

EDA Zestawu danych - Titanic

O Danych

Dane o pasażerach Titanica

Zbiór danych zawiera informacje o pasażerach RMS Titanic, który zatonął 15 kwietnia 1912 roku po zderzeniu z górą lodową. Dane obejmują takie atrybuty jak klasa podróży, wiek, płeć, liczba rodzeństwa/małżonków na pokładzie, liczba rodziców/dzieci na pokładzie, cena biletu oraz miejsce zaokrętowania.

Zbiór zawiera także informację o tym, czy pasażer przeżył katastrofę.

Titanic przewoził ponad 2,200 osób, z czego ponad 1,500 zginęło, co czyni tę katastrofę jedną z najbardziej tragicznych w historii morskiej.

Kolumny:

- **pclass** - Klasa biletu
- **survived** - Czy pasażer przeżył katastrofę
- **name** - Imię i nazwisko pasażera
- **sex** - Płeć pasażera
- **age** - Wiek pasażera
- **sibsp** - Liczba rodzeństwa/małżonków na pokładzie
- **parch** - Liczba rodziców/dzieci na pokładzie
- **ticket** - Numer biletu
- **fare** - Cena biletu
- **cabin** - Numer kabiny
- **embarked** - Port, w którym pasażer wszedł na pokład (C = Cherbourg, Q = Queenstown, S = Southampton)
- **boat** - Numer łodzi ratunkowej
- **body** - Numer ciała (jeśli pasażer nie przeżył i ciało zostało odnalezione)
- **home.dest** - Miejsce docelowe

Zapraszam na moją analizę danych dotyczącą pasażerów Titanica.

Spróbuję odpowiedzieć na pytanie o przeżywalność pasażerów rejsu.



1) Wstępna inspekcja danych:

Dane zawierają **1310** wierszy i **14** kolumn.

Ogólny przegląd:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked
1206	3.0	0.0	Skoog, Master. Harald	male	4.0000	3.0	2.0	347088	27.9000	NaN	
55	1.0	1.0	Carter, Miss. Lucile Polk	female	14.0000	1.0	2.0	113760	120.0000	B96 B98	
64	1.0	1.0	Chambers, Mr. Norman Campbell	male	27.0000	1.0	0.0	113806	53.1000	E8	
306	1.0	0.0	White, Mr. Percival Wayland	male	54.0000	0.0	1.0	35281	77.2875	D26	
784	3.0	0.0	Dyker, Mr. Adolf Fredrik	male	23.0000	1.0	0.0	347072	13.9000	NaN	
519	2.0	0.0	Norman, Mr. Robert Douglas	male	28.0000	0.0	0.0	218629	13.5000	NaN	
213	1.0	1.0	Newell, Miss. Madeleine	female	31.0000	1.0	0.0	35273	113.2750	D36	
1174	3.0	0.0	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8.0	2.0	CA. 2343	69.5500	NaN	
1162	3.0	1.0	Ryan, Mr. Edward	male	NaN	0.0	0.0	383162	7.7500	NaN	
534	2.0	1.0	Phillips, Miss. Kate Florence ("Mrs Kate Louis...	female	19.0000	0.0	0.0	250655	26.0000	NaN	
300	1.0	0.0	Van der hoef, Mr. Wyckoff	male	61.0000	0.0	0.0	111240	33.5000	B19	
590	2.0	1.0	West, Miss. Barbara J	female	0.9167	1.0	2.0	C.A. 34651	27.7500	NaN	
230	1.0	1.0	Perreault, Miss. Anne	female	30.0000	0.0	0.0	12749	93.5000	B73	
769	3.0	0.0	Dennis, Mr. Samuel	male	22.0000	0.0	0.0	A/5 21172	7.2500	NaN	

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked
515	2.0	1.0	Navratil, Master. Michel M	male	3.0000	1.0	1.0	230080	26.0000	F2	
585	2.0	0.0	Weisz, Mr. Leopold	male	27.0000	1.0	0.0	228414	26.0000	NaN	
128	1.0	0.0	Gee, Mr. Arthur H	male	47.0000	0.0	0.0	111320	38.5000	E63	
78	1.0	1.0	Compton, Mrs. Alexander Taylor (Mary Eliza Ing...	female	64.0000	0.0	2.0	PC 17756	83.1583	E45	
866	3.0	1.0	Hirvonen, Miss. Hildur E	female	2.0000	0.0	1.0	3101298	12.2875	NaN	
1152	3.0	0.0	Robins, Mr. Alexander A	male	50.0000	1.0	0.0	A/5. 3337	14.5000	NaN	

