

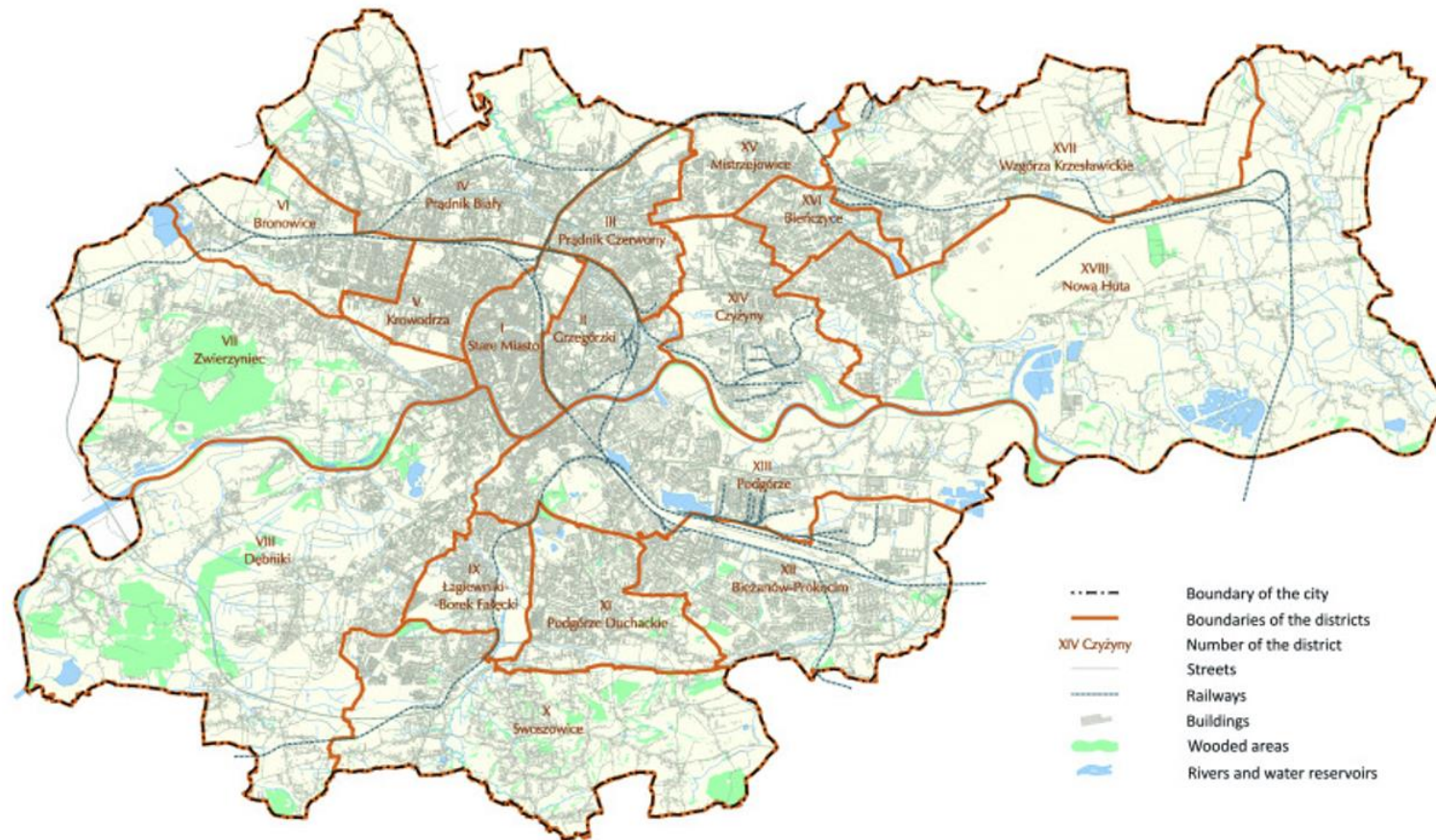
Wykrywanie obserwacji odstających w modelach regresyjnych

Przykład zastosowania w analizie rynku nieruchomości

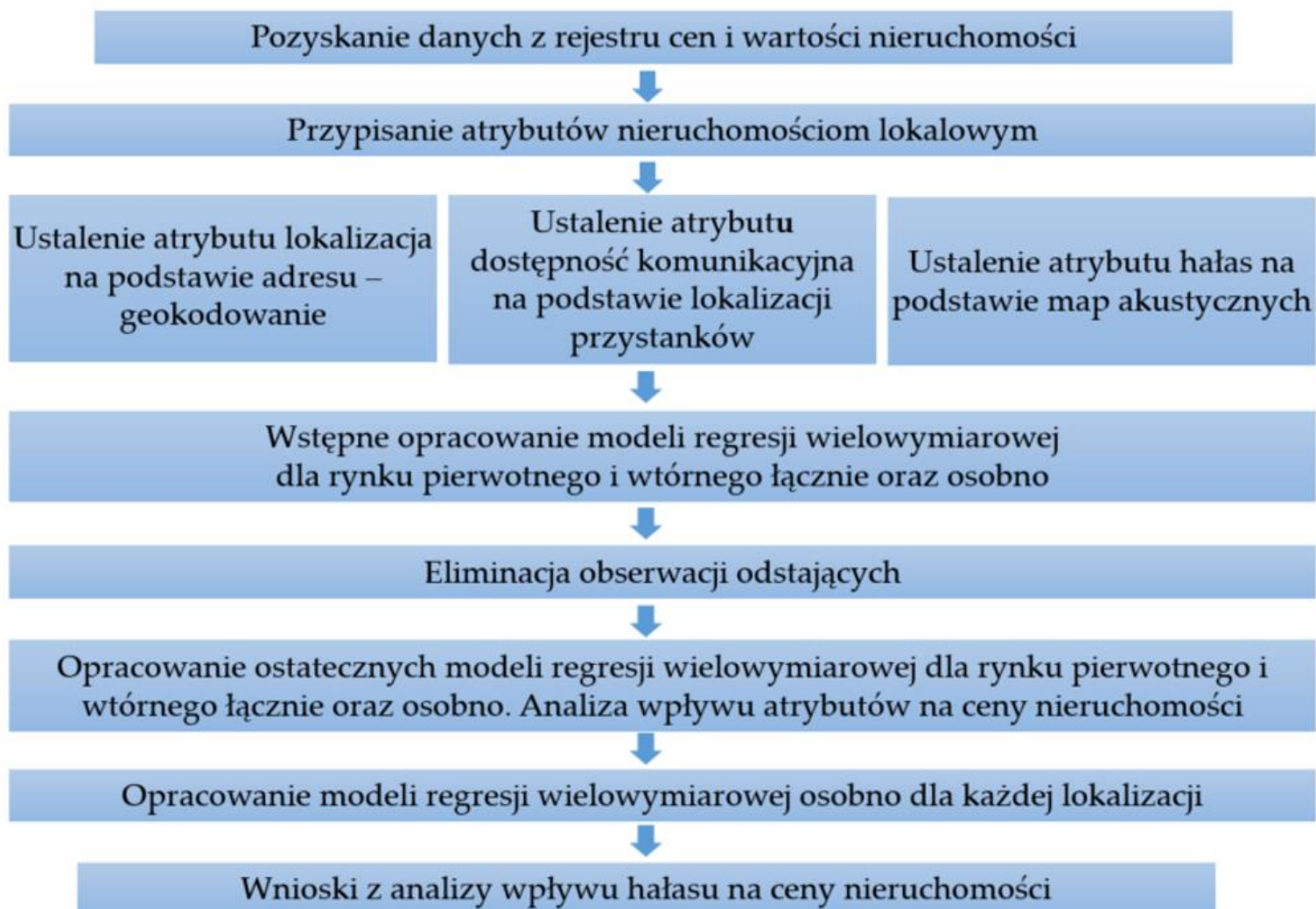
Edward Preweda

Przykładowy problem:

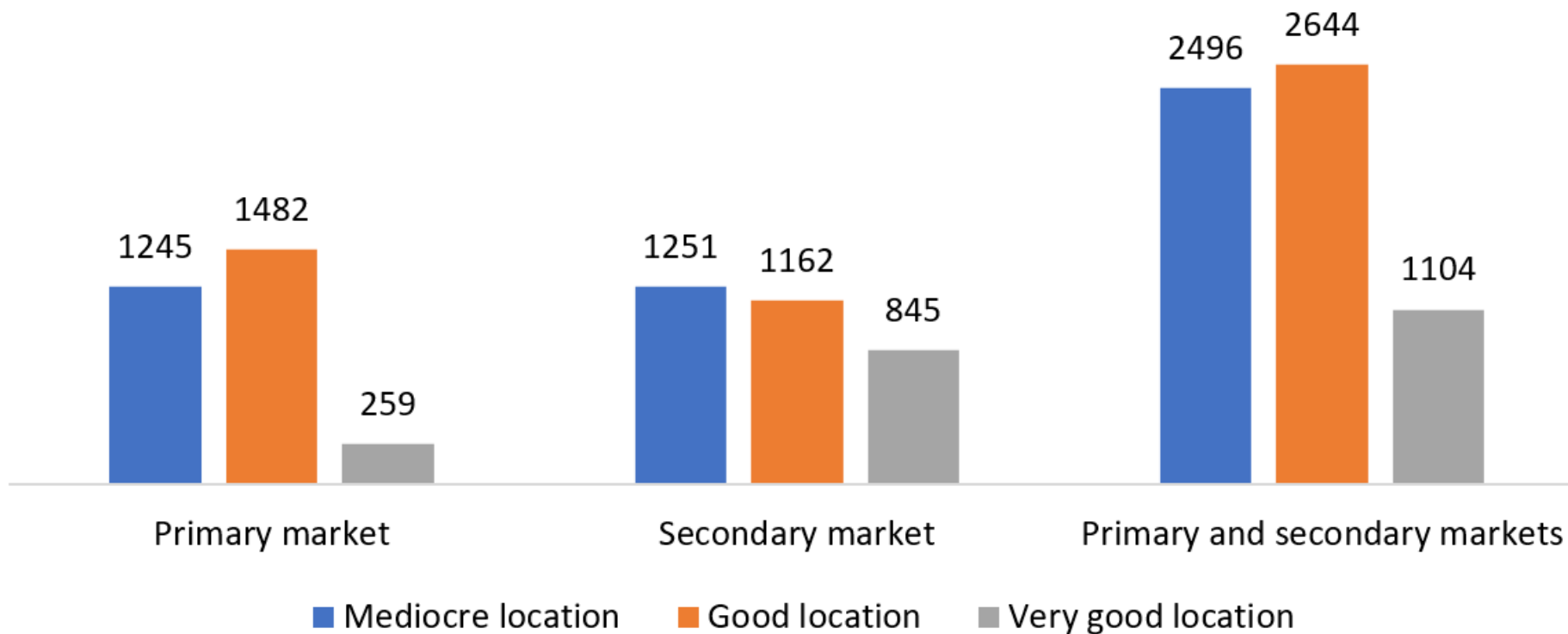
Zbadać wpływ atrybutów na ceny nieruchomości



ustalić wpływ hałasu na wartość nieruchomości lokalowych



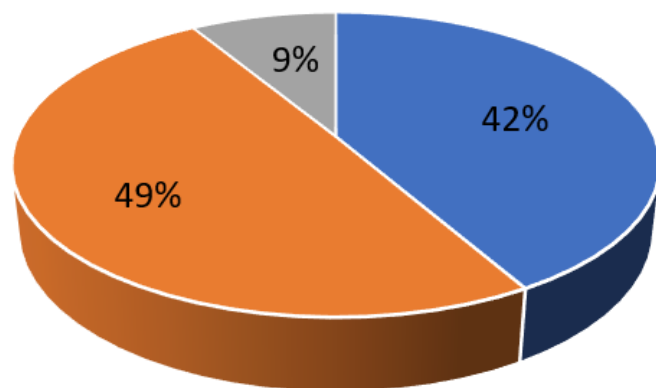
Pozyskanie danych



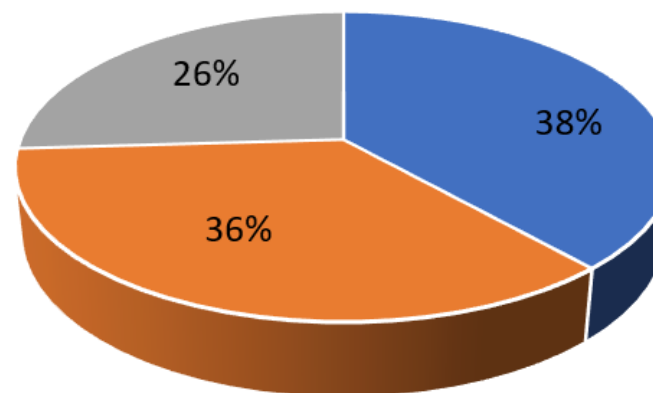
Number of analyzed properties on the primary and secondary market

