Analiza ofert sprzedaży mieszkań oraz przewidywanie ich cen w miastach wojewódzkich w Polsce

Dawid Chudzicki 272715

9 czerwca 2024

Spis treści

1 Wstęp

Rynek nieruchomości w miastach wojewódzkich w Polsce oraz w innych obszarach miejskich jest skomplikowany i dynamiczny. Aby potencjalni najemcy i inwestorzy mogli podejmować świadome decyzje, potrzebują szczegółowych i precyzyjnych informacji.

Celem tego raportu jest analiza czynników, które mają największy wpływ na cenę sprzedaży oraz stworzenie narzędzia do przewidywania cen mieszkań w miastach wojewódzkich w Polsce. Analiza obejmie różnorodne czynniki, takie jak miasto, metraż, wykończenie nieruchomości, typ zabudowy, rynek, piętro oraz odległość do centrum miasta.

Ze względu na subiektywny charakter problemu, można oczekiwać, że wyniki analizy będą podlegać wpływom osobistych przekonań. Dlatego też, ważnym aspektem raportu będzie również identyfikacja i minimalizacja tych subiektywnych wpływów, aby zapewnić jak najbardziej obiektywną ocenę atrakcyjności ofert mieszkaniowych.

2 Pozyskanie danych

Dane były pozyskiwane od kwietnia do końca maja 2024 roku, przy użyciu zewnętrznego niepublicznego API serwisu OLX, dwa razy w każdym tygodniu, z wykorzystaniem skryptów Pythona. Łącznie dla wszystkim miast zebrano około 28 tysięcy ofert mieszkań.

3 Przetwarzanie danych

3.1 Dane kategoryczne

Przekształcenie danych kategorycznych obejmowało:

- Miasto
- Piętro
- Umeblowanie
- Rynek (pierwotny lub wtórny)
- Typ zabudowy
- Liczbę pokoi

3.2 Dane numeryczne

Przekształcenie danych numerycznych obejmowało:

- Cenę
- Metraż
- Odległość do centrum miasta

3.3 Przyjęte centrum dla poszczególnych miast

Miasto	Centrum
Białystok	Ratusz na Rynku Kościuszki
Bydgoszcz	Zegar z czasem bydgoskim
Gdańsk	Brama złota
Gorzów Wielkopolski	Katedra na starym rynku
Katowice	Spodek Arena
Kielce	Pałac Biskupów Krakowskich
Kraków	Sukiennice
Lublin	Plac Litewski
Łódź	Kościół Najświętszego Imienia Jezus
Olsztyn	Zamek Kapituły Warmińskiej
Opole	Ratusz na rynku
Poznań	Studnia Bamberki
Rzeszów	Pomnik Tadeusza Kościuszki
Szczecin	Plac Grunwaldzki
Toruń	Ratusz Staromiejski
Warszawa	Pałac Kultury i Nauki
Wrocław	Galeria Dominikańska
Zielona Góra	Ratusz na starym rynku

Tablica 1: Przyjęte centrum dla poszczególnych miast

3.4 Wstępne oczyszczenie danych i obliczanie odległości do centrum miasta

Podczas odczyty danych z pliku .json został dokonany wstępny preprocessing polegający na wywołaniu funkcji obliczającej odległość z odczytanej lokalizacji (przyjęto odległość dzielnicy mieszkania od centrum) oraz odrzuceniu wierszy z pustymi rekordami – gdyż w skali całego pliku jedynie około 200-300 rekordów posiadało

wartość null co w skali 28 tysięcy danych jest znikomą ilością (około 1% danych), dlatego usunięcie tych rekordów nie wpłynie negatywnie na nauczanie modelu.

3.5 Kodowanie danych

Dane kategoryczne zostały zakodowane przy użyciu kodowania one-hot z biblioteki Scikit-learn w klasie Preprocessor. Dzięki temu kodowaniu dane kategoryczne są zamieniane na wektory binarne, co pozwala na lepsze zrozumienie danych przez model.

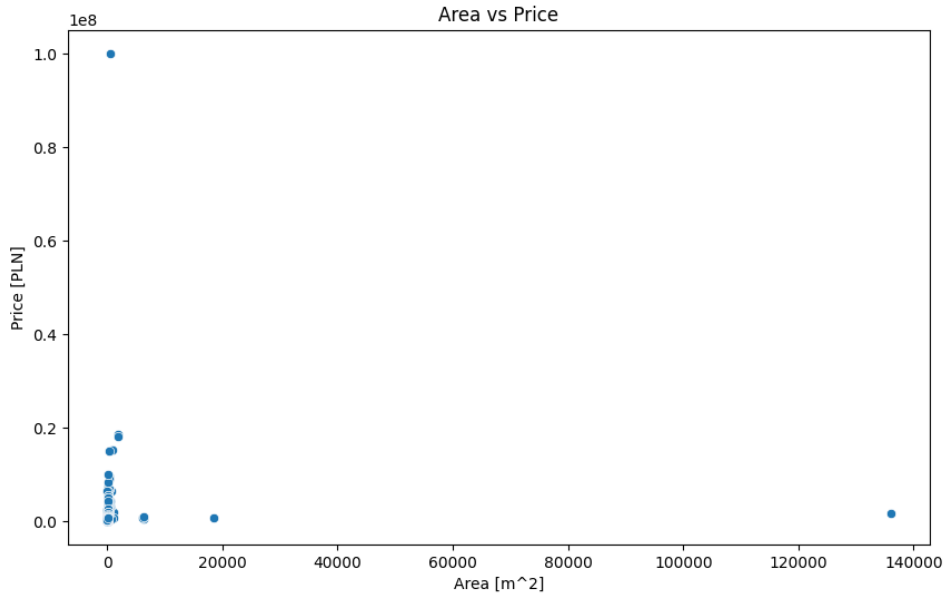
Przekształcenie danych kategorycznych jest wymagane przez większość algorytmów uczenia maszynowego. Dzięki temu zabiegowi, model może lepiej interpretować i analizować różnorodne kategorie, co zwiększa dokładność przewidywań i poprawia interpretowalność wyników. Dodatkowo, pozwala uniknąć wprowadzenia fałszywego porządku między kategoriami, co mogłoby zniekształcić wyniki analizy.

3.6 Wstępna analiza danych

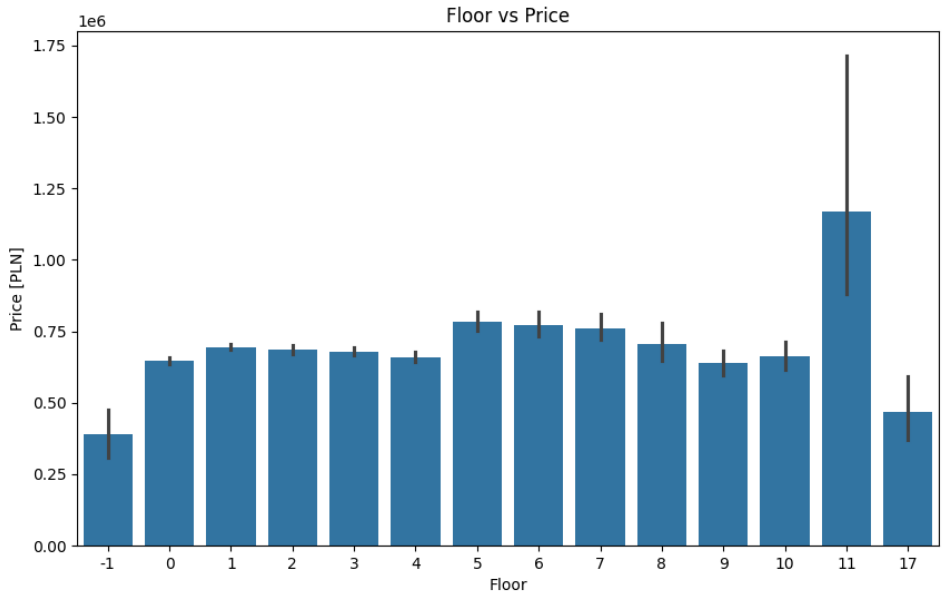
Do określenia jakości danych, zostały narysowane wykresy dla ceny zależnej od powierzchni, piętra i odległości od centrum miasta. Na wykresach 1, 2 i 3 widać, że dane są zanieczyszczone wartościami skrajnymi:

