

Analiza osób chorych na depresję

Przemysław Rewiś, Łukasz Wajda

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE
AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY



Wprowadzenie

- Celem naszych badań była dokładna analiza informacji dotyczących osób borykających się z depresją, z celem identyfikacji relacji oraz charakterystyk wskazujących na kluczowe czynniki przyczyniające się do występowania tej choroby, zwłaszcza w zróżnicowanych grupach wiekowych i o zróżnicowanym statusie ekonomicznym.
- Dane te pochodzą z badania o charakterze warunków życia ludności wiejskiej przeprowadzonego w 2015 r. przez Centrum Busara w hrabstwie Siaya, w pobliżu Jeziora Wiktorii w zachodniej Kenii.
- Badanie obejmuje ponad 70 elementów z czego w wykorzystywanym zbiorze danych (<https://www.kaggle.com/datasets/diegobabativa/depression> udostępniony na licencji GPL 2) zawarto 23 atrybuty ekonomiczne i społeczne oraz 1429 rekordy (1191 dane osób zdrowych oraz 238 osób chorych).

Zbiór danych - przykładowe rekordy

	Survey_id	Ville_id	sex	Age	Married	Number_children	education_level	total_members	gained_asset	durable_asset	...
0	926	91	1	28	1	4	10	5	28912201	22861940	...
1	747	57	1	23	1	3	8	5	28912201	22861940	...
2	1190	115	1	22	1	3	9	5	28912201	22861940	...
3	1065	97	1	27	1	2	10	4	52667108	19698904	...
4	806	42	0	59	0	4	10	6	82606287	17352654	...

incoming_no_business	incoming_agricultural	farm_expenses	labor_primary	lasting_investment	no_lasting_investmen	depressed
0	30028818	31363432	0	28411718	28292707.0	0
0	30028818	31363432	0	28411718	28292707.0	1
0	30028818	31363432	0	28411718	28292707.0	0
1	22288055	18751329	0	7781123	69219765.0	0
0	53384566	20731006	1	20100562	43419447.0	0

Zbiór danych - analiza wstępna

Dla każdego atrybutu:

- Liczba nie pustych rekordów
- Średnia arytmetyczna
- Odchylenie standardowe
- Minimum
- Percentyl 50%
- Maksimum

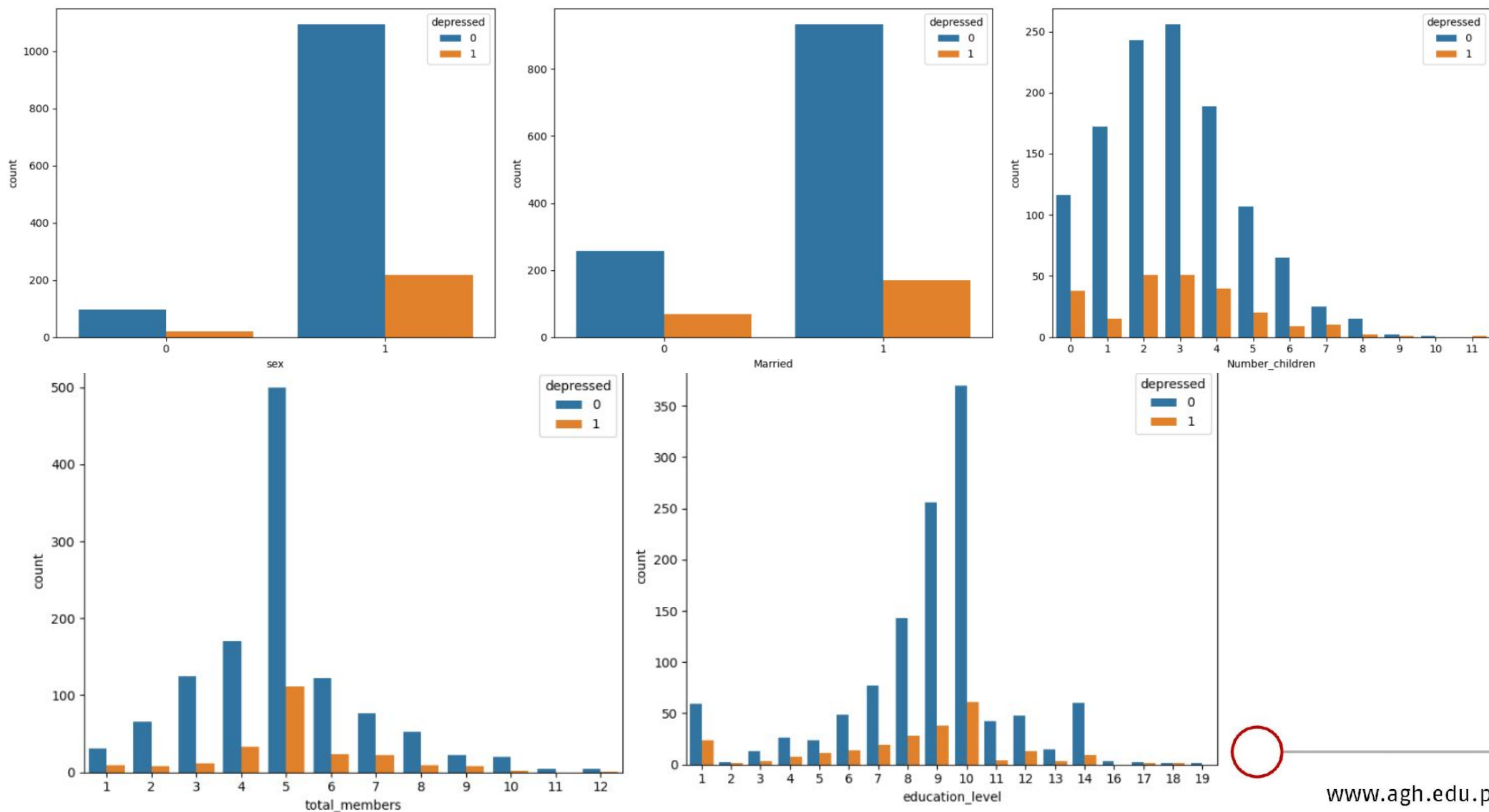
	Survey_id	Ville_id	sex	Age	Married	Number_children	education_level	total_members	gained_asset	durable_asset
count	1429.00000	1429.000000	1429.000000	1429.000000	1429.000000	1429.000000	1429.000000	1429.000000	1.429000e+03	1.429000e+03
mean	715.00000	76.286214	0.918125	34.777467	0.772568	2.883135	8.687194	4.969209	3.363448e+07	2.717296e+07
std	412.66108	66.444012	0.274271	13.986219	0.419320	1.874472	2.923532	1.786317	2.003854e+07	1.815672e+07
min	1.00000	1.000000	0.000000	17.000000	0.000000	0.000000	1.000000	1.000000	3.251120e+05	1.625560e+05
50%	715.00000	57.000000	1.000000	30.000000	1.000000	3.000000	9.000000	5.000000	2.891220e+07	2.286194e+07
max	1429.00000	292.000000	1.000000	91.000000	1.000000	11.000000	19.000000	12.000000	9.912755e+07	9.961560e+07

incoming_no_business	incoming_agricultural	farm_expenses	labor_primary	lasting_investment	no_lasting_investmen	depressed
1429.000000	1.429000e+03	1.429000e+03	1429.000000	1.429000e+03	1.409000e+03	1429.000000
0.260322	3.451039e+07	3.549153e+07	0.213436	3.299222e+07	3.360385e+07	0.166550
0.438964	2.077846e+07	2.112372e+07	0.409876	2.121621e+07	2.160228e+07	0.372704
0.000000	3.251120e+05	2.715050e+05	0.000000	7.429200e+04	1.263120e+05	0.000000
0.000000	3.002882e+07	3.136343e+07	0.000000	2.841172e+07	2.829271e+07	0.000000
1.000000	9.978910e+07	9.965119e+07	1.000000	9.944667e+07	9.965119e+07	1.000000