Pozyskanie danych

Primary market



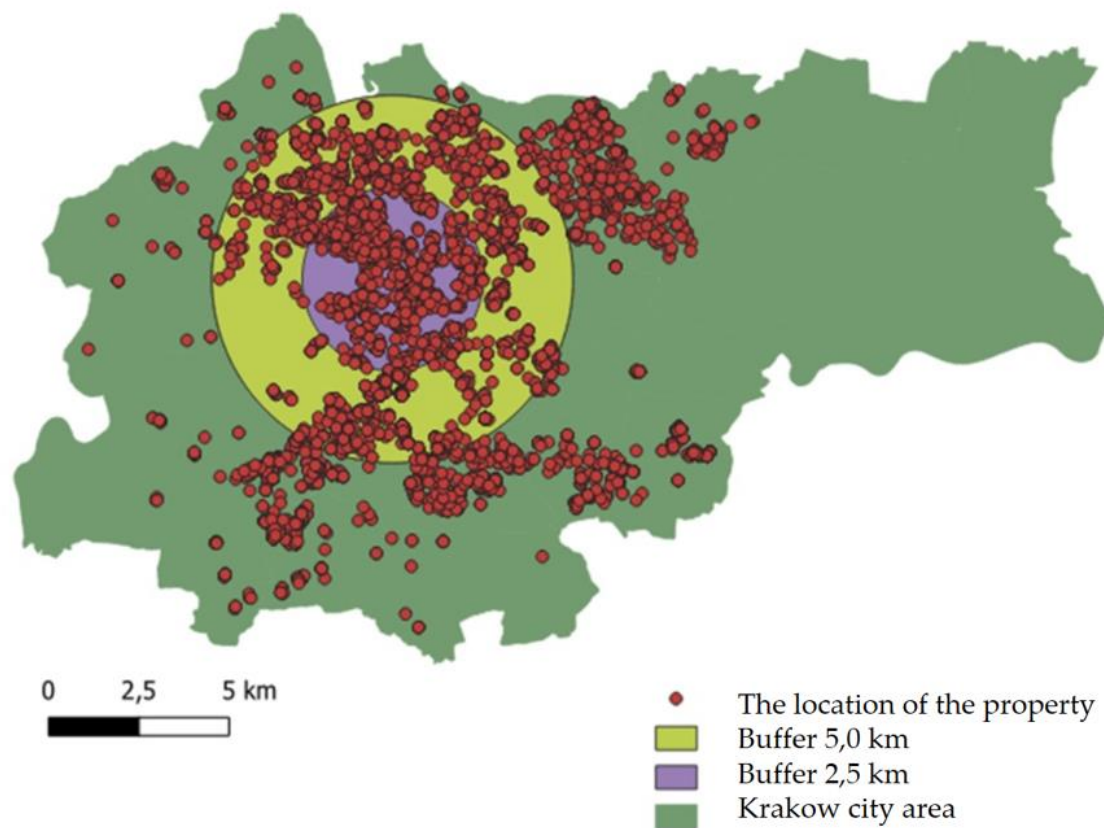
Secondary market



■ Mediocre location ■ Good location

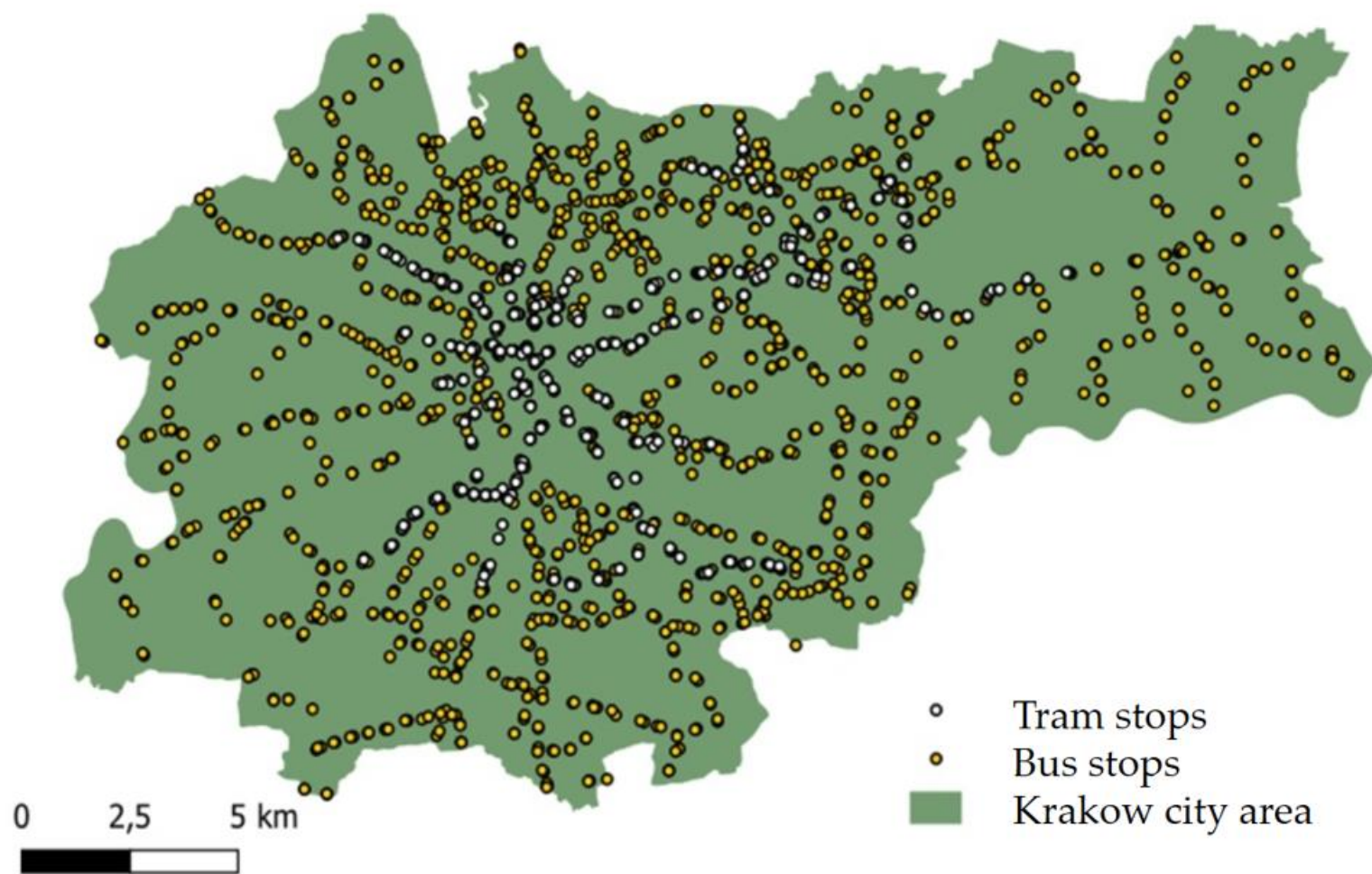
■ Very good location

Przypisanie atrybutów

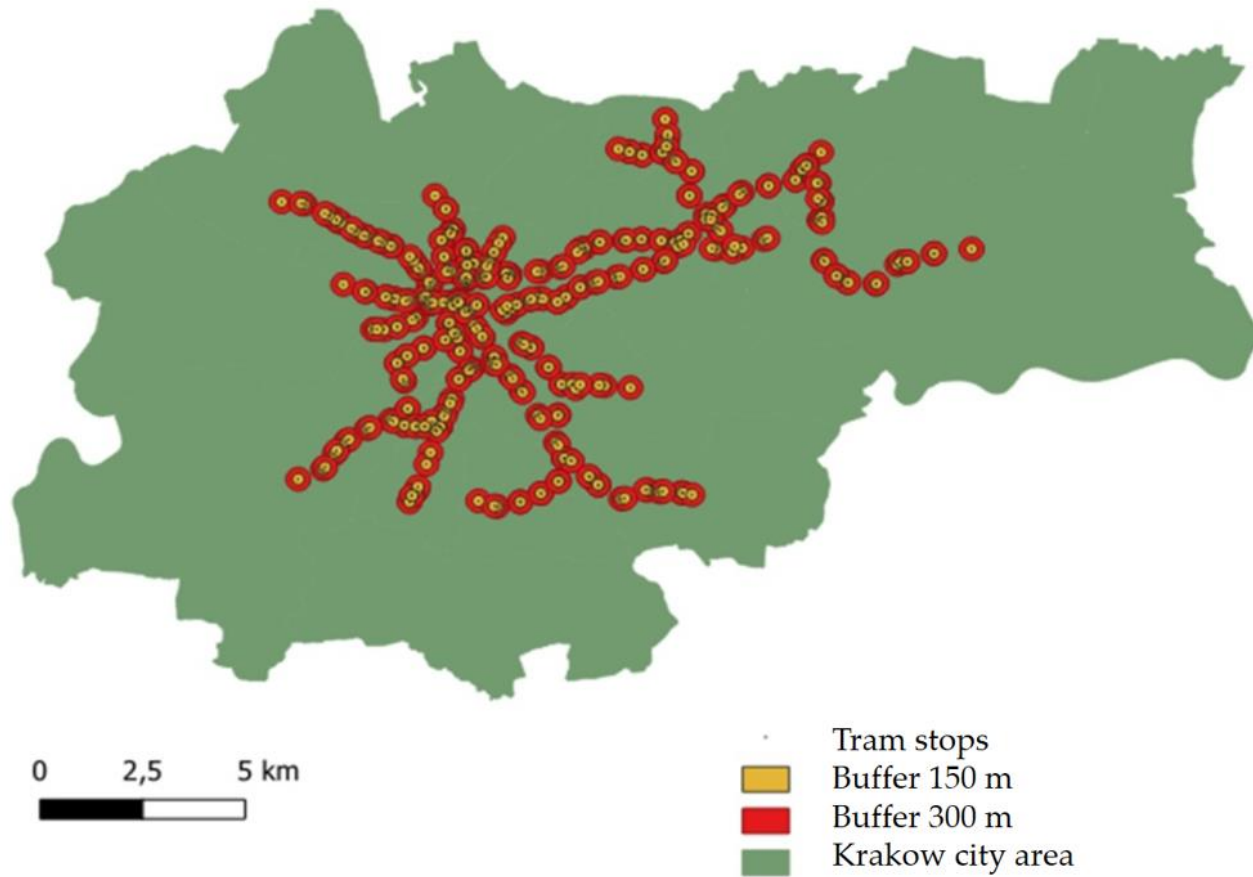


Np. na podstawie portalu MSIP Obserwatorium, serwisu Gis-Support

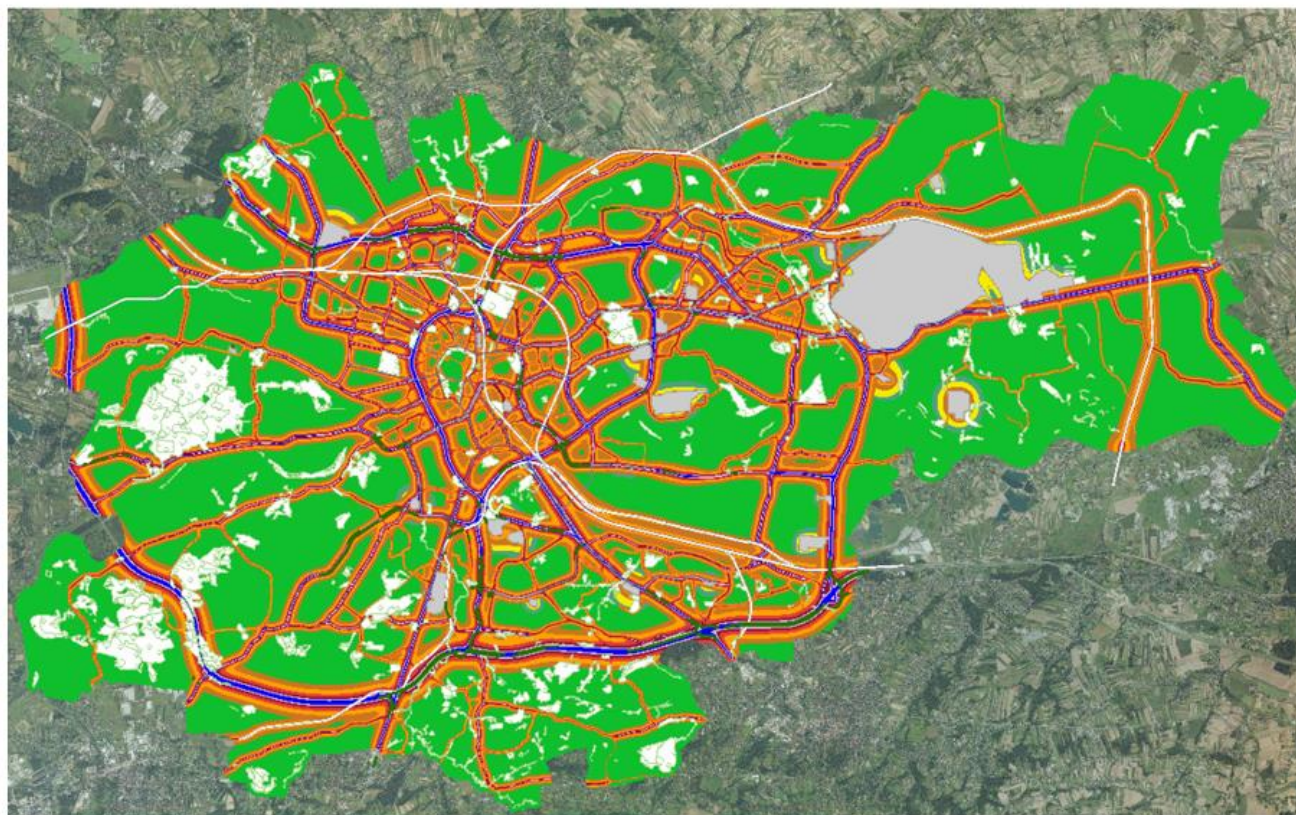
Przypisanie atrybutów



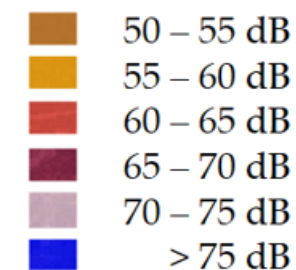
Przypisanie atrybutów



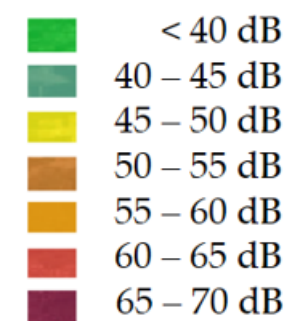
Przypisanie atrybutów



LDWN road noise



LDWN industrial noise

