

1. Opis Danych i Struktura

Zbiór danych składa się z **6690** obserwacji (zwierząt) i **15** kolumn (cech).

1.1 Typy Danych

Większość kolumn z datami (`Outcome Date` , `Date of Birth` , `Intake Date`) jest załadowana jako **typ object (ciąg tekstowy)**. `AnimalID` , `isAdopted` , `Days in shelter` jest numeryczna.

1.2 Statystyki Numeryczne

Kolumna `Animal ID` jest zmienną identyfikacyjną i nie ma wartości predykcyjnej.

Kolumna `is_adopted` (nasza binarna zmienna docelowa) wykazuje następujący rozkład:

- Średnia (`mean`): **0.567**
- Wniosek: Około **56.7%** zwierząt w zbiorze danych zostało zaadaptowanych. Jest to rozkład umiarkowanie zbalansowany, co jest dobrą podstawą do budowy modelu klasyfikacyjnego.

2. Analiza Jakości Danych (Braki Danych)

Poniższa tabela podsumowuje kolumny, w których wykryto braki danych (NaN):

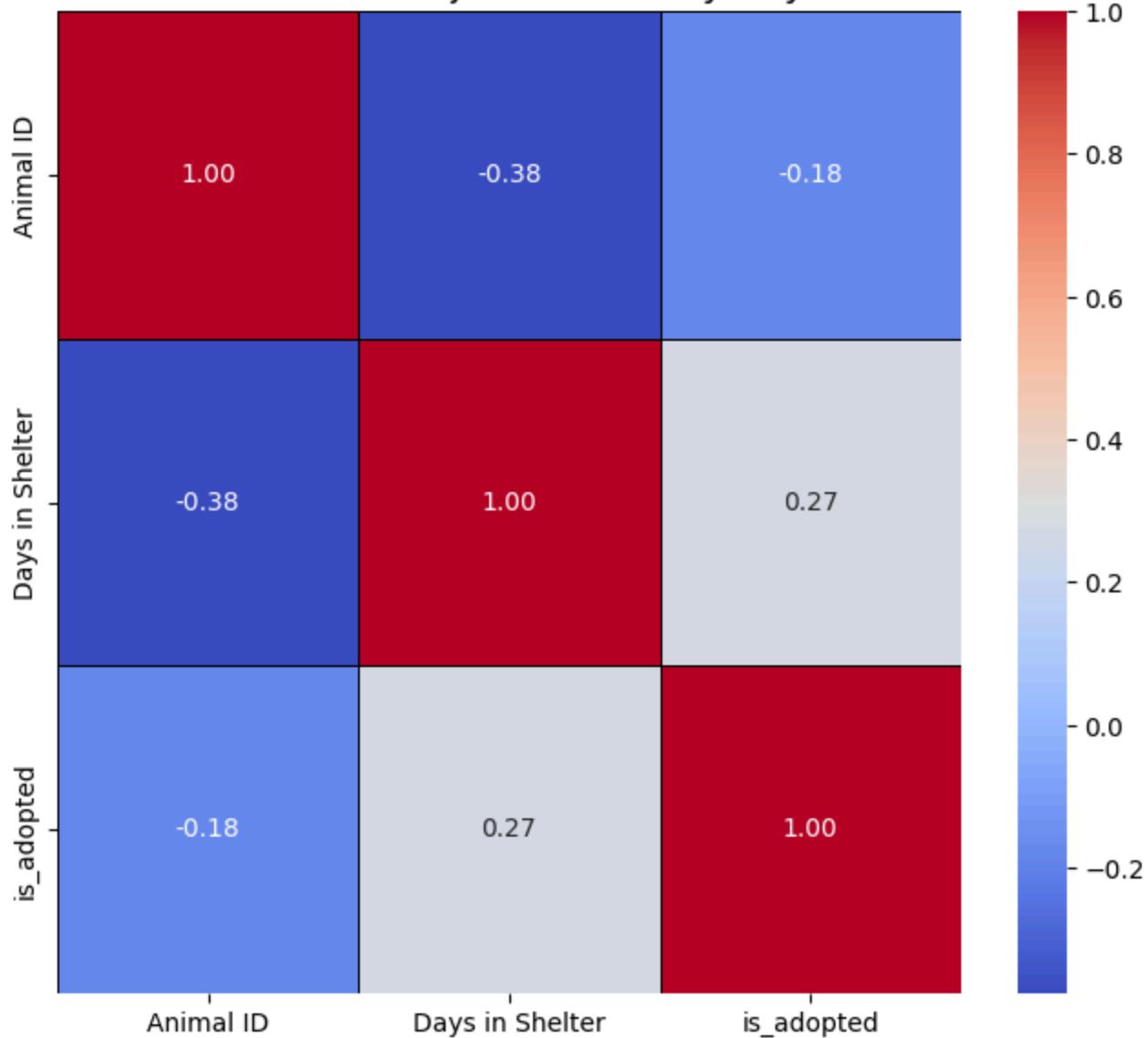
Kolumna	Brakujące Wartości	Procent Braków	Wnioskowana Akcja
<code>Euthanasia Reason</code>	6286	93.96%	Usunięcie kolumny (zbyt wysoki odsetek braków, brak wartości predykcyjnej).
<code>Secondary Color</code>	3149	47.07%	Imputacja (uzupełnienie) wartością: " Brak " lub " None ".
<code>Name</code>	2791	41.72%	Imputacja wartością: " No Name ". Następnie stworzenie cechy binarnej: <code>has_name</code> (czy zwierzę ma imię).
<code>Date of Birth</code>	385	5.75%	Imputacja medianą (lub średnią) wieku po konwersji daty.
<code>Spayed/Neutered</code>	11	0.16%	Usunięcie wierszy lub imputacja wartością dominującą (mode).