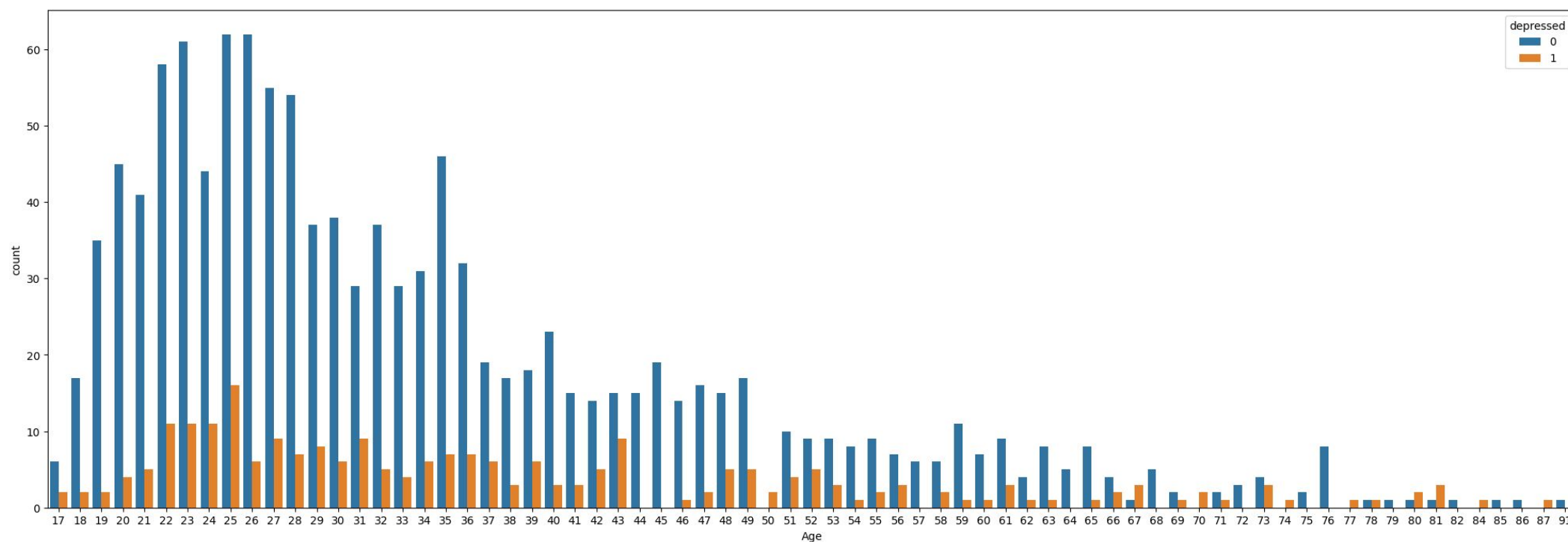
Macierz korelacji - najbardziej skorelowane atrybuty



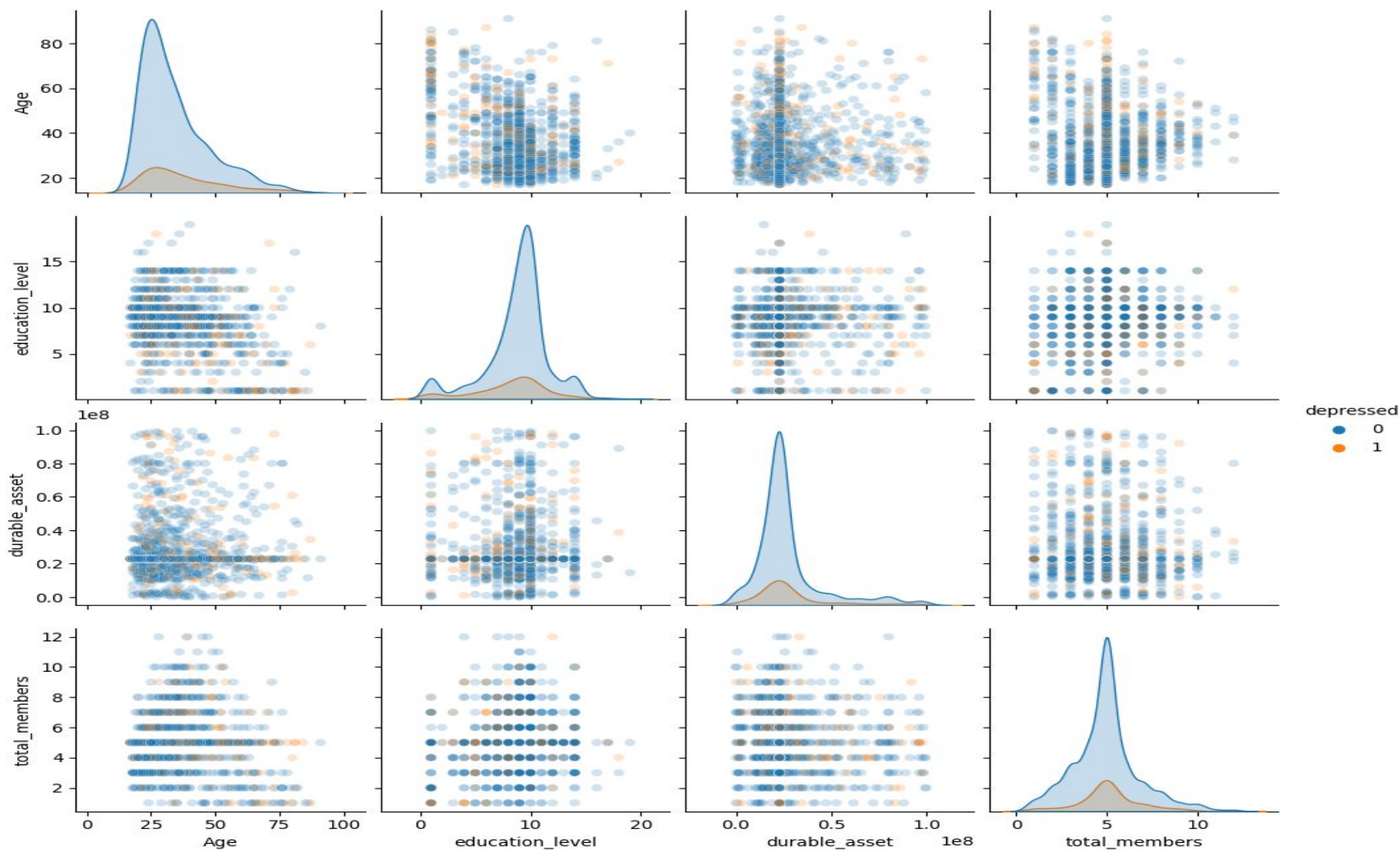
Wykresy wybranych atrybutów



Wykresy wybranych atrybutów



Wykres relacji par atrybutów

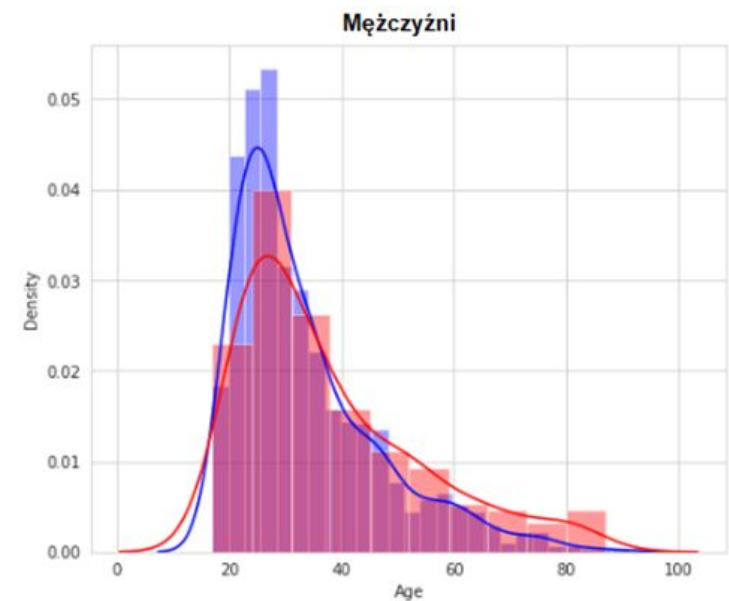
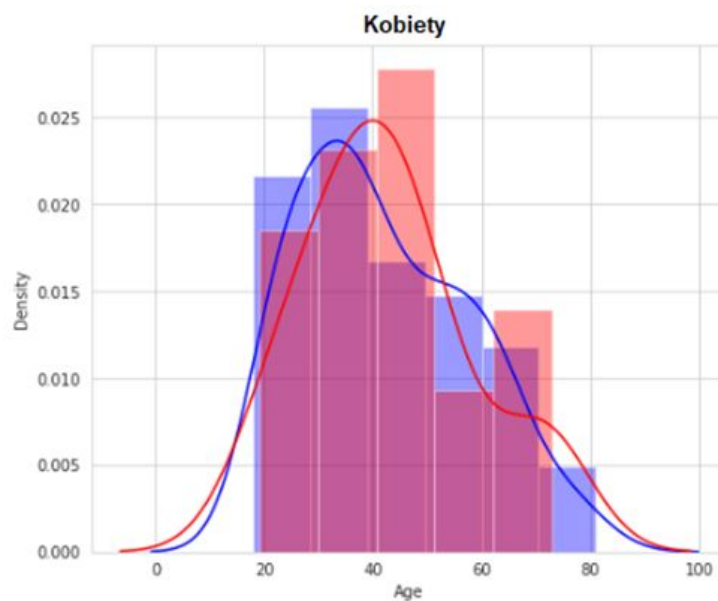
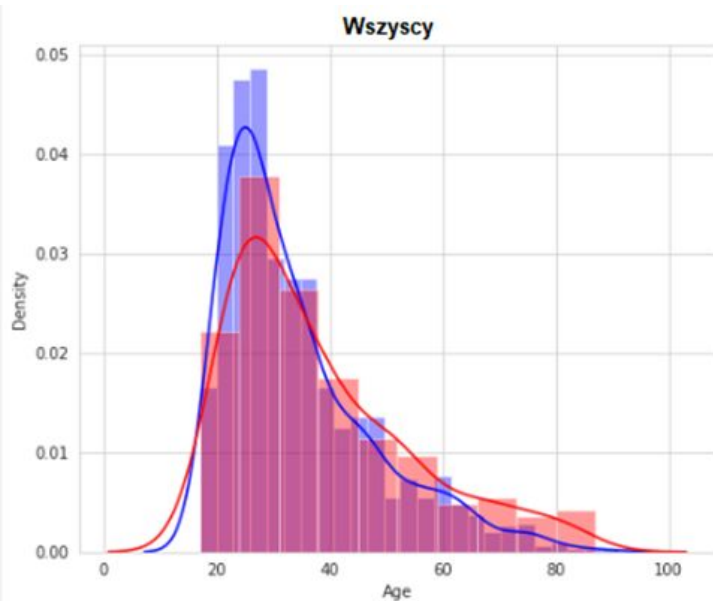


