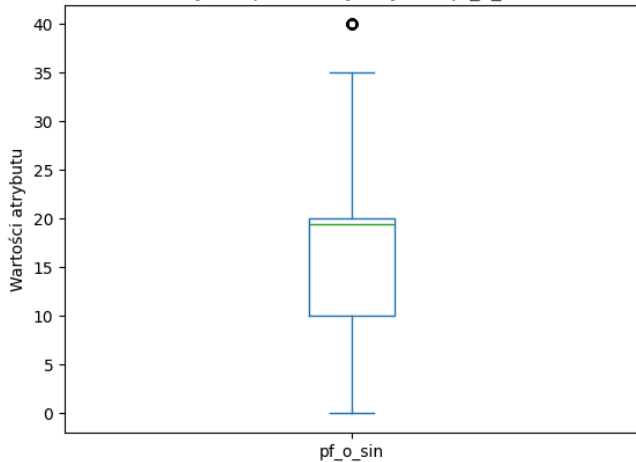
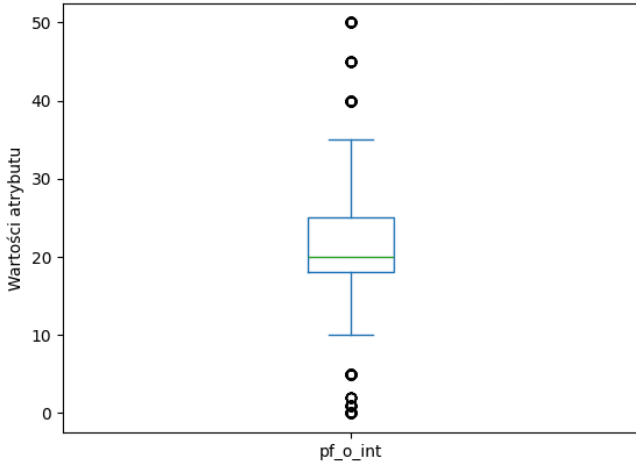
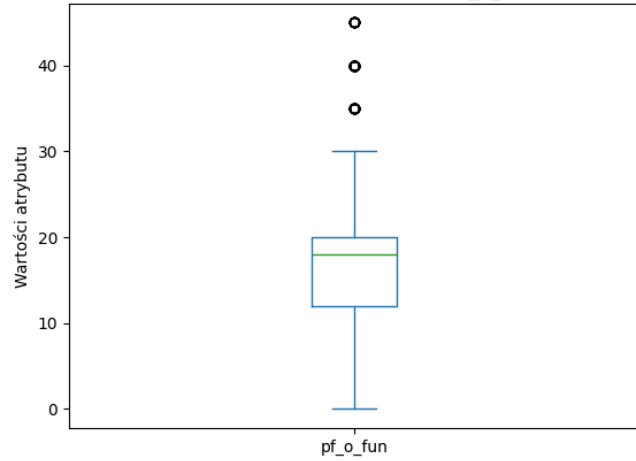
intel3_1	<p>Wykres pudełkowy atrybutu intel3_1</p>  <p>intel3_1</p>	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [8.0; 9.0]</p> <p>Liczba punktów oddalonych: 1343</p>
fun3_1	<p>Wykres pudełkowy atrybutu fun3_1</p>  <p>fun3_1</p>	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [7.0; 9.0]</p> <p>Liczba punktów oddalonych: 1244</p>
amb3_1	<p>Wykres pudełkowy atrybutu amb3_1</p>  <p>amb3_1</p>	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [6.0; 9.0]</p> <p>Liczba punktów oddalonych: 1084</p>

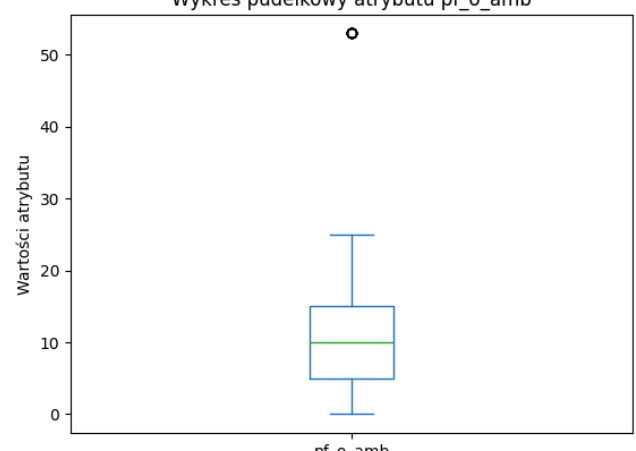
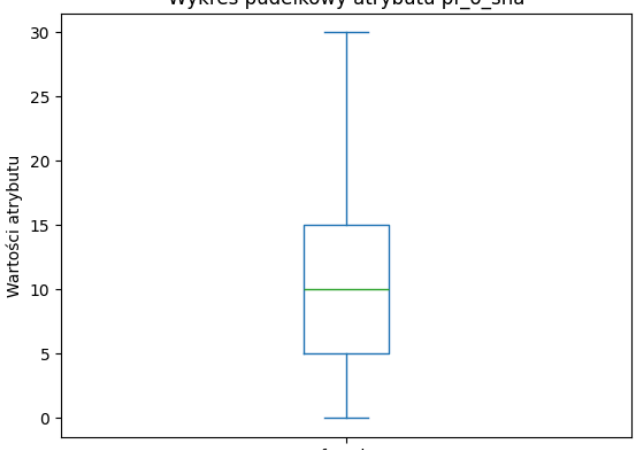
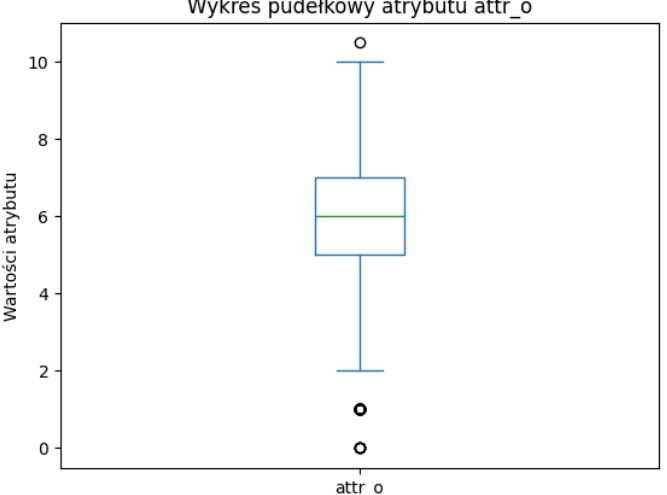
attr1_2	<p>Wykres pudełkowy atrybutu attr1_2</p>  <p>attr1_2</p>	<p>Mediana: 25.0</p> <p>Przedział wartości występujących najczęściej: [20.0; 30.0]</p> <p>Liczba punktów oddalonych: 1645</p>
sinc1_2	<p>Wykres pudełkowy atrybutu sinc1_2</p>  <p>sinc1_2</p>	<p>Mediana: 15.0</p> <p>Przedział wartości występujących najczęściej: [10.0; 20.0]</p> <p>Liczba punktów oddalonych: 864</p>
intel1_2	<p>Wykres pudełkowy atrybutu intel1_2</p>  <p>intel1_2</p>	<p>Mediana: 20.0</p> <p>Przedział wartości występujących najczęściej: [15.0; 20.0]</p> <p>Liczba punktów oddalonych: 1465</p>

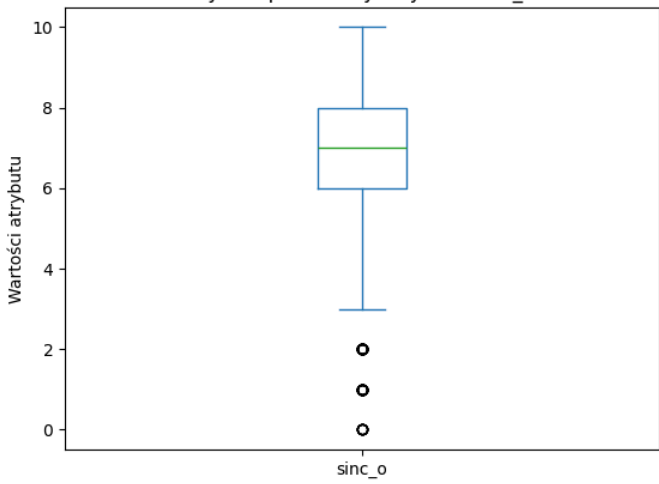
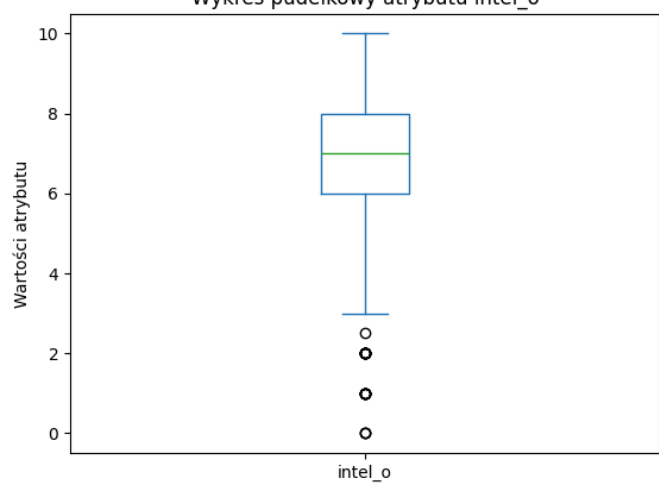
fun1_2	<p data-bbox="395 277 750 304">Wykres pudełkowy atrybutu fun1_2</p> 	<p data-bbox="1161 235 1331 262">Mediana: 20.0</p> <p data-bbox="971 288 1525 358">Przedział wartości występujących najczęściej: [15.0; 20.0]</p> <p data-bbox="1040 383 1453 409">Liczba punktów oddalonych: 1400</p>
amb1_2	<p data-bbox="395 909 758 936">Wykres pudełkowy atrybutu amb1_2</p> 	<p data-bbox="1161 866 1331 893">Mediana: 10.0</p> <p data-bbox="971 920 1525 990">Przedział wartości występujących najczęściej: [5.0; 10.0]</p> <p data-bbox="1040 1014 1453 1041">Liczba punktów oddalonych: 1366</p>
shar1_2	<p data-bbox="395 1532 758 1559">Wykres pudełkowy atrybutu shar1_2</p> 	<p data-bbox="1161 1489 1331 1516">Mediana: 10.0</p> <p data-bbox="971 1543 1525 1612">Przedział wartości występujących najczęściej: [10.0; 16.0]</p> <p data-bbox="1040 1637 1453 1664">Liczba punktów oddalonych: 1731</p>

