

# WB-XAI-1

## Grupa *California*

Sebastian Deręgowski  
Maciej Gryszkiewicz  
Paweł Morgen

17 marca 2021r.

## 1 Opis

Celem naszego badania jest wyznaczenie mediany wartości domu w obrębie kwartału (bloku). Do określenia tej mediany zastosujemy regresyjny model predykcyjny.

Model ten opierać się będzie na zbiorze danych z serwisu Kaggle.com. Owa ramka danych opisuje nieruchomości w stanie Kalifornia. Dane pochodzą ze spisu ludności z roku 1990. składa się ona z 20640 obserwacji i 10 zmiennych objaśniających:

1. **longitude** - długość geograficzna
2. **latitude** - szerokość geograficzna
3. **housing\_median\_age** - mediana wieku domu w obrębie kwartału
4. **total\_rooms** - liczba pokoi w bloku
5. **total\_bedrooms** - liczba sypialni w bloku
6. **population** - populacja w bloku
7. **households** - liczba gospodarstw domowych w bloku
8. **median\_income** - mediana przychodów gospodarstwa domowego (mierzona w dziesiątkach tysięcy dolarów)
9. **median\_house\_value** - mediana wartości gospodarstwa domowego (mierzona w dolarach)
10. **ocean\_proximity** - odległość domu od morza/oceanu (zmienna opisowa przyjmująca pięć wartości)

## 2 Analiza danych

Przy pomocy `pandas_profiling` wygenerowaliśmy raport, który pomógł nam poznać lepiej dane. Poniżej prezentujemy kilka najciekawszych naszym zdaniem korelacji między poszczególnymi zmiennymi.

`total_bedrooms`

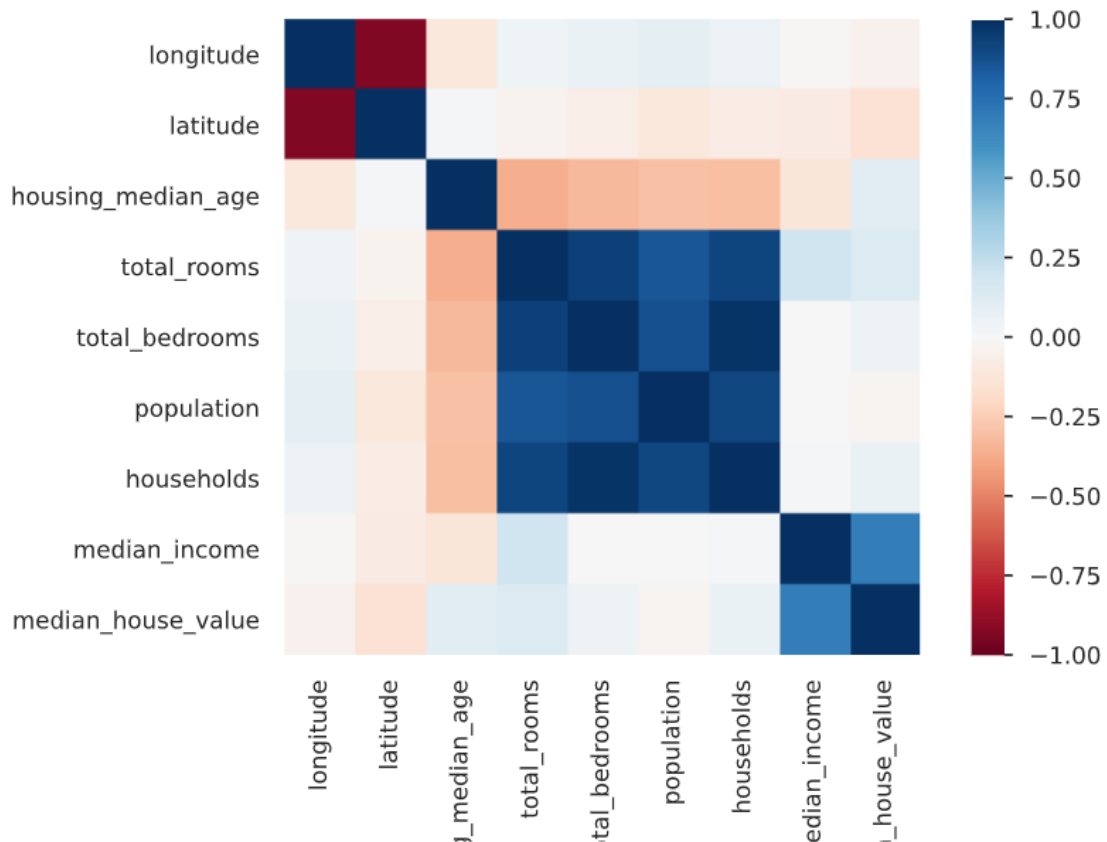
Real number ( $\mathbb{R}_{\geq 0}$ )

