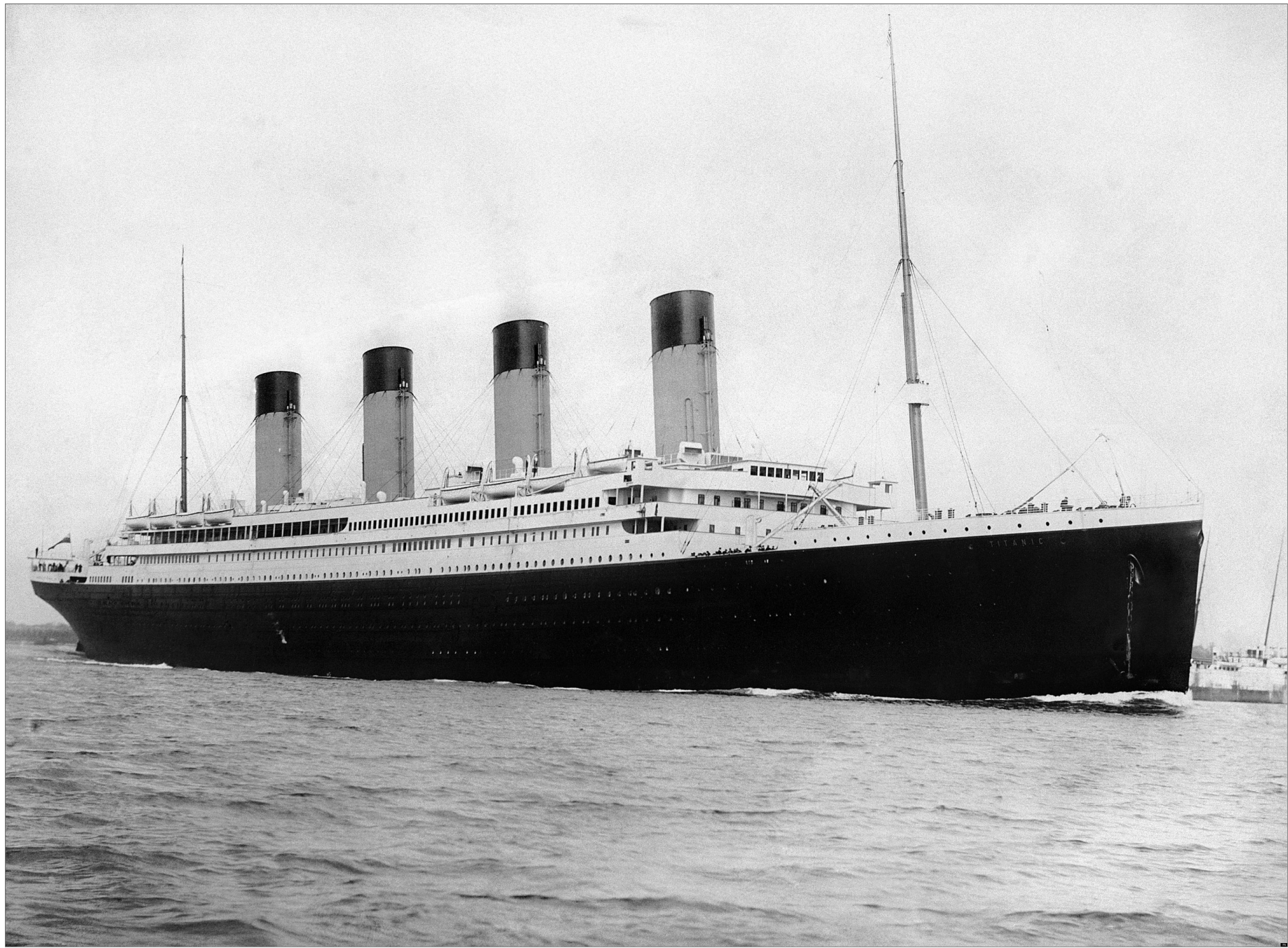


## RMS Titanic (Royal Mail Ship Titanic)

BRYTYJSKI TRANSATLANTYK TYPU OLYMPIC W NOCY Z **14 NA 15 KWIETNIA 1912** ROKU, PODCZAS DZIEWICZEGO REJSU NA TRASIE SOUTHAMPTON — CHERBOURG — QUEENSTOWN — NOWY JORK, ZDERZYŁ SIĘ Z GÓRĄ LODOWĄ I ZATONAŁ. JEGO KATASTROFA SPOWODOWAŁA NOWELIZACJĘ ZASAD BEZPIECZEŃSTWA MORSKIEGO. DANE O OFIARACH SĄ NIEJEDNOZNACZNE. SPOŚRÓD OKOŁO **2208** PASAŻERÓW I ZAŁOGI „TITANICA” ZGINEŁO OKOŁO **1496** OSÓB. PRZEŻYŁO KATASTROFĘ TYLKO OKOŁO **712** (ŹRÓDŁO).