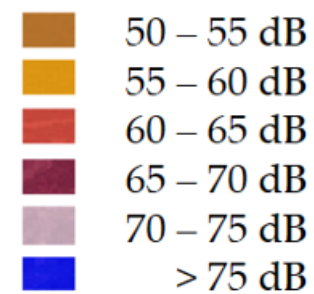


Mapy emisyjne hałasu drogowego i przemysłowego

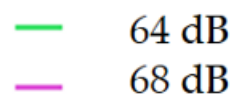
Przypisanie atrybutów



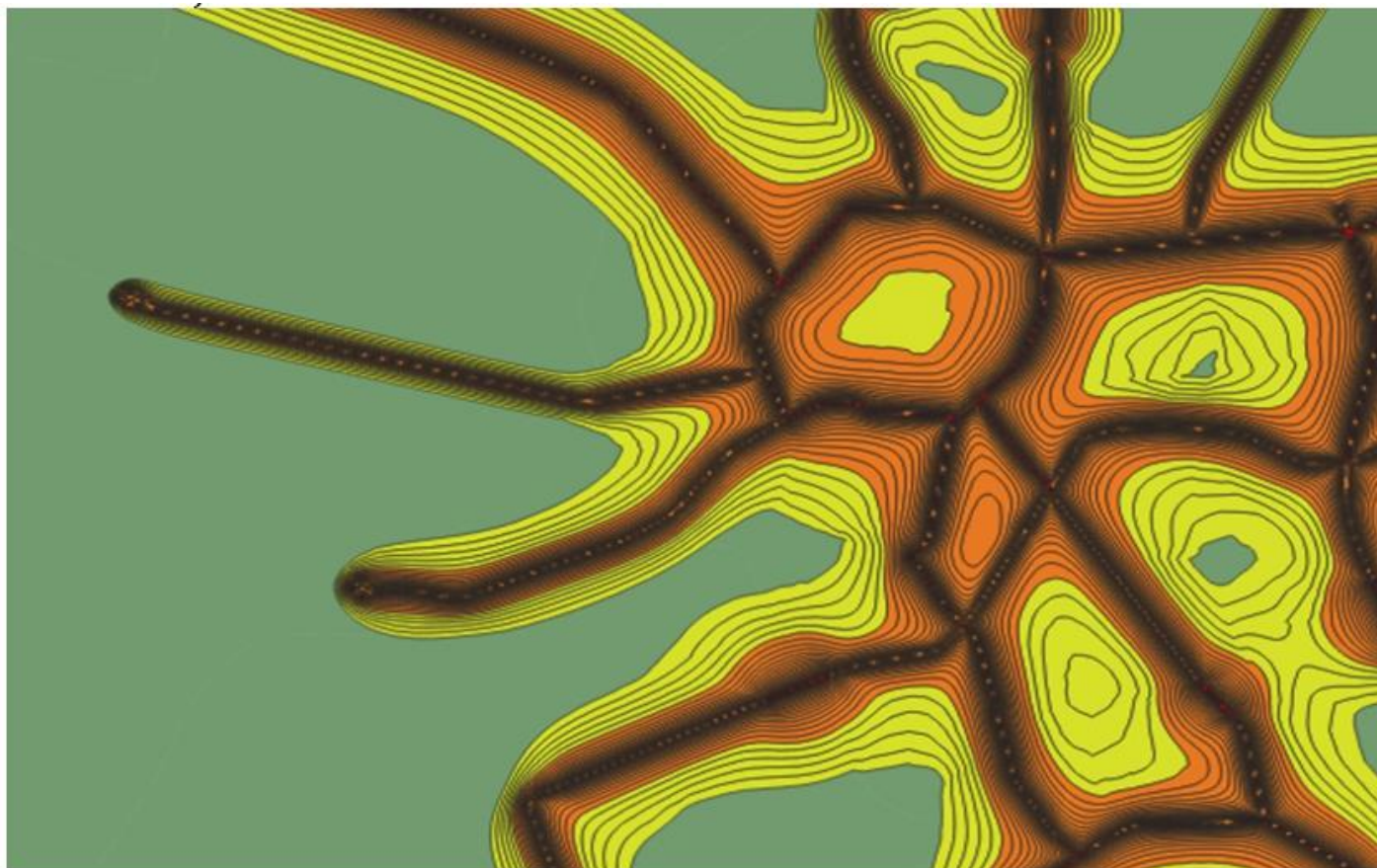
LDWN road noise



Isophones of permissible road noise



Przypisanie atrybutów



■ Krakow city area

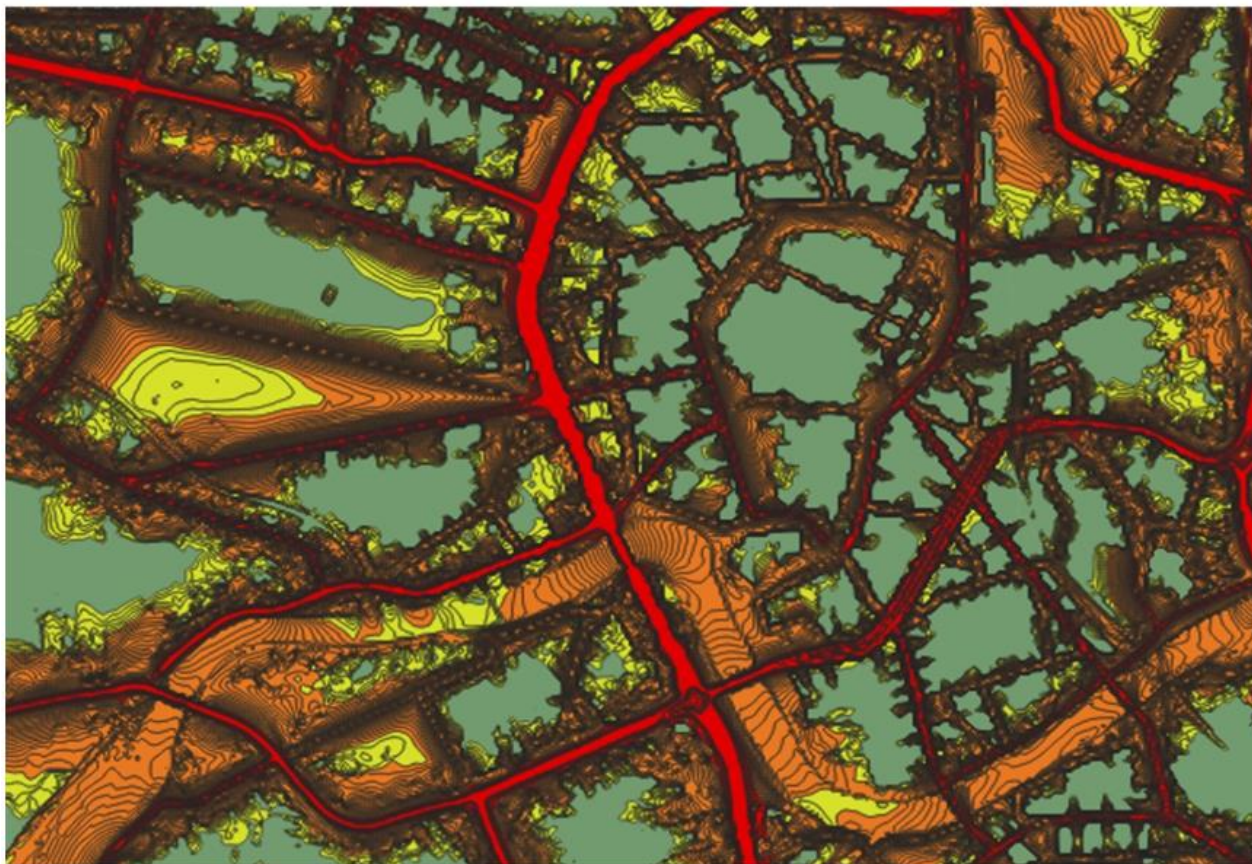
Tram nois

■ > 68 dB

■ < 50 dB

■ 50 - 68 dB

Przypisanie atrybutów



■ Krakow city area

Road, rail and industrial noise

■ > 68 db

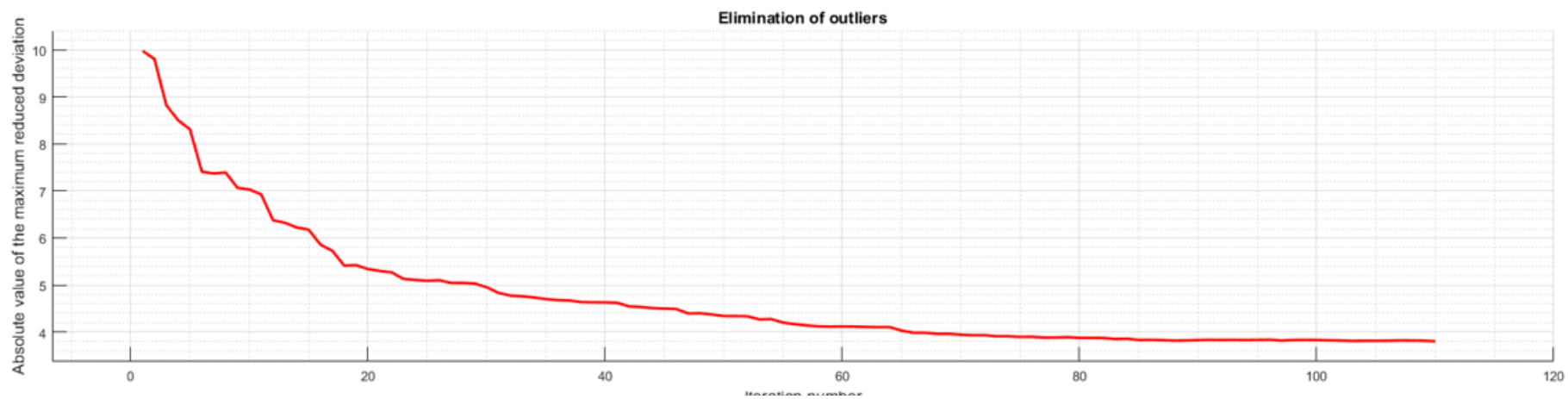
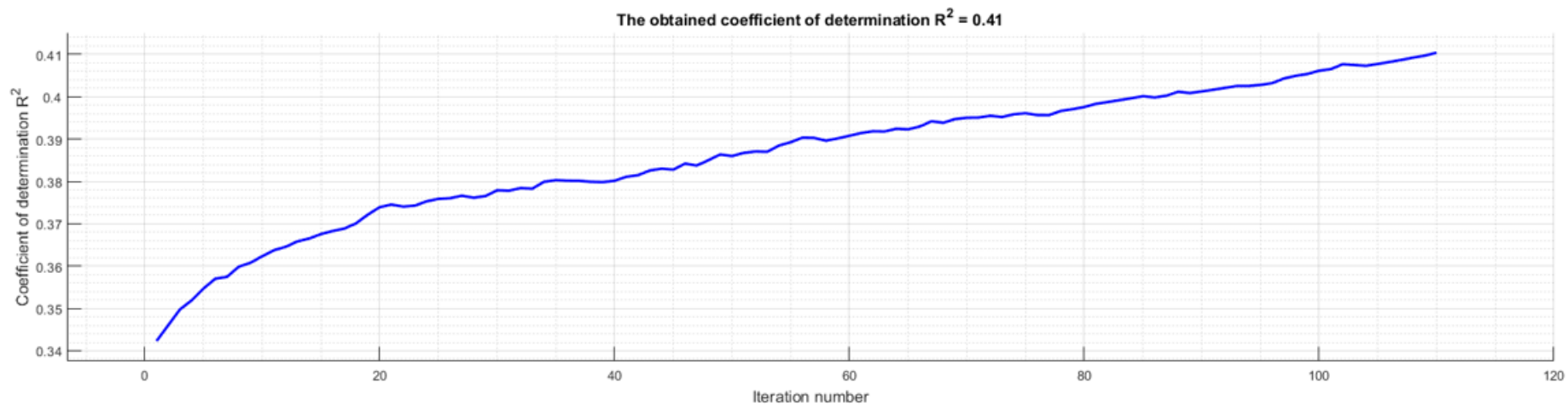
■ < 50 db