Wartości unikalne & typy danych:

pclass	3
survived	2
name	1307
sex	2
age	98
sibsp	7
parch	8
ticket	929
fare	281
cabin	186
embarked	3
boat	27
body	2
home.dest	369
dtype: int64	

```
pclass      float64
survived     float64
name         object
sex          object
age          float64
sibsp        float64
parch        float64
ticket       object
fare         float64
cabin        object
embarked     object
boat         object
body         float64
home.dest    object
dtype: object
```

2) Czyszczenie i transformacja danych:

```
pclass      0
survived     0
name         0
sex          0
age          263
sibsp        0
parch        0
ticket       0
fare         1
cabin        1014
embarked     2
boat         823
body         1188
home.dest    564
dtype: int64
```

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat
906	3	False	Jonsson, Mr. Nils Hilding	male	27.0	0	0	350408	7.8542	NaN	S	1
707	3	False	Carlsson, Mr. August Sigfrid	male	28.0	0	0	350042	7.7958	NaN	S	1
479	2	True	Laroche, Miss. Simonne Marie Anne Andree	female	3.0	1	2	SC/Paris 2123	41.5792	NaN	C	
126	1	False	Futrelle, Mr. Jacques Heath	male	37.0	1	0	113803	53.1000	C123	S	1
1083	3	False	Olsen, Mr. Henry Margido	male	28.0	0	0	C 4001	22.5250	NaN	S	1



```

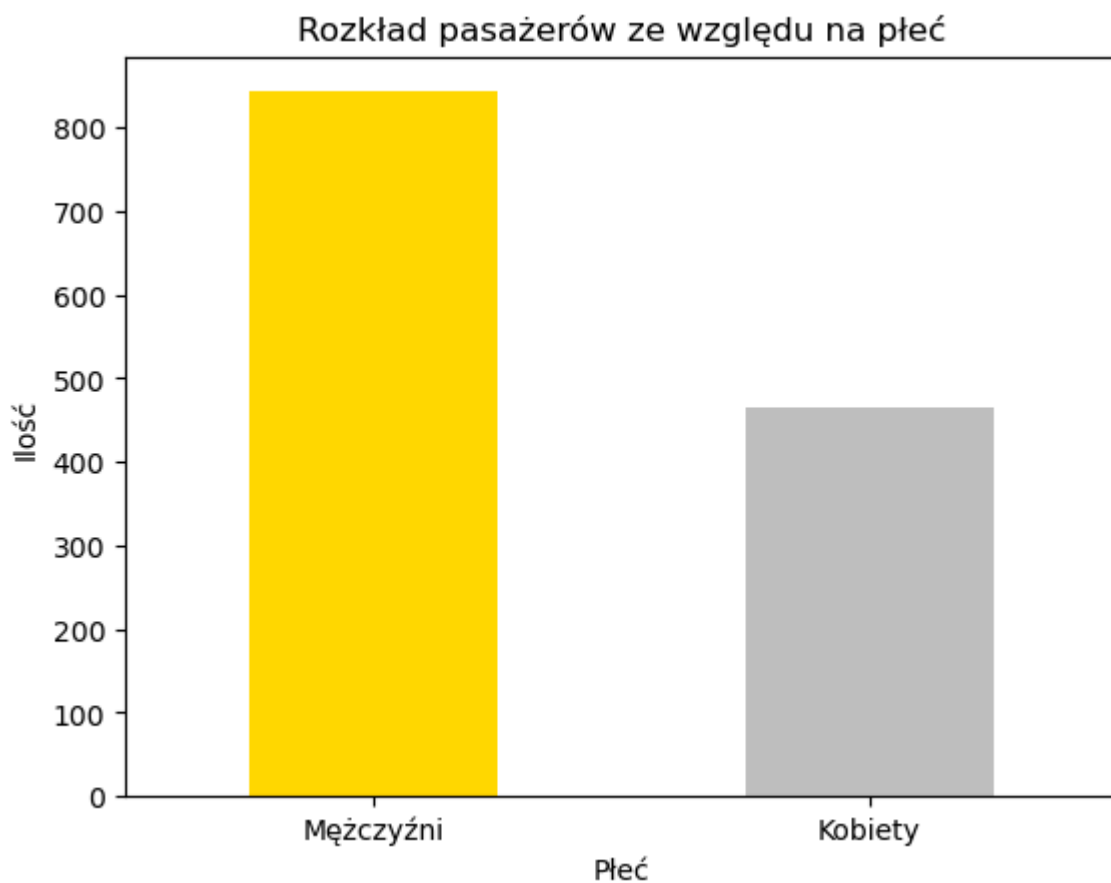
pclass      int32
survived     bool
name        object
sex         object
age         float64
sibsp       int32
parch       int32
ticket      object
fare        float64
cabin       object
embarked    object
boat        object
body        bool
home.dest   object
dtype: object

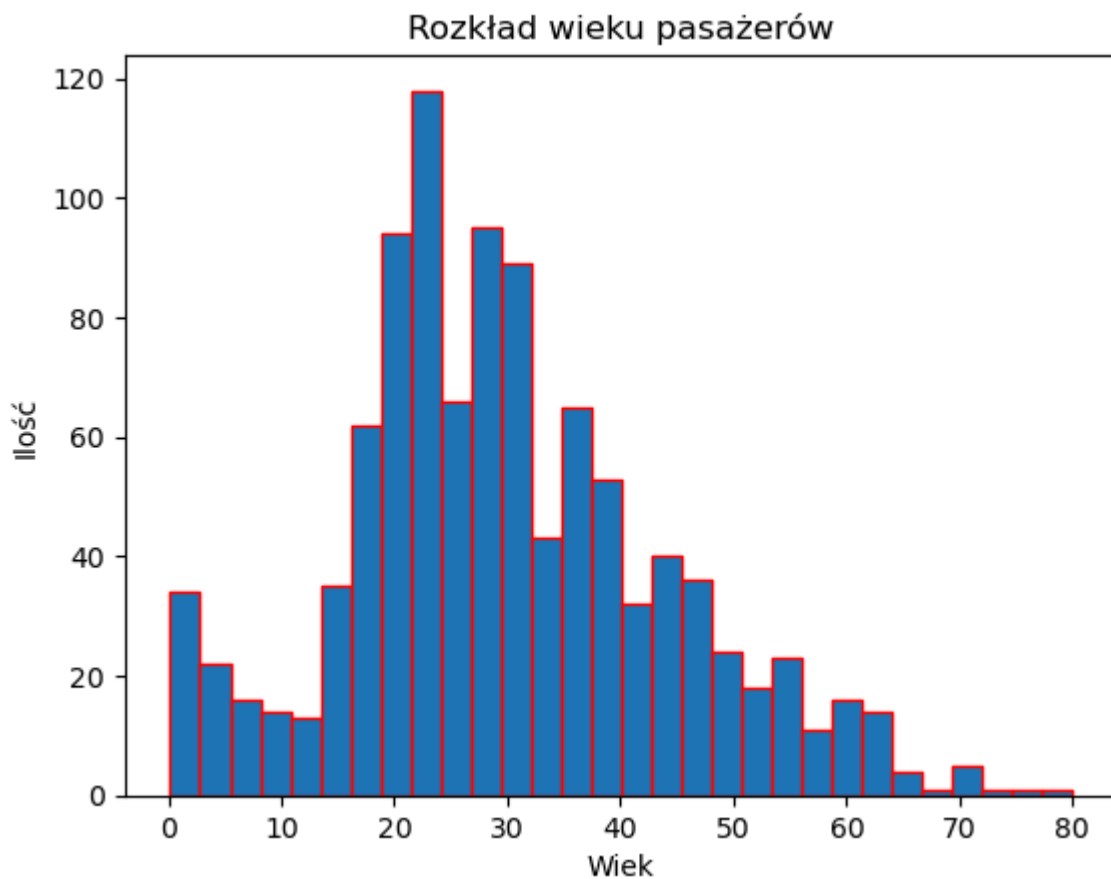
```

3) Analiza jednowymiarowa:

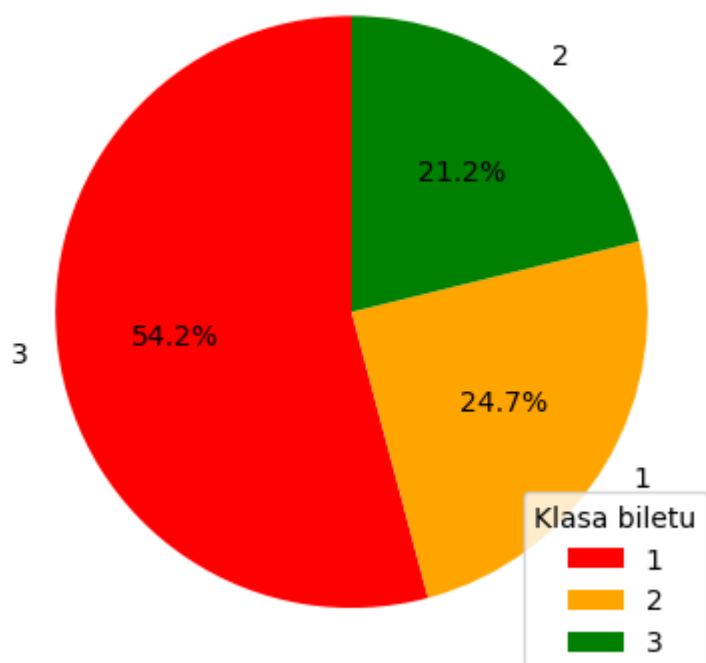
	pclass	age	sibsp	parch	fare
count	1309.000000	1046.000000	1309.000000	1309.000000	1308.000000
mean	2.294882	29.881135	0.498854	0.385027	33.295479
std	0.837836	14.413500	1.041658	0.865560	51.758668
min	1.000000	0.166700	0.000000	0.000000	0.000000
25%	2.000000	21.000000	0.000000	0.000000	7.895800
50%	3.000000	28.000000	0.000000	0.000000	14.454200
75%	3.000000	39.000000	1.000000	0.000000	31.275000
max	3.000000	80.000000	8.000000	9.000000	512.329200

Wykresy:



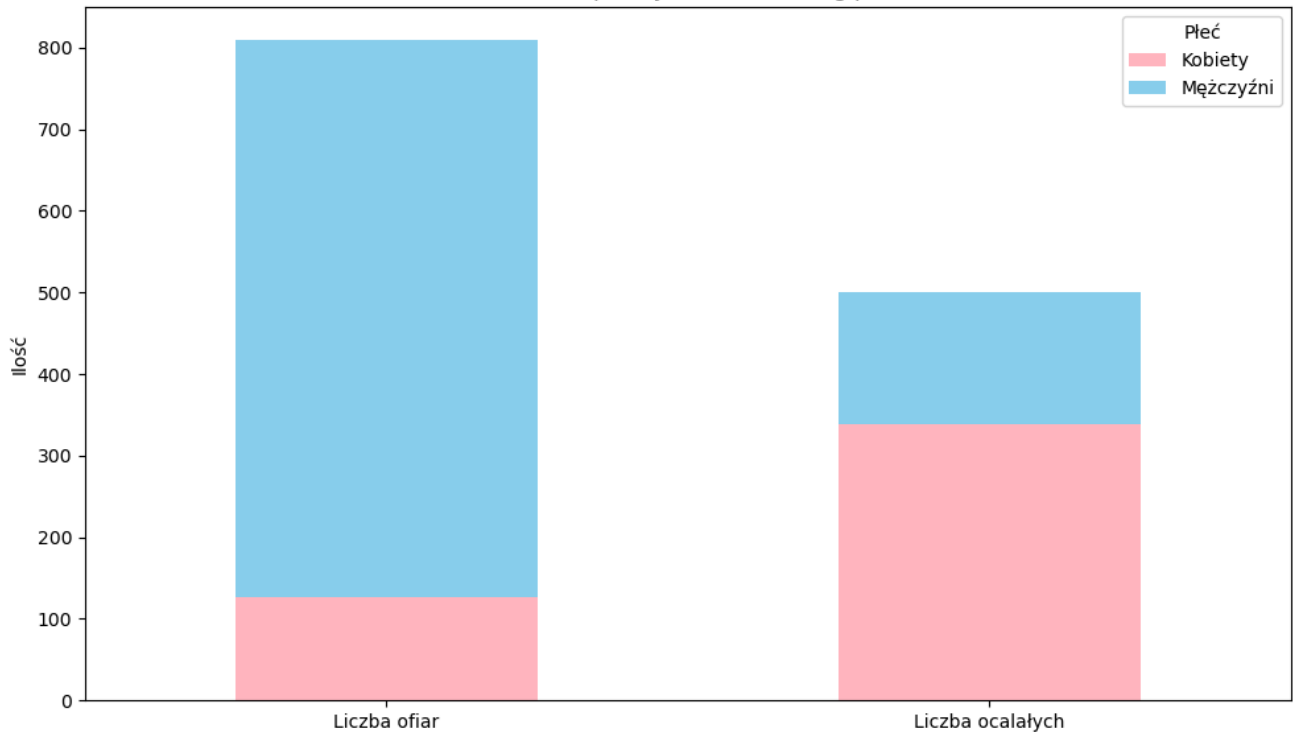


Rozkład pasażerów ze względu na klasę biletu

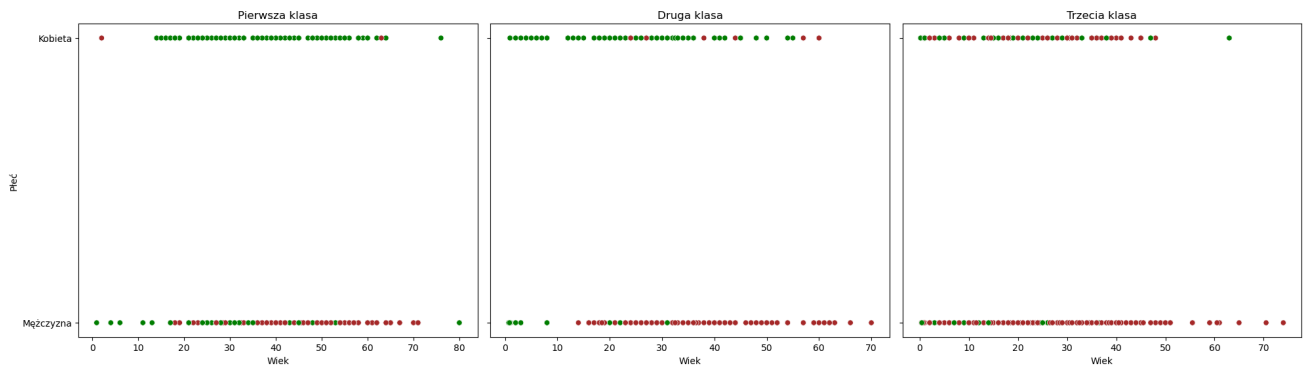
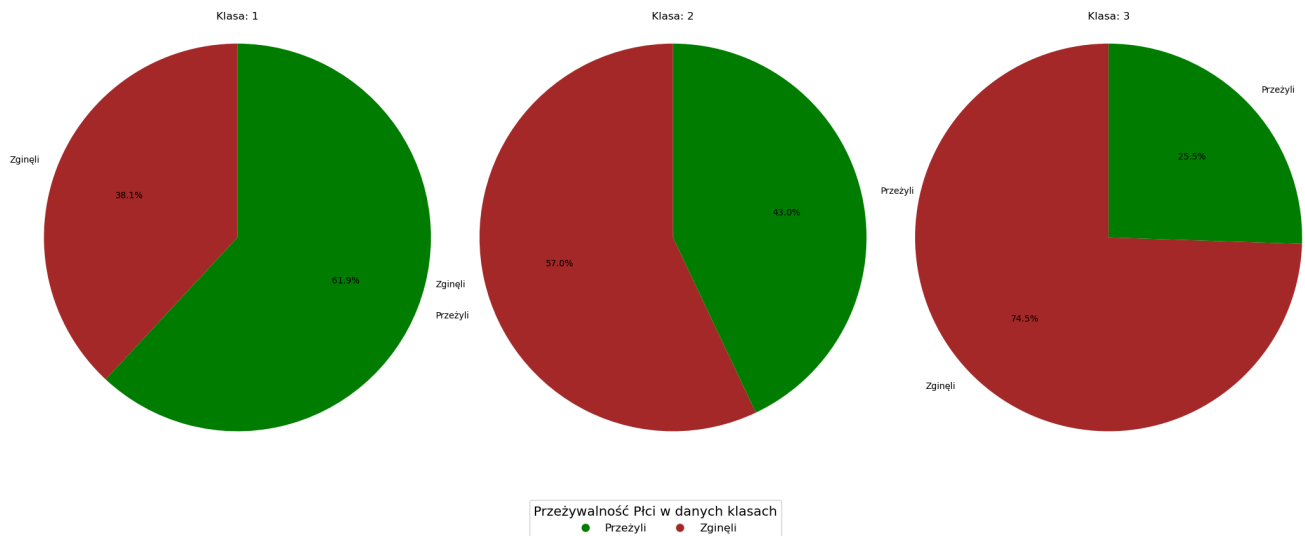