Poniższa analiza została przeprowadzona na próbie liczącej 1310 pasażerów.



# 1. Ogólny przegląd danych

PRZYKŁAD 5 LOSOWYCH WIERSZY

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
1016	3.0	0.0	Mernagh, Mr. Robert	male	NaN	0.0	0.0	368703	7.7500	NaN	Q	NaN	NaN	NaN
198	1.0	1.0	Marvin, Mrs. Daniel Warner (Mary Graham Carmichael Farquarson)	female	18.0	1.0	0.0	113773	53.1000	D30	S	10	NaN	New York, NY
1151	3.0	0.0	Risien, Mrs. Samuel (Emma)	female	NaN	0.0	0.0	364498	14.5000	NaN	S	NaN	NaN	NaN
586	2.0	1.0	Weisz, Mrs. Leopold (Mathilde Francoise Pede)	female	29.0	1.0	0.0	228414	26.0000	NaN	S	10	NaN	Bromsgrove, England / Montreal, PQ
945	3.0	1.0	Lam, Mr. Ali	male	NaN	0.0	0.0	1601	56.4958	NaN	S	C	NaN	NaN

## INFORMACJA O TYPIE DANYCH DLA POSZCZEGÓLNYCH KOLUMN ORAZ IŁOŚCI WIERSZY BEZ BRAKÓW

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1310 entries, 0 to 1309
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0   pclass      1309 non-null   float64
1   survived    1309 non-null   float64
2   name        1309 non-null   object
3   sex         1309 non-null   object
4   age         1046 non-null   float64
5   sibsp       1309 non-null   float64
6   parch       1309 non-null   float64
7   ticket      1309 non-null   object
8   fare        1308 non-null   float64
9   cabin       295 non-null    object
10  embarked    1307 non-null   object
11  boat        486 non-null    object
12  body        121 non-null    float64
13  home.dest    745 non-null    object
dtypes: float64(7), object(7)
memory usage: 143.4+ KB
```

## WARTOŚCI UNIKATOWE W POSZCZEGÓLNYCH KOLUMNACH

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	121
home.dest	369
dtype: int64	

## 2. Analiza brakujących wartości

	brakujące rzędy	braki wyrażone w procentach
pclass	1	0.076336
survived	1	0.076336
name	1	0.076336
sex	1	0.076336
age	264	20.152672
sibsp	1	0.076336
parch	1	0.076336
ticket	1	0.076336
fare	2	0.152672
cabin	1015	77.480916
embarked	3	0.229008
boat	824	62.900763
body	1189	90.763359
home.dest	565	43.129771

DO ANALIZY PRZEŻYCIA W KATASTROFIE TITANICA, KLUCZOWE SĄ INFORMACJE O:

- przeżyciu,
- płci,
- klasie biletu,
- wieku.

DLATEGO POSTANOWIONO USUNĄĆ WSZYSTKIE WERSZE Z BRAKAMI W NAJWAŻNIEJSZYCH DO ANALIZY KOLUMNACH. W WYNIKU ZABIEGU OTRZYMANO PRÓBĘ LICZĄCĄ **1046 PRZYPADKÓW** .

### 3. Transformacja danych

Polegała na:

- utworzeniu nowej kolumny '**płeć**' z wartościami '**kobieta**'/'**mężczyzna**' na podstawie kolumny '**sex**';
- utworzeniu kolumny '**przeżycie str**' jako string;
- **zmianie nazw kolumn;**
- **uporządkowaniu kolejności** kolumn;
- sprawdzeniu czy dane zawierają **duplikaty** (nie wykryto).



W WYNIKU TRANSFORMACJI OTRZYMANO DANE GOTOWE DO ANALIZY STATYSTYCZNEJ.