■ 50 - 68 dB

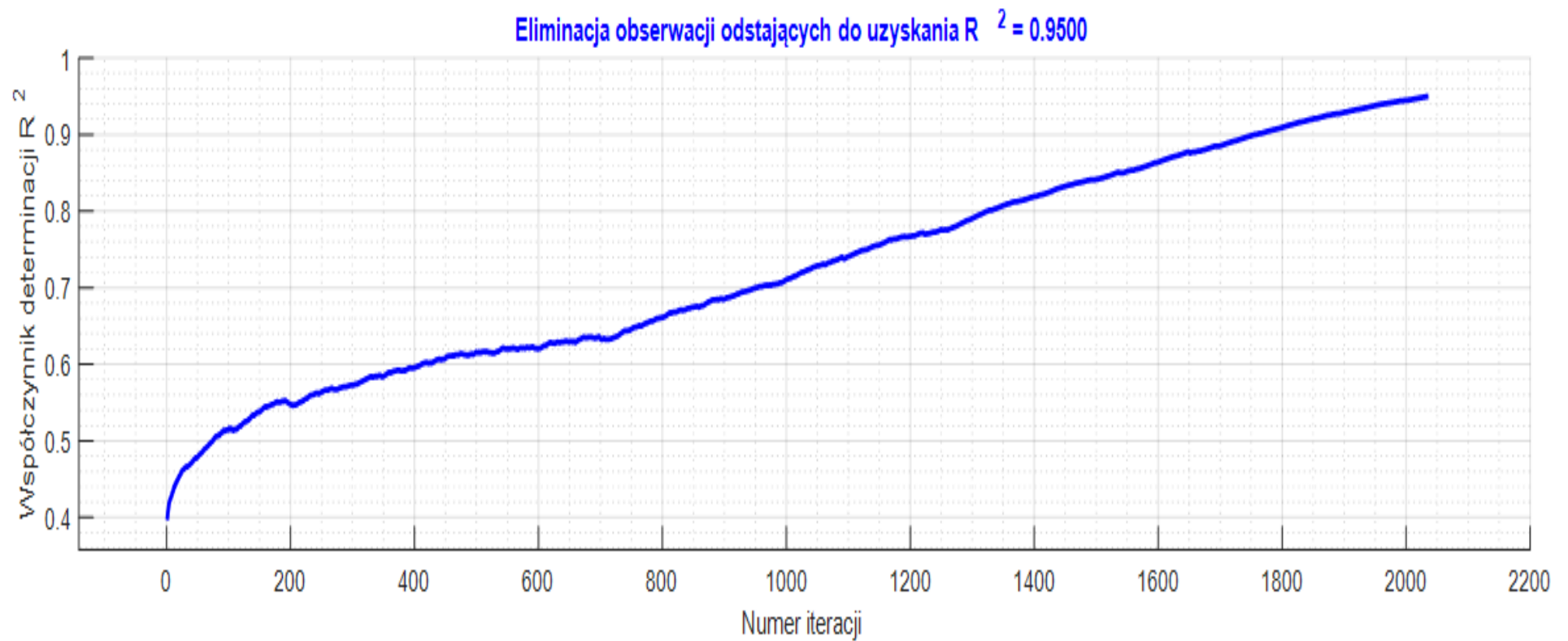
Sformułowanie modelu

Na przykład regresja wielowymiarowa

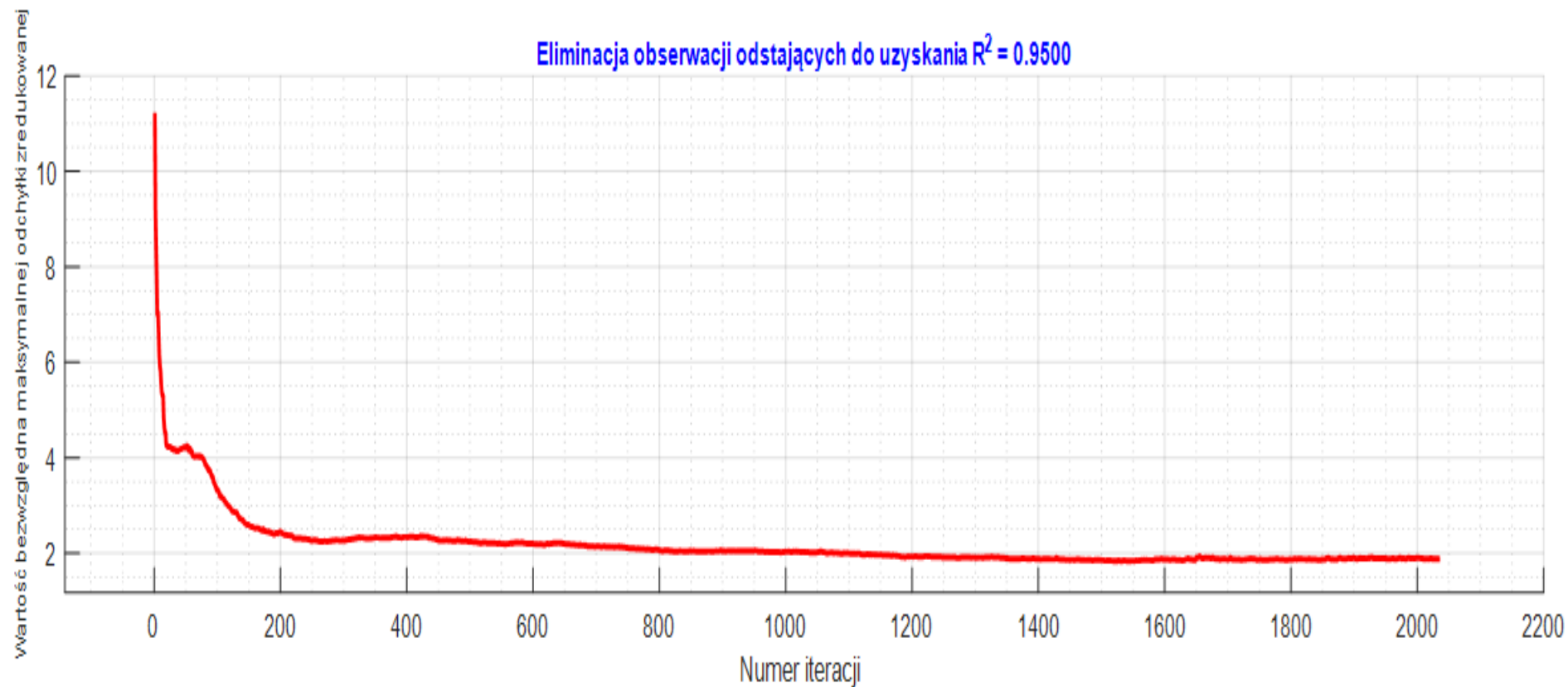
Eliminacja obserwacji odstających



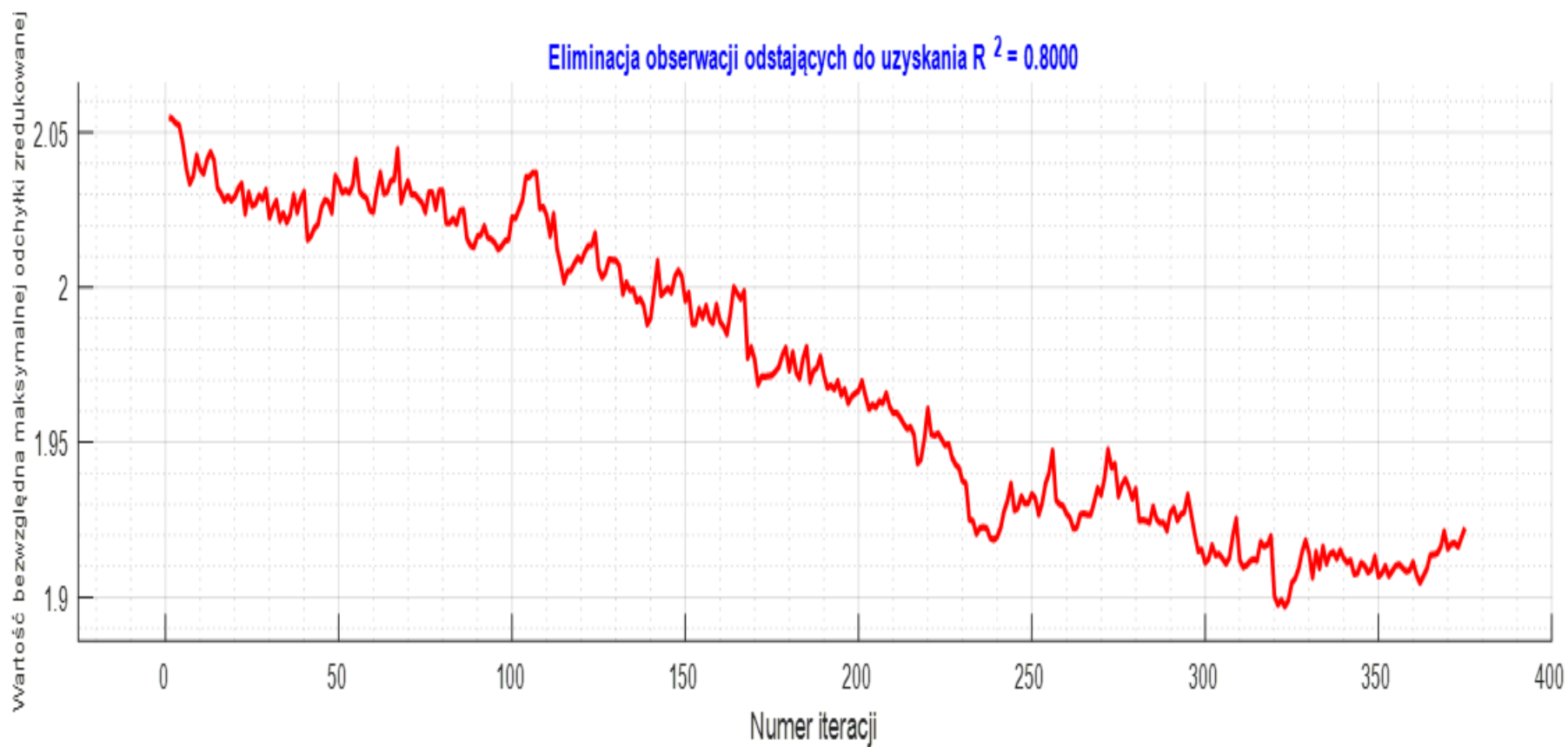
Eliminacja obserwacji odstających



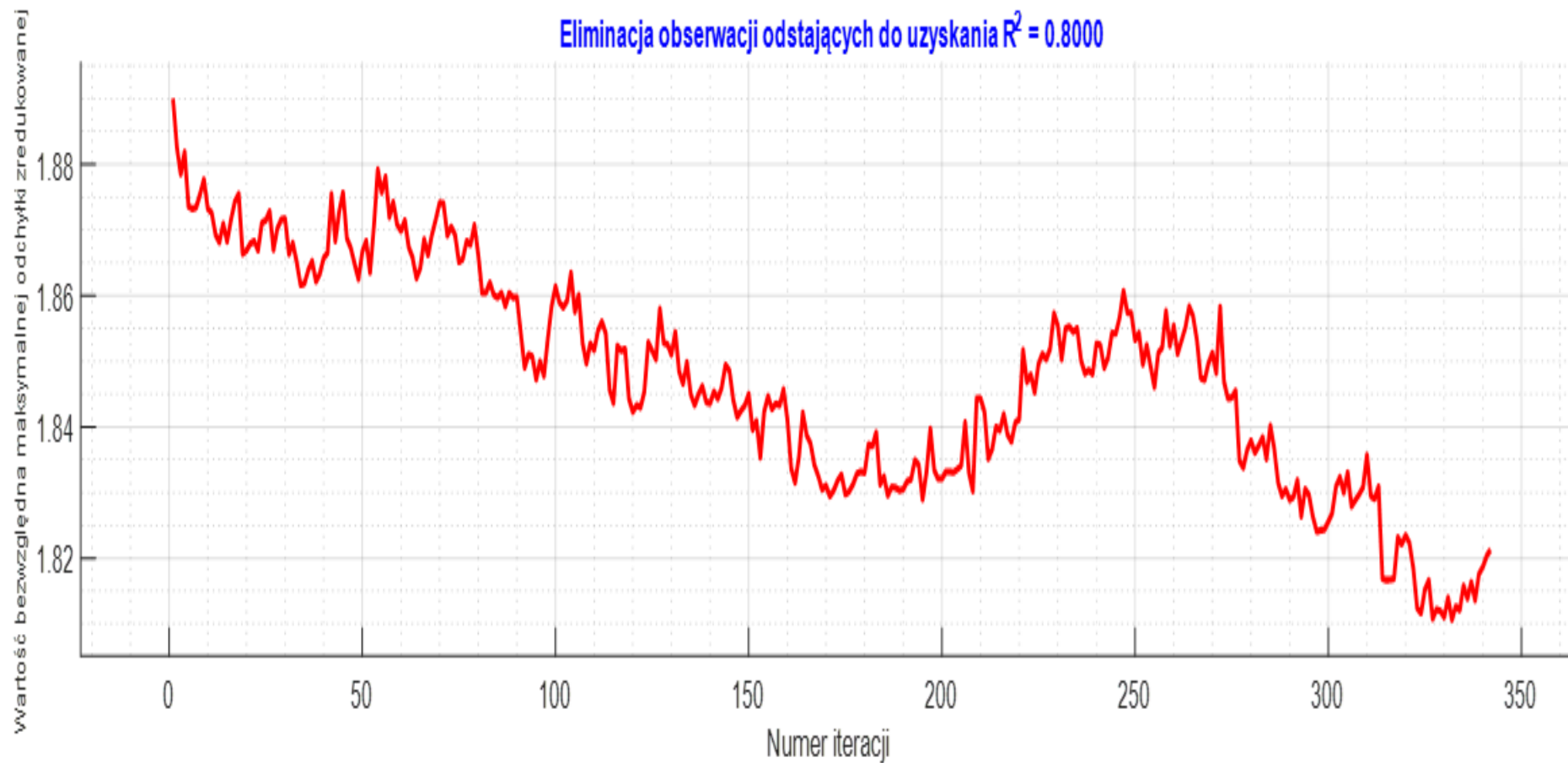
Eliminacja obserwacji odstających



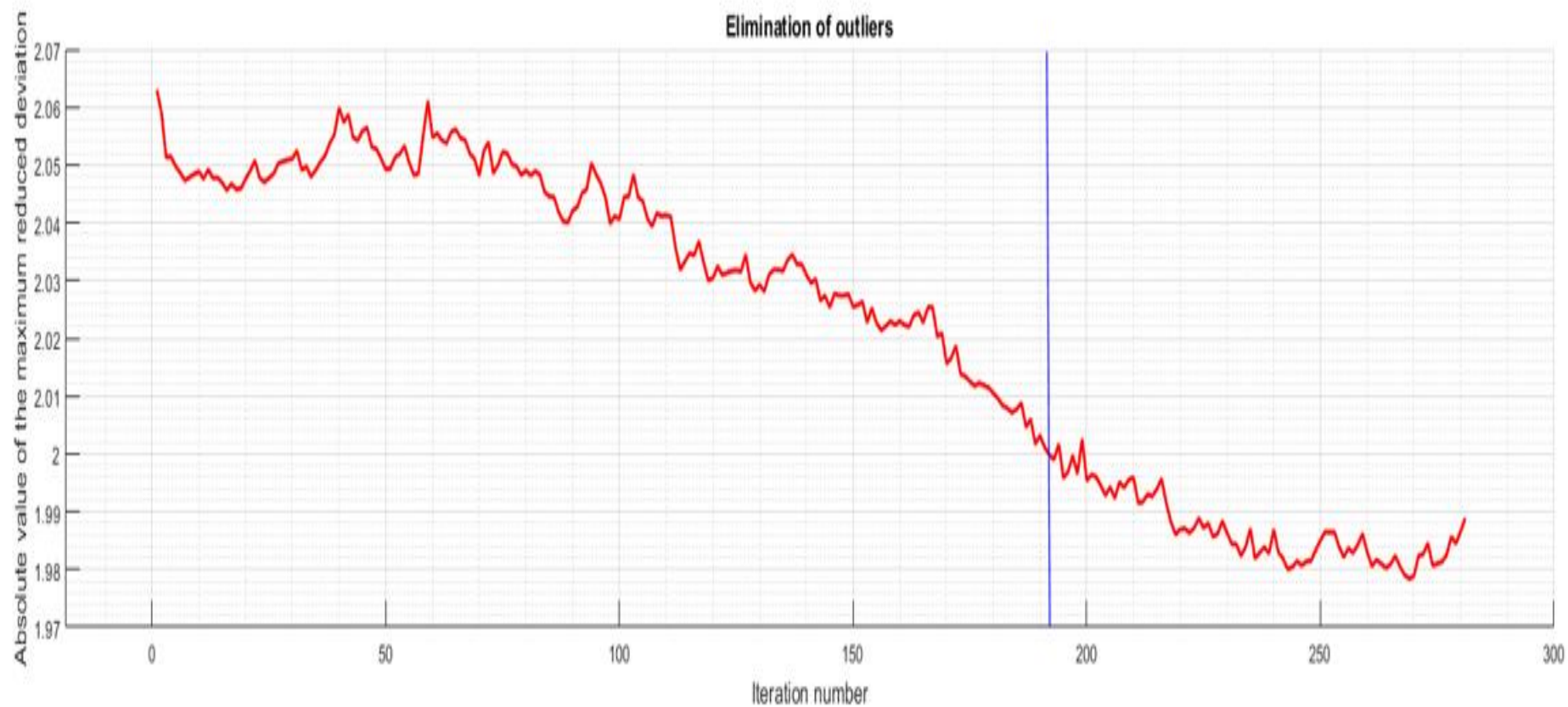