4) Analiza dwuwymiarowa:

Wskaźnik przeżywalności według płci



Przeżywalność pasażerów z biletami różnych klas



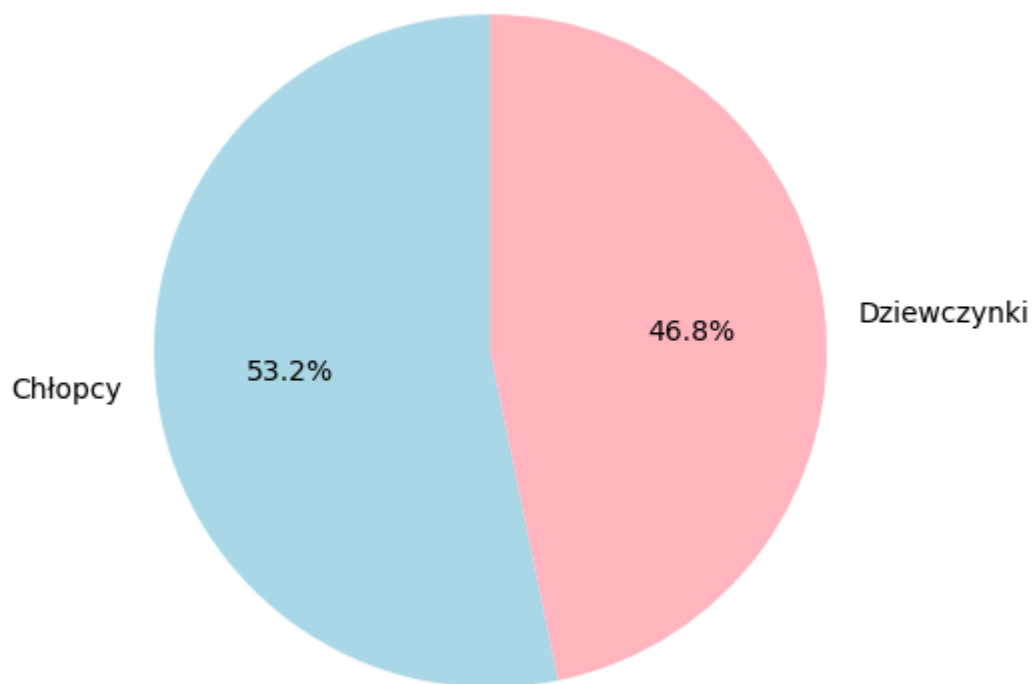
Sprawdzenie przeżywalności dla najmłodszych pasażerów:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	bo
1	1	True	Allison, Master. Hudson Trevor	male	0.9167	1	2	113781	151.55	C22 C26	S	
2	1	False	Allison, Miss. Helen Lorraine	female	2.0000	1	2	113781	151.55	C22 C26	S	N
53	1	False	Carrau, Mr. Jose Pedro	male	17.0000	0	0	113059	47.10	NaN	S	N
54	1	True	Carter, Master. William Thornton II	male	11.0000	1	2	113760	120.00	B96 B98	S	
55	1	True	Carter, Miss. Lucile Polk	female	14.0000	1	2	113760	120.00	B96 B98	S	