- Ceny mieszkań o wartościach mniej niż 10 tys. zł i większych niż 10 mln zł
- Mieszkania o powierzchni 0 lub tysięcy m^2
- Ujemne piętra
- Odległość od centrum w okolicy promienia ziemi

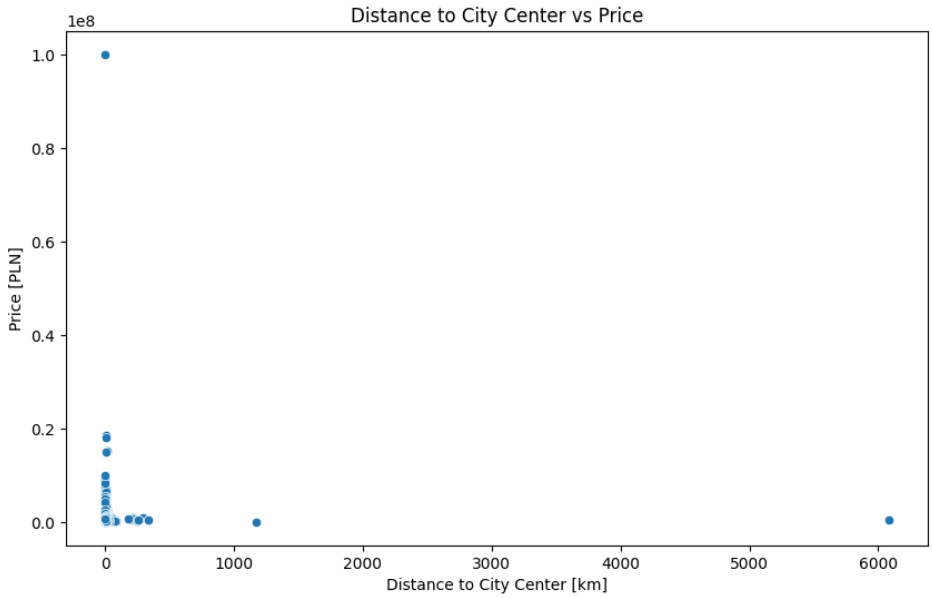
Dane wymagają oczyszczenia, aby uniknąć zniekształceń wyników analizy.



Rysunek 1: Zależność ceny mieszkania od powierzchni przed oczyszczeniem danych



Rysunek 2: Zależność ceny mieszkania od piętra przed oczyszczeniem danych



Rysunek 3: Zależność ceny mieszkania od odległości od centrum przed oczyszczeniem danych

3.7 Oczyszczenie danych

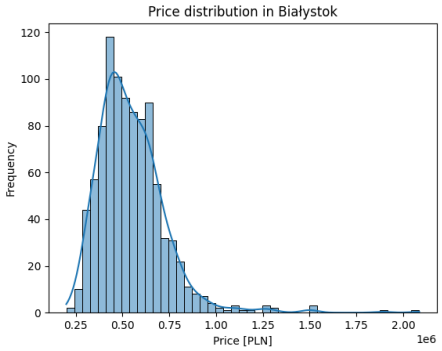
Aby oczyścić dane, zastosowano następujące filtry:

- 10 tys. zł <= cena <= 5 mln zł
- 10 m^2 <= powierzchnia <= 250 m^2
- 0 <= piętro
- odległość od centrum <= 20 km

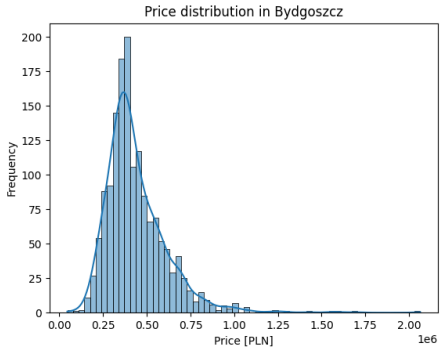
4 Analiza danych

4.1 Rozkład cen mieszkań w miastach wojewódzkich

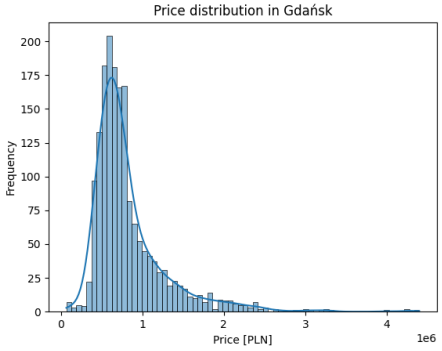
Najbardziej interesującą daną jest cena danego mieszkania dlatego został dla niej narysowany rozkład w poszczególnych miastach. Na wykresach: 4 widać, że rozkład cen mieszkań jest zbliżony do prawostronnie skośnego rozkładu normalnego. Najwięcej ofert mieszkań we wszystkich miastach wojewódzkich w Polsce mieści się w przedziale cenowym od 400 tys. zł do 600 tys. zł.



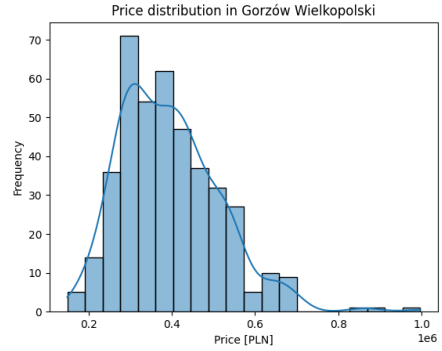
Rysunek 4: Rozkład cen mieszkań w Białymstoku



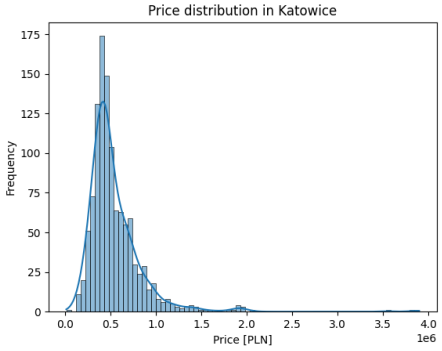
Rysunek 5: Rozkład cen mieszkań w Bydgoszczy



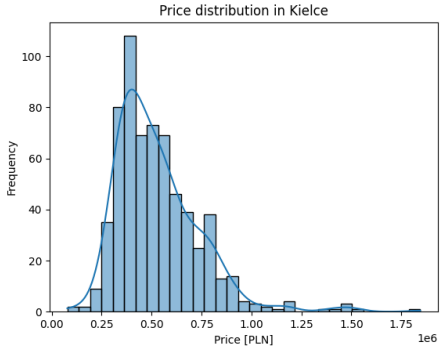
Rysunek 6: Rozkład cen mieszkań w Gdańsku



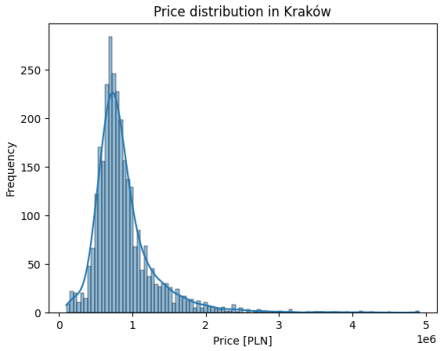
Rysunek 7: Rozkład cen mieszkań w Gorzowie Wielkopolskim



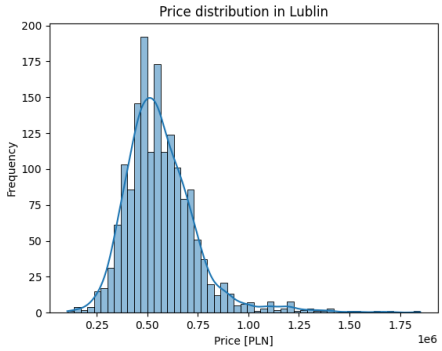
Rysunek 8: Rozkład cen mieszkań w Katowicach