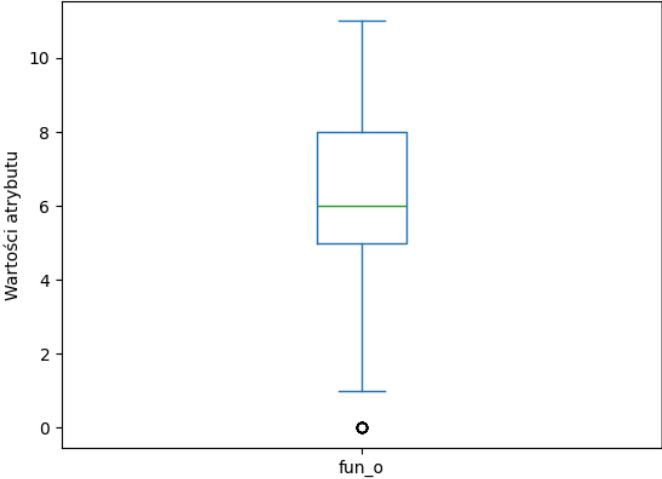
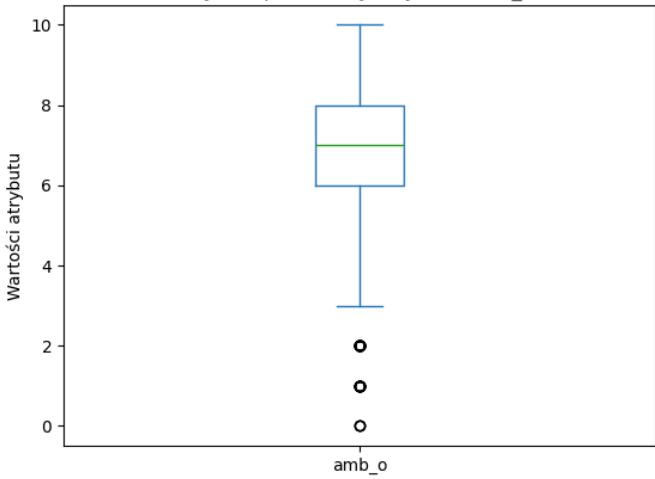
attr3_2	<p>Wykres pudełkowy atrybutu attr3_2</p>  <p>attr3_2</p>	<p>Mediana: 7.0</p> <p>Przedział wartości występujących najczęściej: [6.0; 8.0]</p> <p>Liczba punktów oddalonych: 857</p>
sinc3_2	<p>Wykres pudełkowy atrybutu sinc3_2</p>  <p>sinc3_2</p>	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [7.0; 9.0]</p> <p>Liczba punktów oddalonych: 859</p>
intel3_2	<p>Wykres pudełkowy atrybutu intel3_2</p>  <p>intel3_2</p>	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [8.0; 9.0]</p> <p>Liczba punktów oddalonych: 1165</p>

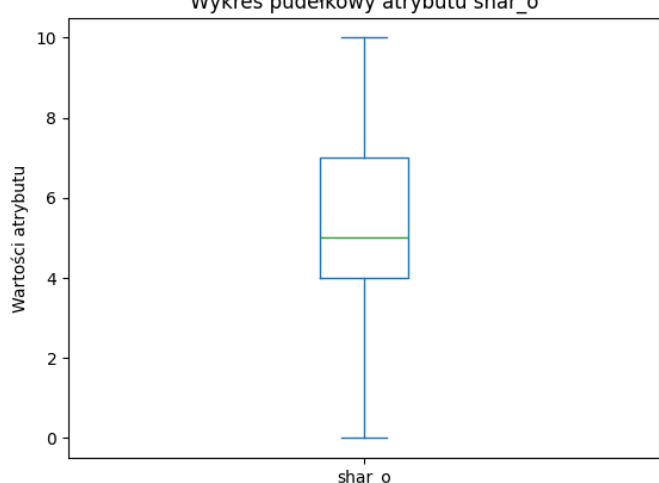
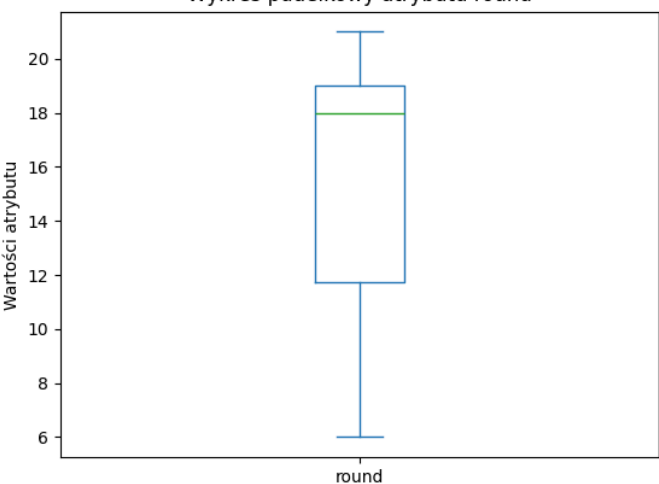
fun3_2	<p>Wykres pudełkowy atrybutu fun3_2</p> 	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [7.0; 9.0]</p> <p>Liczba punktów oddalonych: 948</p>
amb3_2	<p>Wykres pudełkowy atrybutu amb3_2</p> 	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [7.0; 9.0]</p> <p>Liczba punktów oddalonych: 1175</p>
pf_o_att	<p>Wykres pudełkowy atrybutu pf_o_att</p> 	<p>Mediana: 20.0</p> <p>Przedział wartości występujących najczęściej: [15.0; 30.0]</p> <p>Liczba punktów oddalonych: 1259</p>

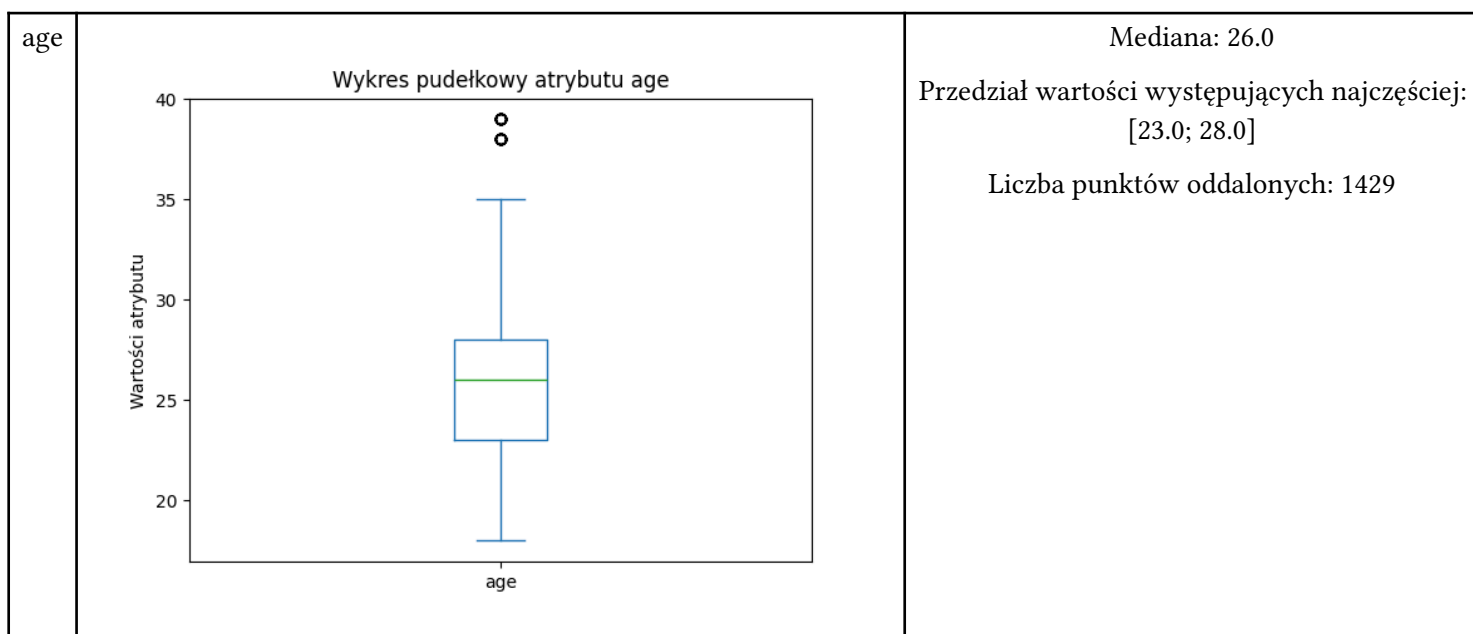
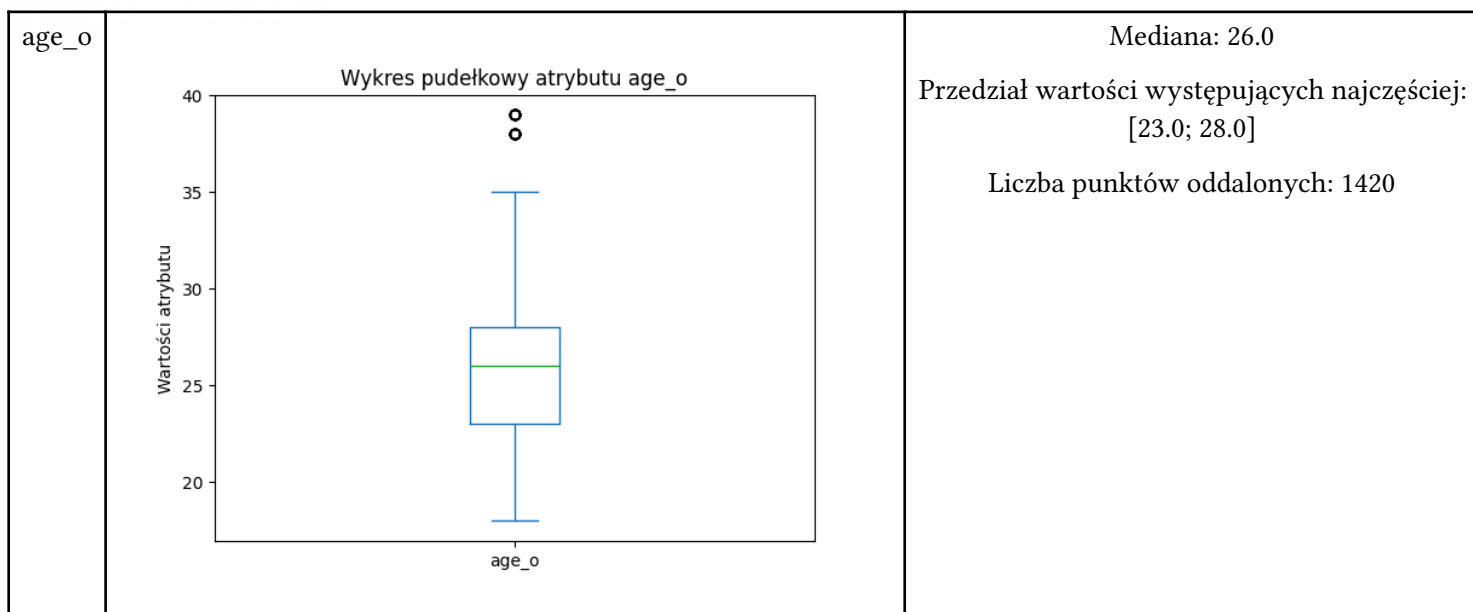
pf_o_sin	<p>Wykres pudełkowy atrybutu pf_o_sin</p> 	<p>Mediana: 19.5</p> <p>Przedział wartości występujących najczęściej: [10.0; 20.0]</p> <p>Liczba punktów oddalonych: 1129</p>
pf_o_int	<p>Wykres pudełkowy atrybutu pf_o_int</p> 	<p>Mediana: 20.0</p> <p>Przedział wartości występujących najczęściej: [18.0; 25.0]</p> <p>Liczba punktów oddalonych: 1509</p>
pf_o_fun	<p>Wykres pudełkowy atrybutu pf_o_fun</p> 	<p>Mediana: 18.0</p> <p>Przedział wartości występujących najczęściej: [12.0; 20.0]</p> <p>Liczba punktów oddalonych: 1691</p>