Depresja w kontekście wieku i płci

Wiek	Kobiety	Mężczyźni
0-20	0 %	15.3%
20-30	13.0%	14.0%
30-40	17.2%	16.6%
40-50	27.8%	33.1%
50+	15.8%	30.5%

Depresja w kontekście wieku i płci

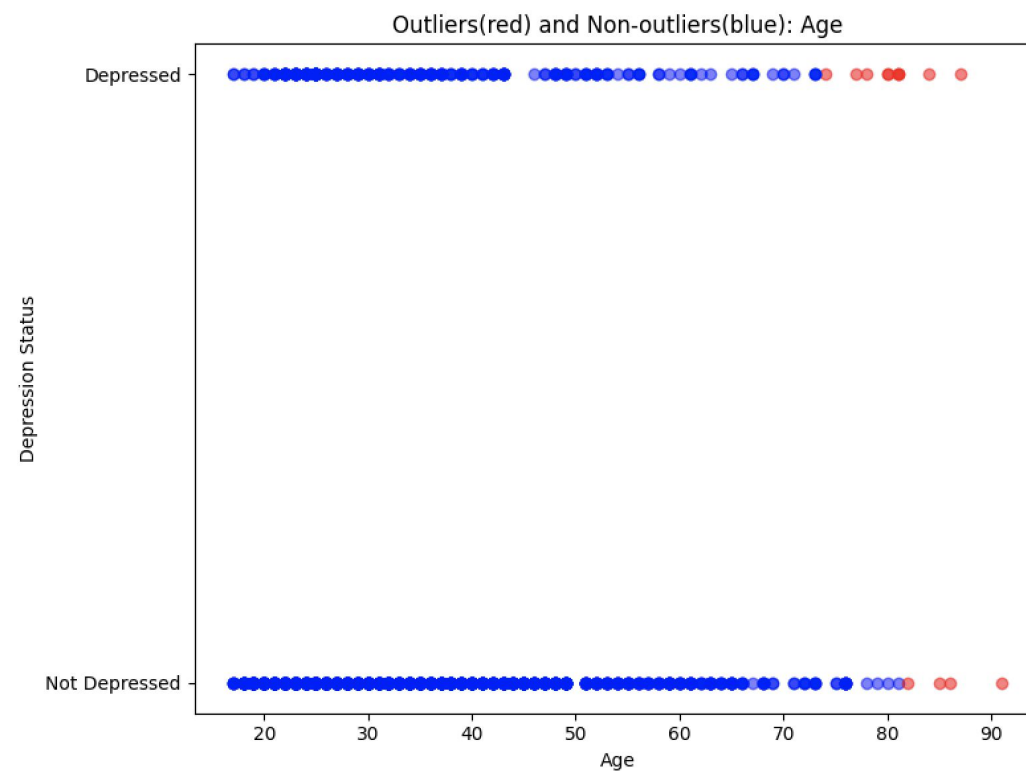
Interpretacja wykresów estymatorów jądrowych i histogramów pokazuje iż w niektórych grupach wiekowych jest większa szansa na zachorowanie na depresję niż w innych.



Elementy nietypowe - wiek

Outliers for Age:

Index	Age	education_level	durable_asset	total_members	Depressed
1	84	1	22861940	5	1
2	74	1	22861940	5	1
3	85	3	22861940	1	0
4	81	1	22861940	5	1
5	86	1	11531066	1	0
6	91	8	22861940	5	0
7	80	1	60858407	4	1
8	78	1	22861940	5	1
9	81	1	22861940	5	1
10	87	6	34433044	1	1
11	81	1	26105053	2	1
12	82	1	43721962	2	0
13	80	1	2642536	1	1
14	77	4	22861940	1	1





Outliers(red) and Non-outliers(blue): education_level

Depression Status

Depressed

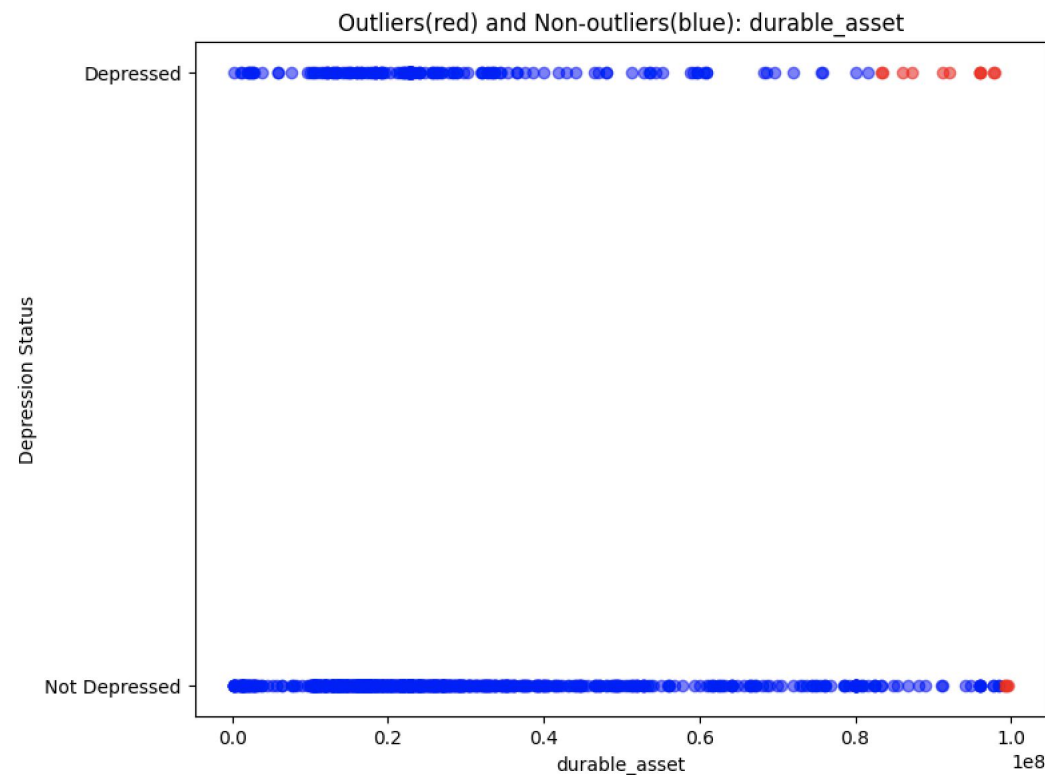
Not Depressed

education_level

Elementy nietypowe - wartość aktywów trwałych

Outliers for durable_asset:

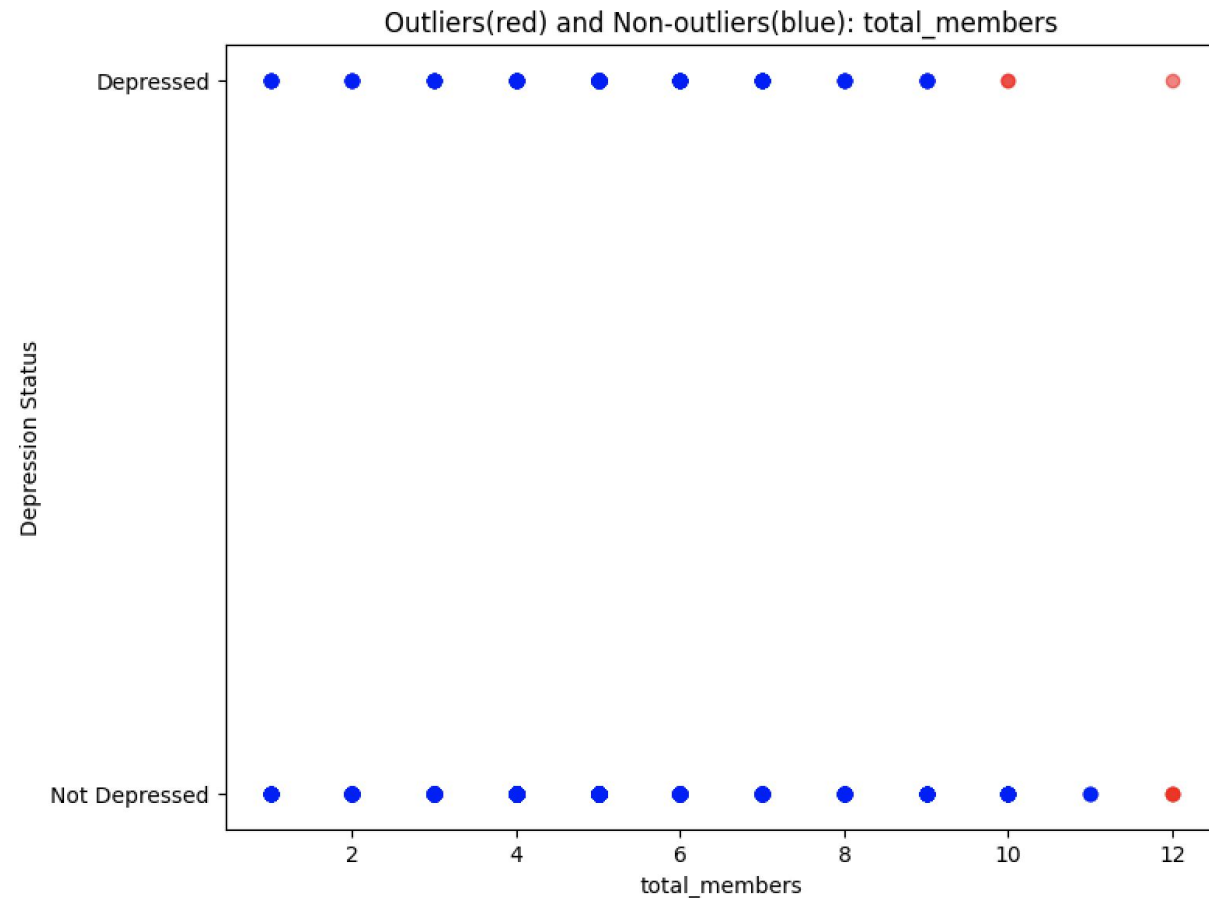
Index	Age	education_level	durable_asset	total_members	Depressed
1	55	6	83440079	1	1
2	48	1	86162689	2	1
3	32	7	87283768	5	1
4	25	12	91287605	6	1
5	35	7	99455444	2	0
6	27	8	99135139	4	0
7	41	10	99295296	4	0
8	58	1	99615601	2	0
9	23	10	96092216	5	1
10	24	5	96092216	5	1
11	29	10	97853912	3	1
12	65	5	83279922	2	1
13	34	6	92088379	9	1
14	73	12	97693758	5	1
15	32	10	96092216	5	1



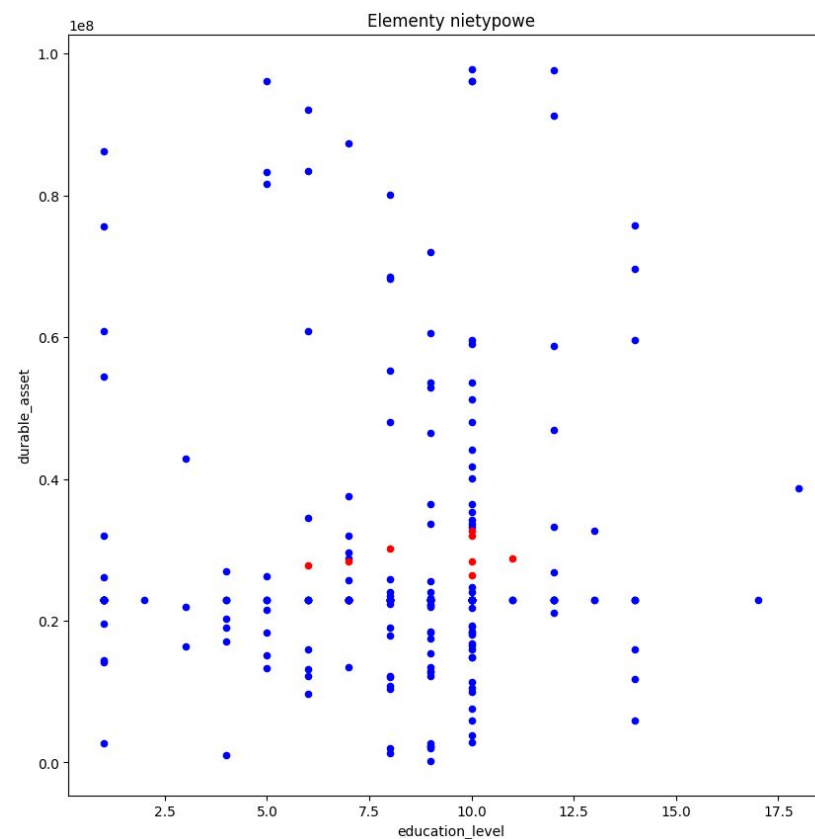
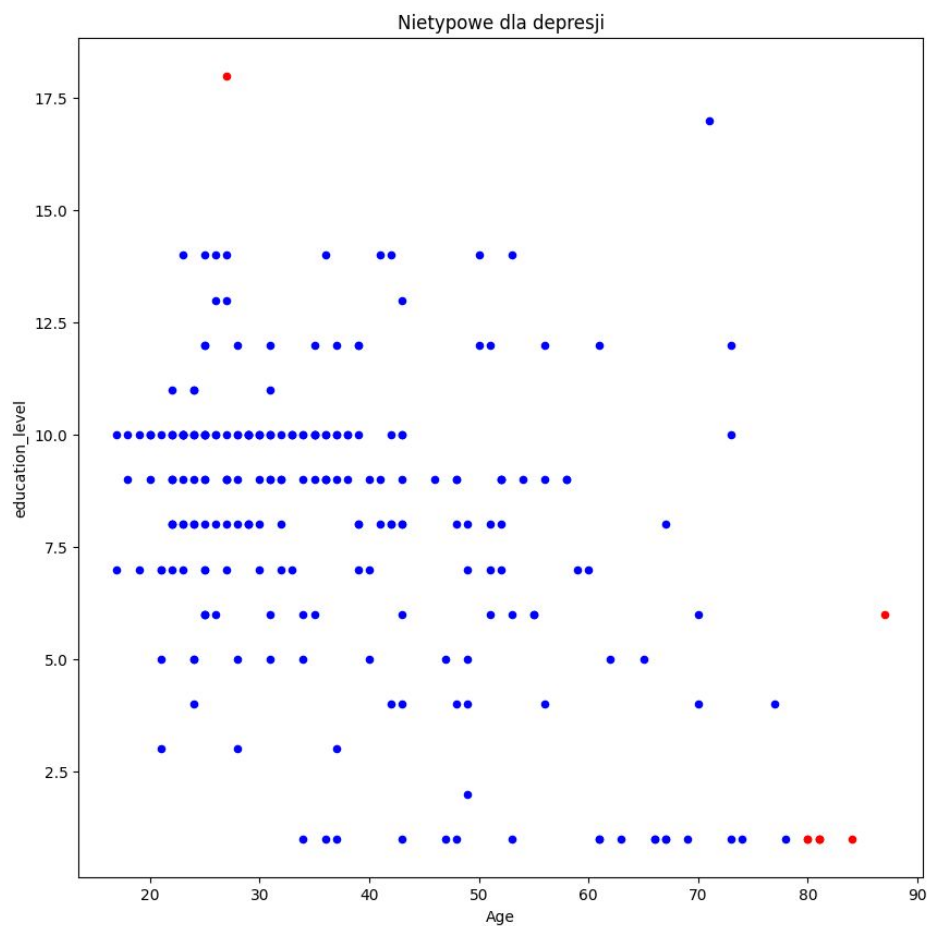
Elementy nietypowe - łączna liczba osób w gospodarstwie domowym