Kolumna	Brakujące Wartości	Procent Braków	Wnioskowana Akcja
Sex	7	0.10%	Usunięcie wierszy (marginalna ilość).

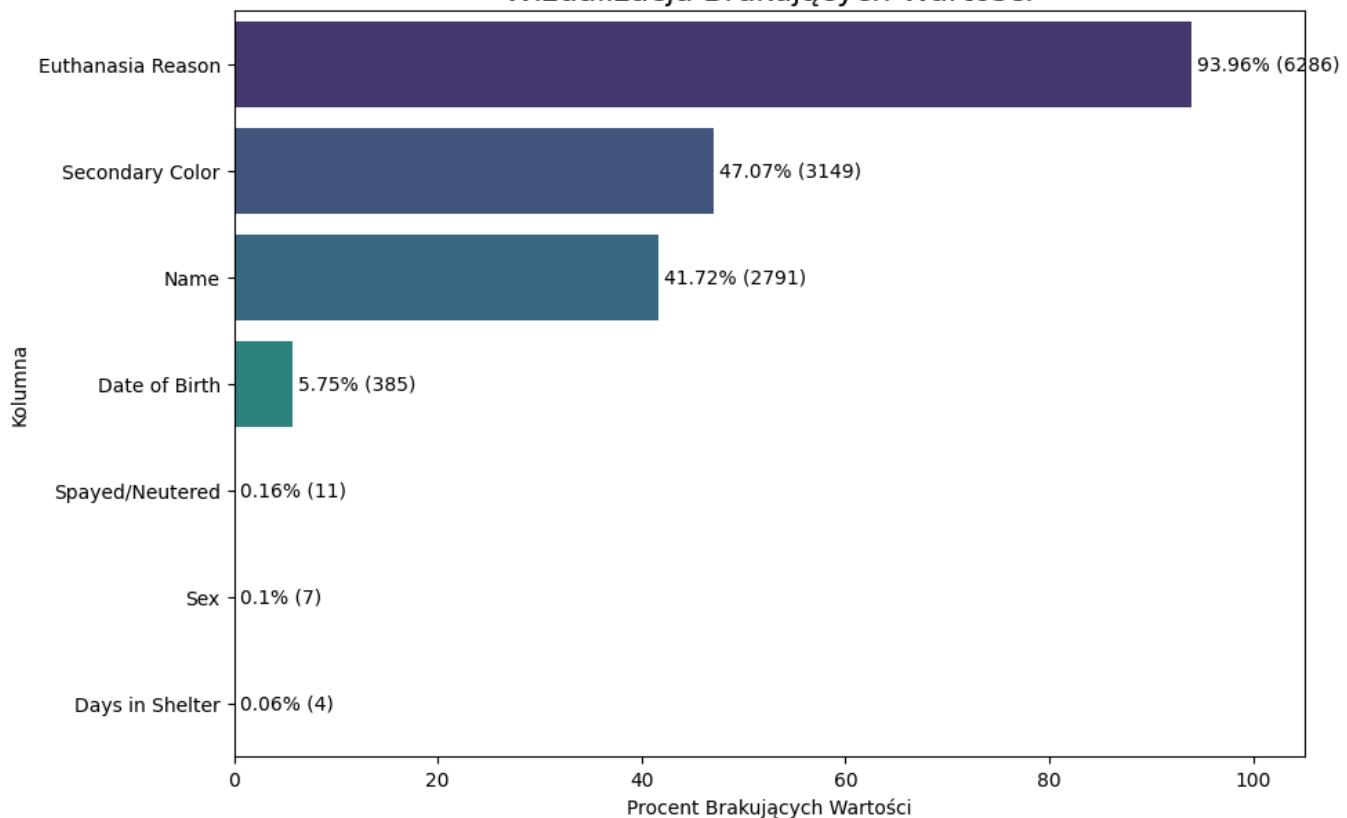
Kluczowe Obserwacje:

- Dane do usunięcia:** Kolumna `Euthanasia Reason` nie nadaje się do modelowania.
- Konwersja Dat:** Konwersja kolumn: `Outcome Date`, `Date of Birth`, `Intake Date` na typ `datetime`.
- Obliczenie Wieku zwierzęcia w dniach/latach na moment przyjęcia (`Intake Date - Date of Birth`).
- Kodowanie Zmiennych:** Zastosowanie kodowania (np. One-Hot Encoding lub Label Encoding) dla kluczowych zmiennych kategorycznych, takich jak: `Primary Breed`, `Type`, `Sex`, `Spayed/Neutered`, oraz `Primary Color`.

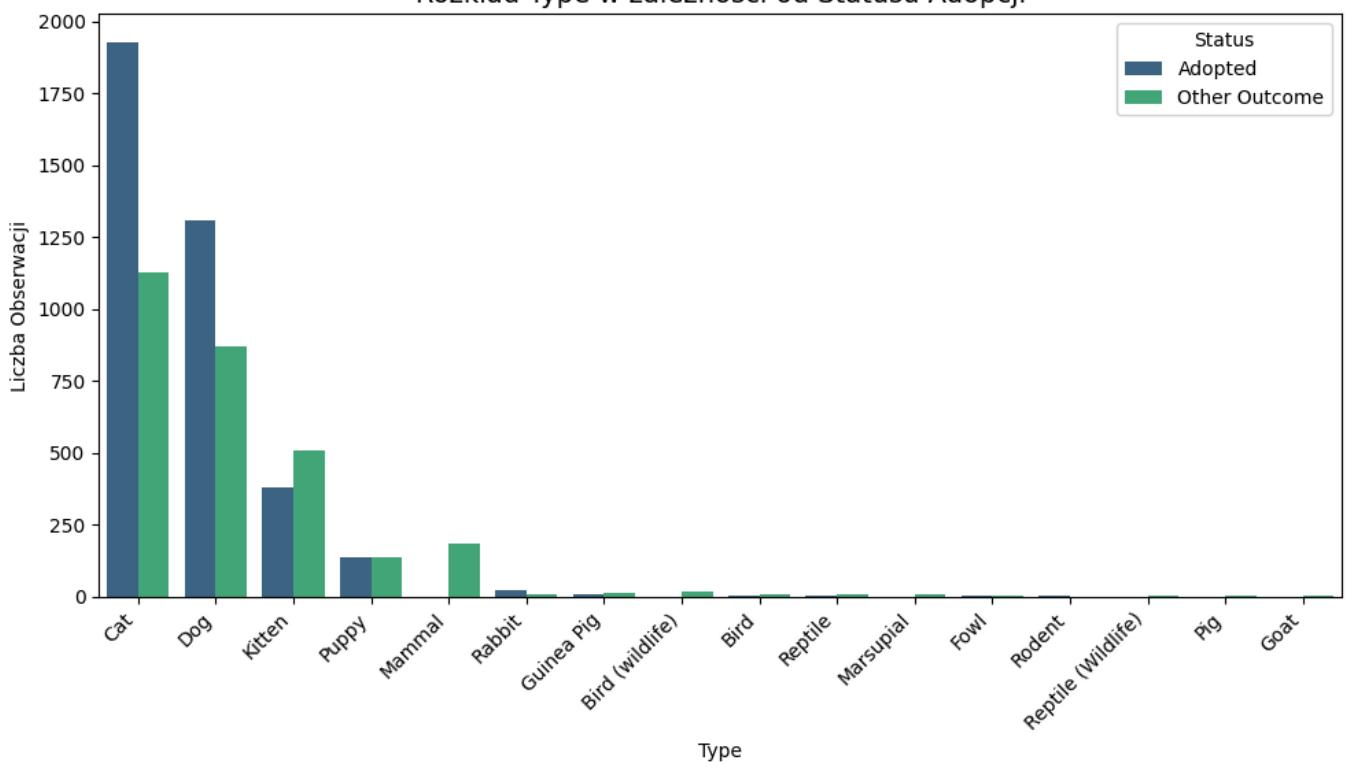