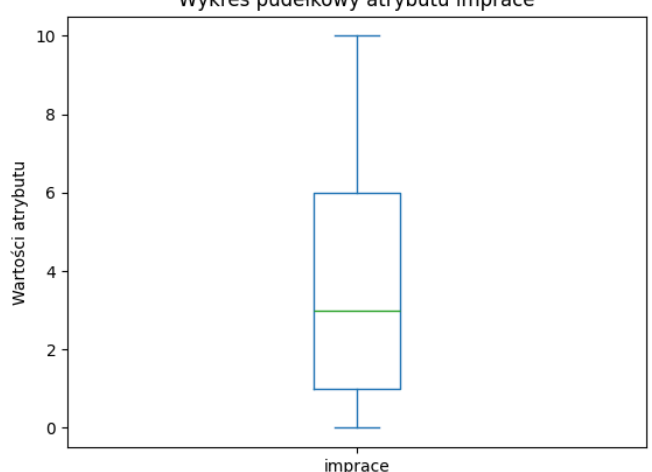
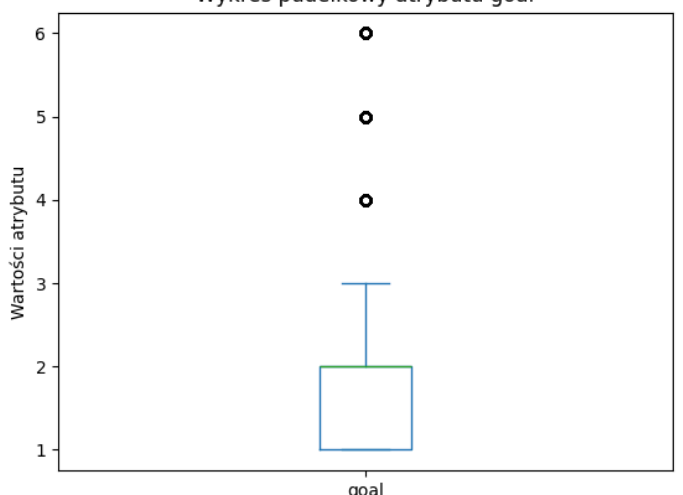
pf_o_amb	<p>Wykres pudełkowy atrybutu pf_o_amb</p> 	Mediana: 10.0 Przedział wartości występujących najczęściej: [5.0; 15.0] Liczba punktów oddalonych: 1125
pf_o_sha	<p>Wykres pudełkowy atrybutu pf_o_sha</p> 	Mediana: 10.0 Przedział wartości występujących najczęściej: [5.0; 15.0] Liczba punktów oddalonych: 1309
attr_o	<p>Wykres pudełkowy atrybutu attr_o</p> 	Mediana: 6.0 Przedział wartości występujących najczęściej: [5.0; 7.0] Liczba punktów oddalonych: 1651

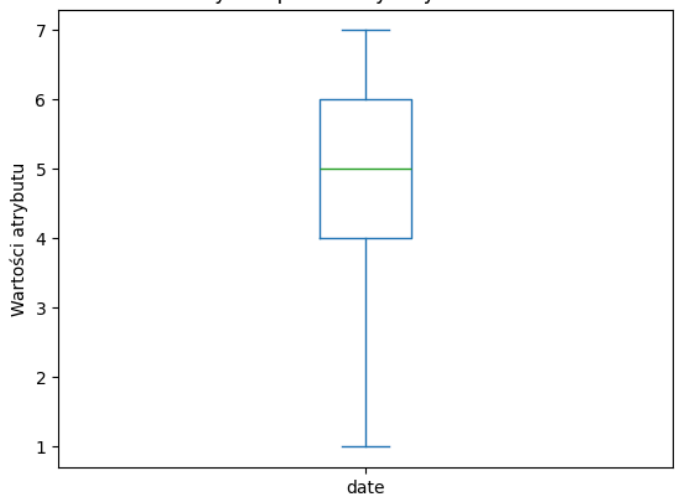
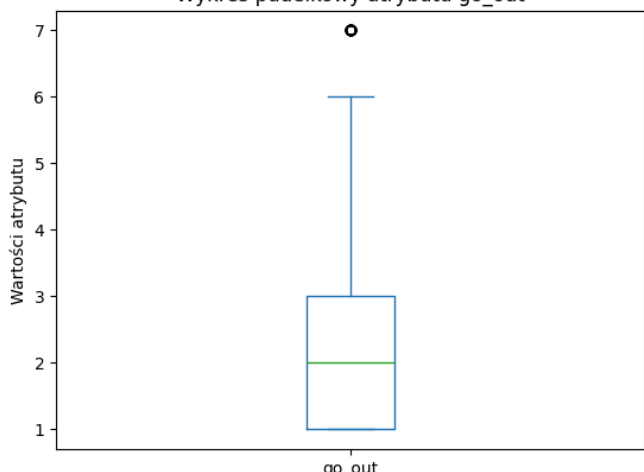
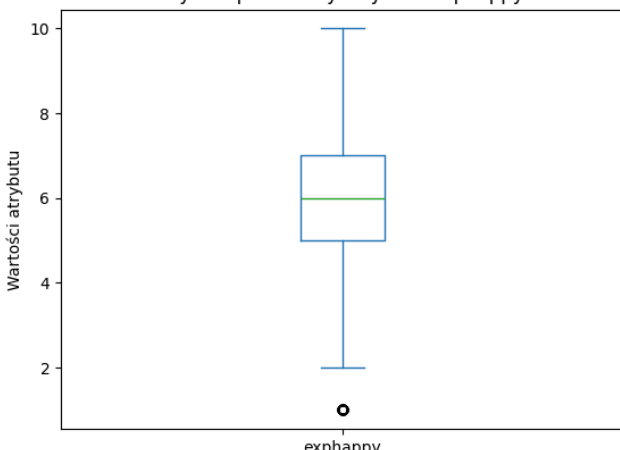
sinc_o	<p>Wykres pudełkowy atrybutu sinc_o</p>  <p>Wartości atrybutu</p> <p>sinc_o</p>	<p>Mediana: 7.0</p> <p>Przedział wartości występujących najczęściej: [6.0; 8.0]</p> <p>Liczba punktów oddalonych: 1404</p>
intel_o	<p>Wykres pudełkowy atrybutu intel_o</p>  <p>Wartości atrybutu</p> <p>intel_o</p>	<p>Mediana: 7.0</p> <p>Przedział wartości występujących najczęściej: [6.0; 8.0]</p> <p>Liczba punktów oddalonych: 1273</p>

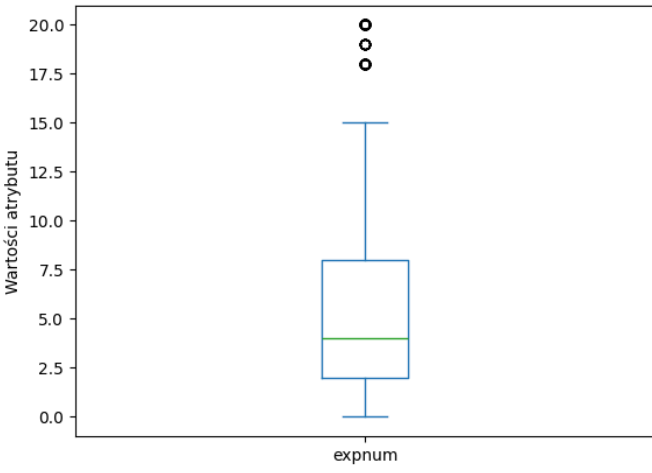
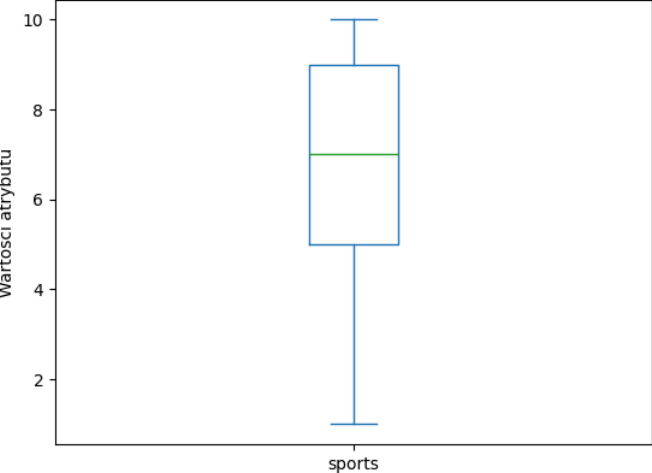
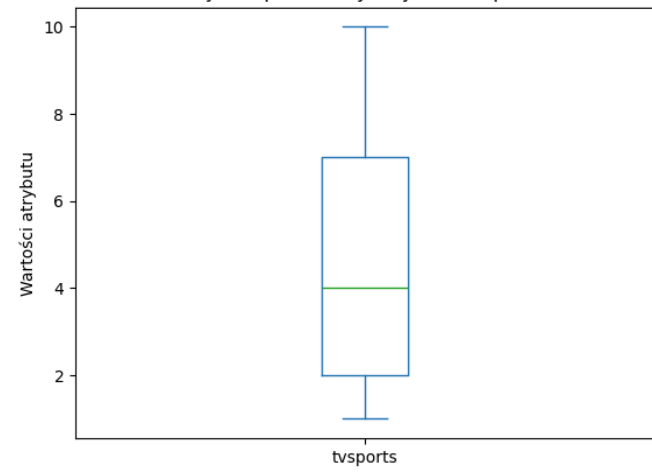
fun_o	<p data-bbox="395 277 738 304">Wykres pudełkowy atrybutu fun_o</p>  <p data-bbox="544 763 592 786">fun_o</p>	<p data-bbox="1171 237 1326 264">Mediana: 6.0</p> <p data-bbox="973 293 1525 360">Przedział wartości występujących najczęściej: [5.0; 8.0]</p> <p data-bbox="1042 383 1453 409">Liczba punktów oddalonych: 1082</p>
amb_o	<p data-bbox="395 916 748 943">Wykres pudełkowy atrybutu amb_o</p>  <p data-bbox="544 1397 600 1420">amb_o</p>	<p data-bbox="1171 875 1326 902">Mediana: 7.0</p> <p data-bbox="973 931 1525 999">Przedział wartości występujących najczęściej: [6.0; 8.0]</p> <p data-bbox="1042 1021 1453 1048">Liczba punktów oddalonych: 1488</p>