Outliers for total_members:

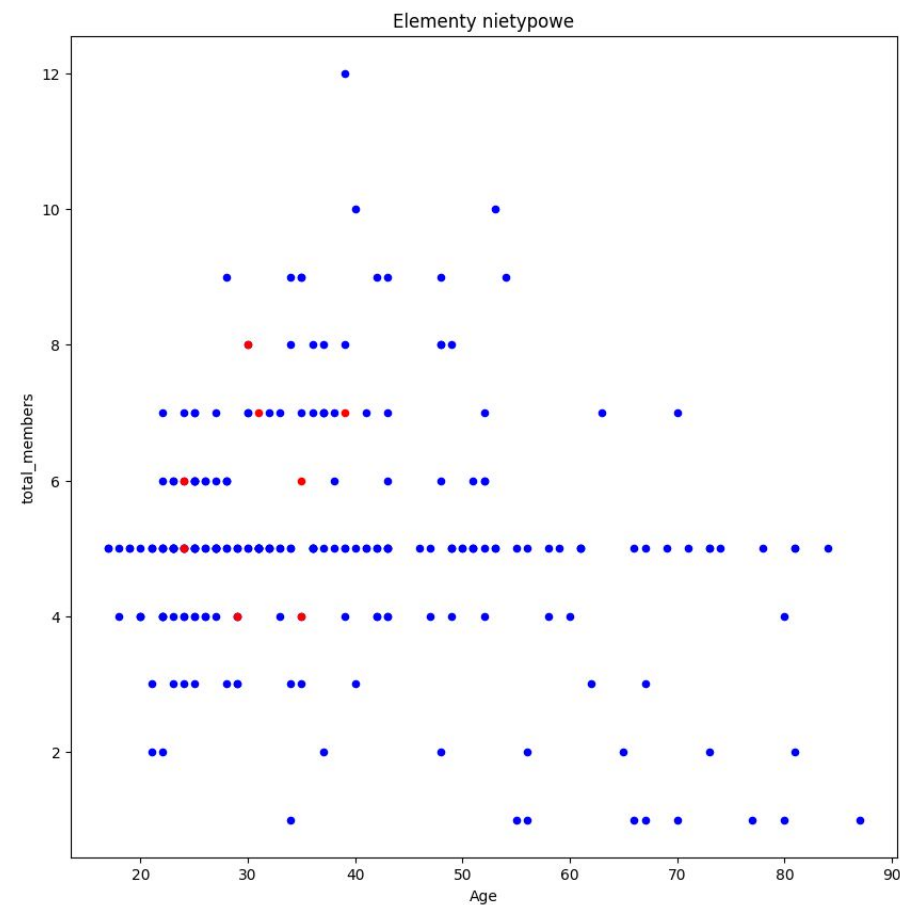
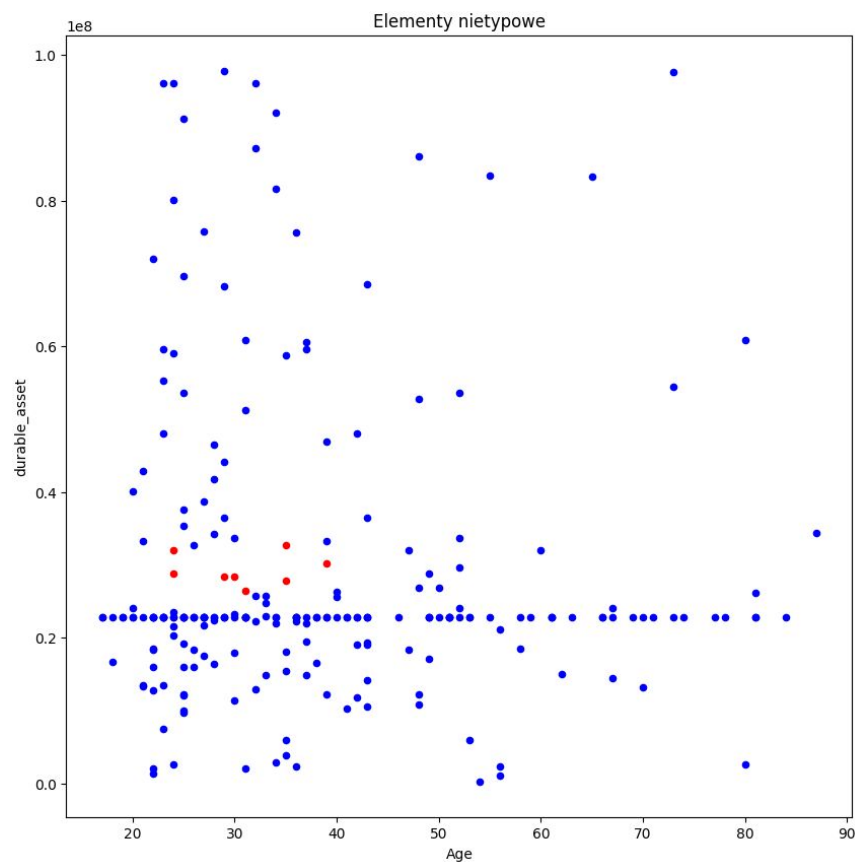
Index	Age	education_level	durable_asset	total_members	Depressed
1	47	4	23975009	12	0
2	39	12	33231894	12	1
3	28	8	21620749	12	0
4	53	14	5877641	10	1
5	45	7	26569498	12	0
6	39	9	80076847	12	0
7	40	9	25624592	10	1