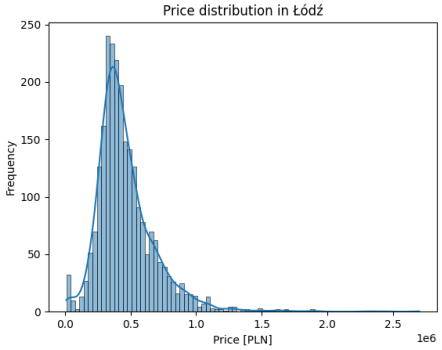
Rysunek 9: Rozkład cen mieszkań w Kielcach



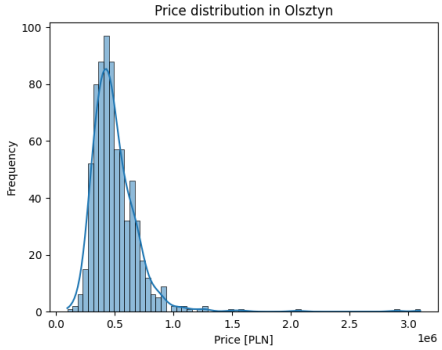
Rysunek 10: Rozkład cen mieszkań w Krakowie



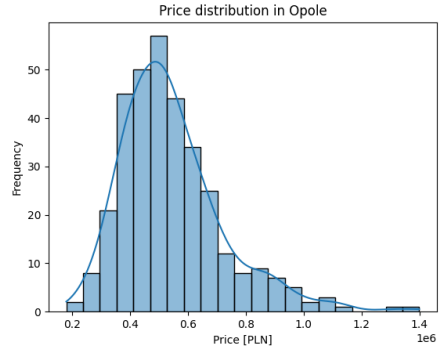
Rysunek 11: Rozkład cen mieszkań w Lublinie



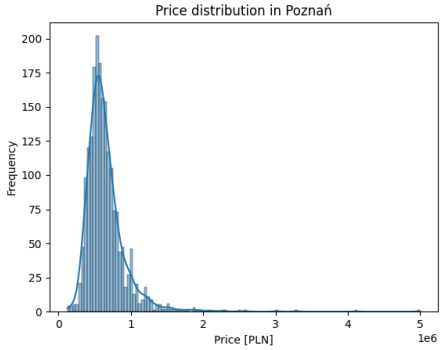
Rysunek 12: Rozkład cen mieszkań w Łodzi



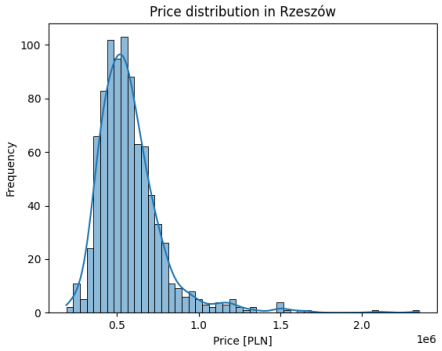
Rysunek 13: Rozkład cen mieszkań w Olsztynie



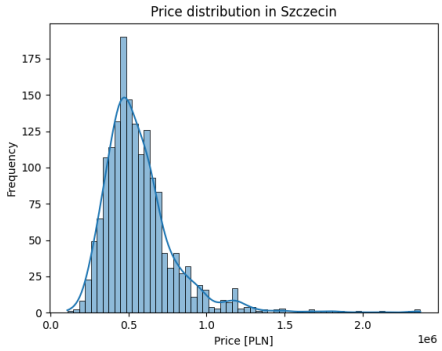
Rysunek 14: Rozkład cen mieszkań w Opolu



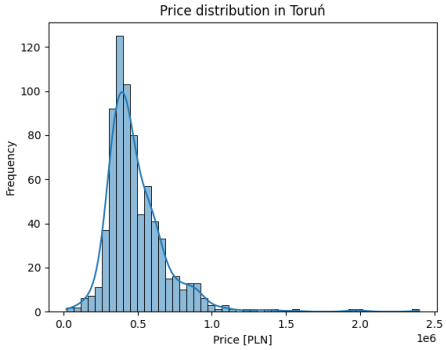
Rysunek 15: Rozkład cen mieszkań w Poznaniu



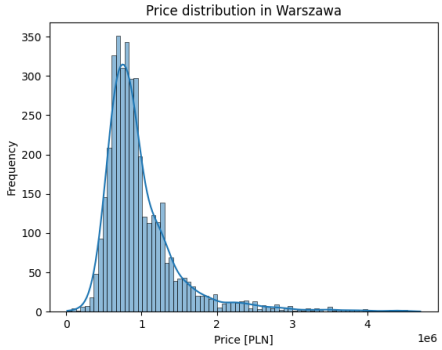
Rysunek 16: Rozkład cen mieszkań w Rzeszowie



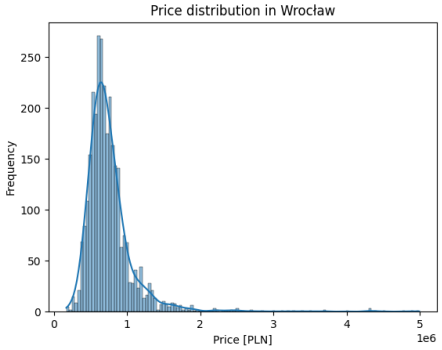
Rysunek 17: Rozkład cen mieszkań w Szczecinie



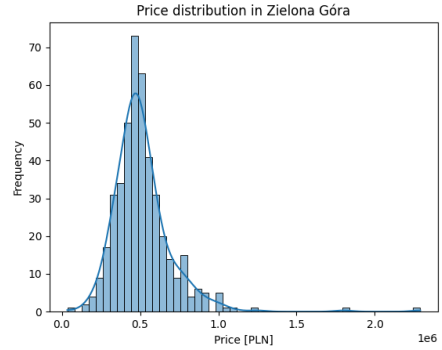
Rysunek 18: Rozkład cen mieszkań w Toruniu



Rysunek 19: Rozkład cen mieszkań w Warszawie



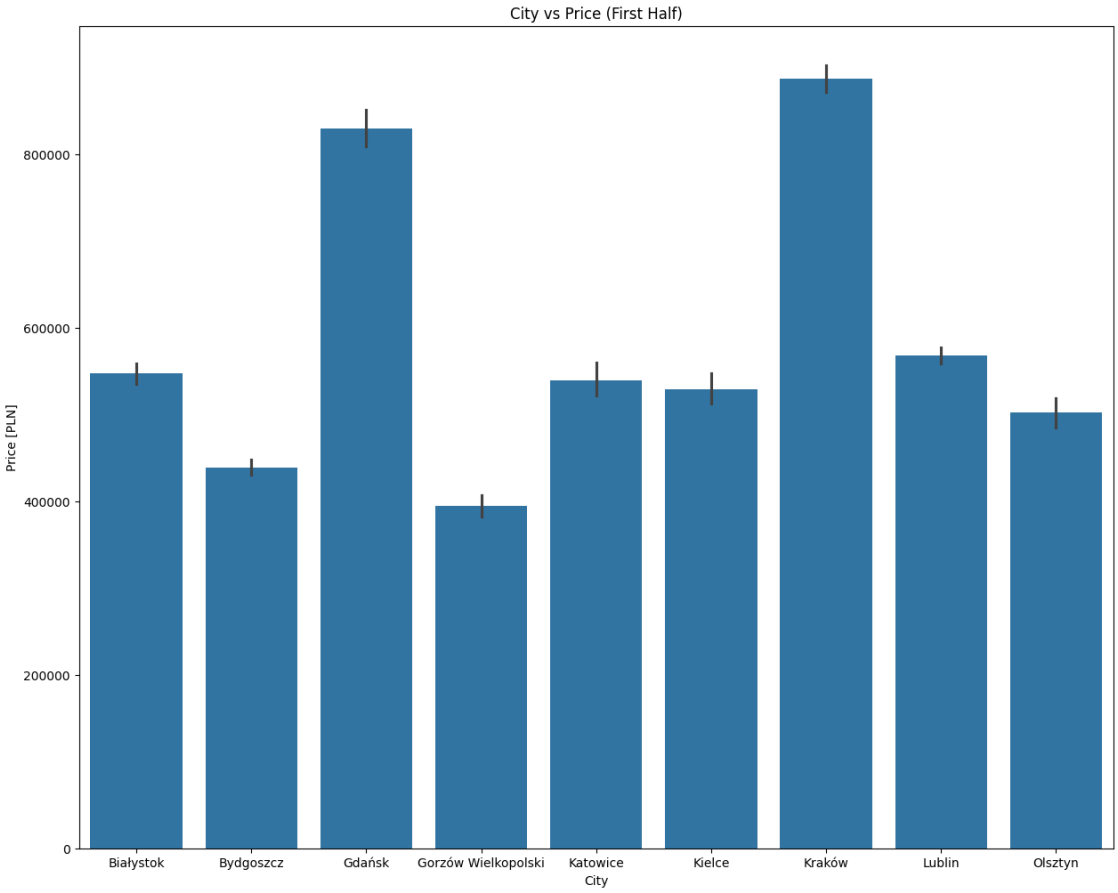
Rysunek 20: Rozkład cen mieszkań we Wrocławiu