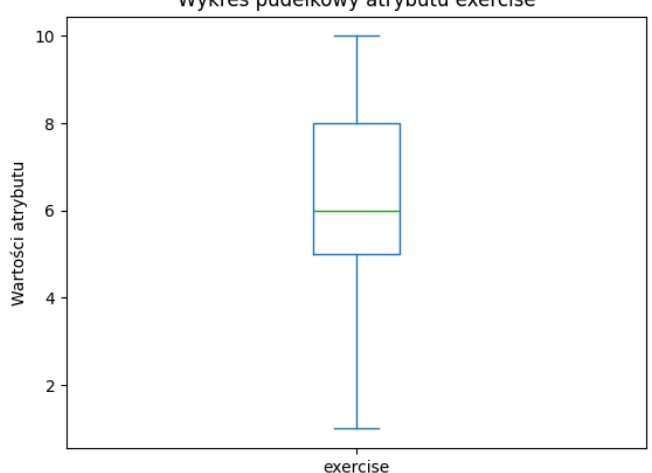
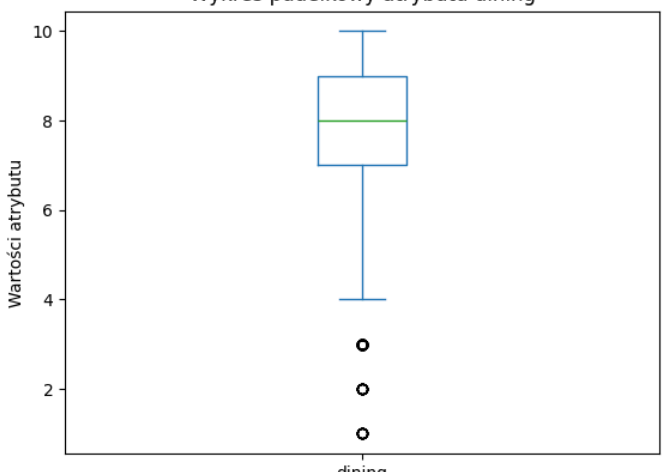
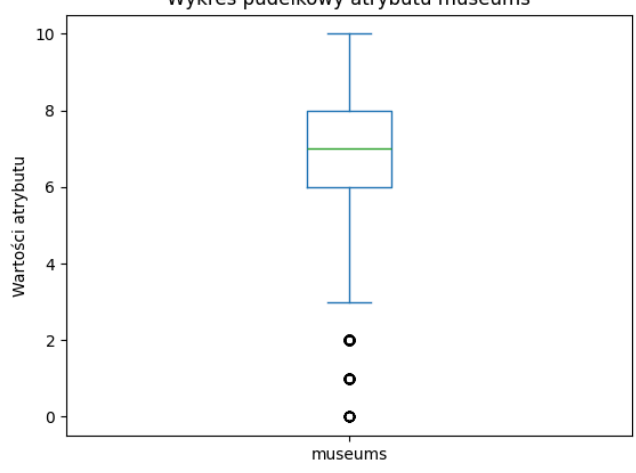
shar_o	<p>Wykres pudełkowy atrybutu shar_o</p> 	<p>Mediana: 5.0</p> <p>Przedział wartości występujących najczęściej: [4.0; 7.0]</p> <p>Liczba punktów oddalonych: 1253</p>
round	<p>Wykres pudełkowy atrybutu round</p> 	<p>Mediana: 18.0</p> <p>Przedział wartości występujących najczęściej: [11.75; 19.0]</p> <p>Liczba punktów oddalonych: 1870</p>

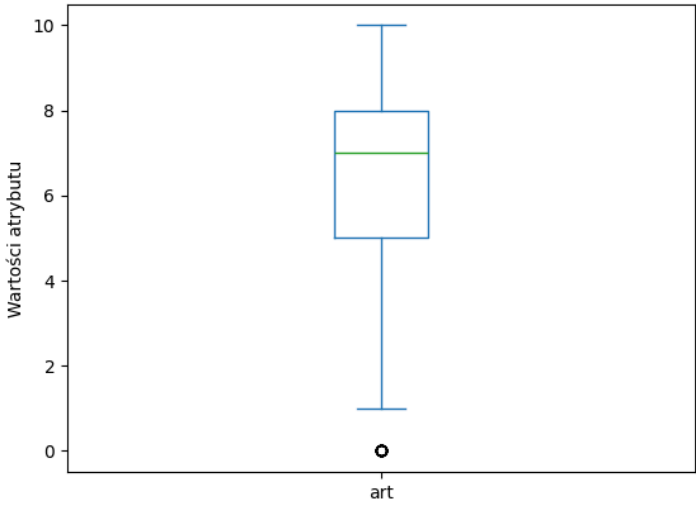
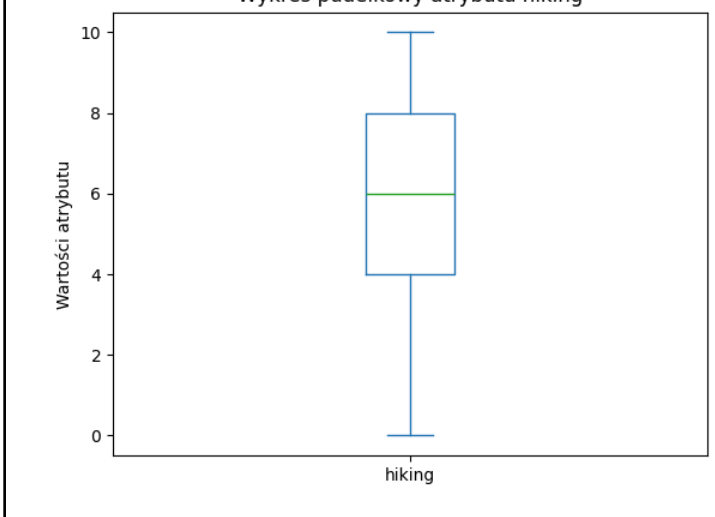



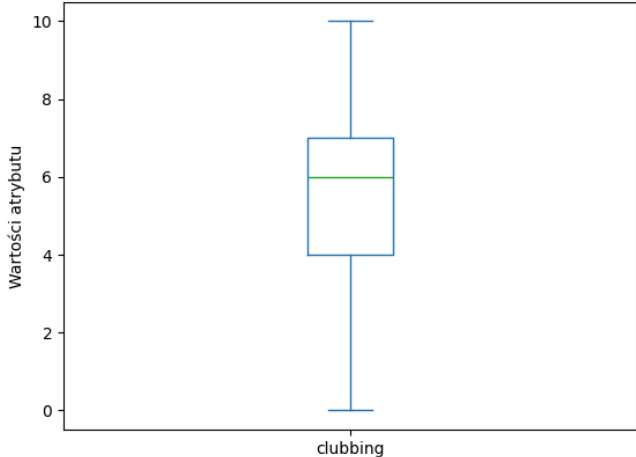
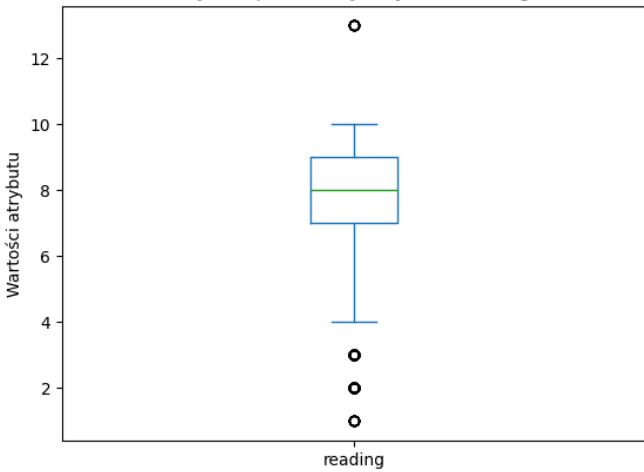
imprace	<p>Wykres pudełkowy atrybutu imprace</p>  <p>Wartości atrybutu</p> <p>imprace</p>	<p>Mediana: 3.0</p> <p>Przedział wartości występujących najczęściej: [1.0; 6.0]</p> <p>Liczba punktów oddalonych: 799</p>
goal	<p>Wykres pudełkowy atrybutu goal</p>  <p>Wartości atrybutu</p> <p>goal</p>	<p>Mediana: 2.0</p> <p>Przedział wartości występujących najczęściej: [1.0; 2.0]</p> <p>Liczba punktów oddalonych: 955</p>

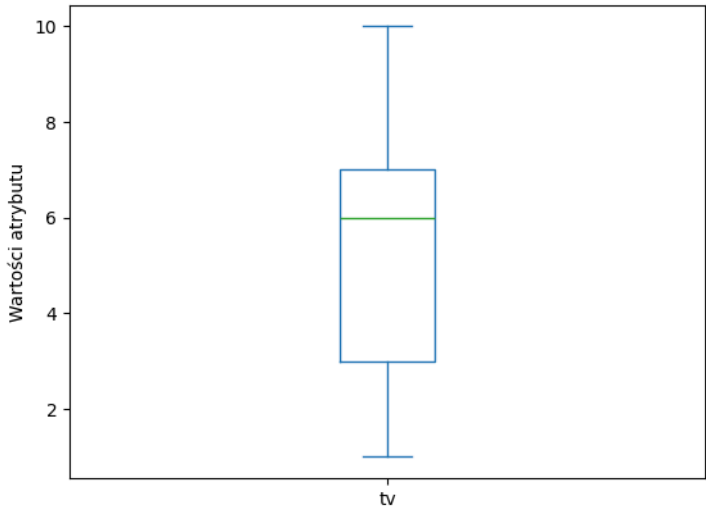
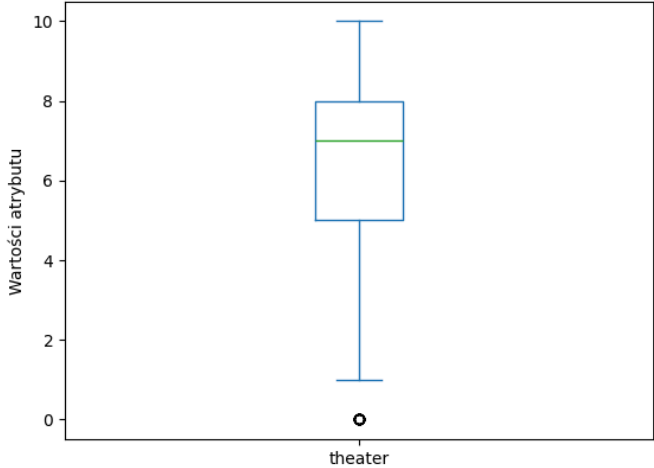
date	<p>Wykres pudełkowy atrybutu date</p> 	<p>Mediana: 5.0</p> <p>Przedział wartości występujących najczęściej: [4.0; 6.0]</p> <p>Liczba punktów oddalonych: 1257</p>
go_out	<p>Wykres pudełkowy atrybutu go_out</p> 	<p>Mediana: 2.0</p> <p>Przedział wartości występujących najczęściej: [1.0; 3.0]</p> <p>Liczba punktów oddalonych: 414</p>
exphappy	<p>Wykres pudełkowy atrybutu exphappy</p> 	<p>Mediana: 6.0</p> <p>Przedział wartości występujących najczęściej: [5.0; 7.0]</p> <p>Liczba punktów oddalonych: 1302</p>