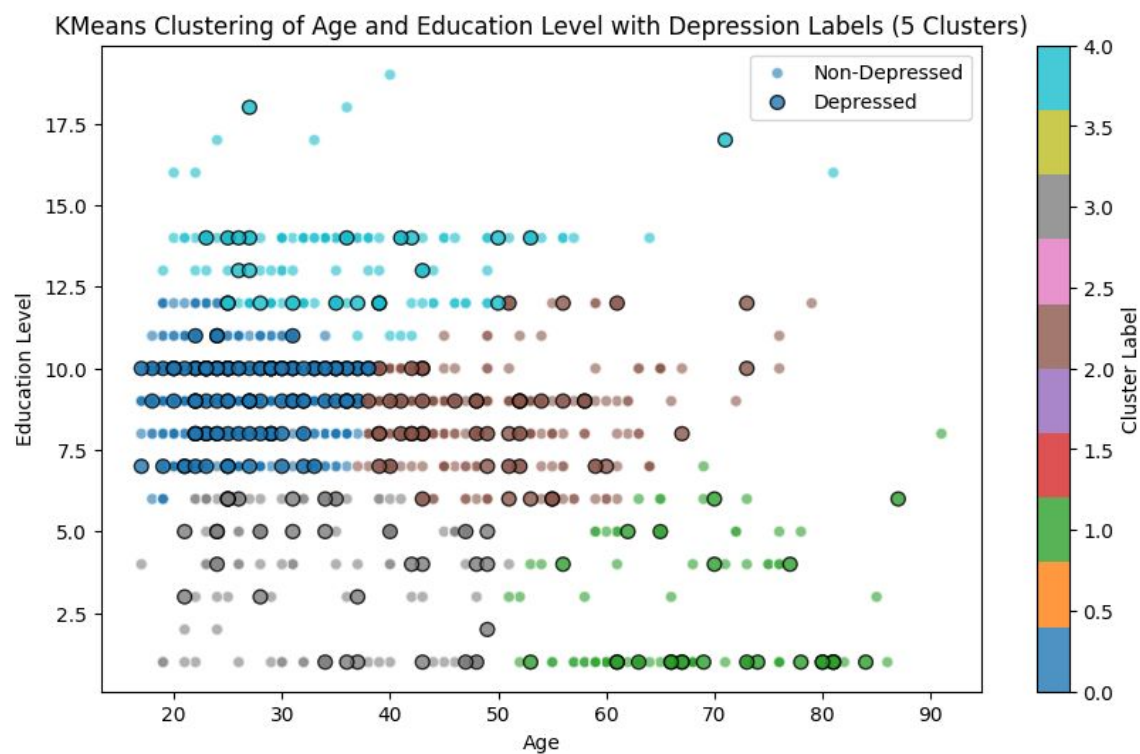
Elementy nietypowe 2D



Elementy nietypowe 2D



Klasteryzacja wiek i wykształcenie

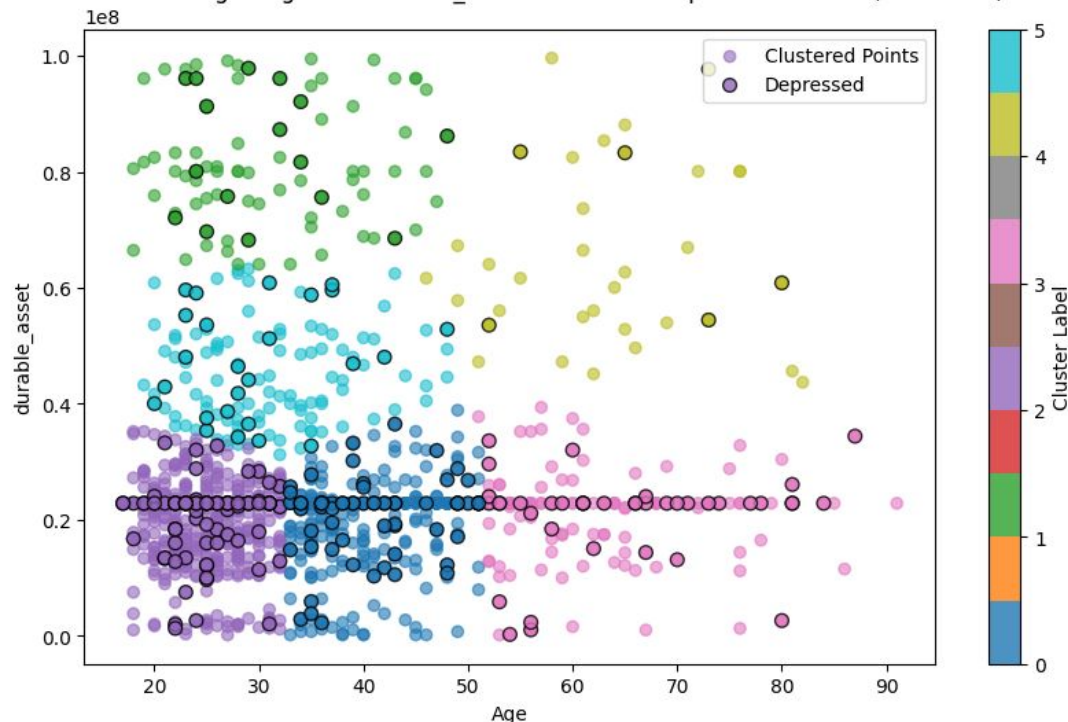


Klastry:

- ludzie dobrze wykształceni
- słabo wykształceni ludzie młodzi i w średnim wieku
- słabo wykształcone osoby starsze
- **średnio wykształceni ludzie młodzi**
- słabo wykształceni ludzie w średnim wieku i osoby starsze

Klasteryzacja wiek i wartość aktywów trwałych

KMeans Clustering of Age and durable_asset Level with Depression Labels (6 Clusters)

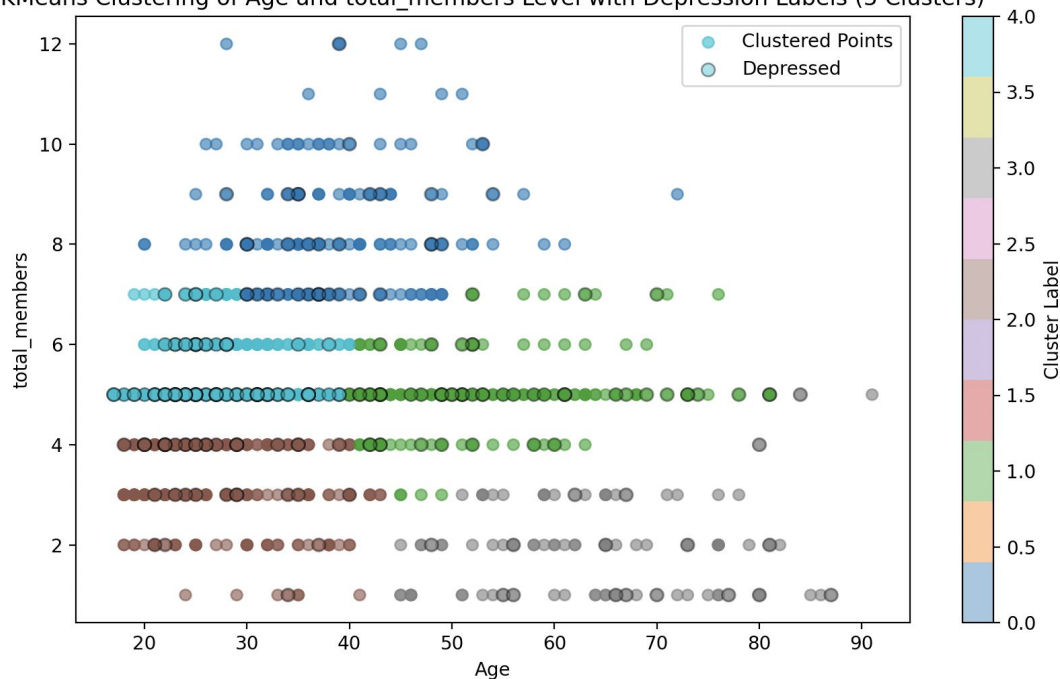


Klastry:

- bogaci ludzie młodzi i w średnim wieku
- średnio bogaci ludzie młodzi i w średnim wieku
- **biedni ludzie młodzi**
- **biedni ludzie w średnim wieku**
- **biedne osoby starsze**
- bogate i średnio bogate osoby starsze

Klasteryzacja wiek i łączna liczba osób w gospodarstwie domowym

KMeans Clustering of Age and total_members Level with Depression Labels (5 Clusters)

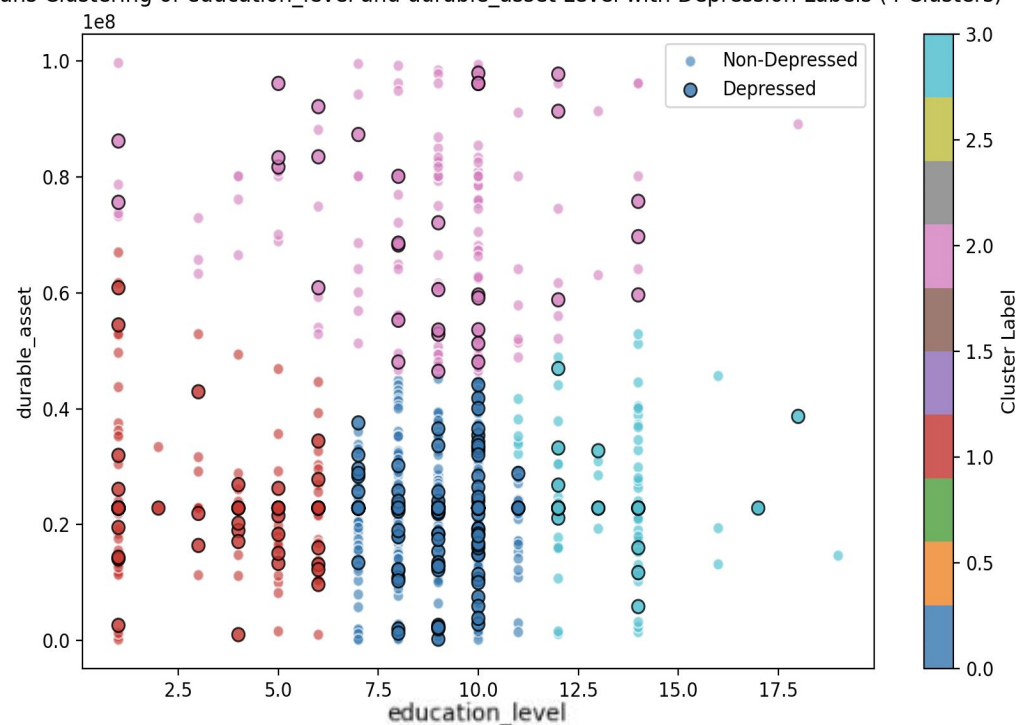


Klastry:

- ludzie młodzi mieszkający w mało licznych gospodarstwach
- ludzie młodzi mieszkający w średnio licznych gospodarstwach
- ludzie mieszkający w bardzo licznych gospodarstwach
- osoby starsze i w średnim wieku mieszkający w średnio licznych gospodarstwach
- osoby starsze i w średnim wieku mieszkający w mało licznych gospodarstwach