HIGH CORRELATION

MISSING

Distinct	1923
Distinct (%)	9.4%
Missing	207
Missing (%)	1.0%
Infinite	0
Infinite (%)	0.0%

Rysunek 1: Brakujące wartości

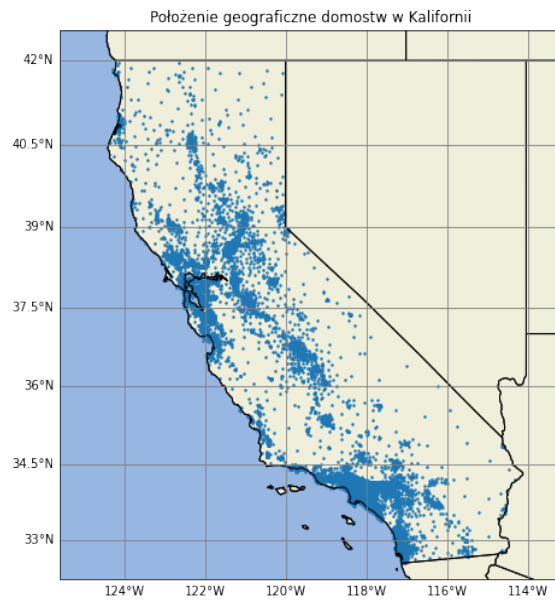


Rysunek 2: Korelacje między zmiennymi

Przyjrzyjmy się bliżej korelacji między długością a szerokością geograficzną. By ją wyjaśnić, wystarczy spojrzeć na mapę:

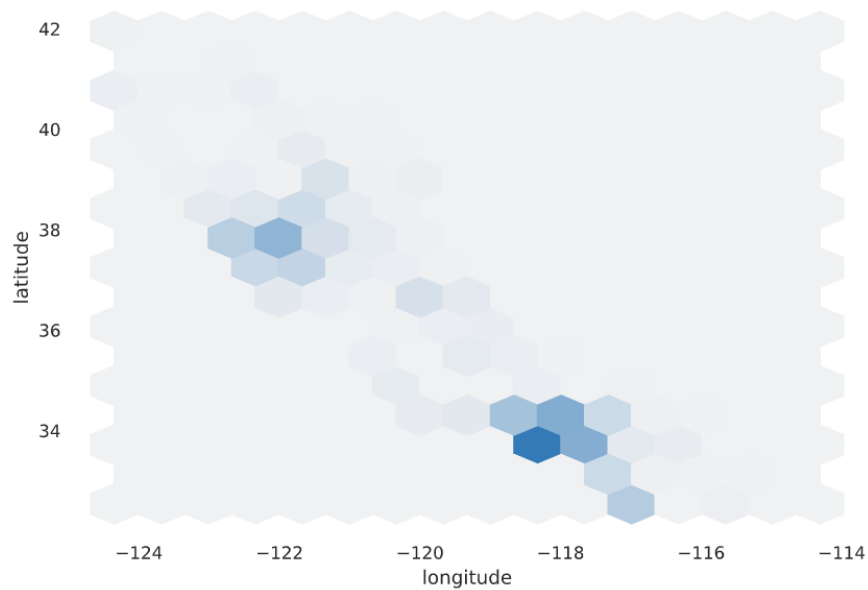


(a) Mapa

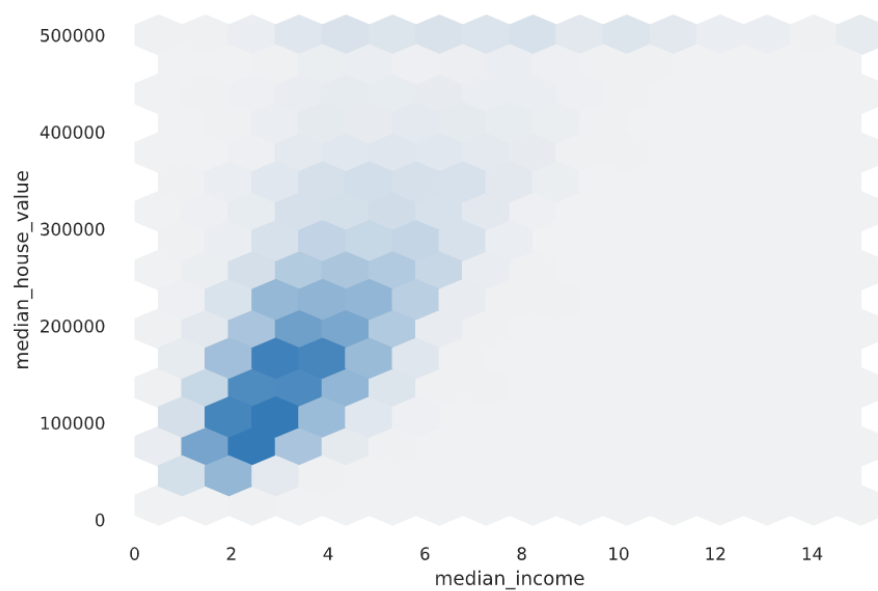


(b) Wykres

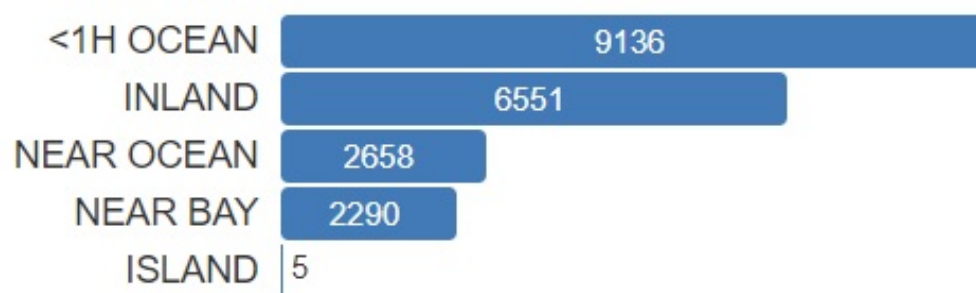
Rysunek 3: Mapa Kalifornii i wykres punktowy danych