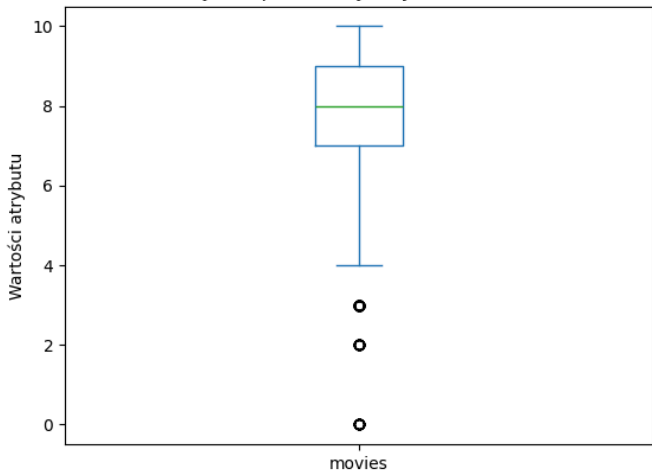
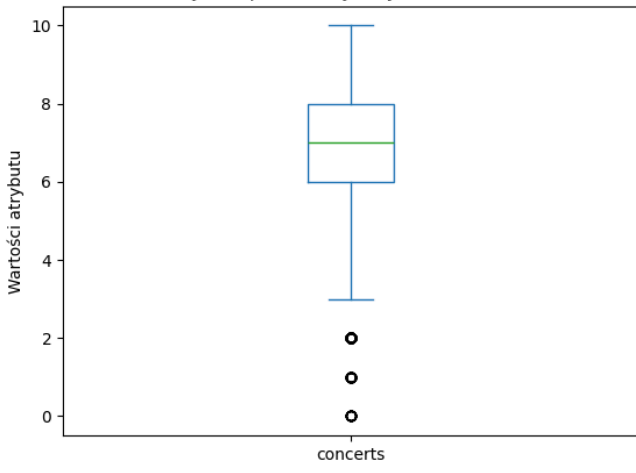
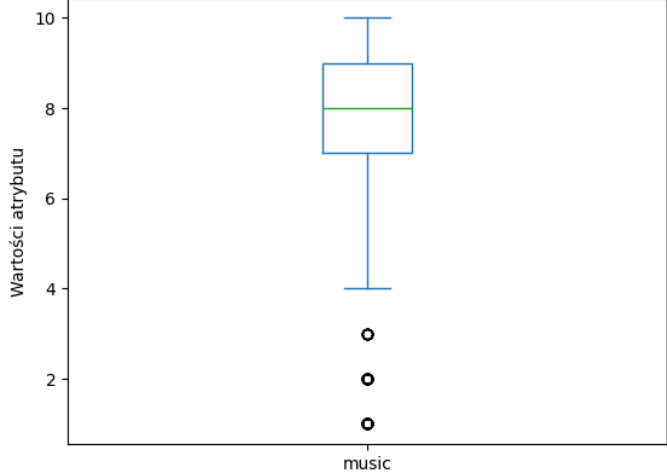
expnum	<p>Wykres pudełkowy atrybutu expnum</p> 	<p>Mediana: 4.0</p> <p>Przedział wartości występujących najczęściej: [2.0; 8.0]</p> <p>Liczba punktów oddalonych: 655</p>
sports	<p>Wykres pudełkowy atrybutu sports</p> 	<p>Mediana: 7.0</p> <p>Przedział wartości występujących najczęściej: [5.0; 9.0]</p> <p>Liczba punktów oddalonych: 1441</p>
tvsports	<p>Wykres pudełkowy atrybutu tvsports</p> 	<p>Mediana: 4.0</p> <p>Przedział wartości występujących najczęściej: [2.0; 7.0]</p> <p>Liczba punktów oddalonych: 1478</p>

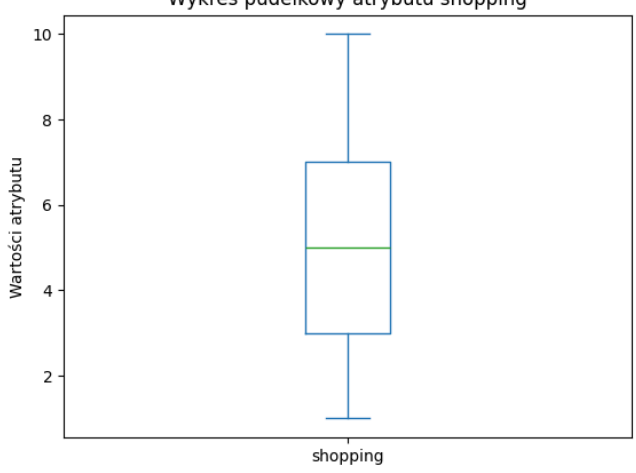
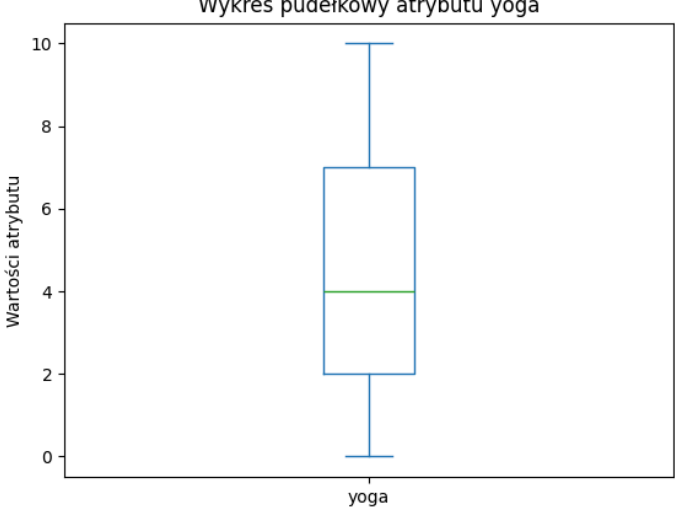
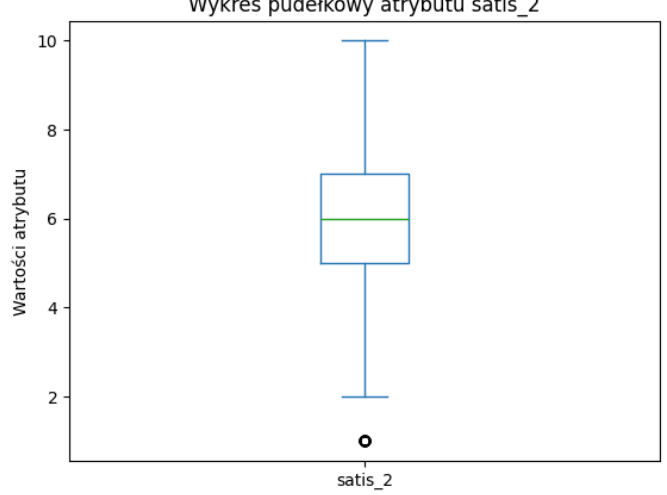
exercise	<p>Wykres pudełkowy atrybutu exercise</p> 	<p>Mediana: 6.0</p> <p>Przedział wartości występujących najczęściej: [5.0; 8.0]</p> <p>Liczba punktów oddalonych: 1545</p>
dining	<p>Wykres pudełkowy atrybutu dining</p> 	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [7.0; 9.0]</p> <p>Liczba punktów oddalonych: 1546</p>
museums	<p>Wykres pudełkowy atrybutu museums</p> 	<p>Mediana: 7.0</p> <p>Przedział wartości występujących najczęściej: [6.0; 8.0]</p> <p>Liczba punktów oddalonych: 1792</p>

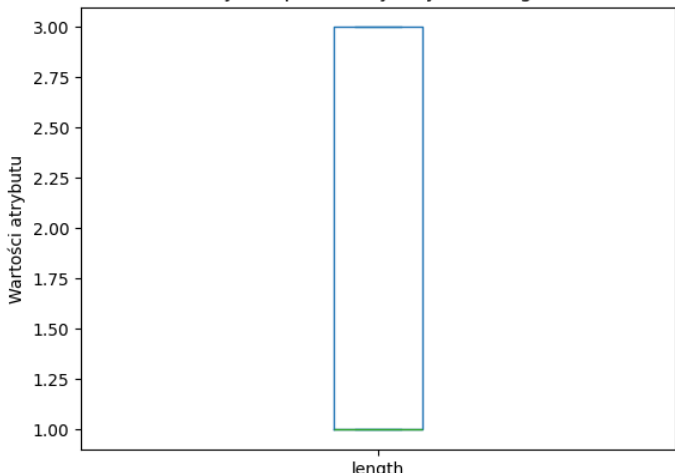
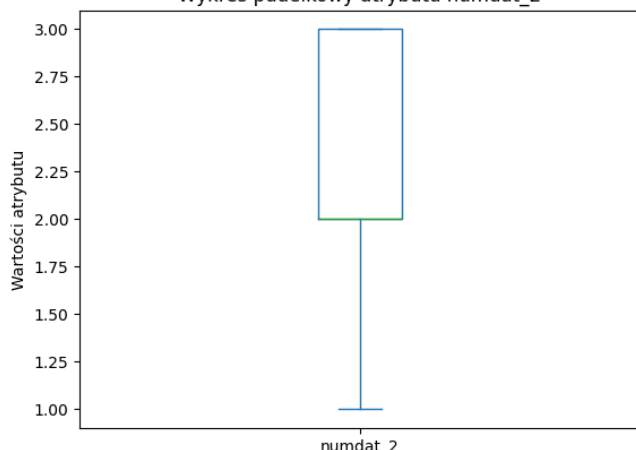
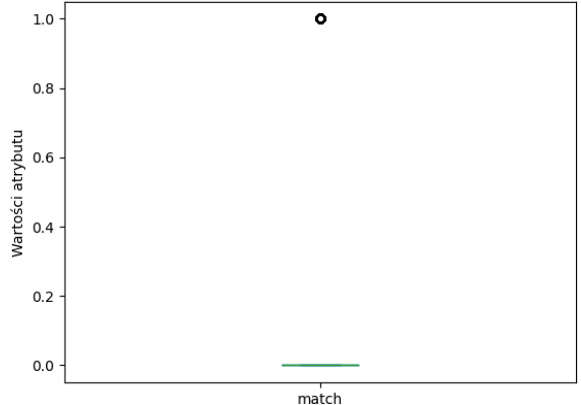
art	<p data-bbox="379 280 715 309">Wykres pudełkowy atrybutu art</p>  <p data-bbox="539 790 563 813">art</p>	<p data-bbox="1169 235 1324 264">Mediana: 7.0</p> <p data-bbox="970 291 1525 360">Przedział wartości występujących najczęściej: [5.0; 8.0]</p> <p data-bbox="1038 383 1453 412">Liczba punktów oddalonych: 1613</p>
hiking	<p data-bbox="395 945 746 974">Wykres pudełkowy atrybutu hiking</p>  <p data-bbox="545 1429 593 1451">hiking</p>	<p data-bbox="1169 900 1324 929">Mediana: 6.0</p> <p data-bbox="970 956 1525 1025">Przedział wartości występujących najczęściej: [4.0; 8.0]</p> <p data-bbox="1038 1048 1453 1077">Liczba punktów oddalonych: 1522</p>