Eliminacja obserwacji odstających



Eliminacja obserwacji odstających

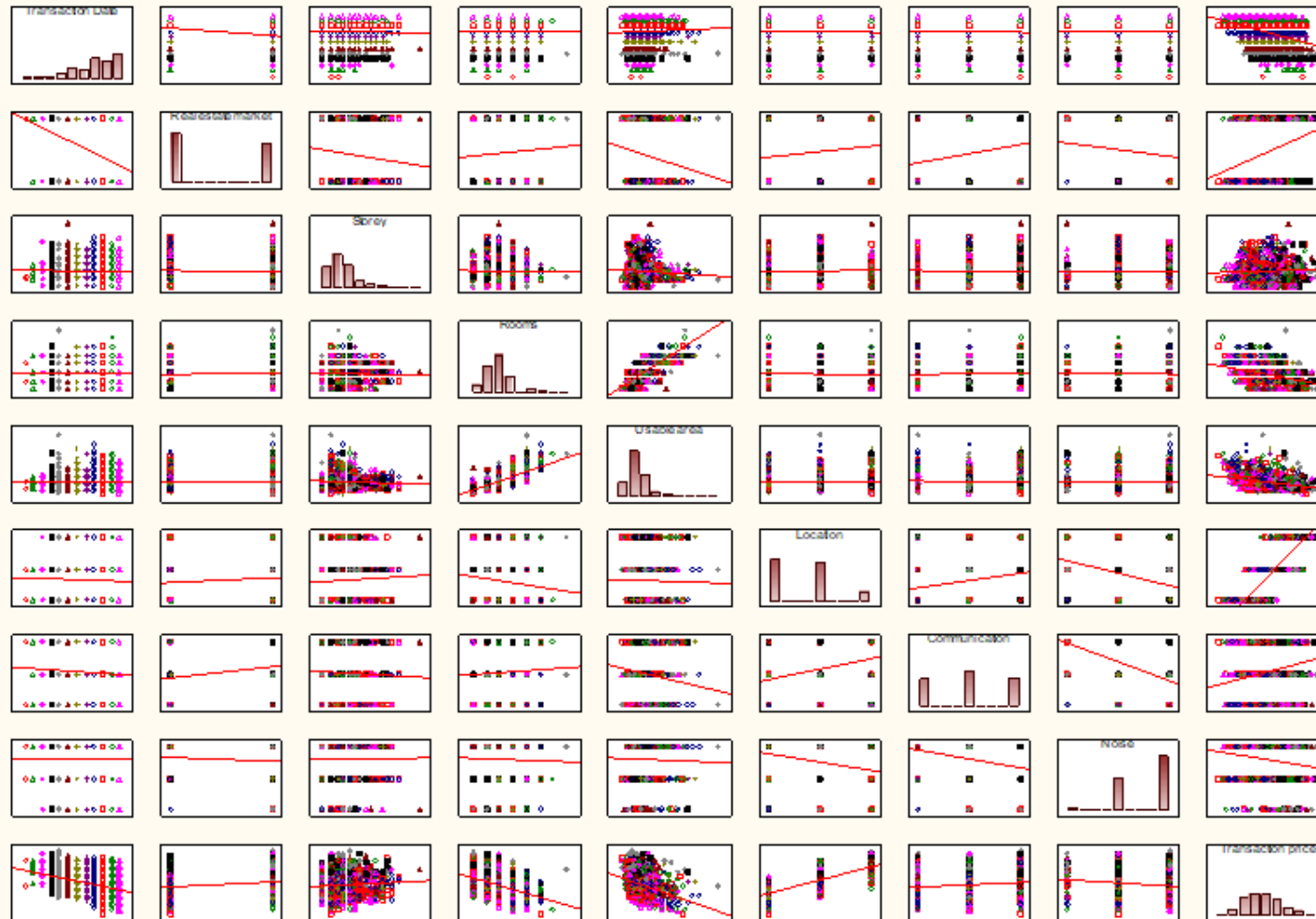


Eliminacja obserwacji odstających



Analiza zależności

Pearson correlation plot



Wycena wg modelu

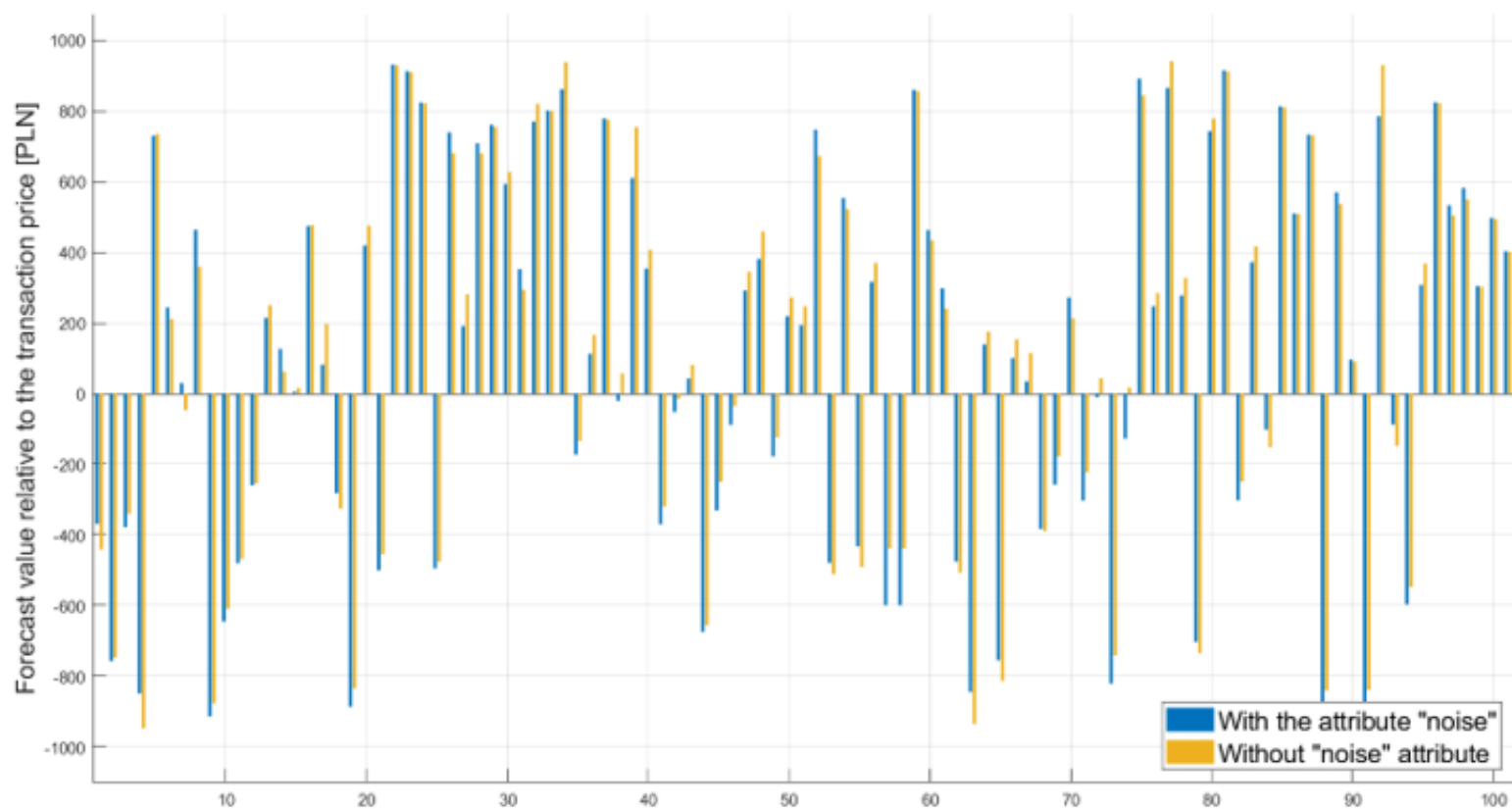


Figure 16. Różnice pomiędzy wartości prognozowanymi a cenami transakcyjnymi dla 100 nieruchomości - rynek pierwotny i wtórny – całe miasto