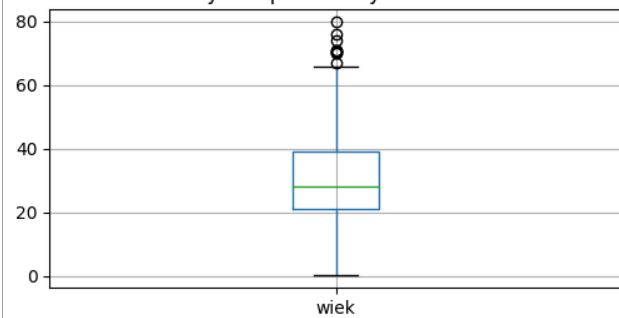
imię i nazwisko	przeżycie str	przeżycie int	płeć	wiek	liczba rodzeństwa/małżonka	liczba rodziców/dzieci	klasa biletu	cena biletu	nr biletu	nr kabiny	port początkowy	nr łodzi ratunkowej
Nicola-Yarred, Miss. Jamila	tak	1	kobieta	14	1	0	3	11	2651	nan	C	C
Maybery, Mr. Frank Hubert	nie	0	mężczyzna	40	0	0	2	16	239059	nan	S	nan
Mellinger, Miss. Madeleine Violet	tak	1	kobieta	13	0	1	2	20	250644	nan	S	14
Aldworth, Mr. Charles Augustus	nie	0	mężczyzna	30	0	0	2	13	248744	nan	S	nan
Karlsson, Mr. Julius Konrad Eugen	nie	0	mężczyzna	33	0	0	3	8	347465	nan	S	nan

## 4. Analiza statystyczna

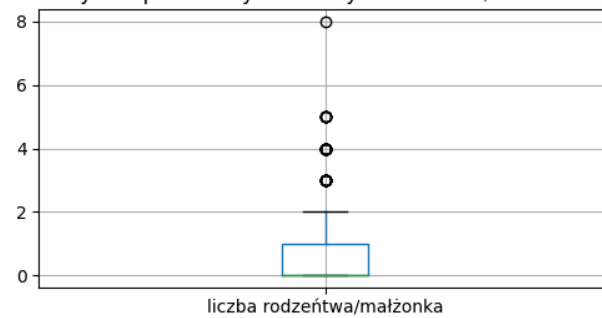
### 4.1. Ogólne cechy danych liczbowych

	przeżycie int	wiek	liczba rodzeństwa/małżonka	liczba rodziców/dzieci	klasa biletu	cena biletu	nr ciała
count	1046.000000	1046.000000	1046.000000	1046.000000	1046.000000	1045.000000	120.000000