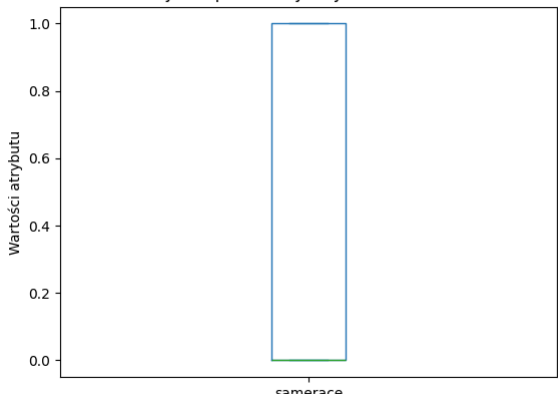
gaming	<p>Wykres pudełkowy atrybutu gaming</p>  <p>The box plot for the 'gaming' attribute shows a median of 4.0. The interquartile range (IQR) is from 2.0 to 6.0. The whiskers extend from 1.0 to 10.0. There is one outlier at 14.0. The y-axis is labeled 'Wartości atrybutu' and ranges from 0 to 14. The x-axis is labeled 'gaming'.</p>	<p>Mediana: 4.0</p> <p>Przedział wartości występujących najczęściej: [2.0; 6.0]</p> <p>Liczba punktów oddalonych: 1582</p>
clubbing	<p>Wykres pudełkowy atrybutu clubbing</p>  <p>The box plot for the 'clubbing' attribute shows a median of 6.0. The interquartile range (IQR) is from 4.0 to 7.0. The whiskers extend from 0.0 to 10.0. There are no outliers. The y-axis is labeled 'Wartości atrybutu' and ranges from 0 to 10. The x-axis is labeled 'clubbing'.</p>	<p>Mediana: 6.0</p> <p>Przedział wartości występujących najczęściej: [4.0; 7.0]</p> <p>Liczba punktów oddalonych: 1870</p>
reading	<p>Wykres pudełkowy atrybutu reading</p>  <p>The box plot for the 'reading' attribute shows a median of 8.0. The interquartile range (IQR) is from 7.0 to 9.0. The whiskers extend from 4.0 to 10.0. There are four outliers at 1.0, 2.0, 3.0, and 13.0. The y-axis is labeled 'Wartości atrybutu' and ranges from 0 to 12. The x-axis is labeled 'reading'.</p>	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [7.0; 9.0]</p> <p>Liczba punktów oddalonych: 1642</p>

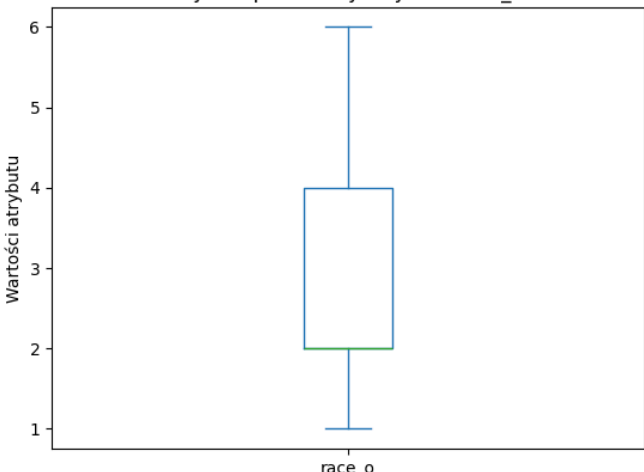
tv	<p>Wykres pudełkowy atrybutu tv</p> 	<p>Mediana: 6.0</p> <p>Przedział wartości występujących najczęściej: [3.0; 7.0]</p> <p>Liczba punktów oddalonych: 1488</p>
theater	<p>Wykres pudełkowy atrybutu theater</p> 	<p>Mediana: 7.0</p> <p>Przedział wartości występujących najczęściej: [5.0; 8.0]</p> <p>Liczba punktów oddalonych: 1549</p>

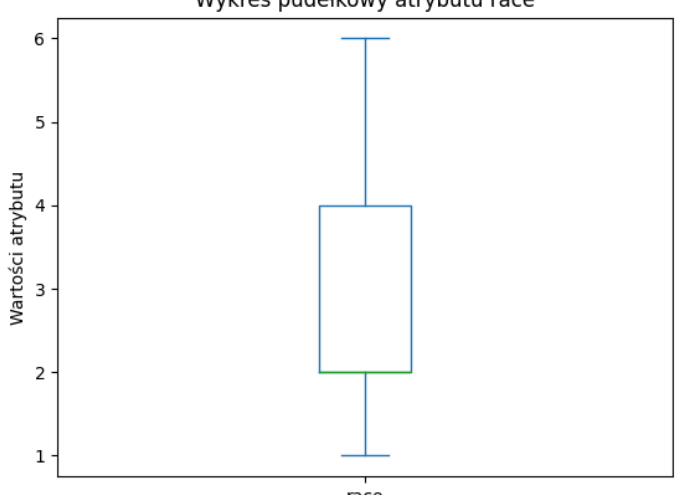
movies	<p>Wykres pudełkowy atrybutu movies</p> 	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [7.0; 9.0]</p> <p>Liczba punktów oddalonych: 1366</p>
concerts	<p>Wykres pudełkowy atrybutu concerts</p> 	<p>Mediana: 7.0</p> <p>Przedział wartości występujących najczęściej: [6.0; 8.0]</p> <p>Liczba punktów oddalonych: 1770</p>
music	<p>Wykres pudełkowy atrybutu music</p> 	<p>Mediana: 8.0</p> <p>Przedział wartości występujących najczęściej: [7.0; 9.0]</p> <p>Liczba punktów oddalonych: 1669</p>

shopping	<p>Wykres pudełkowy atrybutu shopping</p> 	<p>Mediana: 5.0</p> <p>Przedział wartości występujących najczęściej: [3.0; 7.0]</p> <p>Liczba punktów oddalonych: 1709</p>
yoga	<p>Wykres pudełkowy atrybutu yoga</p> 	<p>Mediana: 4.0</p> <p>Przedział wartości występujących najczęściej: [2.0; 7.0]</p> <p>Liczba punktów oddalonych: 1387</p>
satis_2	<p>Wykres pudełkowy atrybutu satis_2</p> 	<p>Mediana: 6.0</p> <p>Przedział wartości występujących najczęściej: [5.0; 7.0]</p> <p>Liczba punktów oddalonych: 1335</p>

length	<p>Wykres pudełkowy atrybutu length</p> 	<p>Mediana: 1.0</p> <p>Przedział wartości występujących najczęściej: [1.0; 3.0]</p> <p>Liczba punktów oddalonych: 0</p>
numdat_2	<p>Wykres pudełkowy atrybutu numdat_2</p> 	<p>Mediana: 2.0</p> <p>Przedział wartości występujących najczęściej: [2.0; 3.0]</p> <p>Liczba punktów oddalonych: 315</p>
match	<p>Wykres pudełkowy atrybutu match</p> 	<p>Mediana: 0.0</p> <p>Przedział wartości występujących najczęściej: [0.0; 0.0]</p> <p>Liczba punktów oddalonych: 690</p> <p>Atrybut binarny, wykres pudełkowy analizować ostrożnie.</p>

samerace	<p>Wykres pudełkowy atrybutu samerace</p> 	<p>Mediana: 0.0</p> <p>Przedział wartości występujących najczęściej: [0.0; 1.0]</p> <p>Liczba punktów oddalonych: 0</p> <p>Atrybut binarny, wykres pudełkowy analizować ostrożnie.</p>
----------	---	--

race_o	<p>Wykres pudełkowy atrybutu race_o</p> 	<p>Mediana: 2.0</p> <p>Przedział wartości występujących najczęściej: [2.0; 4.0]</p> <p>Liczba punktów oddalonych: 456</p>
--------	--	---

race	<p>Wykres pudełkowy atrybutu race</p> 	<p>Mediana: 2.0</p> <p>Przedział wartości występujących najczęściej: [2.0; 4.0]</p> <p>Liczba punktów oddalonych: 457</p>
------	---	---

6. Wnioski

6.1. Wnioski dotyczące danych

1. Wiek uczestników

- Większość uczestników mieściła się w przedziale wiekowym 20-30 lat, co jest charakterystyczne dla studentów MBA. Ten przedział wiekowy może wpływać na preferencje i priorytety romantyczne, a także na ogólną dynamikę eksperymentu.

2. Ambicja jako atrybut

- Przed wydarzeniem histogram oceny ambicji wskazywał, że około 2/3 uczestników oceniło ambicję partnerów na mniej niż 10 punktów w skali 0-100. Był to rozkład lewoskośny, z pozostałą częścią ocen rozłożoną między 10 a 20 punktów.
- Dzień po wydarzeniu znaczenie ambicji u partnerów wzrosło — ponad połowa uczestników uznała ją za istotną, a wykres zmienił charakter na wycentrowany, z dwoma wyraźnymi szczytami w końcowych przedziałach 0-20 punktów (większość ocen powyżej 10).
- Ocena własnej ambicji uczestników pozostała stabilna, z prawoskośnym rozkładem wskazującym na wysoką samoocenę w tej kategorii.

3. Zainteresowania i aktywności

- Rozkład zainteresowań uczestników w większości był prawostronnie skierowany (więcej ocen na niższym poziomie).
- Wyjątkami były kategorie filmów (wysokie zainteresowanie, centralny rozkład) oraz gier komputerowych (gaming), które cechowały się niskim poziomem zainteresowania wśród uczestników.

4. Atrakcyjność jako atrybut

- Przed randkami uczestnicy oceniali atrakcyjność partnerów w skali 1–10. Rozkład był nieco prawoskośny, z ocenami najczęściej wynoszącymi 5, 6, 7 lub 8, a średnia wynosiła 6,14 punktów.
- W skali 0–100 rozkład punktów był lewoskośny, z większością ocen koncentrujących się w przedziale 15-30 punktów. Atrybut atrakcyjności często otrzymywał najwięcej punktów w ramach dostępnej puli. Średnia wynosiła 24,12 punktów, a w skrajnych przypadkach osiągała nawet 100 punktów.
- Po wydarzeniu rozkład znaczenia atrakcyjności uległ zmianie. Skrajne przypadki wzrosły do około 80 punktów, a minimalne znaczenie podniosło się z 0 do 10 punktów. Średnia wzrosła do 28,64 punktów.

5. Szczerość i clubbing

- Histogramy cech takich jak szczerość oraz zainteresowanie clubbingiem zawierały najbardziej odległe punkty w rozkładzie. - Większość uczestników była zainteresowana clubbingiem, co pokazał wyraźny szczyt rozkładu w wysokich przedziałach. Tylko niewielki odsetek uczestników ocenił swoje zainteresowanie tą aktywnością na minimalnym poziomie (1 w skali 0–10).

6. Częstotliwość randek

- Najczęstsze odpowiedzi uczestników na pytanie o częstotliwość randek mieściły się w przedziale od „dwa razy w miesiącu” do „kilka razy w roku”. Średnia oscylowała wokół wartości „raz w miesiącu”.