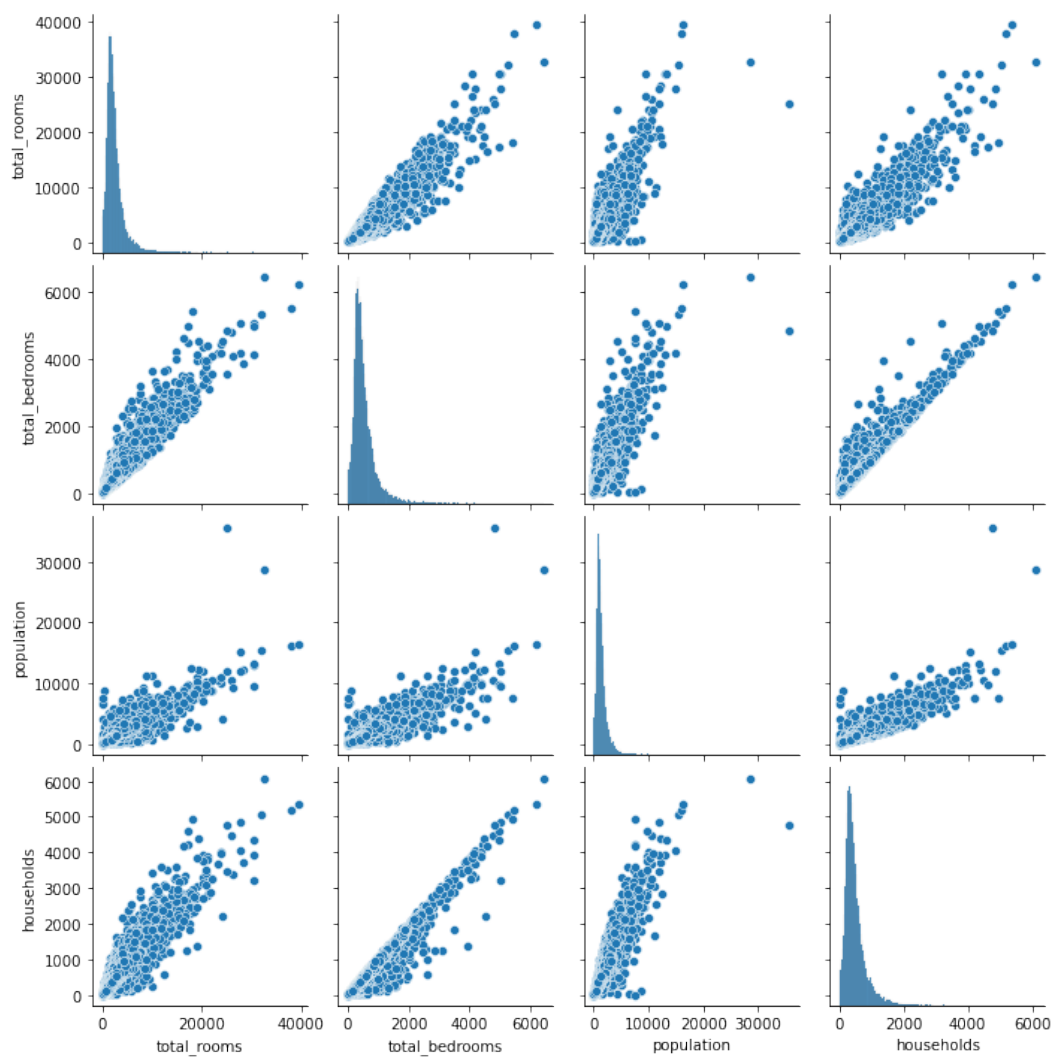
Rysunek 4: Rozmieszczenie mieszkań



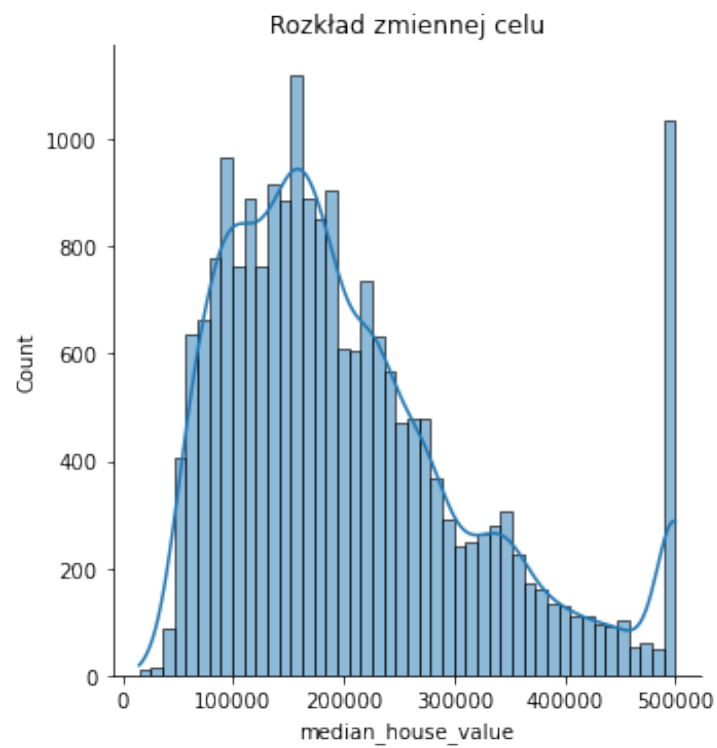
Rysunek 5: Wartość mieszkania a przychody mieszkańców



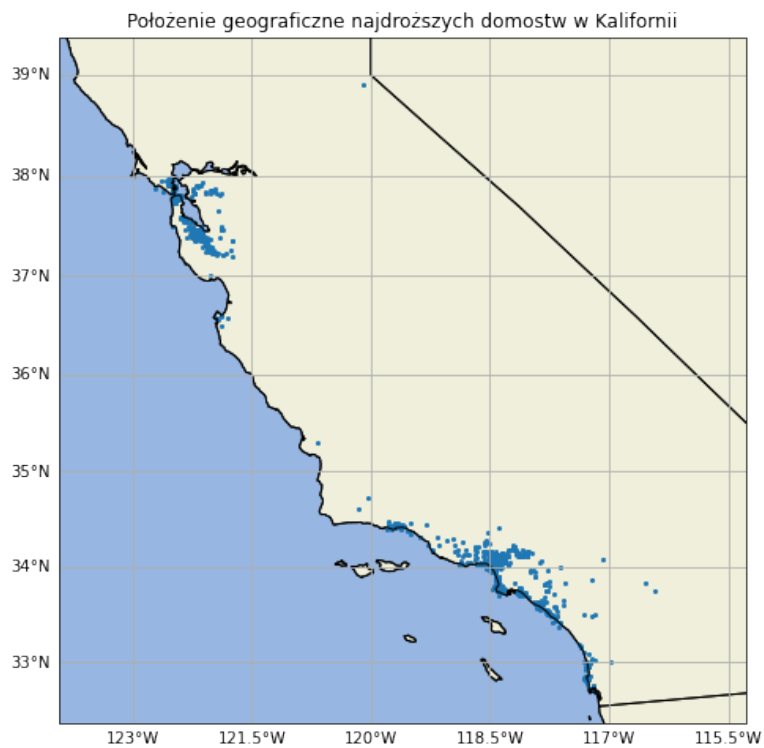
Rysunek 6: Odległość od oceanu



Rysunek 7: Wykresy skorelowanych zmiennych



Rysunek 8: Rozkład zmiennej celu



Rysunek 9: Lokalizacja najdroższych nieruchomości