mean	0.408222	29.881135	0.502868	0.42065	2.207457	36.686080	160.87500
std	0.491740	14.413500	0.912167	0.83975	0.841497	55.732533	98.10392
min	0.000000	0.166700	0.000000	0.00000	1.000000	0.000000	1.00000
25%	0.000000	21.000000	0.000000	0.00000	1.000000	8.050000	71.50000
50%	0.000000	28.000000	0.000000	0.00000	2.000000	15.750000	155.50000
75%	1.000000	39.000000	1.000000	1.00000	3.000000	35.500000	256.50000
max	1.000000	80.000000	8.000000	6.00000	3.000000	512.329200	328.00000

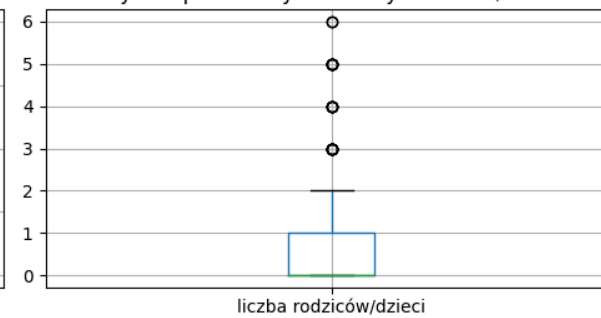
Wykres pudełkowy dla wieku



Wykres pudełkowy dla liczby rodzeństwa/małżonka



Wykres pudełkowy dla liczby rodziców/dzieci



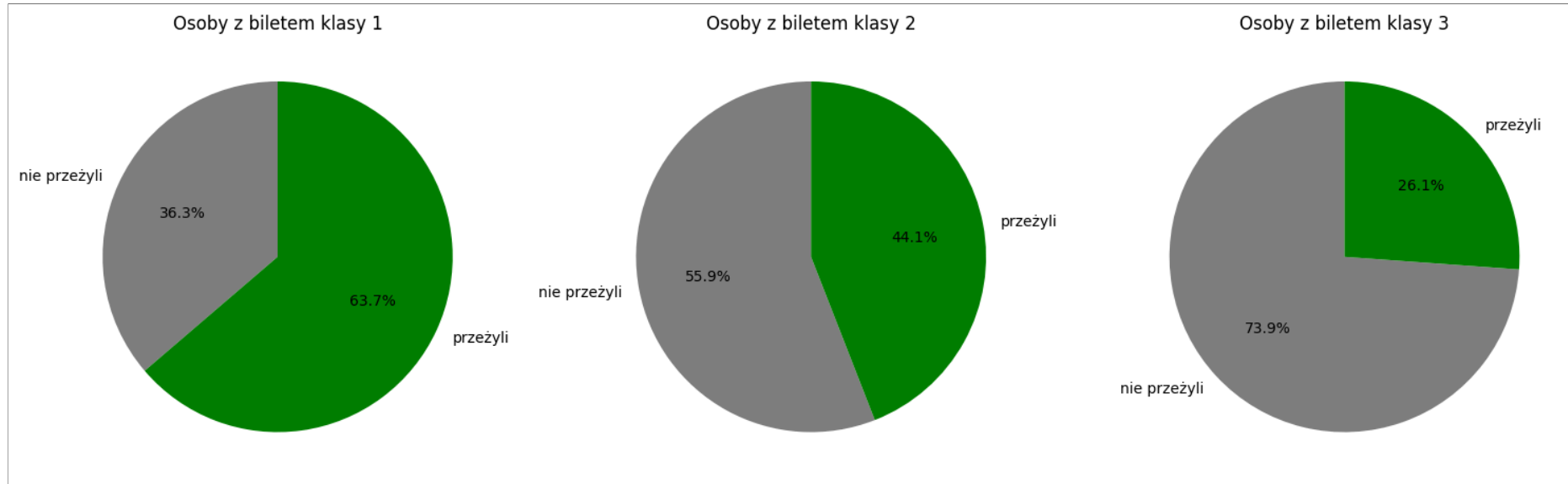
## Wnioski

- **Wiek pasażerów** w badanej próbie mieści się w przedziale od **0,17** do **80** lat, przy czym **osoby najstarsze (powyżej 66 lat)** stanowią **pojedyncze przypadki** (9 przypadków, co stanowi **0,86%** wszystkich badanych pasażerów).