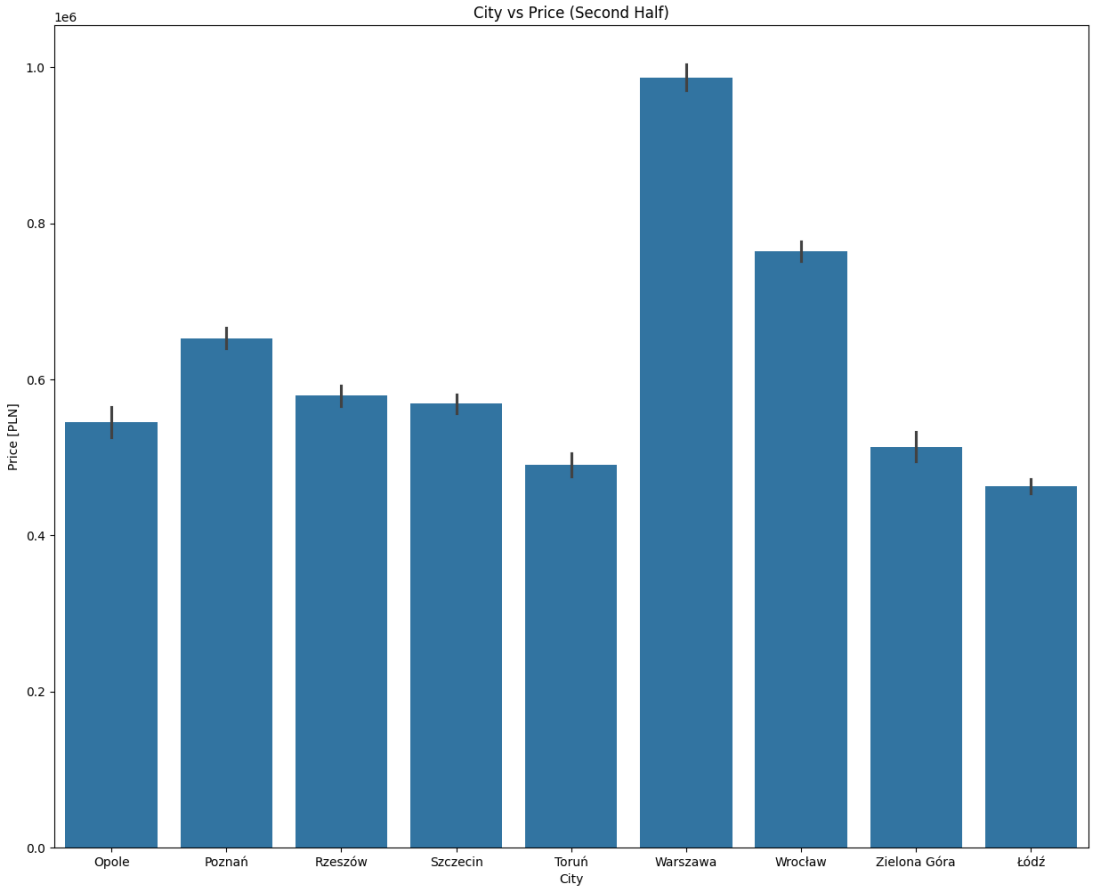
Rysunek 21: Rozkład cen mieszkań w Zielonej Górze

4.2 Analiza wpływu danych kategorycznych na cenę mieszkań

4.2.1 Zależność ceny mieszkania od miasta



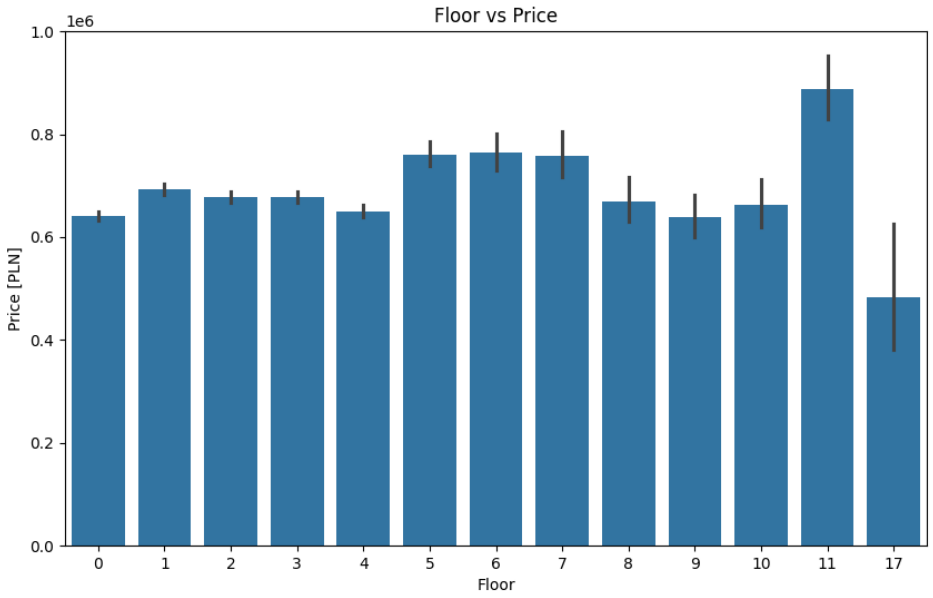
Rysunek 22: Zależność ceny mieszkania od miasta (pierwsza połowa miast)



Rysunek 23: Zależność ceny mieszkania od miasta (druga połowa miast)

Na wykresie 22 i 23 widać, że ceny mieszkań w poszczególnych miastach wojewódzkich w Polsce różnią się od siebie. Miasta takie jak Kraków, Warszawa, Gdańsk i Poznań są najdroższe pod względem nieruchomości, co może wynikać z ich znaczenia gospodarczo-kulturalnego oraz wysokiego popytu na mieszkania. Natomiast niższe ceny w miastach takich jak Bydgoszcz czy Gorzów Wielkopolski mogą świadczyć o mniejszym popycie lub innych czynnikach ekonomicznych i demograficznych.