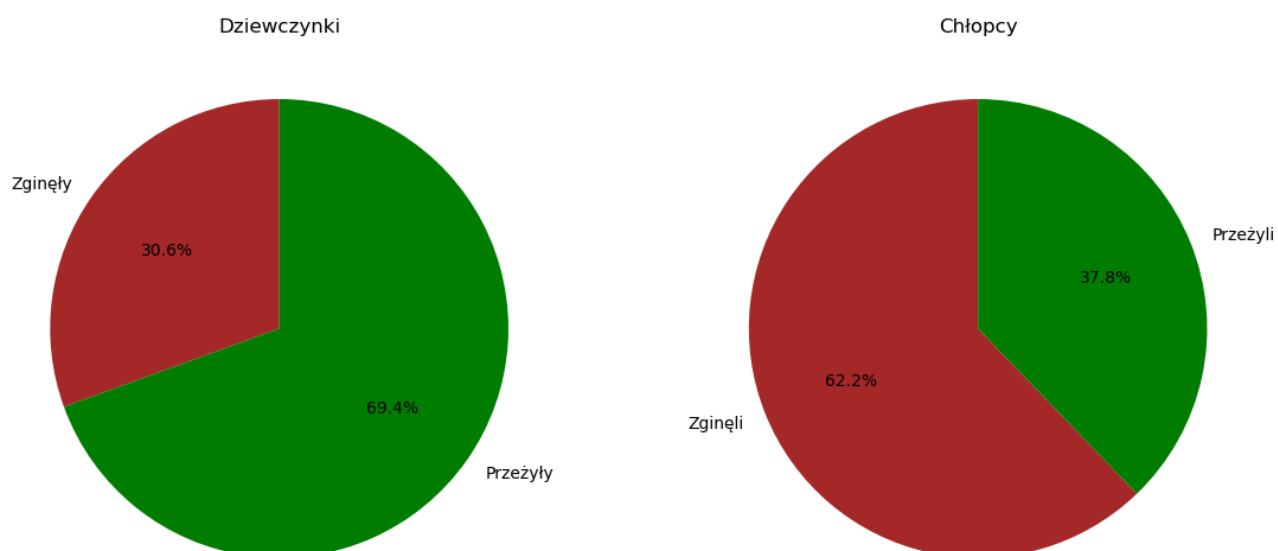


```
sex
male      82
female    72
Name: count, dtype: int64
```