Analiza wybranego atrybutu

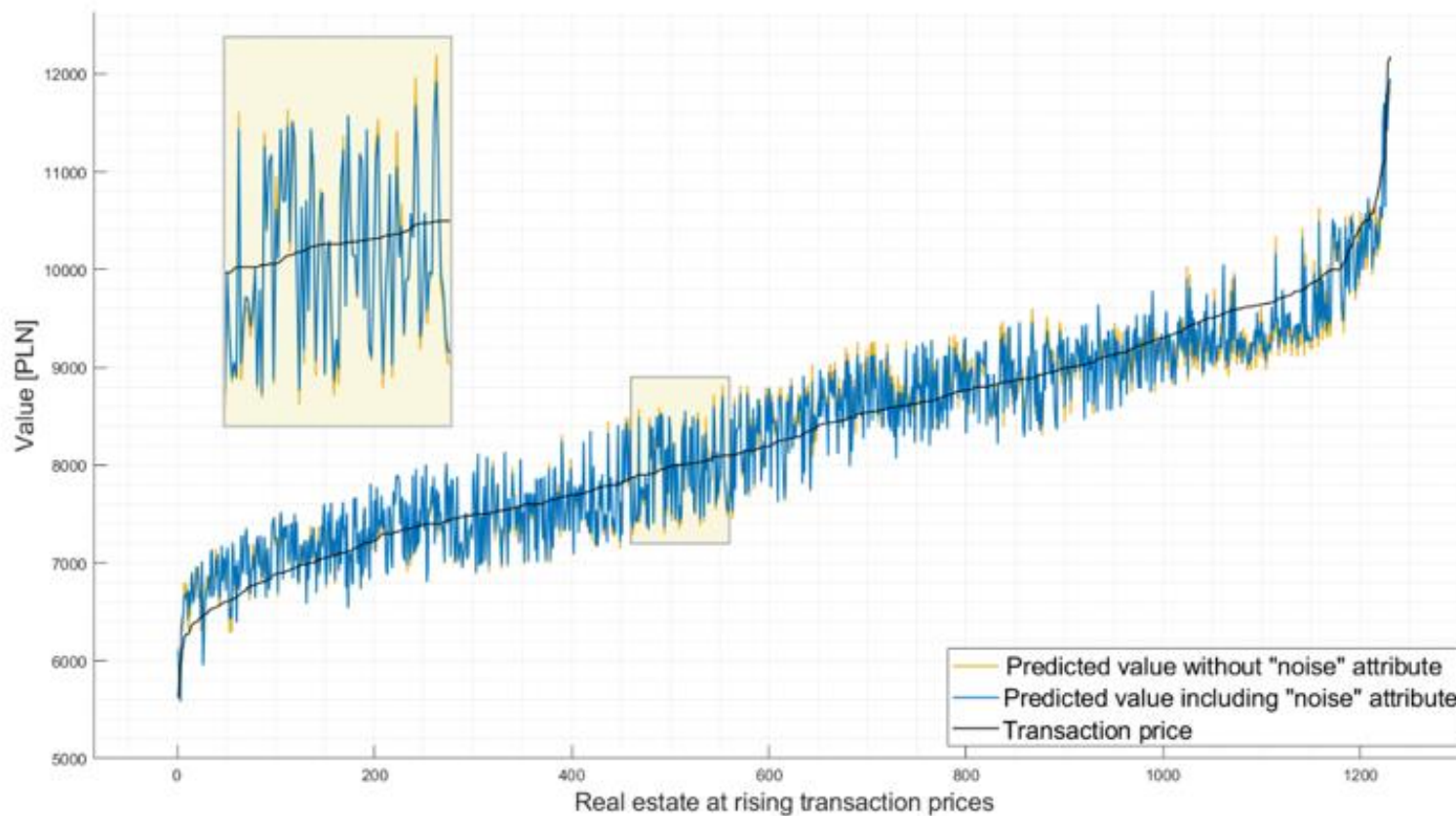


Figure 17. Wartości modelowe z uwzględnieniem i bez uwzględnienia hałasu na rynku pierwotnym – całe miasto

Analiza wybranego atrybutu

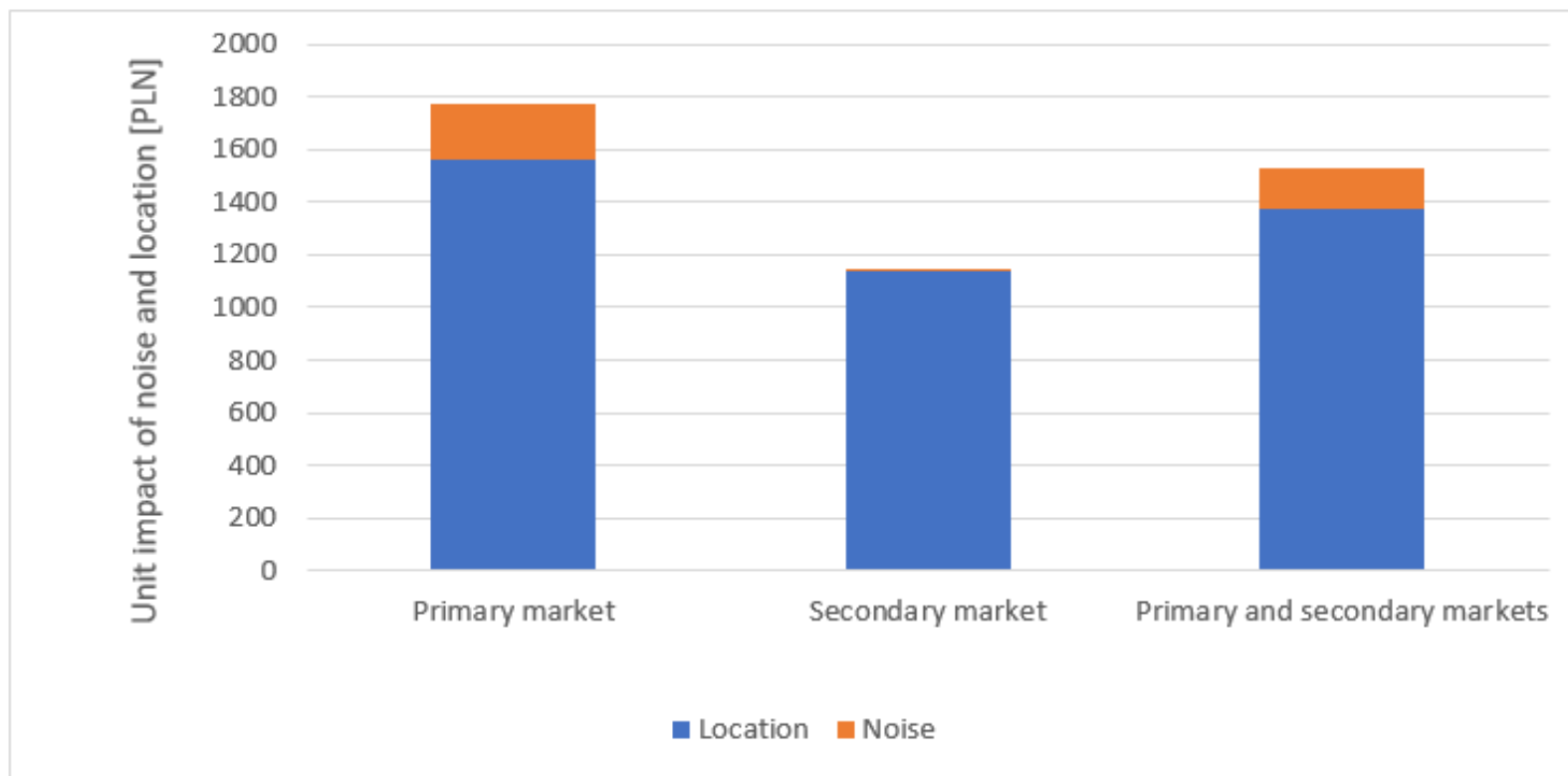


Figure 18. Jednostkowy wpływ lokalizacji i hałasu na krakowskim rynku nieruchomości lokalowych