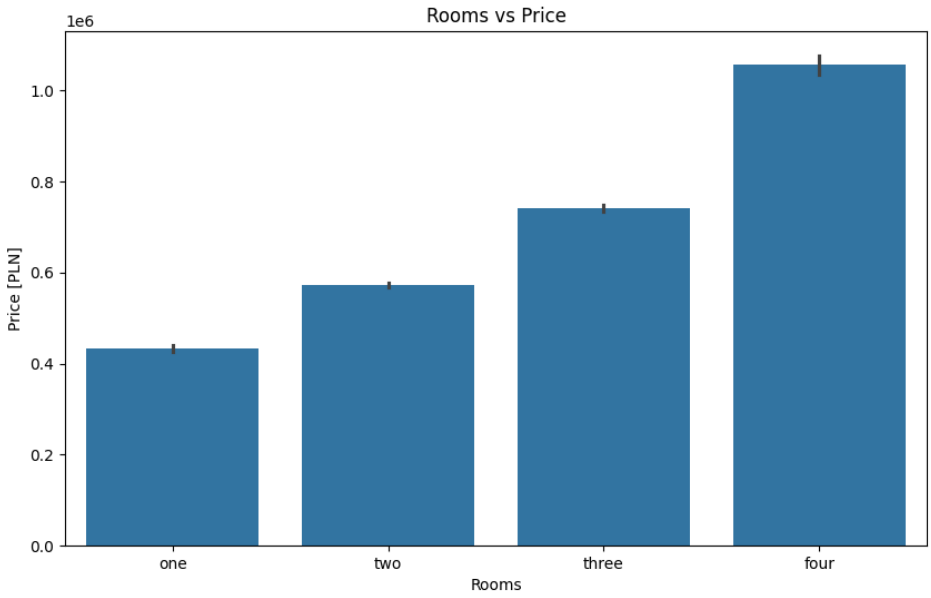
4.2.2 Zależność ceny mieszkania od piętra



Rysunek 24: Zależność ceny mieszkania od piętra

Analiza wykresu 24 wskazuje na pewne trendy w zależności cen mieszkań od piętra. Mieszkania na wyższych piętrach (szczególnie 5-7 oraz 11 piętro) charakteryzują się wyższymi cenami, natomiast skrajne piętra (parter i siedemnaste piętro) mają tendencję do niższych cen. Rozbieżności w cenach mogą wynikać z różnych czynników, takich jak widok z okna, dostęp do światła dziennego, hałas, czy atrakcyjność lokalizacji. Na wyższych piętrach znajdują się też zazwyczaj mieszkania o większym metrażu, bardziej ekskluzywne i z lepszym widokiem, co może wpływać na cenę na wyższych piętrach.

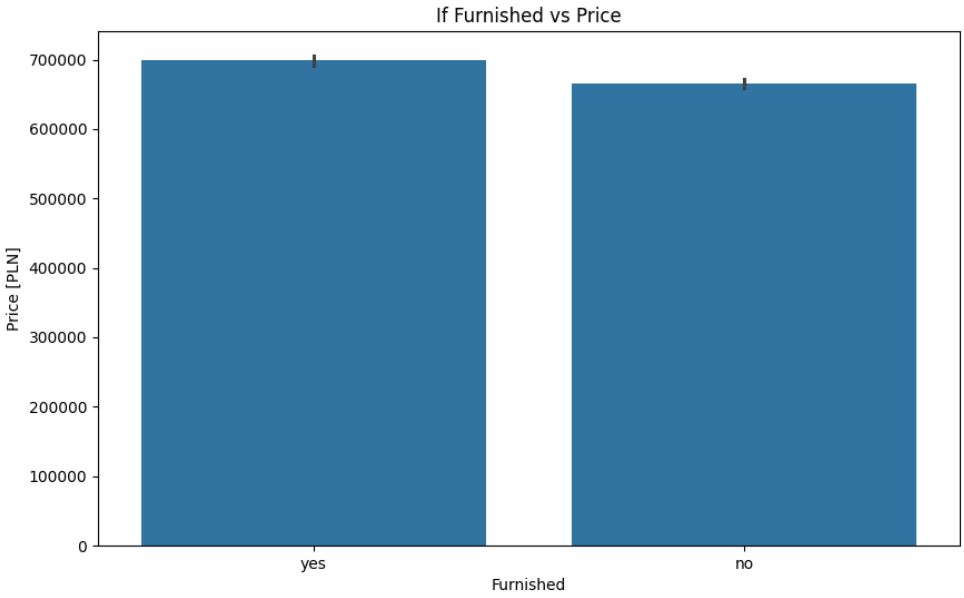
4.2.3 Zależność ceny mieszkania od liczby pokoi



Rysunek 25: Zależność ceny mieszkania od liczby pokoi

Analiza wykresu 25 przedstawiającego zależność ceny mieszkań od liczby pokoi wskazuje, że liczba pokoi ma istotny wpływ na cenę nieruchomości. Mieszkania jednopokojowe mają średnią cenę około 450,000 PLN, podczas gdy mieszkania czteropokojowe osiągają średnio około 1,000,000 PLN. Każde dodatkowe pomieszczenie znacząco zwiększa wartość mieszkania, co sugeruje, że większe mieszkania są bardziej pożądane na rynku i mogą oferować większy komfort oraz przestrzeń życiową dla mieszkańców.

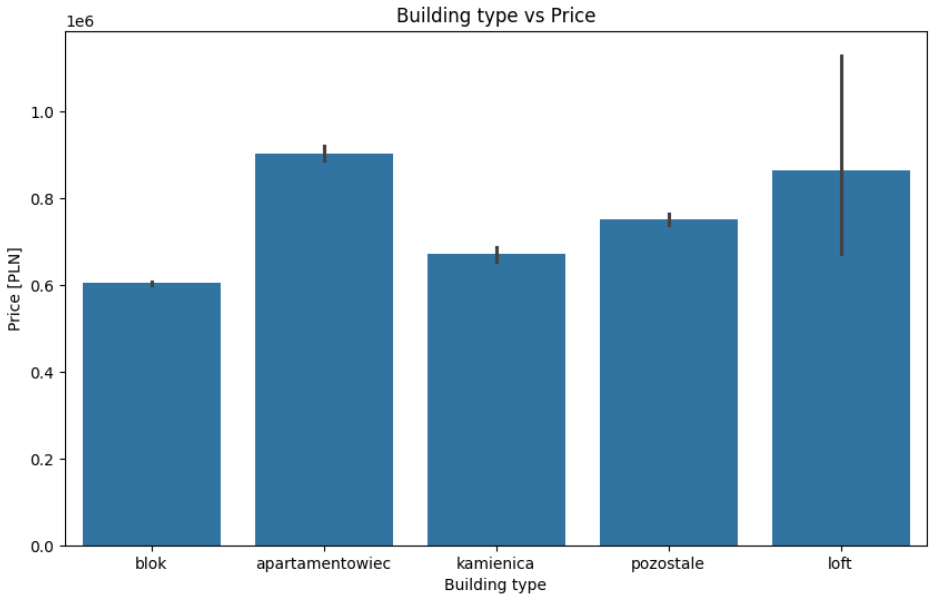
4.2.4 Zależność ceny mieszkania od umeblowania



Rysunek 26: Zależność ceny mieszkania od umeblowania

Wykres 26 pokazujący wpływ umeblowania na cenę mieszkań sugeruje, że obecność mebli nie ma znaczącego wpływu na wartość nieruchomości. Zarówno umeblowane, jak i nieumeblowane mieszkania mają średnią cenę na poziomie około 700,000 PLN. Oznacza to, że decyzja o zakupie umeblowanego mieszkania jest bardziej związana z osobistymi preferencjami nabywców niż z oczekiwaną różnicą w cenie.'

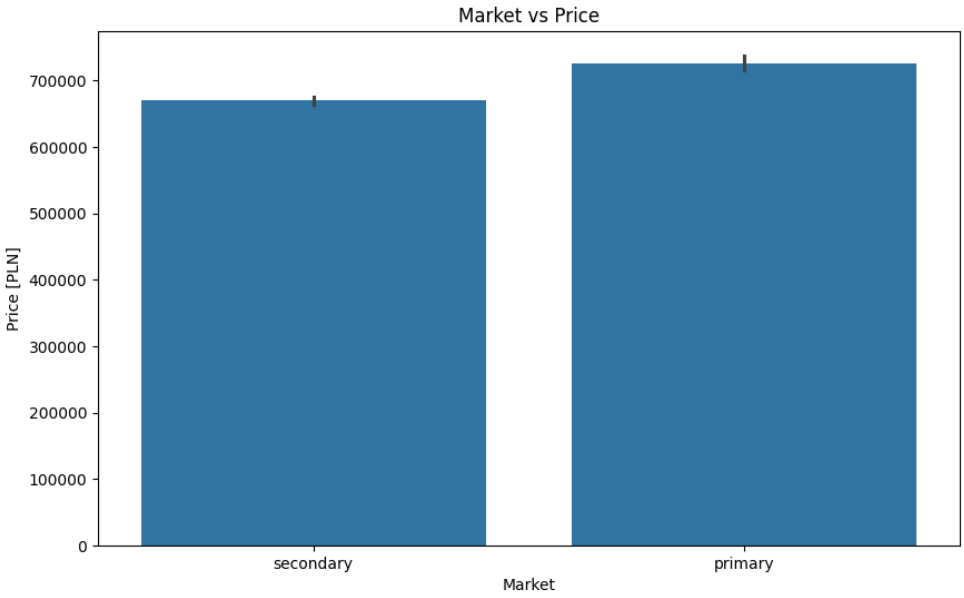
4.2.5 Zależność ceny mieszkania od typu zabudowy