- **W większości** przypadków **pasażerami były osoby samotne**, tzn. podróżujące bez małżonka, dzieci czy rodziców (tylko 34,51% badanych pasażerów podróżowało z małżonkiem lub rodzeństwem, a 4,3% osób miało więcej niż 2 rodzeństwa).
- Z **dwójką rodziców** lub z **dwójką dzieci** podróżowało **9,27%** pasażerów.
- Korelacja surowych danych **nie wykazała** zależności między zmiennymi.

## 4.2. Zależność przeżycia od klasy biletu



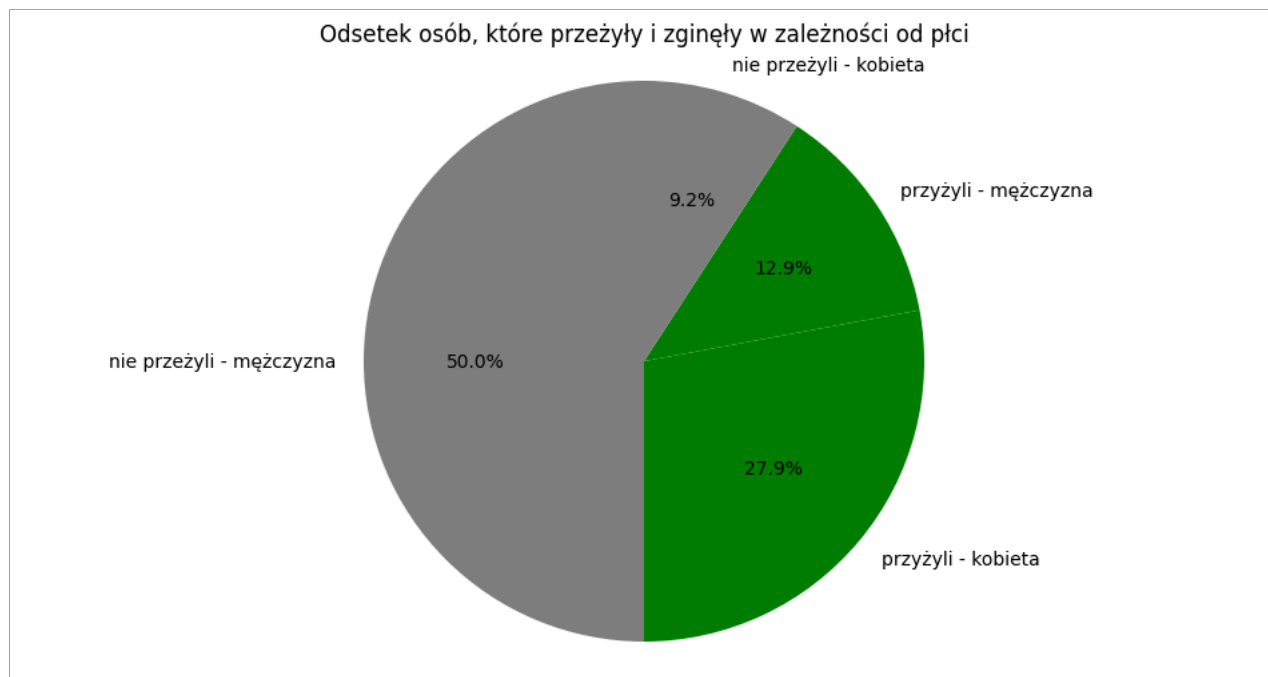
Odsetek osób z biletem w danej klasie, które przeżyły lub zginęły

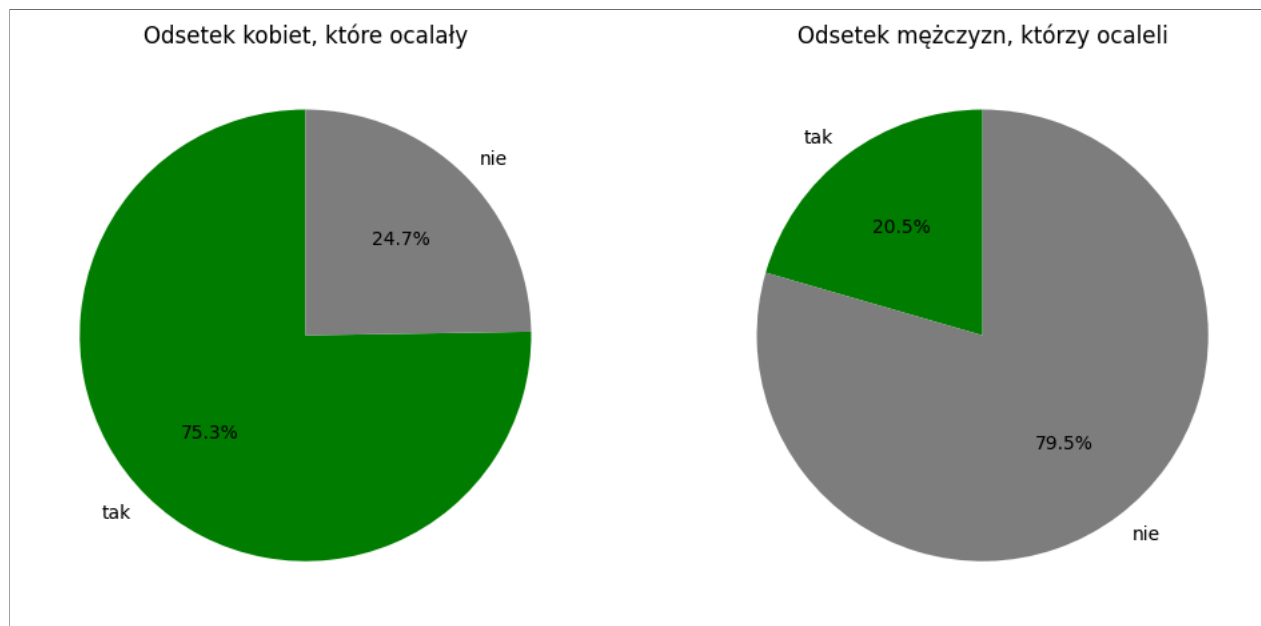


## Wnioski

- W badanej próbie **najwięcej** było **osób z biletem klasy 3**.
- **Prawie 60%** wszystkich pasażerów **nie przeżyło**, przy czym zdecydowaną większość (35,4% wszystkich pasażerów) stanowili pasażerowie z biletami klasy 3.
- **Spośród 40%** pasażerów, którzy **ocaleli**, największy udział miały osoby z **biletem klasy 1**.