Podział dzieci ze względu na płeć

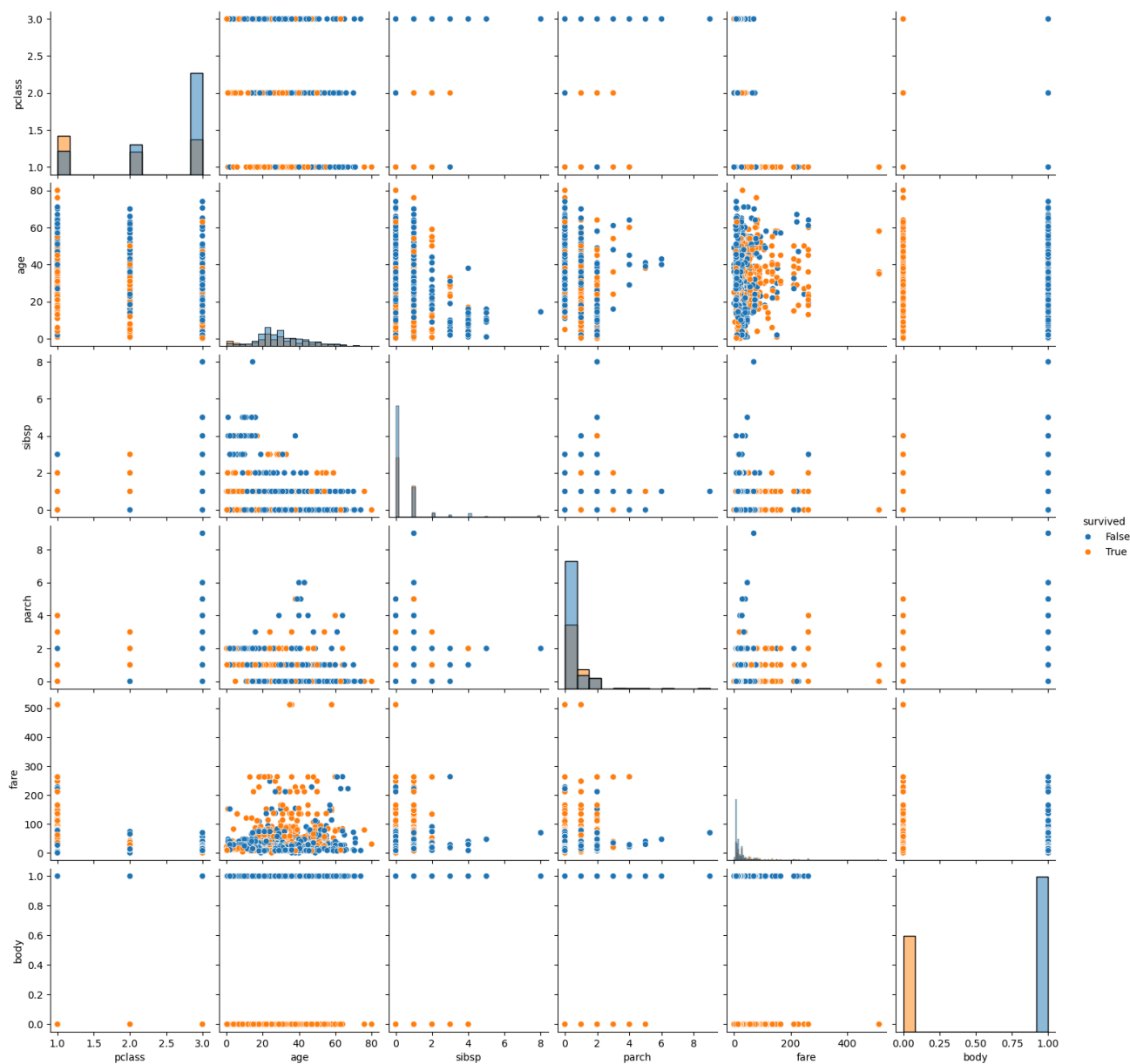


Podział dzieci wg przetrwania i płci



Sprawdzenie rozkładów i wykrycie wzorców w zbiorze

<seaborn.axisgrid.PairGrid at 0x17792334290>



Sprawdzenie korelacji między kolumnami:

	pclass	survived	age	sibsp	parch	fare	body
pclass	1.000000	-0.319979	-0.411086	0.047746	0.017685	-0.565255	0.319979
survived	-0.319979	1.000000	-0.053958	-0.012657	0.114091	0.249164	-1.000000
age	-0.411086	-0.053958	1.000000	-0.243139	-0.150241	0.178739	0.053958
sibsp	0.047746	-0.012657	-0.243139	1.000000	0.374291	0.141184	0.012657
parch	0.017685	0.114091	-0.150241	0.374291	1.000000	0.216723	-0.114091
fare	-0.565255	0.249164	0.178739	0.141184	0.216723	1.000000	-0.249164
body	0.319979	-1.000000	0.053958	0.012657	-0.114091	-0.249164	1.000000

<Axes: >



Analiza danych dotyczących pasażerów ujawnia kilka istotnych spostrzeżeń i wniosków:

Niekompletność Danych:

Znaczące braki danych występują w kolumnach dotyczących kabin, łodzi ratunkowych i ciał,

co utrudnia pełną rekonstrukcję warunków katastrofy.

Dane Demograficzne:

Na statku znajdowało się 466 kobiet i 843 mężczyzn. Większość ofiar stanowili mężczyźni – przeżyło zaledwie 161 z 843, podczas gdy ze 466 kobiet przeżyło 339.

Wiek Pasażerów:

Średnia wieku pasażerów wynosiła 29 lat, co wskazuje na młody profil demograficzny podróżnych.

Podział Klasy:

Pasażerowie byli rozmieszczeni pomiędzy trzy klasy: 24.7% w pierwszej, 21.2% w drugiej oraz znacząca większość,

54.2%, w trzeciej klasie.

Przeżywalność:

Ogólnie rzecz biorąc, większość pasażerów nie przeżyła katastrofy. Szczególnie narażeni byli mężczyźni oraz pasażerowie trzeciej klasy.

Kobiety, a zwłaszcza te z pierwszej i drugiej klasy, miały wyższą szansę na przeżycie.

Dzieci:

Na pokładzie znajdowało się 82 chłopców i 72 dziewczynki poniżej 18 roku życia. Przeżywalność wśród dzieci była

wyższa w przypadku dziewczynek (69.4%) niż chłopców (37.8%).

Wpływy na Przeżywalność:

Analiza korelacji wykazała, że na przeżywalność wpływały czynniki takie jak obecność rodziny na pokładzie,

cena biletów (które były wysoko skorelowane z przeżywalnością) oraz klasa biletu.

Podsumowując, analiza danych pozwala na zrozumienie nie tylko demografii pasażerów,

ale także kluczowych czynników wpływających na przeżywalność w katastrofie.