Rysunek 27: Zależność ceny mieszkania od typu zabudowy

Rodzaj budynku, w którym znajduje się mieszkanie, wyraźnie wpływa na jego cenę. Mieszkania w apartamentowcach mają najwyższą średnią cenę około 900,000 PLN, natomiast mieszkania w blokach są najtańsze, z ceną około 600,000 PLN. Lofty, pomimo wysokiej średniej ceny około 850,000 PLN, charakteryzują się dużą zmiennością cen - najprawdopodobniej wynika to z tego, że w dużej części loftów znajdują się lukusowe mieszkania, które wpływają na zawyżanie cen. Mieszkania w kamienicach i innych typach budynków mają średnią cenę około 700,000 PLN. Wybór typu budynku wpływa więc znacząco na wartość nieruchomości, z apartamentowcami i loftami na czele jako najdroższymi opcjami.

4.2.6 Zależność ceny mieszkania od rynku

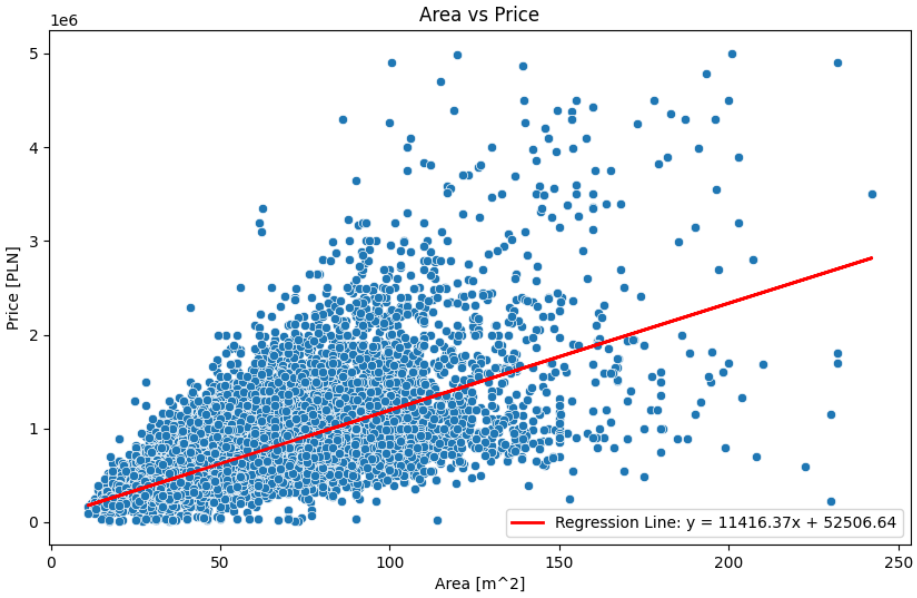


Rysunek 28: Zależność ceny mieszkania od rynku

Analiza wykresu 28 pokazuje minimalne różnice między cenami mieszkań na rynku pierwotnym i wtórnym. Średnia cena mieszkań na rynku wtórnym wynosi około 700,000 PLN, podobnie jak na rynku pierwotnym. Stabilność cen w obu segmentach sugeruje, że nabywcy nie powinni oczekiwać znaczących różnic w wartości nieruchomości w zależności od tego, czy kupują mieszkanie nowe, czy z drugiej ręki. Ostateczna decyzja może zatem opierać się na preferencjach dotyczących stanu technicznego mieszkania oraz dodatkowych korzyści oferowanych przez deweloperów na rynku pierwotnym.

4.3 Analiza wpływu danych numerycznych na cenę mieszkań

4.3.1 Zależność ceny mieszkania od metrażu



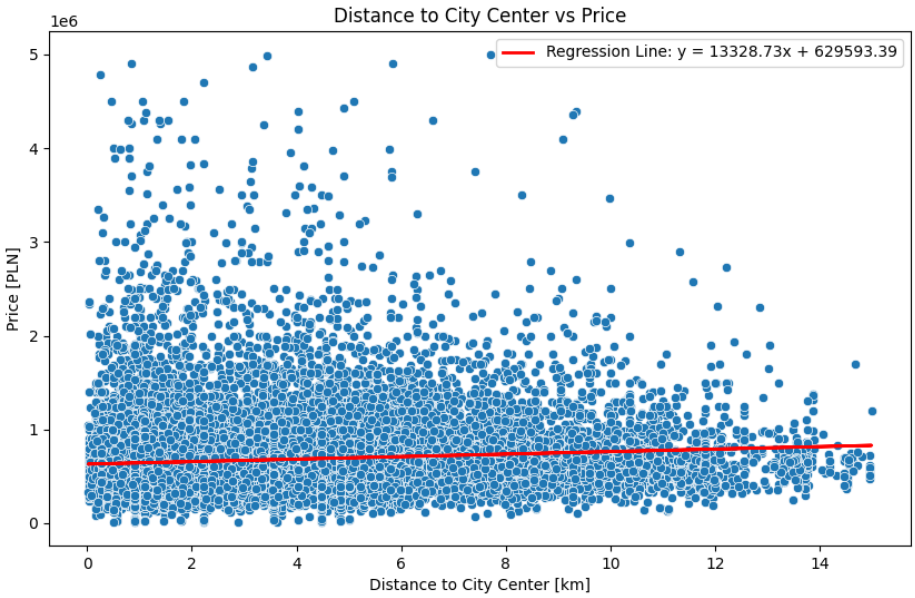
Rysunek 29: Zależność ceny mieszkania od metrażu

Analiza wykresu przedstawiającego zależność ceny mieszkań od ich powierzchni pokazuje wyraźną korelację między tymi dwoma zmiennymi. Wykres przedstawia linię regresji o równaniu:

$$y = 11416.37x + 52506.64, \tag{1}$$

co sugeruje, że każdy dodatkowy metr kwadratowy powierzchni mieszkania zwiększa jego cenę średnio o 11,416.37 PLN. Jest to istotny wskaźnik, który pokazuje, że większa powierzchnia mieszkania bezpośrednio przekłada się na wyższą cenę, co jest zgodne z oczekiwaniami rynkowymi.

4.3.2 Zależność ceny mieszkania od odległości do centrum miasta



Rysunek 30: Zależność ceny mieszkania od odległości do centrum miasta

Wykres przedstawiający zależność ceny mieszkań od odległości do centrum miasta pokazuje, że ta zależność jest znacznie słabsza w porównaniu do powierzchni mieszkania. Linia regresji o równaniu:

$$y = 13328.73x + 629593.39, \tag{2}$$

sugeruje, że z każdym dodatkowym kilometrem odległości od centrum cena mieszkania wzrasta średnio o 13,328.73 PLN. Jednakże, rozkład punktów na wykresie pokazuje dużą zmienność cen na różnych odległościach, co wskazuje, że cena mieszkania nie jest silnie uzależniona od odległości od centrum. Inne czynniki, takie jak dostęp do infrastruktury, transportu publicznego oraz atrakcyjność dzielnicy, mogą mieć większy wpływ na cenę.