- Wykres "Odsetek osób z biletem w danej klasie, które przeżyły lub zginęły" pokazuje, że **największą szansę przeżycia** mieli pasażerowie z **biletami klasy 1 (63,7%** z tych osób przeżyło), natomiast **najmniejszą szansę przeżycia** miały osoby z **biletami klasy 3 (tylko 26,1%** z nich przeżyło).

### 4.3. Zależność przeżycia od płci



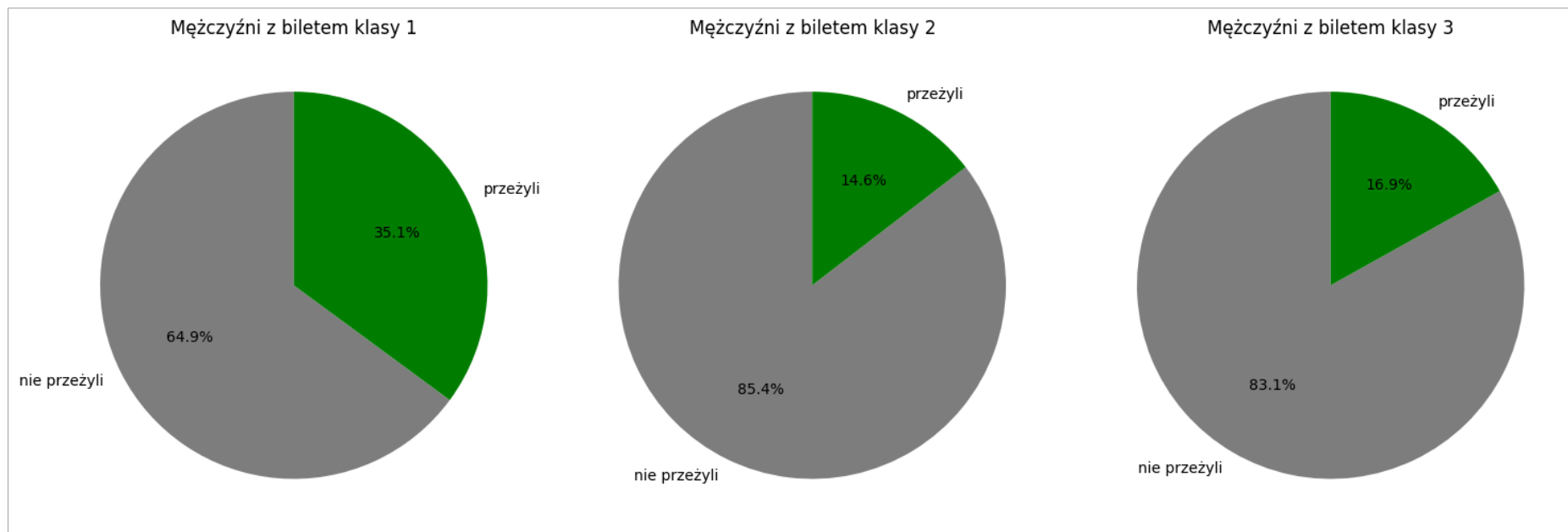


## Wnioski

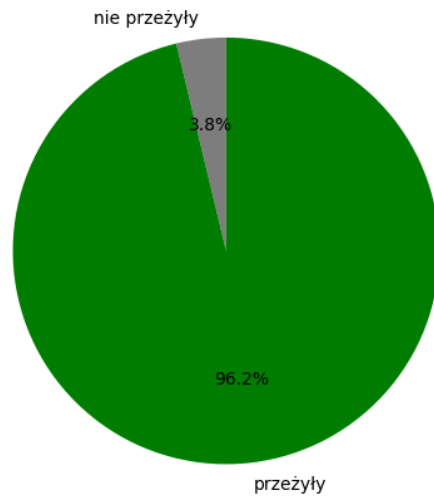
- W badanej próbie **dominowali liczebnie mężczyźni**, jednak **większość mężczyzn (79,5%) nie przeżyła** katastrofy.
- Odmienną zależność można zaobserwować z grupie kobiet, tzn. **większość kobiet przeżyła (75,3% wszystkich kobiet)**.
- **Mężczyźni**, którzy **zginęli** stanowią **50,0%** wszystkich pasażerów, natomiast **kobiety**, które **zginęły 9,2%**.
- Spośród uratowanych osób **kobiety** stanowiły około **2/3 wszystkich uratowanych**.