Analiza wybranego atrybutu

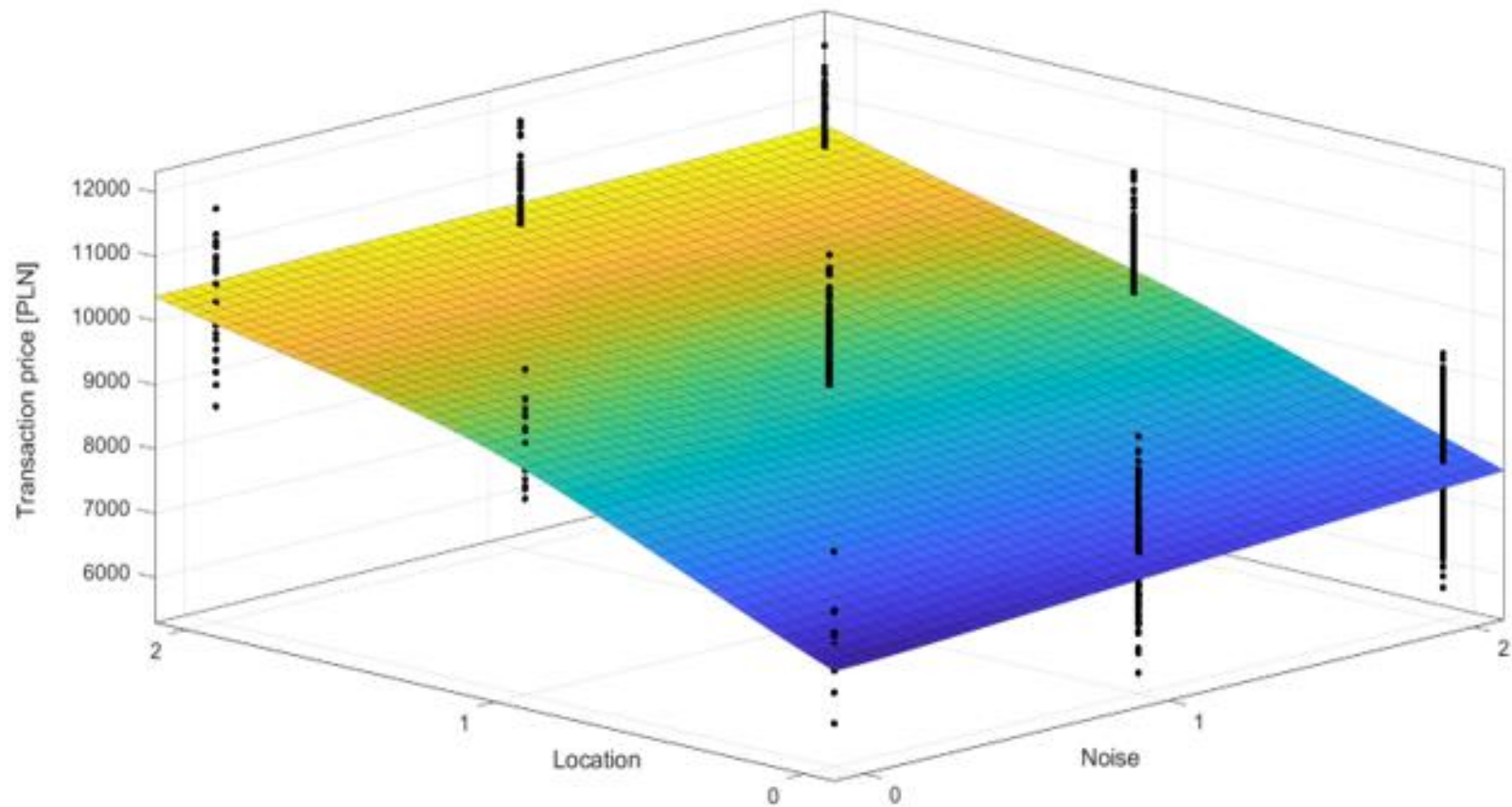


Figure 19. Zależność ceny transakcyjnej od lokalizacji i hałasu

Analiza wybranego atrybutu

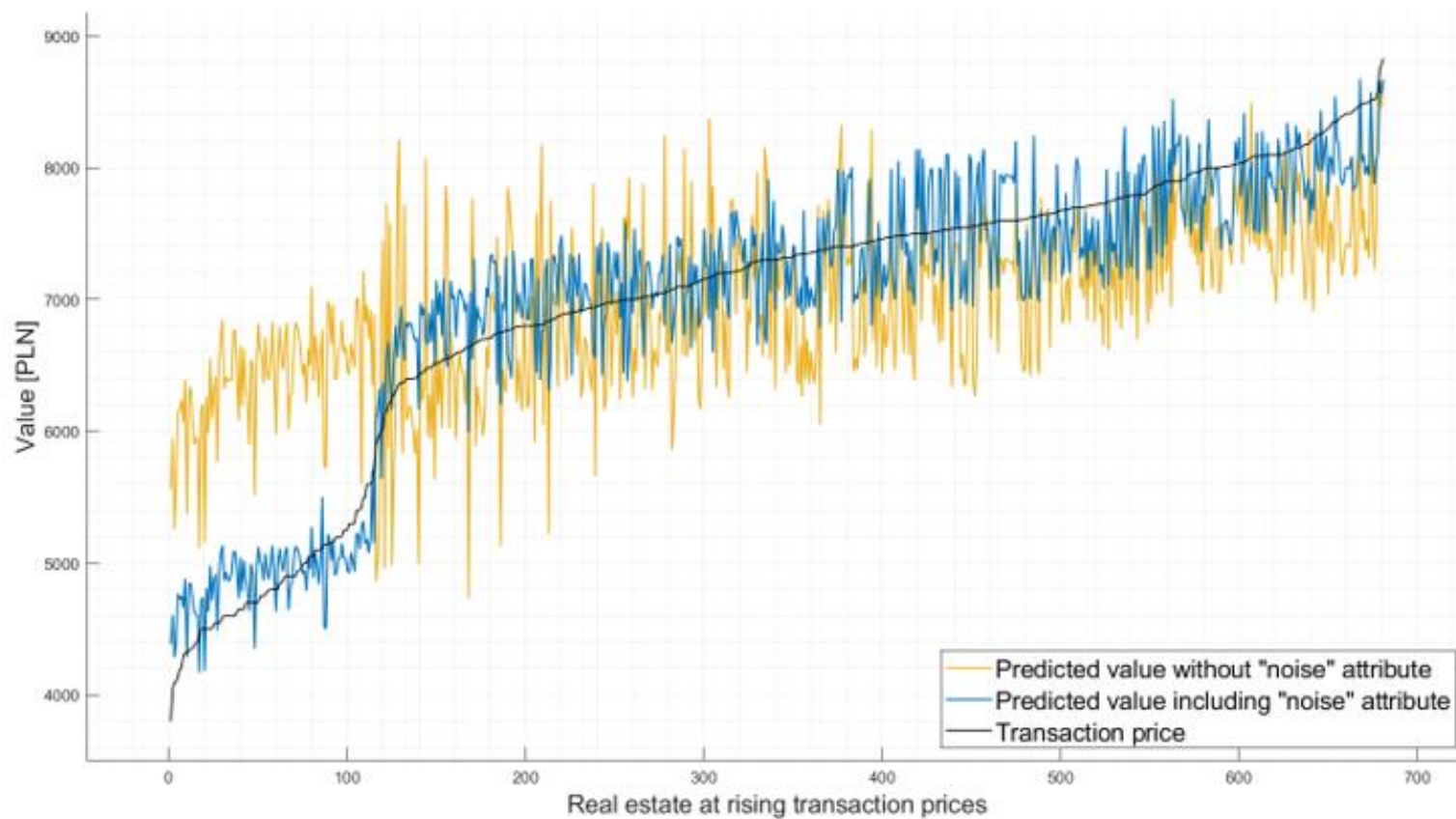


Figure 21. Wartości modelowe z uwzględnieniem i bez uwzględnienia hałasu na rynku pierwotnym – lokalizacja przeciętna.