5 Modelowanie

5.1 Domyślne parametry do modelowania

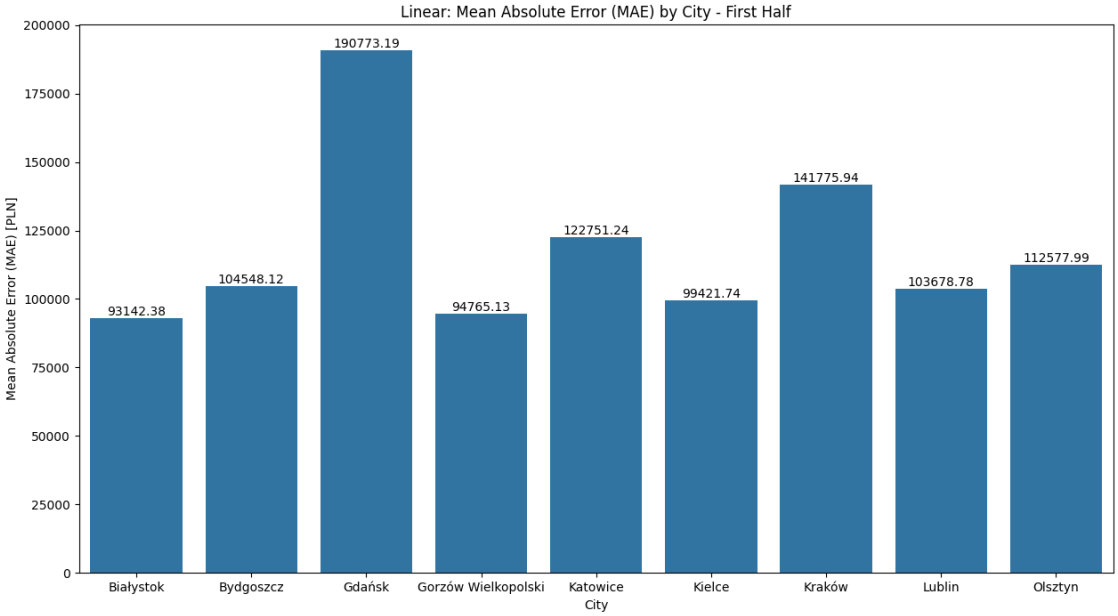
- Miasto: Katowice
- Piętro: 2
- Umeblowanie: 'no'
- Rynek: 'secondary'
- Typ zabudowy: 'blok'
- Powierzchnia: średnia powierzchnia z danych
- Liczba pokoi: 'two'
- Odległość do centrum: 5.0

5.2 Dobranie modelu

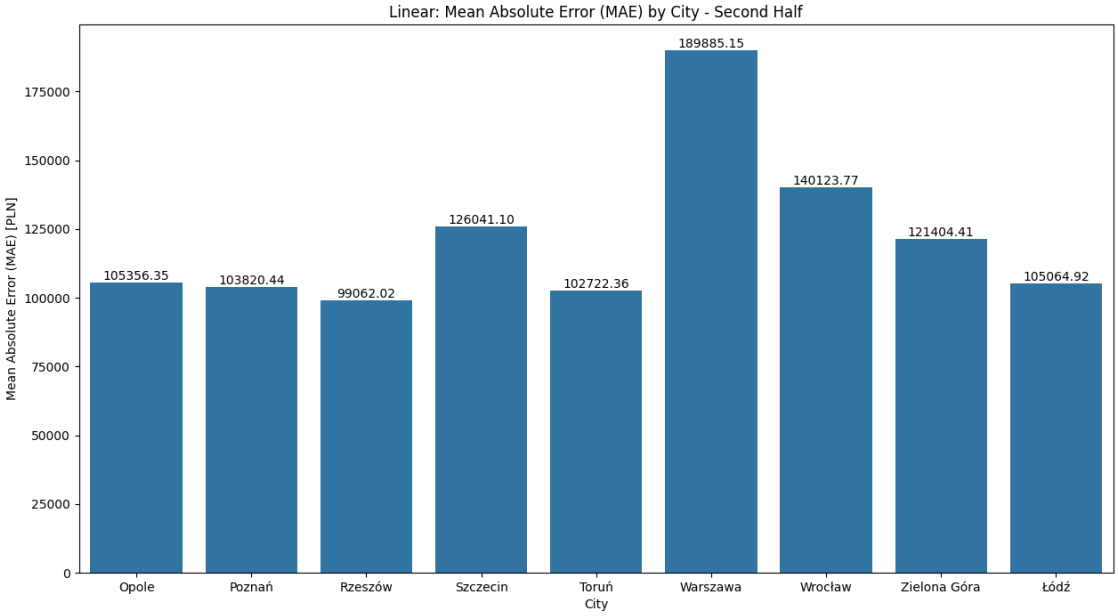
Pierwotnie do przewidywania cen mieszkań w miastach wojewódzkich w Polsce został wybrany model regresji liniowej. Jednakże, po przeprowadzeniu analizy danych, okazało się, że zależności między zmiennymi kategorycznymi a ceną mieszkań są bardziej złożone. Dlatego też, zastosowano model oparty na drzewach decyzyjnych, który jest bardziej elastyczny i potrafi uwzględnić nieliniowe zależności między zmiennymi.

Prównanie modeli regresji liniowej i drzewa decyzyjnego zostało przeprowadzone na podstawie metryk Mean Absolute Error (MAE) oraz na porównaniu wykresów predykcji cen mieszkań dla obu modeli.

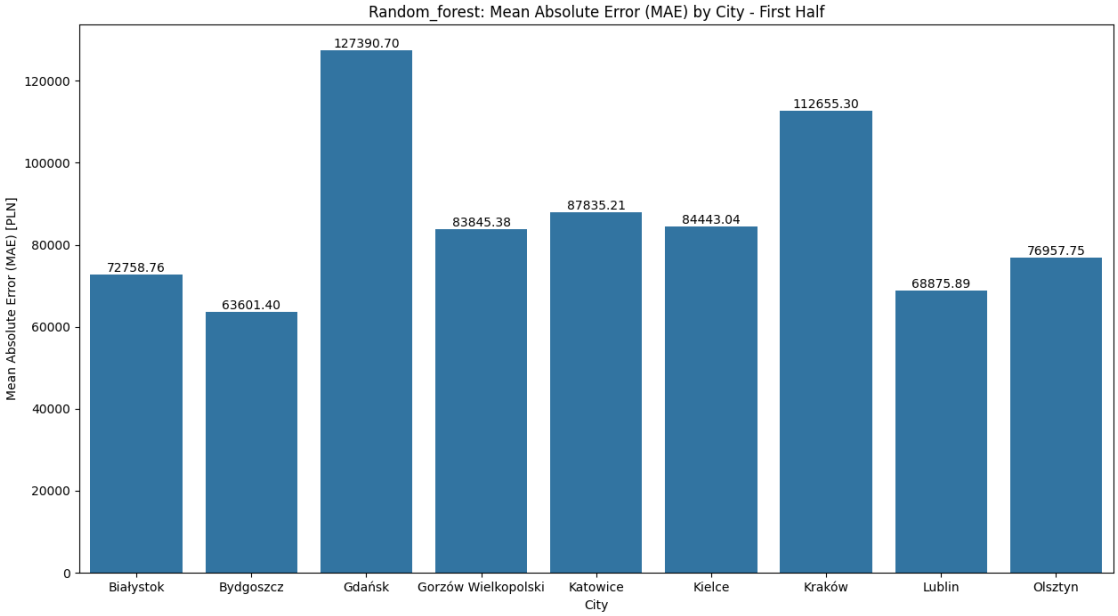
Średni Błąd Absolutny (MAE) został wybrany jako miara błędu ze względu na jego łatwą interpretowalność oraz odporność na wartości odstające, co jest istotne w analizie cen nieruchomości. MAE informuje o przeciętnej wielkości błędu, wyrażonej w złotówkach, co czyni ją intuicyjnie zrozumiałą. Dodatkowo, traktując wszystkie błędy z równą wagą, MAE zapewnia bezstronną ocenę modelu. Prosta metoda obliczania i interpretacji MAE czyni ją odpowiednią miarą dla owego modelu predykcyjnego.