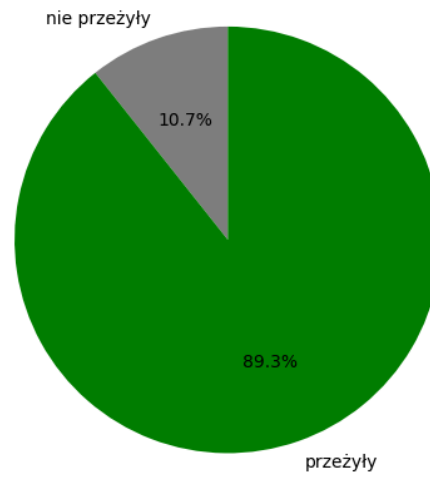
#### 4.4. Zależność przeżycia od płci i klasy biletu



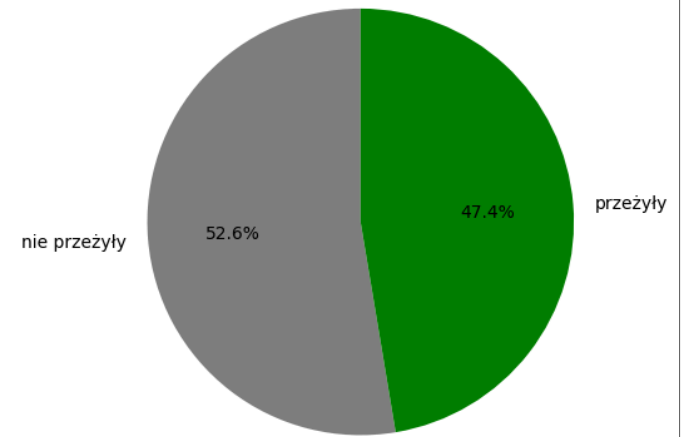
Kobiety z biletem klasy 1



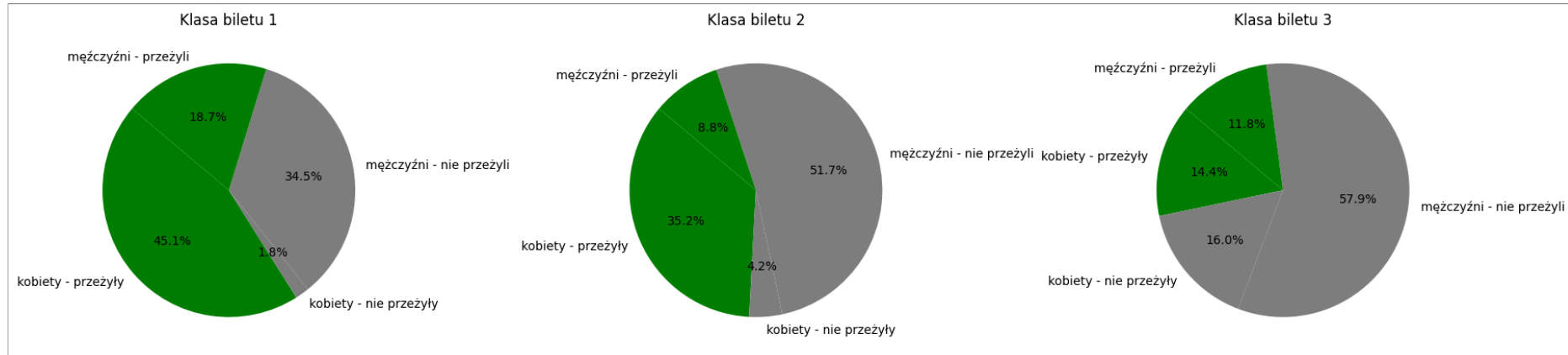
Kobiety z biletem klasy 2



Kobiety z biletem klasy 3



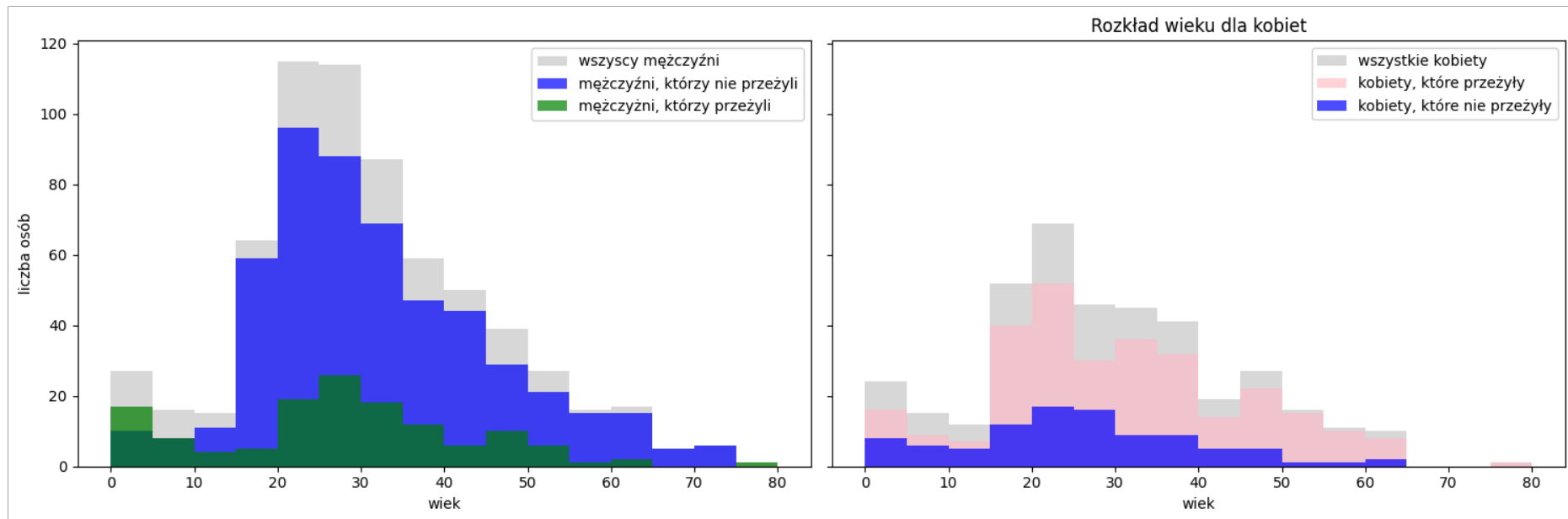
Odsetek osób z biletem w danej klasie, które przeżyły lub zginęły w zależności od płci



## Wnioski

- Na podstawie powyższej analizy, wynika że **w grupie kobiet** wraz ze **wzrostem klasy biletu** znacząco wzrosła **szansa na przeżycie** (kobiety z biletem w klasie 1 w 96,2% przypadkach przeżyły, w klasie 2 w 89,3%, a w klasie 3 już tylko w 47,4% kobiet przeżyło).
- Także w grupie badanych **mężczyzn**, ci z **biletem w klasie 1** mieli **największą szansę na przeżycie** (35,1% mężczyzn z biletem w klasie 1), natomiast **mężczyźni** z biletami w **klasie 2 i 3** mieli **zbliżone szanse** na przeżycie (przeżyło 14,6% mężczyzn spośród tych z biletem w klasie 2, oraz 16,9% mężczyzn spośród tych z biletem w klasie 3).
- Ostatni wykres pt. "Odsetek osób z biletem w danej klasie, które przeżyły lub zginęły w zależności od płci", ukazuje że **dysproporcja kobiet**, które **przeżyły**, względem **mężczyzn**, którzy **przeżyli**, **zmniejsza się wraz ze spadkiem klasy biletu** (por. klasę biletu 1 z 3).