Zestawienie wyników i wnioski

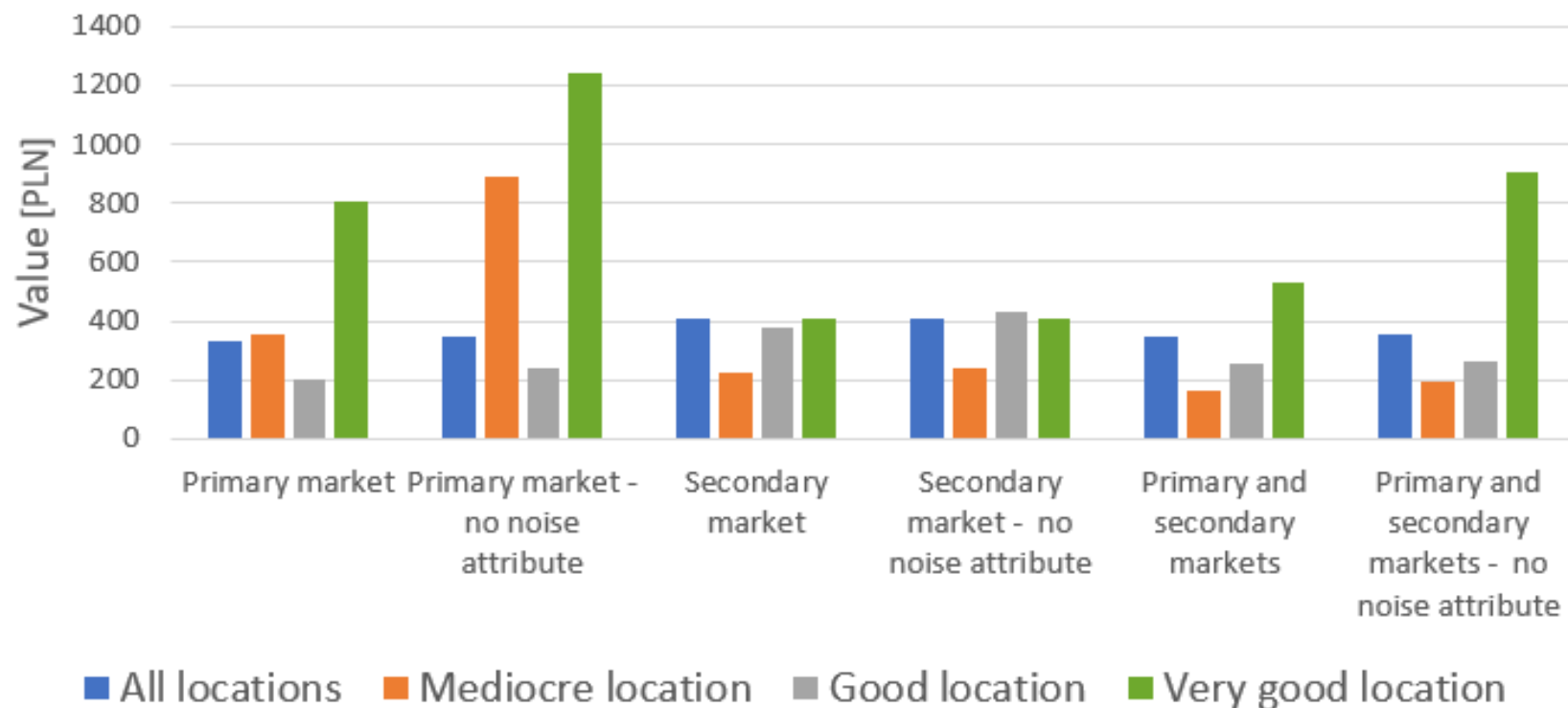


Figure 29. Standard deviation of the multivariate regression model

Zestawienie wyników i wnioski

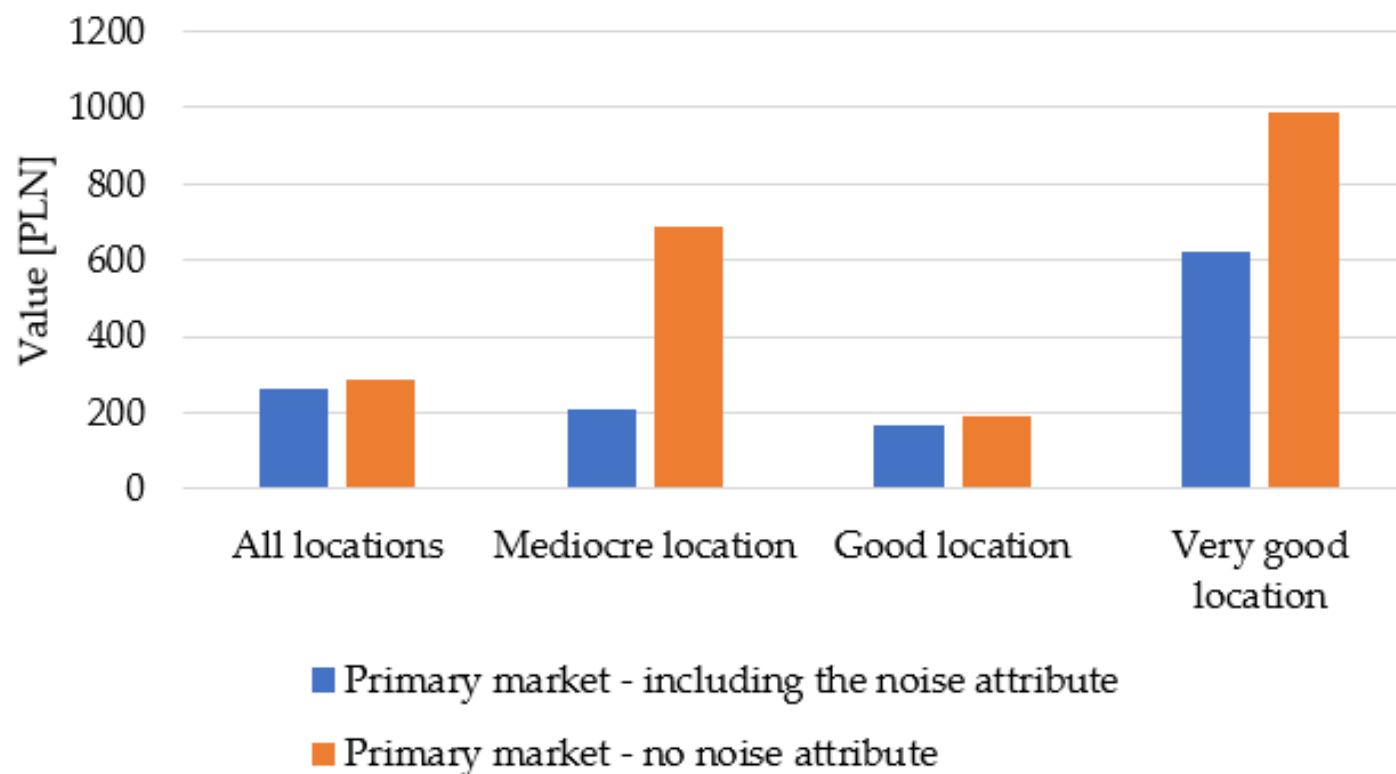


Figure 30. Primary market - mean modulus of residuals

Zestawienie wyników i wnioski

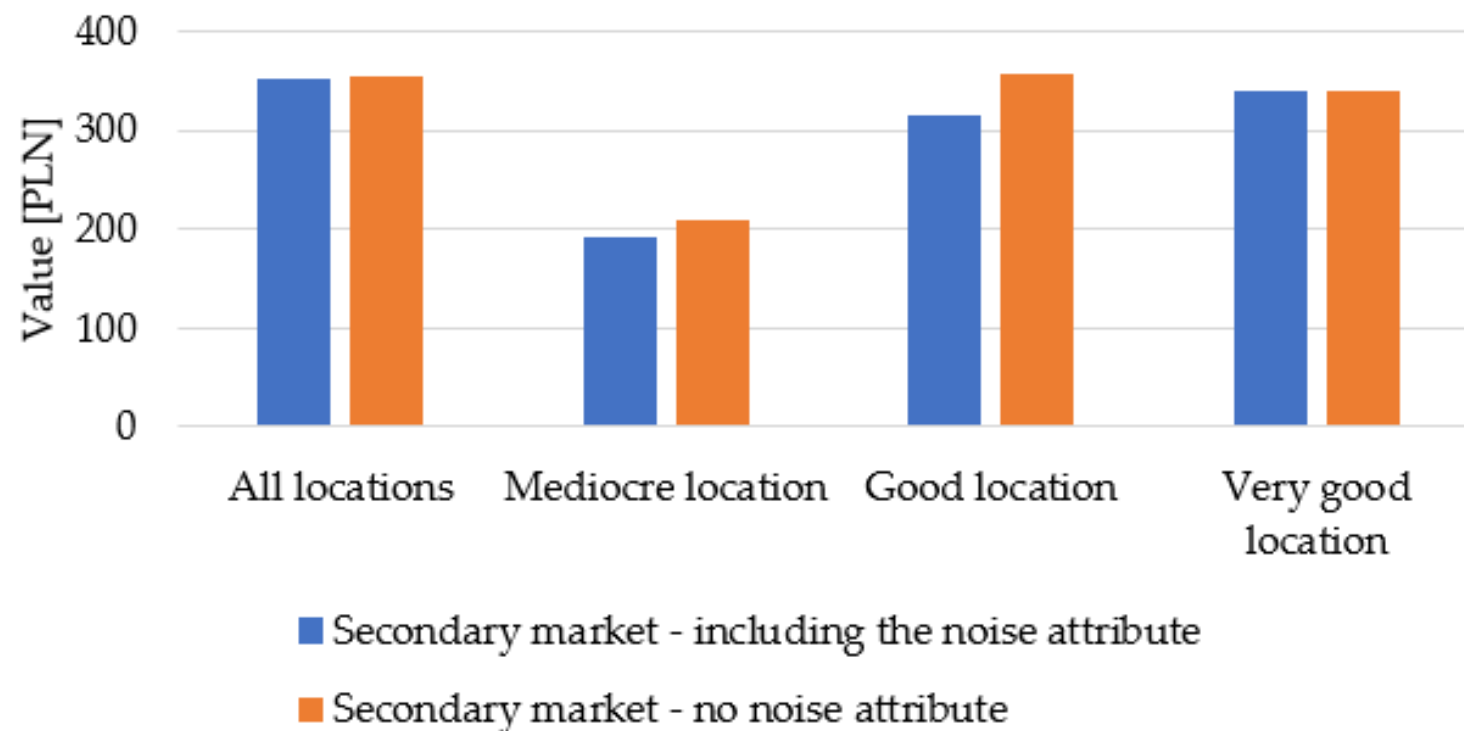


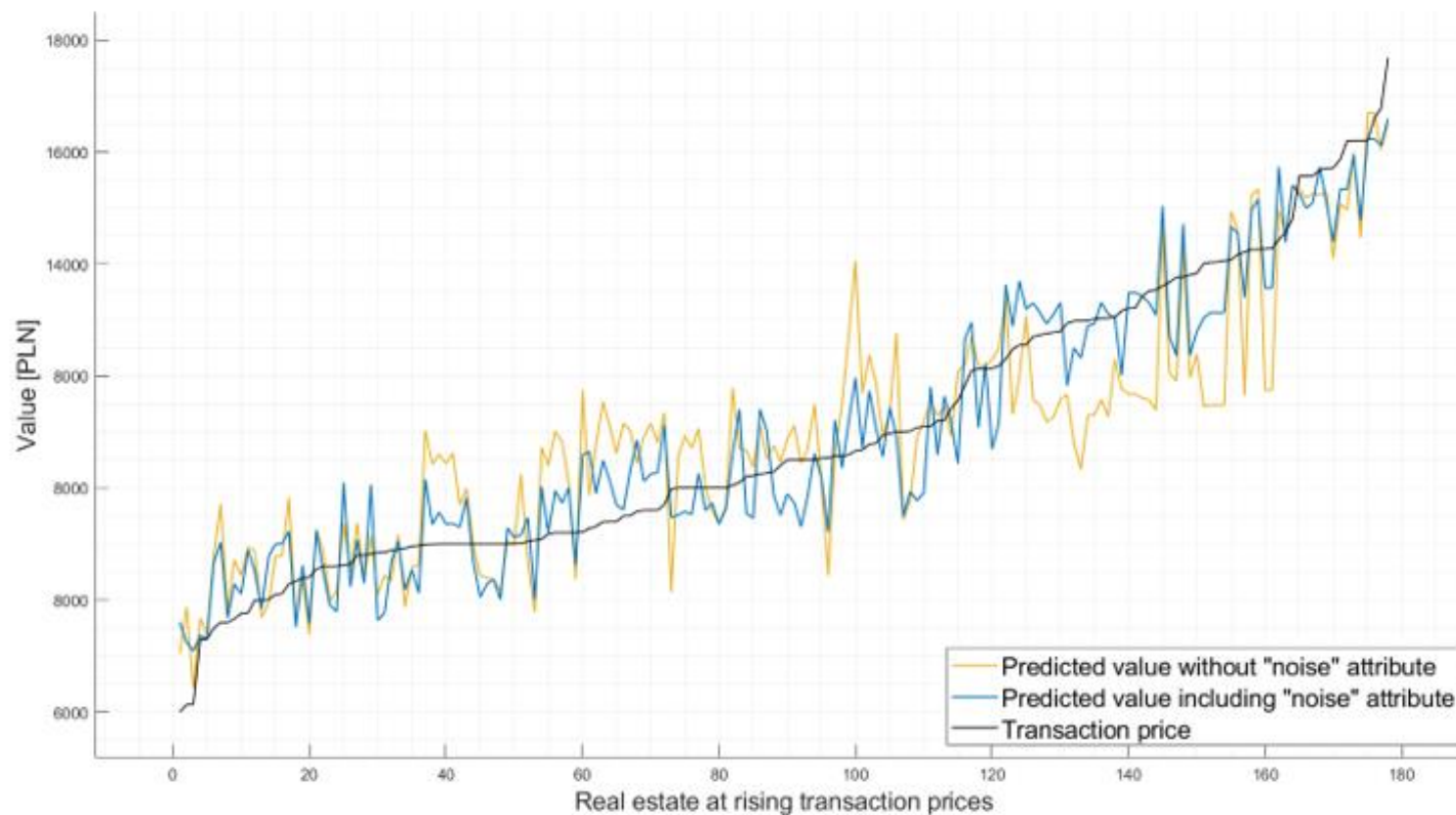
Figure 31. Secondary market - mean modulus of residuals

Zestawienie wyników i wnioski



Figure 31. Secondary market - mean modulus of residuals

Kluczowe jest opracowanie modelu i umiejętna eliminacja wpływu obserwacji odstających



Przykład w Statistica

The screenshot displays the Statistica software interface. The top menu bar includes 'Plik', 'Podstawowe', 'Edycja', 'Widok', 'Format', 'Statystyka', and 'Data Mining'. The 'Statystyka' menu is expanded, showing options like 'Statystyki podstawowe', 'Regresja wieloraka', 'ANOVA', 'Nieparametryczne', 'Dopasowanie rozkładu', 'Rozkłady i symulacja', 'Modele zaawansowane', 'Wielowymiarowe', and 'Analiza mocy testów'. The 'Regresja wieloraka' dialog box is open, showing the following settings:

- Tab: Podstawowe
- Zmienne: Zmienne
- Zależna: Cena
- Niezależne: 4-11 13
- Momenty ważone:
- DF = W-1 N-1
- Usuwanie BD: Przypadkami, Parami, Zastępowanie średnią

Buttons: OK, Anuluj, Opcje, Otwórz dane, SELECT CASES, S, W.

Patrz także moduł Ogólne modele regresji (GRM)

| Baz | Lp. |
|-----|-----|
| | 1 |
| | 2 |
| | 3 |
| | 4 |
| | 5 |
| | 6 |
| | 7 |
| | 8 |
| | 9 |
| | 10 |
| | 11 |
| | 12 |
| | 13 |
| | 14 |
| | 15 |
| | 16 |

Przykład w Statistica

Wyniki regresji wielorakiej: D_1

? X

Wyniki regresji wielorakiej

```
Zmn. zależ. Cena          Wielor. R = ,74453110      F = 13,81995
                          R^2= ,55432655      df = 9,100
Liczba przyp. 110        Popraw. R^2= ,51421594      p = ,000000
                          Błąd standardowy estymacji:56239,356352
Wyr. wolny -202325,1359  Błąd std.: 60855,79  t( 100) = -3,325  p = ,0012
```

```
Uzbrojenie b* = ,070      Dostęp do dro b* = ,132      Dostępność ko b* = ,158
Ukształtowani b* = -,06  Kształt dział b* = ,080      Lokalizacja b* = ,139
Intensywność b* = ,077   Ogrodzenie b* = ,083      Powierzchnia b* = ,696
```

(istotne b^* są podświetlone na czerwono)

Alfa do podświetlania efektów:

Podstawowe | Więcej | Reszty, założenia, predykcja

Wartości przewidywane:

Predykcja zmiennej zależnej

Oblicz granice ufności Alfa:

Oblicz granice predykcji

Przykład w Statistica

Analiza reszt: D_1

? X

Zmn. zależ. Cena Wielor. R : ,74453110 F = 13,81995
R^2: ,55432655 df = 9,100
Liczba przyp. 110 Popraw. R^2: ,51421594 p = ,000000
Błąd standardowy estymacji: 56239,356352
Wyr. wolny -202325,1359 Błąd std.: 60855,79 t(100) = -3,325 p < ,0012



Podstawowe | Więcej | Reszty | Przewid. | Wykr. rozrzutu | Wykr. prawd. | Odstające | Zapisz

Podsumowanie

Wykr. odstających wg przypadków

Typ wartości odstającej

Standaryzowane reszty ($\pm 2-5$ sigma)

100 najbardziej odstających:

Standaryzowane wart. przewidywane

Standaryzowane reszty

Usunięte reszty

Odległości Mahalanobisa

Odległości Cooka

Opcje

Grupami

Anuluj

Przykład w Statistica

| Reszty standaryz.: Cena (D_1) | | | | | | | | | | | | | |
|-------------------------------|-------------------|---|---|---|-----------------|--------------------|--------|-------------------|-----------------|-------------------|-----------------|-----------------|---------------|
| Uporządkowane | | | | | | | | | | | | | |
| Przyp. | Reszty standaryz. | | | | Obserw. Wartość | Przewidyw. Wartość | Reszta | Standard Przewid. | Standard Reszta | Bł. std. W.przew. | Mahaln. Odległ. | Usunięte Reszta | Cooka Odległ. |
| 63 | . | . | . | . | 3,37 | . | . | * | | | | | |
| 3 | . | . | . | . | . | * | . | . | | | | | |
| 31 | * | . | . | . | . | . | . | . | | | | | |
| 48 | . | . | . | . | . | * | . | . | | | | | |
| 33 | . | . | . | . | . | * | . | . | | | | | |
| 38 | . | . | . | . | . | * | . | . | | | | | |
| 89 | . | . | . | . | . | * | . | . | | | | | |
| 43 | * | . | . | . | . | . | . | . | | | | | |
| 67 | . | . | . | . | . | * | . | . | | | | | |
| 2 | . | . | . | . | . | * | . | . | | | | | |
| 4 | . | . | . | . | . | * | . | . | | | | | |
| 60 | . | . | . | . | . | * | . | . | | | | | |
| 10 | . | . | . | . | . | * | . | . | | | | | |
| 86 | * | . | . | . | . | . | . | . | | | | | |
| 1 | . | . | . | . | . | * | . | . | | | | | |
| 62 | . | . | . | . | . | * | . | . | | | | | |
| 78 | . | . | . | . | . | * | . | . | | | | | |
| 21 | * | . | . | . | . | . | . | . | | | | | |
| 98 | * | . | . | . | . | . | . | . | | | | | |
| 82 | * | . | . | . | . | . | . | . | | | | | |
| 39 | * | . | . | . | . | . | . | . | | | | | |
| 75 | . | * | . | . | . | . | . | . | | | | | |
| 84 | . | * | . | . | . | . | . | . | | | | | |
| 109 | . | * | . | . | . | . | . | . | | | | | |
| 22 | . | . | . | . | . | * | . | . | | | | | |

Przykład w Statistica

| Reszty standaryz.: Cena (D_1) | | | | | | | | | | | | | |
|-------------------------------|-------------------|---|---|---|-----------------|--------------------|--------|-------------------|-----------------|-------------------|-----------------|-----------------|---------------|
| Uporządkowane | | | | | | | | | | | | | |
| Przyp. | Reszty standaryz. | | | | Obserw. Wartość | Przewidyw. Wartość | Reszta | Standard Przewid. | Standard Reszta | Bł. std. W.przew. | Mahaln. Odległ. | Usunięte Reszta | Cooka Odległ. |
| 63 | . | . | . | . | 3,37 | . | . | * | | | | | |
| 3 | . | . | . | . | . | * | . | | | | | | |
| 31 | * | . | . | . | . | . | . | | | | | | |
| 48 | . | . | . | . | . | * | . | | | | | | |
| 33 | . | . | . | . | . | * | . | | | | | | |
| 38 | . | . | . | . | . | * | . | | | | | | |
| 89 | . | . | . | . | . | * | . | | | | | | |
| 43 | * | . | . | . | . | . | . | | | | | | |
| 67 | . | . | . | . | . | * | . | | | | | | |
| 2 | . | . | . | . | . | * | . | | | | | | |
| 4 | . | . | . | . | . | * | . | | | | | | |
| 60 | . | . | . | . | . | * | . | | | | | | |
| 10 | . | . | . | . | . | * | . | | | | | | |
| 86 | * | . | . | . | . | . | . | | | | | | |
| 1 | . | . | . | . | . | * | . | | | | | | |
| 62 | . | . | . | . | . | * | . | | | | | | |
| 78 | . | . | . | . | . | * | . | | | | | | |
| 21 | * | . | . | . | . | . | . | | | | | | |
| 98 | * | . | . | . | . | . | . | | | | | | |
| 82 | * | . | . | . | . | . | . | | | | | | |
| 39 | * | . | . | . | . | . | . | | | | | | |
| 75 | . | * | . | . | . | . | . | | | | | | |
| 84 | . | * | . | . | . | . | . | | | | | | |
| 109 | . | * | . | . | . | . | . | | | | | | |
| 22 | . | . | . | . | . | * | . | | | | | | |

Przykład w Statistica

Wyniki regresji wielorakiej: D_1h-end

Wyniki regresji wielorakiej

| | | |
|-------------------------|--|---------------------------|
| Zmn. zależ. Cena | Wielor. R = ,96095408 | F = 44,62353 |
| | R ² = ,92343274 | df = 10,37 |
| Liczba przyp. 48 | Popraw. R ² = ,90273889 | p = ,000000 |
| | Błąd standardowy estymacji: 19093,557989 | |
| Wyr. wolny -272184,2098 | Błąd std.: 41137,40 | t(37) = -6,616 p = ,0000 |

| | | |
|------------------------|------------------------|------------------------|
| Powierzchnia b*=-,04 | Uzbrojenie b*=-,00 | Dostęp do dro b*=-,144 |
| Dostępność ko b*=-,304 | Ukształtowani b*=-,117 | Kształt dział b*=-,235 |
| Lokalizacja b*=-,203 | Intensywność b*=-,009 | Ogrodzenie b*=-,131 |
| Powierzchnia b*=-,913 | | |

(istotne b* są podświetlone na czerwono)

Alfa do podświetlania efektów: .05

Podstawowe | Więcej | Reszty, założenia, predykcja

OK Anuluj

Dyskusja

Dziękuję za uwagę