7. Szczerość jako wartość w relacjach

- Dzień po spotkaniach szczerość była niżej oceniana jako istotna cecha u potencjalnego partnera przeciwnej płci. Średnia ocena spadła z 17,29 punktów do 15,57 punktów w skali 0–100.
- Może to sugerować, że w krótkich interakcjach, takich jak 4-minutowe randki, inne cechy (np. atrakcyjność lub wspólne zainteresowania) zyskują na znaczeniu, przesuwając uwagę uczestników z wartości takich jak szczerość.

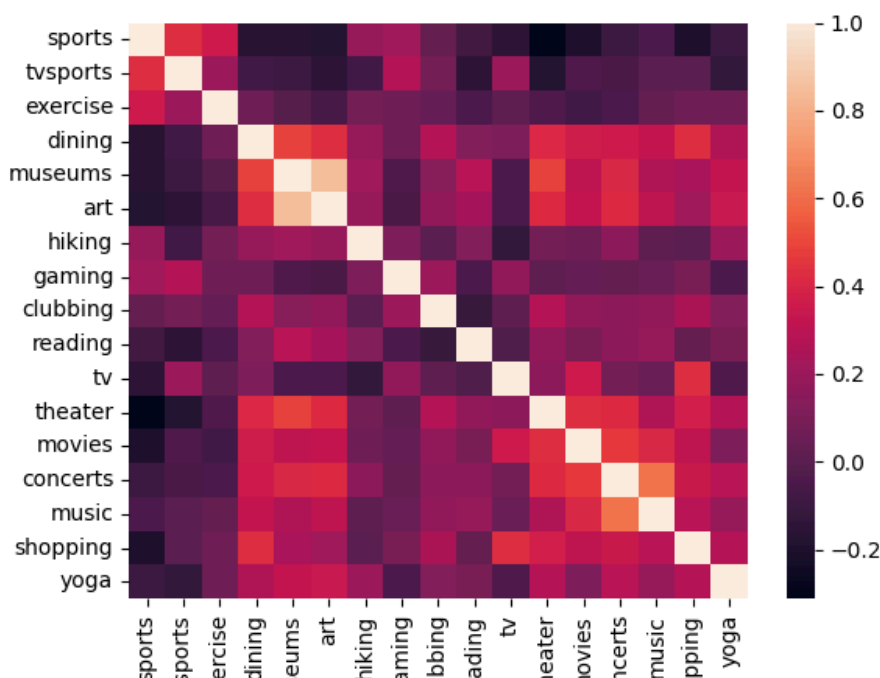
8. Czas trwania spotkań

- Większość uczestników uważała, że czas trwania 4-minutowych randek był zbyt krótki, co potwierdza analiza odpowiedzi. Mediana odpowiedzi wskazywała na potrzebę dłuższych interakcji.
- Drugą najczęściej wybieraną opcją była odpowiedź, że czas randki był „idealny”, co sugeruje pewną grupę uczestników zadowolonych z formatu wydarzenia.
- Jedynie niewielki odsetek uczestników (około 7%) uznał, że spotkania były zbyt długie. Może to wskazywać na różne potrzeby i oczekiwania uczestników wobec formatu randek.

6.2. Korelacje między wartościami atrybutów

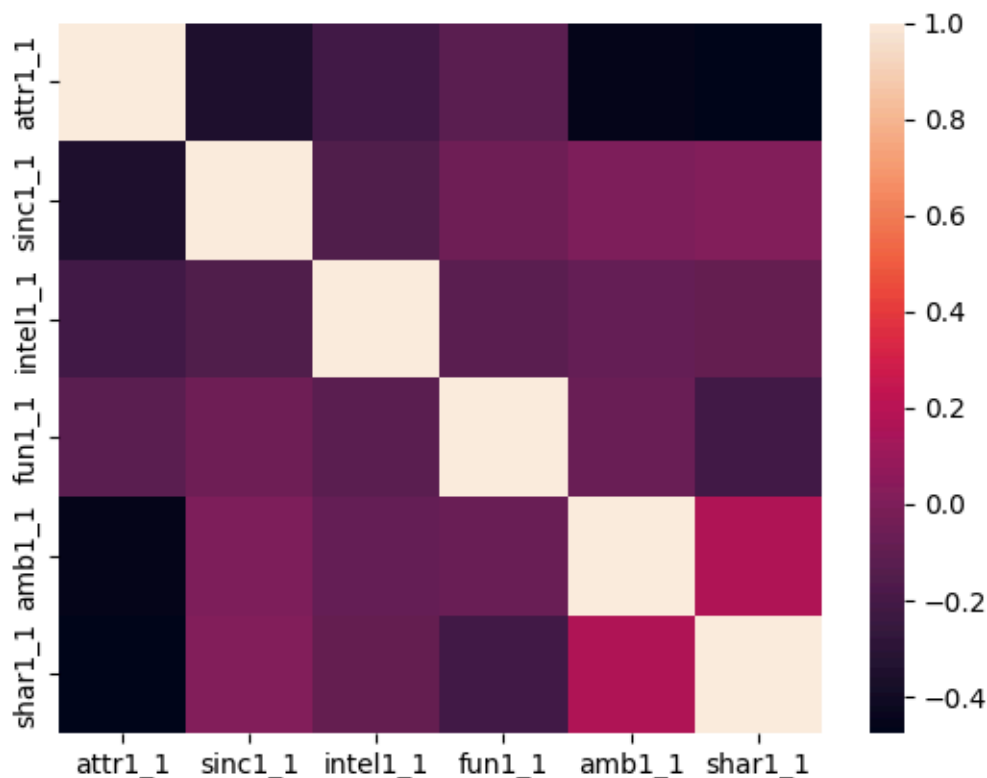
Z uwagi na olbrzymią ilość atrybutów, wygenerowano korelację tylko dla kilku z nich. Mimo to, udało się wyciągnąć z nich kilka ciekawych wniosków (macierze poniżej, opis tylko do ciekawszych wyników).

Macierz korelacji w danych Speed Dating wskazuje na interesujące zależności między preferencjami uczestników. Najczęściej wspólnie cenionymi atrybutami były sztuka i muzea, które wykazują wysoką korelację, podobnie jak koncerty i muzyka, co sugeruje, że osoby zainteresowane jedną z tych dziedzin mają tendencję do doceniania również innych związanych z kulturą i rozrywką. Z kolei sport był najmniej skorelowanym atrybutem, wykazując jedynie słabą korelację z takimi czynnikami jak TV sportowa, ćwiczenia czy gaming. Wskazuje to na stosunkowo izolowane preferencje w tej kategorii, co może odzwierciedlać bardziej wyspecjalizowane zainteresowania. Najwięcej wspólnych korelacji odnotowano wśród kategorii takich jak jedzenie, muzea, sztuka oraz zakupy, co sugeruje, że te aktywności są postrzegane jako uniwersalne elementy stylu życia, które łatwo znajdują wspólny grunt między różnymi uczestnikami. Tego typu dane mogą być szczególnie użyteczne w analizie dopasowania uczestników na podstawie wspólnych zainteresowań, co jest kluczowe w kontekście wydarzeń typu speed dating.



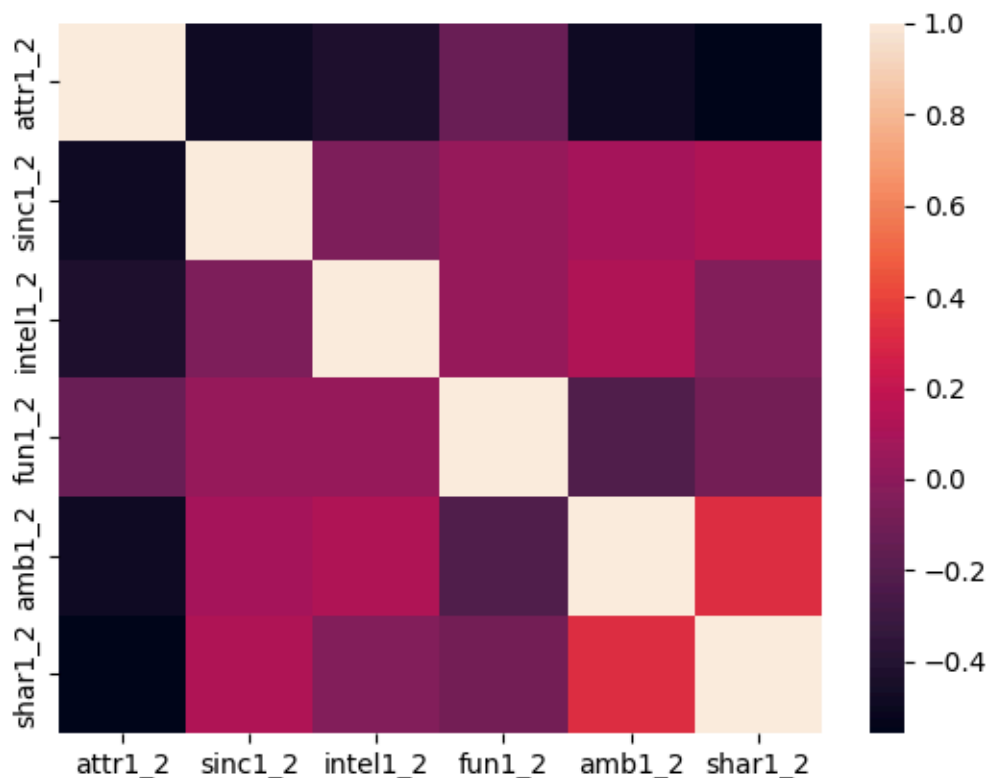
Rysunek 143: Macierz korelacji atrybutów związanych z zainteresowaniami.

Macierz korelacji pokazuje, że nie występują silne zależności pomiędzy analizowanymi cechami – wszystkie współczynniki korelacji poza przekątną są poniżej 0,6. Najwyższe korelacje zaobserwowano między poczuciem humoru a atrakcyjnością fizyczną, ambicją a inteligencją, oraz między poczuciem humoru a ambicją. Z kolei najniższą korelację widać między atrakcyjnością a szczerością, co sugeruje, że osoby uważające się za atrakcyjne niekoniecznie postrzegają siebie jako szczere. Ogólnie cechy te mają względnie niezależny charakter, co oznacza, że uczestnicy różnicowali swoje samooceny w zależności od konkretnego atrybutu.