Macierz Korelacji Cech Numerycznych

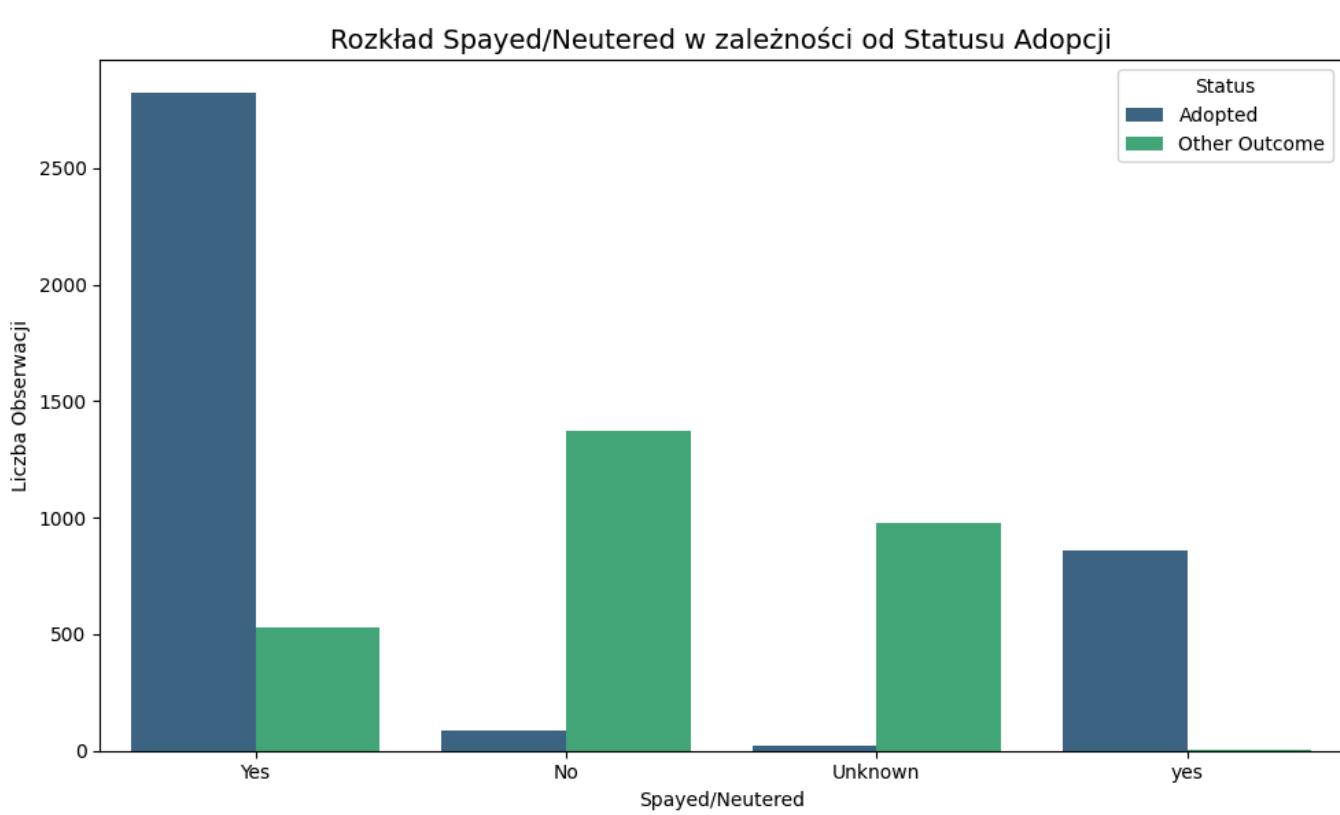
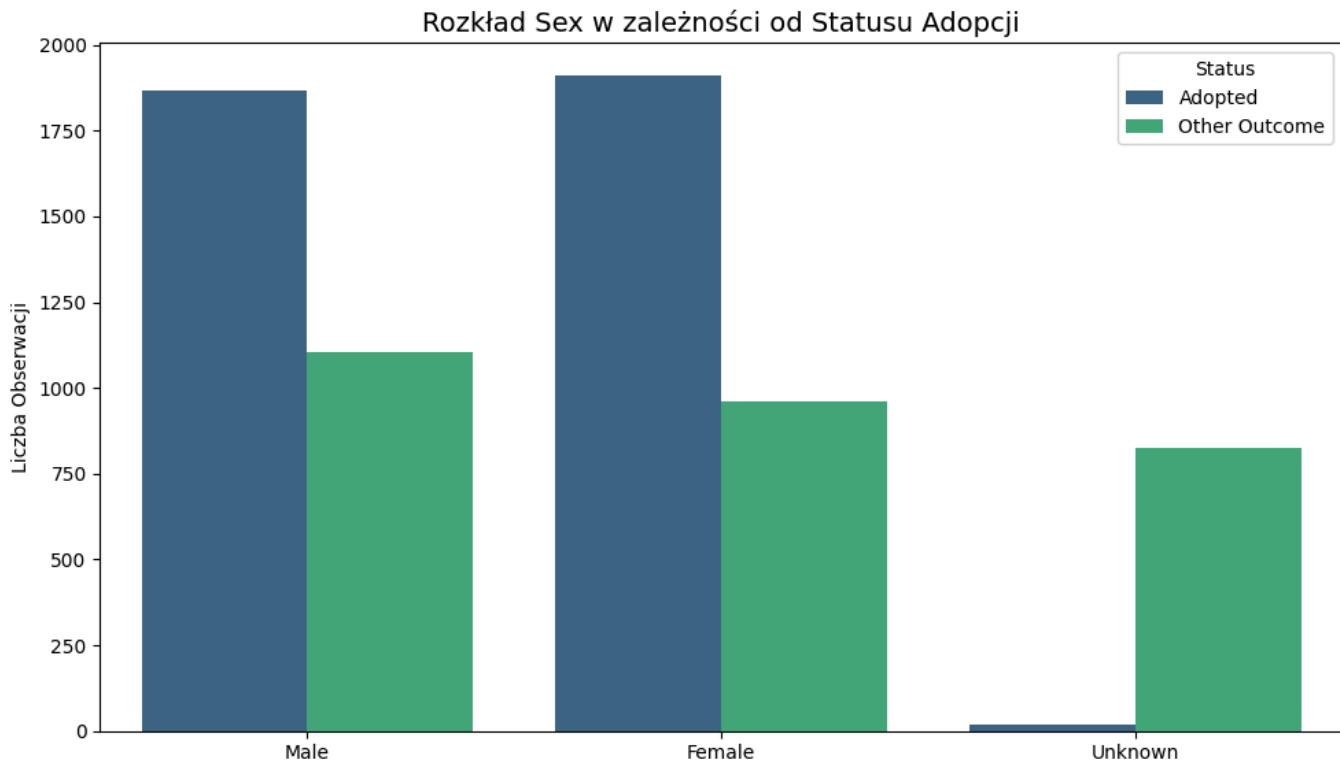


Wizualizacja Brakujących Wartości



Rozkład Type w zależności od Statusu Adopcji





Rozkład Primary Breed w zależności od Statusu Adopcji

