

Zadanie 1

Załaduj ponownie zbiór danych (DSP_6.csv), ale nie wykonuj jeszcze operacji na brakujących wartościach. Poszukaj innego sposobu wyświetlenia informacji o brakujących wartościach - wyświetl liczbę brakujących danych.

Zadanie 2

Dodaj kilka kolejnych wykresów w celu dalszej eksploracji danych. Stwórz wykres, który będzie pokazywać informację dotyczącą liczby członków rodziny na pokładzie. Stwórz wykres, który będzie pokazywać informację dotyczącą opłaty, która została uiszczona przez pasażerów. Z uwagi na to, że dane te zawierają przypadki odstające zmień jego wielkość za pomocą parametru figsize.

Zadanie 3

Spróbuj wytrenować nowy model regresji logistycznej dla tych samych danych. Tym razem zrób wcześniej błąd i nie dokonuj transformacji zmiennych (po prostu załaduj data frame, usuń przypadki odstające i spróbuj je zastosować do modelu). Jaki błąd otrzymujesz? Załaduj ponownie model, tym razem poprawnie, ale wytrenuj model dwukrotnie modyfikując wybrane hiperparametry (test_size, random_state lub max_iter). Czy coś się zmienia w parametrach modelu?

Zadanie 4

Załaduj nowy zbiór danych DSP_2.csv. Wytrenuj nowy model regresji logistycznej w celu określenia czy u danej osoby wystąpi choroba serca (zmienna HeartDisease). Do wytrenowania modelu użyj wszystkich innych zmiennych, które występują w tym arkuszu. Czy model cechuje się dobrymi parametrami? Jeżeli tak, to dlaczego? Czy może zostać wykorzystany w praktyce?