Klasteryzacja wykształcenie i wartość aktywów trwałych

KMeans Clustering of education_level and durable_asset Level with Depression Labels (4 Clusters)

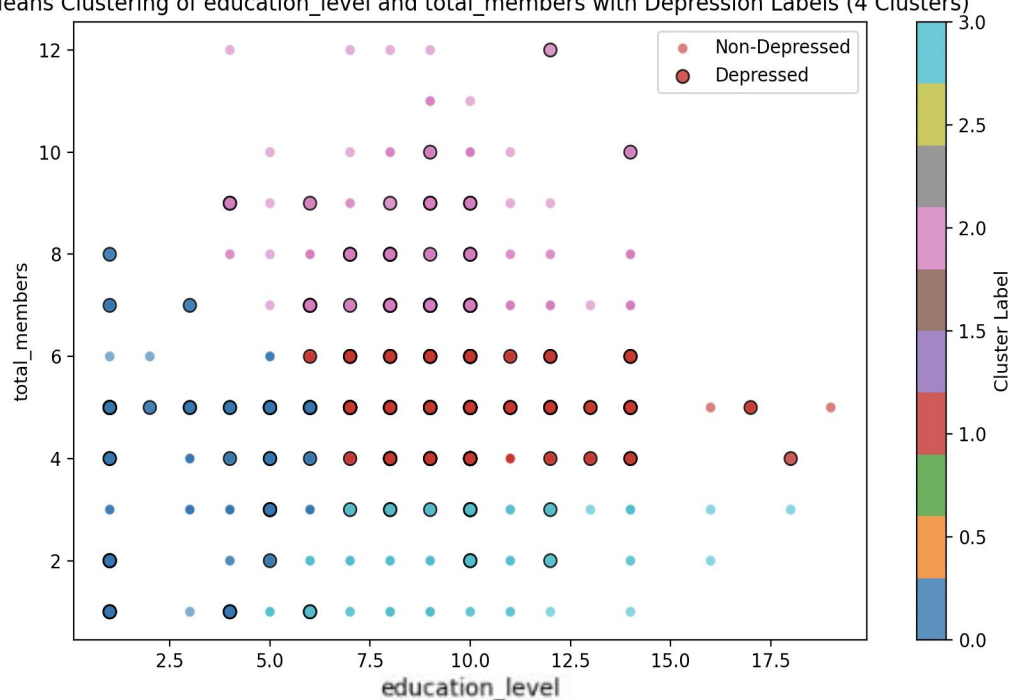


Klastry:

- ludzie biedni i słabo wykształceni
- ludzie bogaci
- ludzie biedni i średnio wykształceni
- ludzie biedni i dobrze wykształceni

Klasteryzacja wykształcenie i łączna liczba osób w gospodarstwie domowym

KMeans Clustering of education_level and total_members with Depression Labels (4 Clusters)

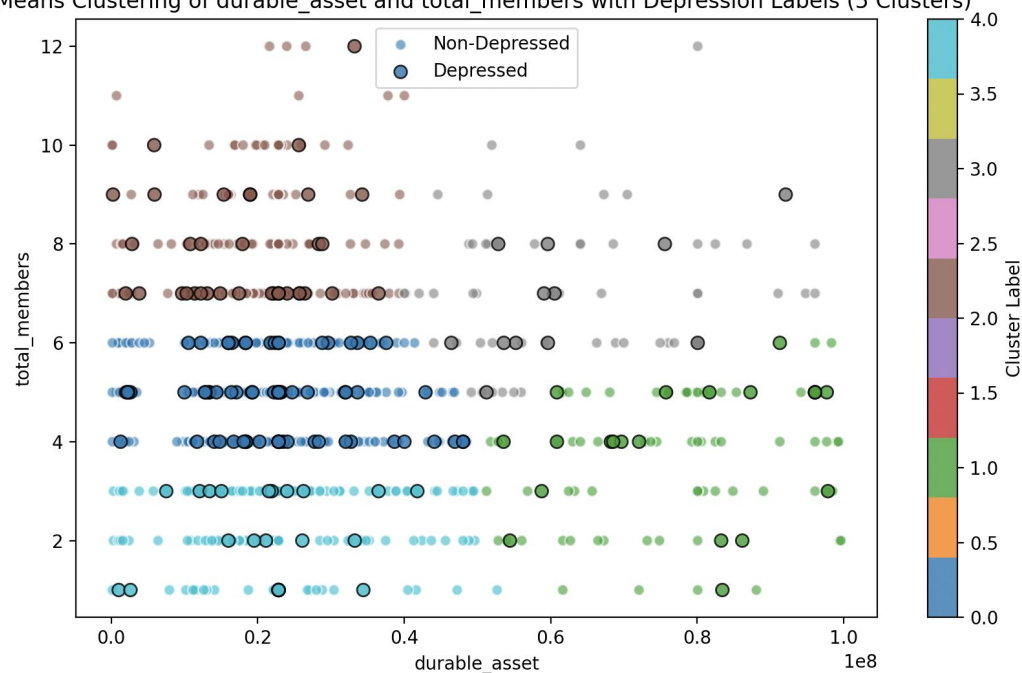


Klastry:

- słabo wykształceni ludzie
- ludzie mieszkający w mało licznych gospodarstwach
- ludzie średnio i dobrze wykształceni mieszkający w średnio licznych gospodarstwach
- ludzie mieszkający w licznych gospodarstwach

Klasteryzacja wartości aktywów trwałych i łączna liczba osób w gospodarstwie domowym

KMeans Clustering of durable_asset and total_members with Depression Labels (5 Clusters)

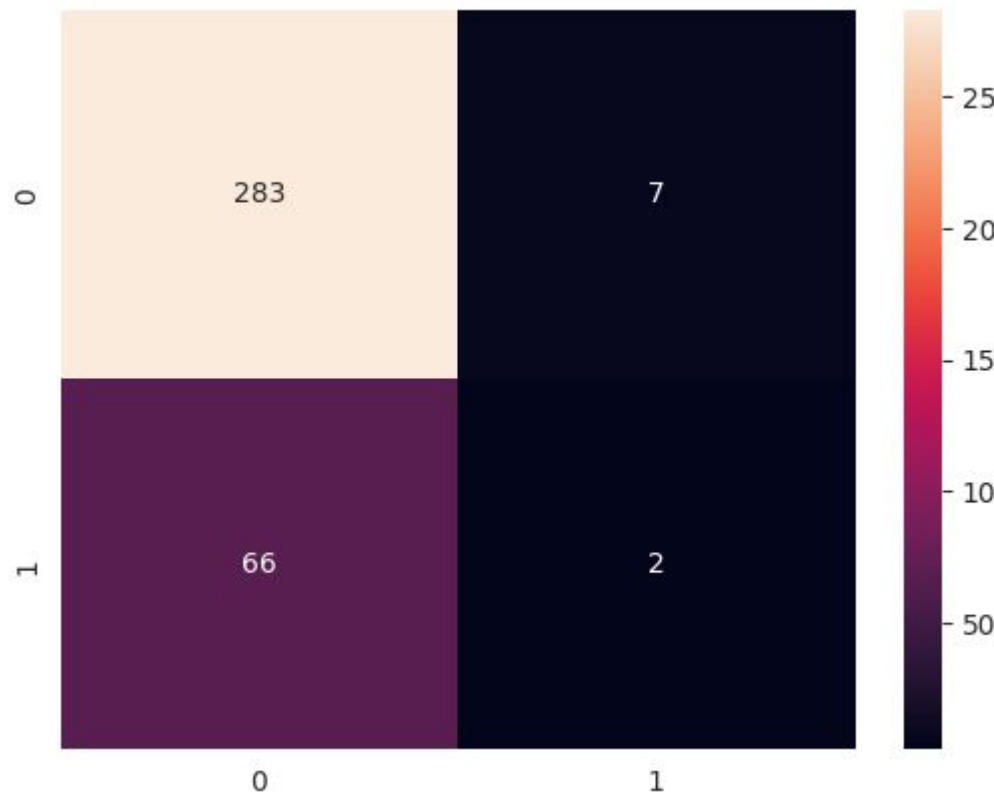


Klastry:

- ludzie biedni mieszkający w licznych gospodarstwach
- ludzie bogaci mieszkający w mało licznych gospodarstwach
- ludzie biedni mieszkający w mało licznych gospodarstwach
- **ludzie biedni mieszkający w średnio licznych gospodarstwach**
- ludzie nie biedni mieszkający w licznych i średnio licznych gospodarstwach