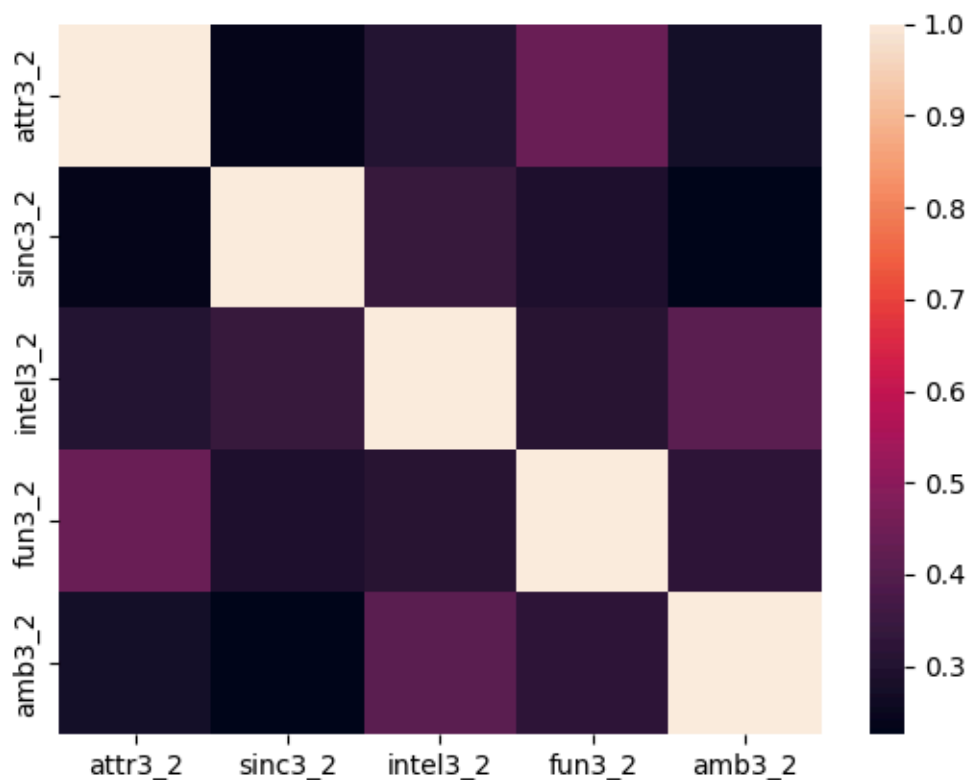


Rysunek 144: Macierz korelacji atrybutów istotności każdej z 6. cech, której dana osoba poszukuje u partnera.

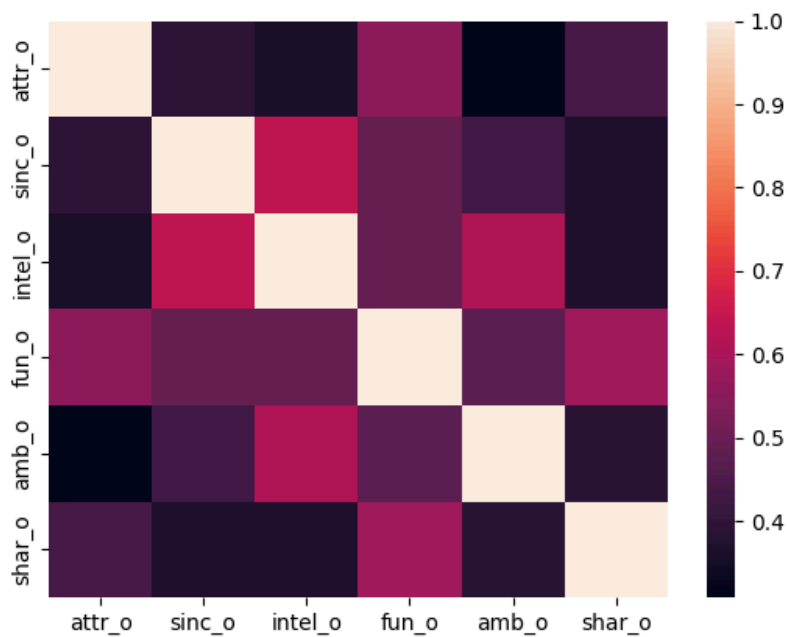
Macierz korelacji wskazuje, że zależności między ocenianymi cechami są słabe lub umiarkowane. Najwyższą korelację można zaobserwować między ambicją a wspólnymi zainteresowaniami, co sugeruje, że osoby przykładające wagę do ambicji partnera często zwracają również uwagę na podobieństwo hobby. Zauważalna jest także umiarkowana dodatnia korelacja między szczerością a większością pozostałych cech. Natomiast atrakcyjność fizyczna oraz wspólne zainteresowania cechują się niską lub nawet ujemną korelacją z pozostałymi cechami, co może świadczyć o ich odmiennym charakterze w ocenie partnera. Warto również zwrócić uwagę, że niektóre korelacje są ujemne, choć niezbyt silne, co wskazuje na zróżnicowane podejście uczestników do wartościowania poszczególnych cech.



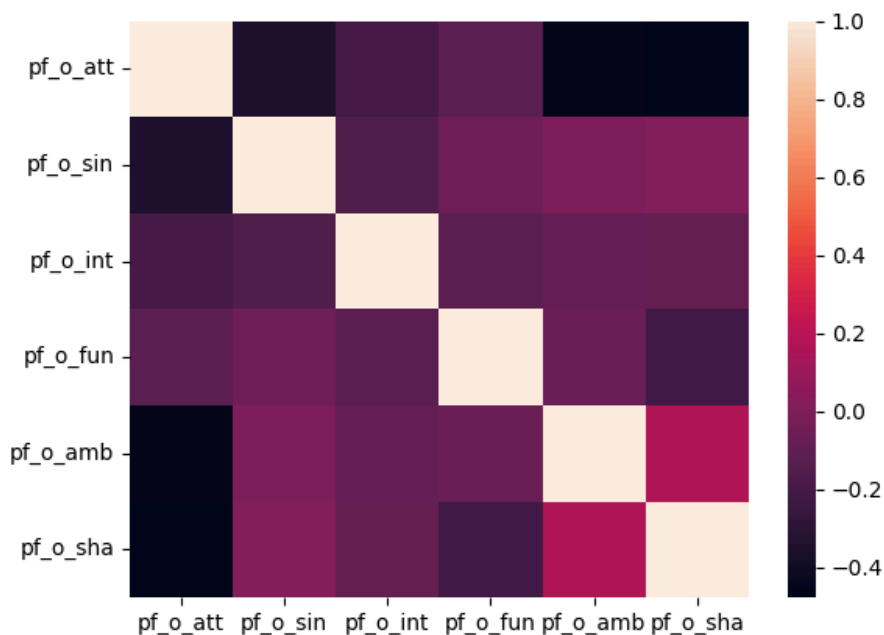
Rysunek 145: Macierz korelacji istotności cech w kontekście: czego uczestnik poszukiwał u partnera płci przeciwnej.



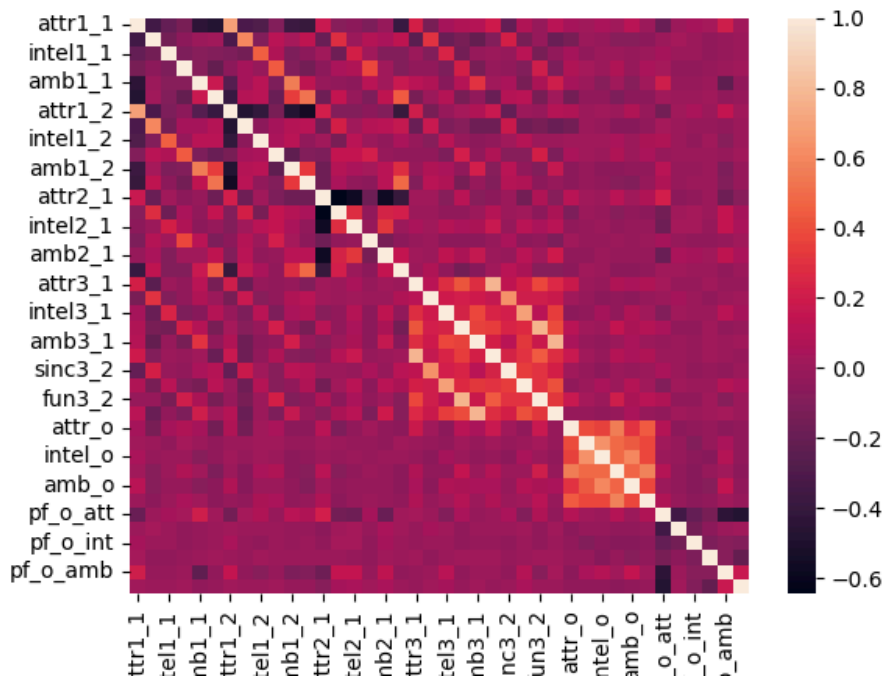
Rysunek 146: Macierz korelacji istotności poszukiwania na randce danej cechy przez płć przeciwną (własna opinia o preferencjach innych).



Rysunek 147: Macierz korelacji istotności danej cechy pod względem preferencji partnera.



Rysunek 148: Macierz korelacji istotności danej cechy pod względem preferencji uczestnika (przed randkowaniem).



Rysunek 149: Macierz korelacji istotności cech - ocena przed wydarzeniem, 1 dzień po wydarzeniu dla partnera jak i uczestnika.

Warto wspomnieć, że niemożliwe jest wyznaczenie korelacji wszystkich atrybutów przez fakt występowania danych tekstowych (zbiór fraz) lub ich mapowania na liczby. W takim przypadku korelacja nie niesie żadnych sensowych informacji.

7. Uwagi na temat jakości danych

7.1. Dane brakujące

W zbiorze danych występują atrybuty, dla których nie ma wartości w całej kolumnie danych (tj. dla każdego uczestnika). Istnieje całkiem dużo atrybutów, dla których pomiarów nie prowadzono w pierwszych wydarzeniach. Atrybuty, których wartości zdobywa się po długim czasie od wydarzenia, również mogą mieć brakujące wartości (np. shar2_2, brak danych dla blisko 25% wszystkich przypadków).

7.2. Dane niespójne

Istnieje dość sporo atrybutów, dla których wartości są niespójne. Spowodowane jest to czasową zmianą skali ocen. Dla przykładu: atrybuty kończące się „1_2” są mierzone inną skalą w wydarzeniach odbywających się w okolicach 1/3 czasu trwania eksperymentu. Warto odnotować, że w tych przypadkach wartości ze skali ocen pokrywają się np. ocena w skali 1-10 i podział 100 punktów. Bez dodatkowej wiedzy nt. spotkań nie sposób odróżnić od siebie obydwie skale oceniania.

Inny problem, utrudniający „przyzywczajenie się” do danych to niespójność mapowania wartości binarnych na wartości liczbowe. Dla części parametrów obowiązuje podział *0 lub 1* a dla innych *1 lub 2*. Dla przykładu: atrybut 'samerace' typu Tak/Nie mapowany do *0 lub 1* oraz „met” typu Tak/Nie mapowany do *1 lub 2*.

7.3. Dane niezrozumiałe

W zbiorze występują atrybuty, których znaczenie nie jest jasne. Przykładem jest atrybut 'prob_o', który nie jest opisany w dokumentacji zbioru danych. Wartości tego atrybutu są liczbami z przedziału [1, 10], ale nie wiadomo, co one oznaczają. Kolejnym przykładem jest atrybut 'met_o', który również nie jest opisany w dokumentacji zbioru danych.

7.4. Punkty oddalone

Prezentowane na wykresach pudełkowych powyżej, w sekcji 5.2.

8. Podsumowanie i wnioski ogólne:

Analiza danych wskazuje, że krótkie interakcje podczas wydarzenia wyraźnie wpływają na ewolucję ocen atrybutów, takich jak atrakcyjność czy ambicja, które zyskują na znaczeniu po bezpośrednim kontakcie. Szczerść była oceniana niżej dzień po spotkaniach, co może wynikać z ograniczonego czasu na głębsze poznanie. Różnorodność preferencji uczestników w zakresie czasu trwania spotkań oraz zainteresowań sugeruje potrzebę lepszego dostosowania formatu wydarzenia, aby uwzględnić oczekiwania różnych grup.