## 4.5. Rozkład wieku dla mężczyzn i kobiet



## Wnioski

- **Średni wiek** pasażerów to w przybliżeniu **30 lat**.
- **Najwięcej** podróżujących **mężczyzn** było w wieku **20-35 lat**.
- **Najwięcej** podróżujących **kobiet** było w wieku **15-40 lat**.
- Z rozkładów wieku dla kobiet i mężczyzn wynika, że **grupami uprzywilejowanymi** do uzyskania pomocy były **osoby najstarsze** (75-80 lat) oraz **kobiety w wieku 50-60 lat** - są to grupy wiekowe, które mają najmniejszy odsetek śmiertelności w katastrofie.

## 5. Podsumowanie

- W analizie przyjęto **ostre kryterium selekcji danych**, tzn. wybrano do analizy tylko te rekordy, które miały pełne dane w kluczowych kolumnach ("przeżycie", "płeć", "wiek"). Ostatecznie statystyki zostały obliczone na podstawie **próby liczącej 1046 przypadków**.
- Uzyskane wyniki pokazały, że **największą szansę na przeżycie w katastrofie Titanica** miały osoby podróżujące w **1 klasie**.
- W badanej próbie **podróżnymi byli w przeważającej części mężczyźni** (62,9%), lecz zdecydowanie **większe szanse na przeżycie miały kobiety** (spośród uratowanych pasażerów, **kobiety stanowiły 2/3**).