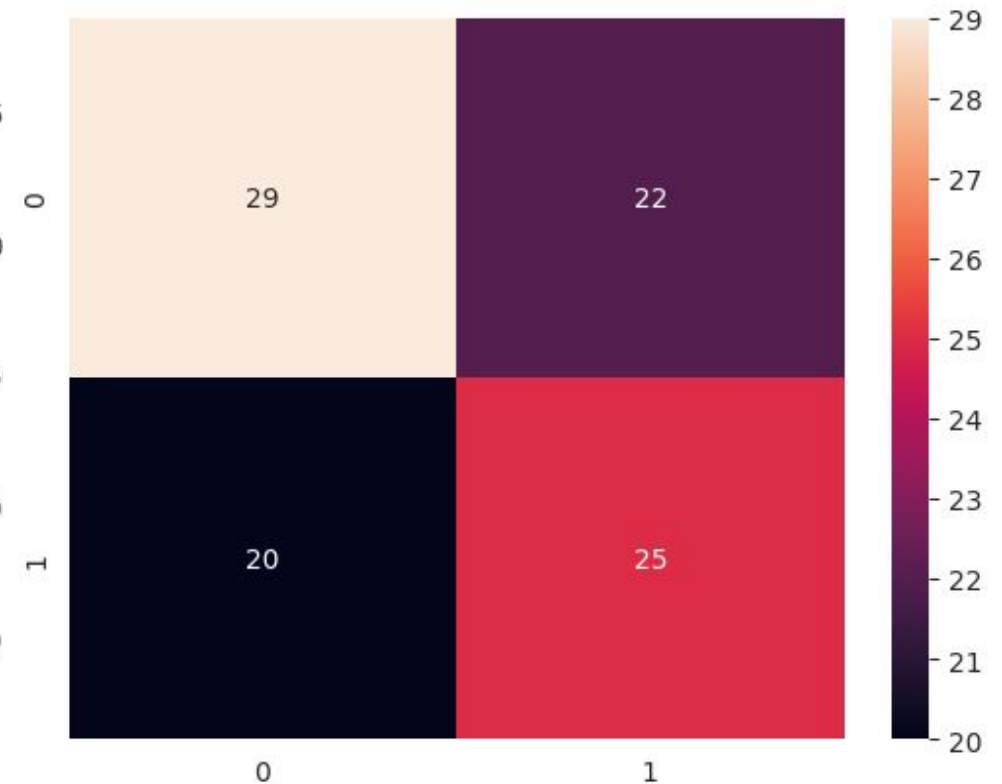
Klasyfikacja, algorytm k najbliższych sąsiadów dla zbiorów

Całego



Accuracy: 0.80

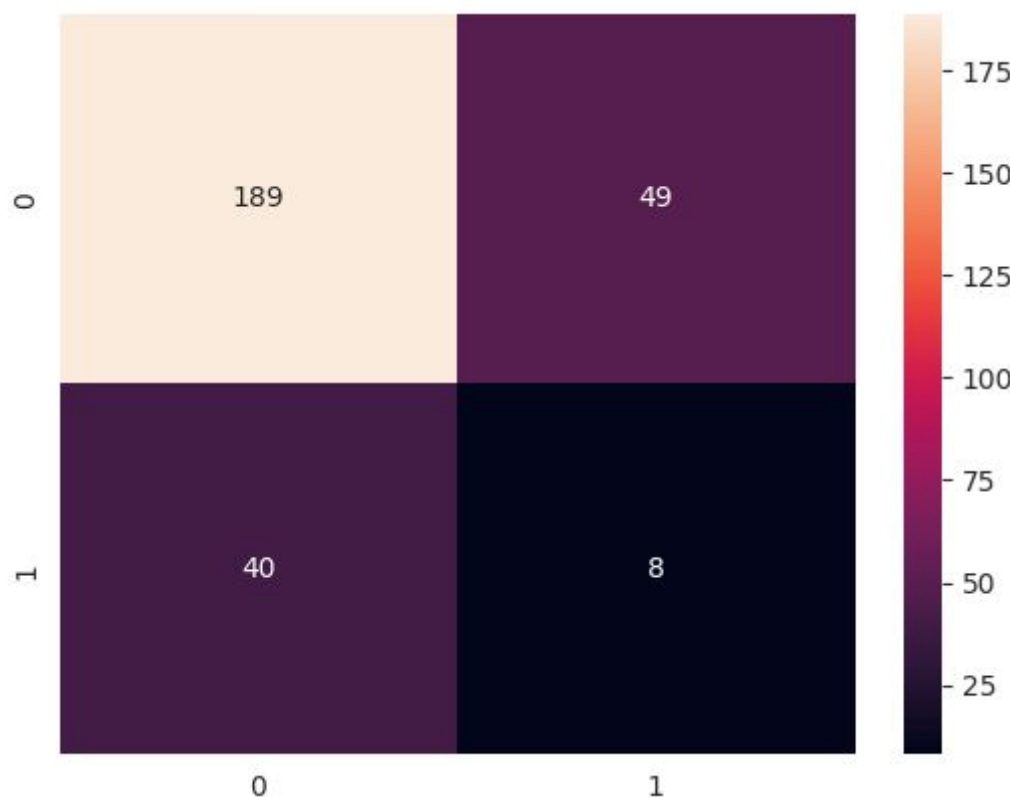
Równolicznego



Accuracy: 0.56

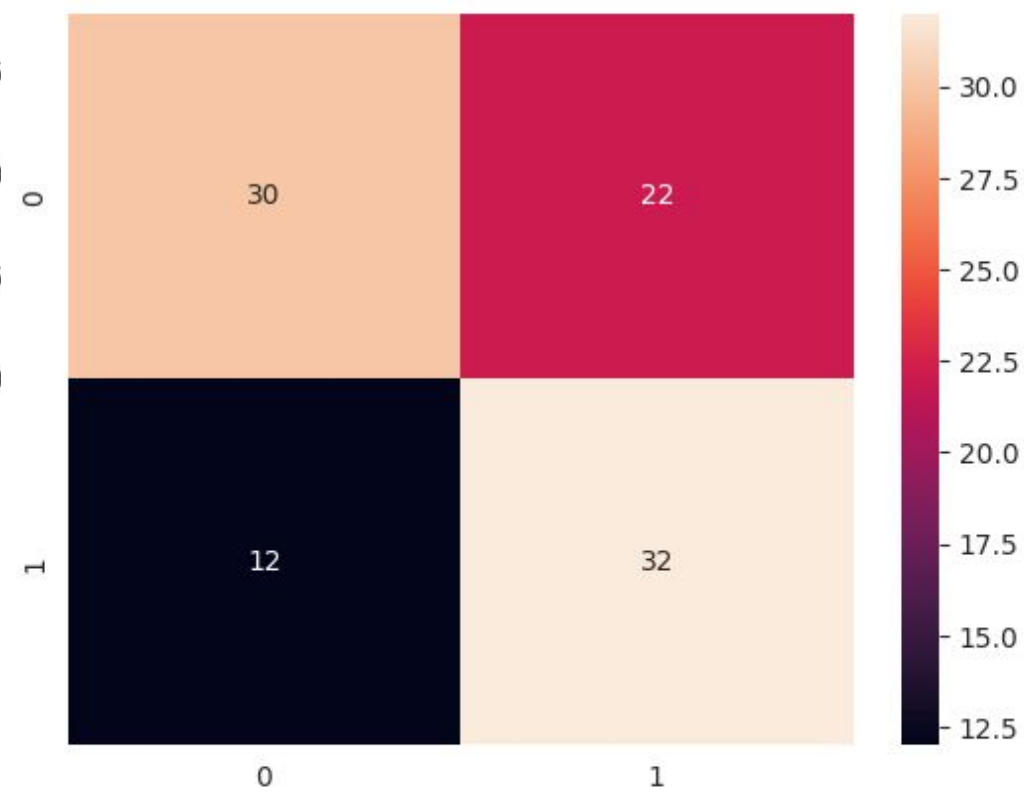
Klasyfikacja, algorytm drzewa decyzyjnego dla zbiorów

Całego



Accuracy: 0.74

Równolicznych



Accuracy: 0.65

Podsumowanie

Można wyróżnić kilka szczególnych grup bardziej narażonych na depresję:

- osoby biedne
- osoby mieszkające w średnio licznych gospodarstwach

Depresja nie wybiera, mogą ją przeżywać ludzie z różnych grup społecznych, niezależnie od wieku czy wykształcenia. Ważne jest zrozumienie, że wpływają na nią różnorodne czynniki, a każdy może być podatny na jej rozwój. Warto otwarcie rozmawiać na ten temat i dążyć do zwiększenia świadomości społecznej w kwestii zdrowia psychicznego



Dziękujemy za uwagę :)

