

Analiza zbioru danych mieszkania

Wielowymiarowa analiza danych

Przemysław Peplinski, Wiktor Galewski, Mikołaj Zalewski

2026-01-05

Spis treści

1	Wstęp	2
1.1	Autorzy	2
1.2	Motywacja i cele	2
1.3	Opis projektu	2
2	Opis danych	2
3	Analiza eksploracyjna	3
3.1	Statystyki opisowe	3
3.2	Analiza rozkładów	3
3.3	Analiza korelacji	3
3.4	Analiza braków danych	3
3.5	Analiza obserwacji odstających	4
4	Czyszczenie danych	4
4.1	Obsługa braków wartości	4
4.2	Obsługa obserwacji odstających	4
5	Obróbka zmiennych	5
5.1	Kodowanie zmiennych kategorycznych	5
5.2	Normalizacja zmiennych numerycznych	5

6	2 metody	5
6.1	Metoda 1	5
6.2	Metoda 2	5
7	Porównanie metod	5
8	Wnioski	5

1 Wstęp

1.1 Autorzy

Autorami są... Wkład w projekt prezentował się następująco...

1.2 Motywacja i cele

Coś tam...

1.3 Opis projektu

Coś tam...

2 Opis danych

Dane pochodzą.../Dane zostały zebrane...

```
## Rows: 106,169
## Columns: 19
## $ Link          <chr> "https://domy.pl/mieszkanie/gdansk-siedlce-malczews~
## $ `Data dodania` <dtm> 2021-02-02 10:24:00, 2021-02-02 11:08:14, 2021-02-~
## $ `Data modyfikacji` <dtm> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ Tytuł         <chr> "Trzypokojowe mieszkanie na sprzedaż:", "Gdańsk, Si~
## $ Opis          <chr> "Młyny Gdańskie Gdańsk, ul. Malczewskiego Młyny Gda~
## $ Cena          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ Powierzchnia   <dbl> 60.02, 89.63, 60.49, 64.89, 104.66, 72.04, 75.21, 6~
## $ `Cena za metr` <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ `Liczba pokoi` <dbl> 3, 4, 3, 3, 4, 4, 4, 3, 3, 3, 4, 3, 1, 2, 3, 3, 3, ~
## $ Piętro         <dbl> 4, 1, 0, 0, 1, 3, 1, 2, 3, 3, 2, 1, 3, 0, 1, 3, 4, ~
## $ `Liczba pięter` <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
```

```
## $ `Rok budowy`      <dbl> 2023, 2023, 2023, 2023, 2023, 2023, 2023, 2023, 202~
## $ Ocena              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ Dzielnica          <chr> "Siedlce", "Siedlce", "Siedlce", "Siedlce", "Siedlc~
## $ `Rodzaj budynku`  <chr> "Apartamentowiec", "Pozostałe", "Pozostałe", "Pozos~
## $ Rynek              <chr> "Pierwotny", "Pierwotny", "Pierwotny", "Pierwotny",~
## $ Ogłoszeniodawca    <chr> "Agencja", "Agencja", "Agencja", "Agencja", "Agencj~
## $ Portal             <chr> "Domy", "Morizon", "Morizon", "Morizon", "Morizon",~
## $ `Numer telefonu`  <chr> NA, "585055401", "585055401", "585055401", "5850554~
```

Coś tam coś tam...

3 Analiza eksploracyjna

3.1 Statystyki opisowe

Jak w nazwie

3.2 Analiza rozkładów

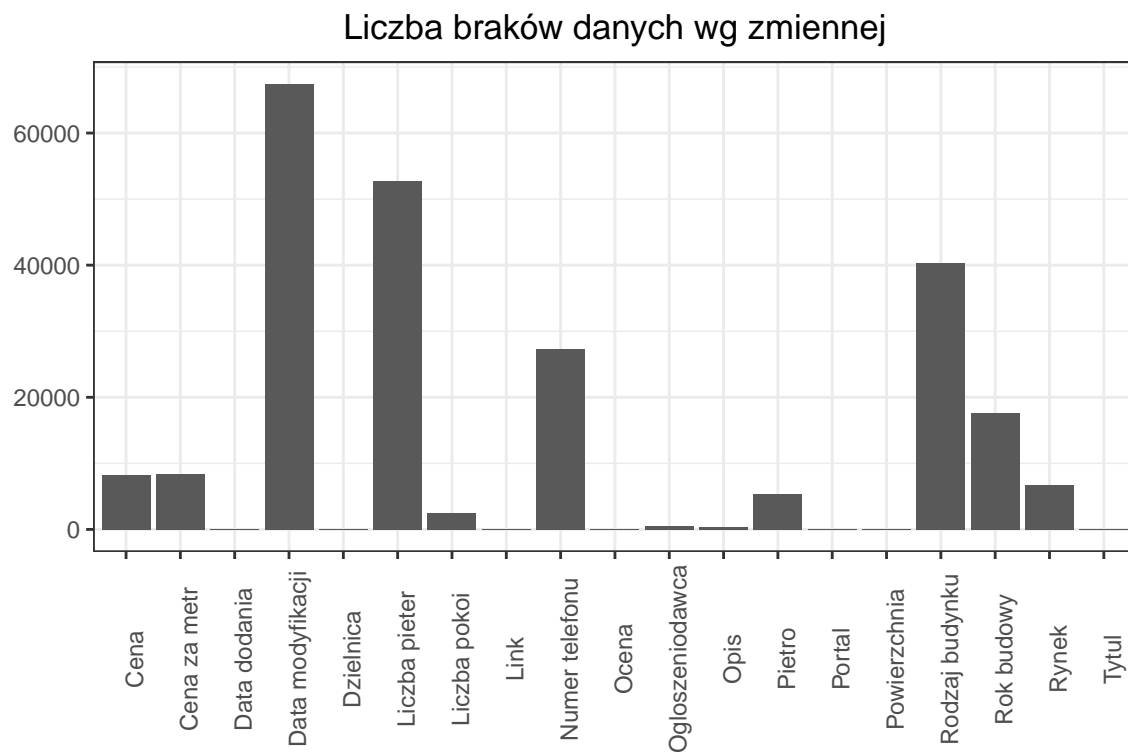
Histogramy zmiennych numerycznych, wykresy kolumnowe zmiennych kategorycznych

3.3 Analiza korelacji

Macierz korelacji

3.4 Analiza braków danych

Krótkie podsumowanie braków danych (wykres słupkowy?)



3.5 Analiza obserwacji odstających

Zrobić boxploty

4 Czyszczenie danych

4.1 Obsługa braków wartości

Nie powinno być

4.2 Obsługa obserwacji odstających

Walidacja, poza tym zostawiamy

5 Obróbka zmiennych

5.1 Kodowanie zmiennych kategoriycznych

5.2 Normalizacja zmiennych numerycznych

6 2 metody

6.1 Metoda 1

6.2 Metoda 2

7 Porównanie metod

Porównanie i wizualizacje

8 Wnioski