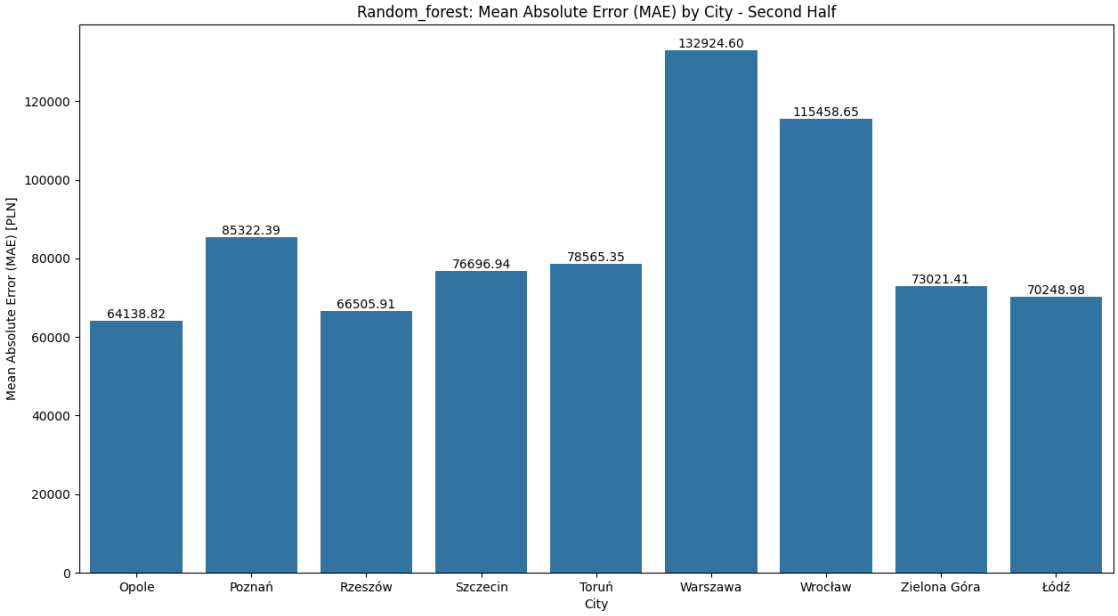
Rysunek 31: Średni Błąd Absolutny dla modelu regresji liniowej (pierwsza połowa miast)



Rysunek 32: Średni Błąd Absolutny dla modelu regresji liniowej (druga połowa miast)

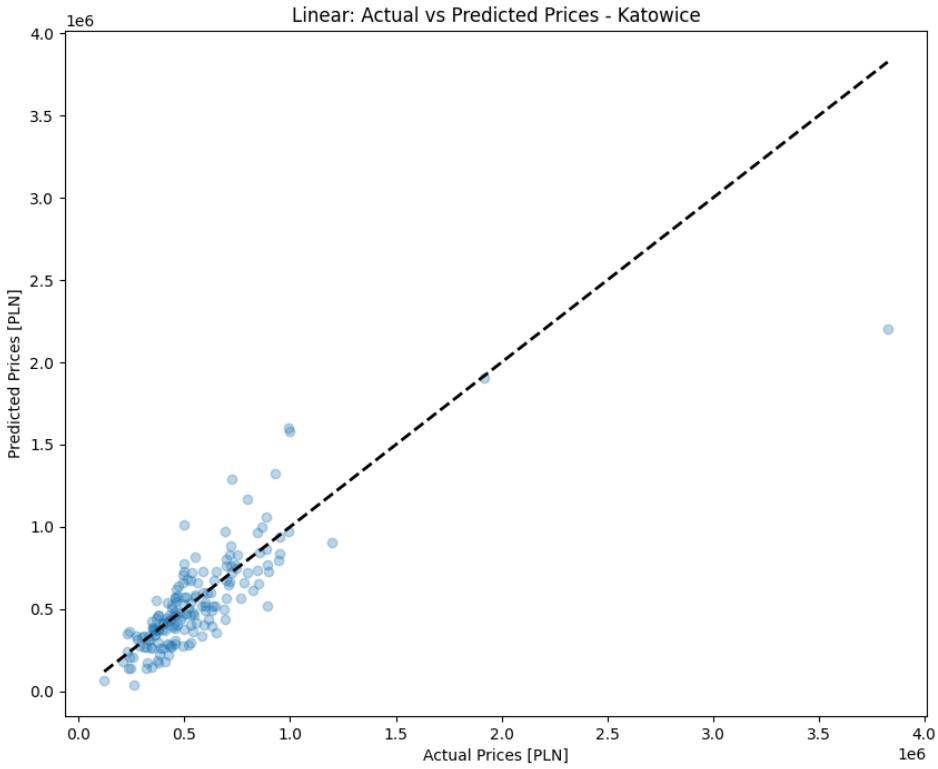


Rysunek 33: Średni Błąd Absolutny dla modelu drzewa decyzyjnego (pierwsza połowa miast)

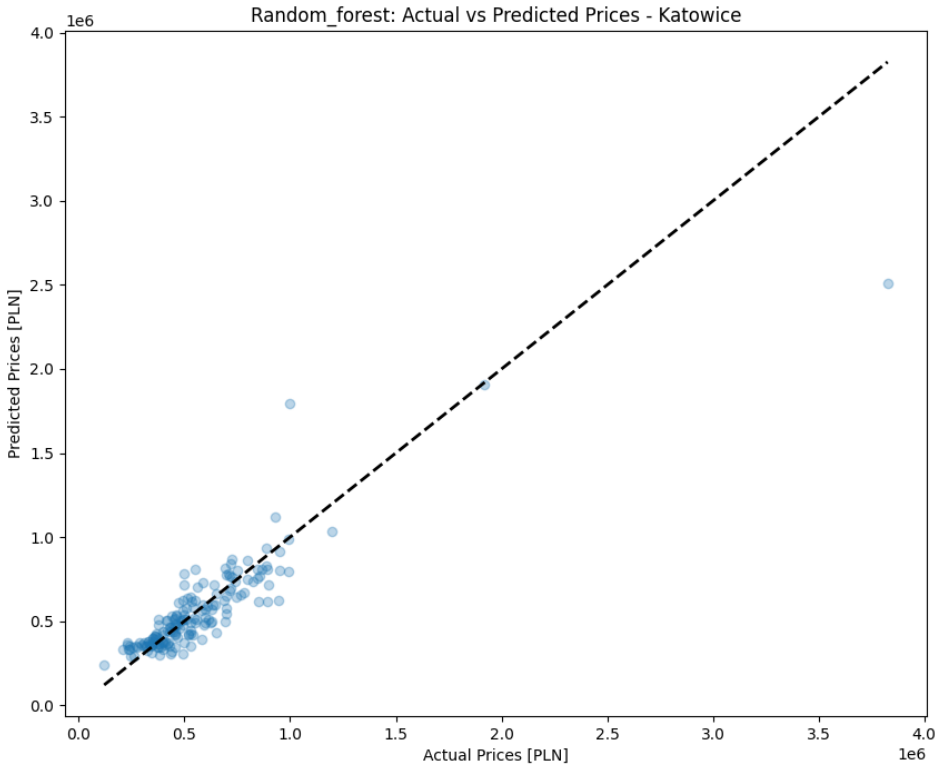


Rysunek 34: Średni Błąd Absolutny dla modelu drzewa decyzyjnego (druga połowa miast)

Jak możemy zauważyć na wykresach 31 i 32, model regresji liniowej osiąga średni błąd absolutny na poziomie około 80,000 PLN - 200,000 PLN, co jest dość wysoką wartością, przez co model nie będzie zbyt użyteczny w praktyce. W przypadku modelu drzewa decyzyjnego 33 i 34, średni błąd absolutny wynosi około 50,000 PLN - 135,000 PLN, co jest znacznie niższą wartością niż dla modelu regresji liniowej.



Rysunek 35: Rzeczywista cena mieszkania vs. przewidywana cena mieszkania dla modelu regresji liniowej w Katowicach



Rysunek 36: Rzeczywista cena mieszkania vs. przewidywana cena mieszkania dla modelu drzewa decyzyjnego w Katowicach

Na wykresach 35 i 36 przedstawiających rzeczywistą cenę mieszkania vs. przewidywaną cenę mieszkania dla modeli regresji liniowej i drzewa decyzyjnego w Katowicach. Jak możemy zauważyć, model drzewa decyzyjnego lepiej przewiduje ceny mieszkań, co potwierdzają wartości na wykresie 36, bliższe prostej o równaniu:

$$y = x, \tag{3}$$

niż wartości dla modelu regresji liniowej.

5.3 Dobranie hiperparametrów

W celu optymalizacji modelu opartego na losowym lesie (Random Forest), przeprowadzona została selekcja hiperparametrów przy użyciu metody Grid Search. Grid Search umożliwia przeszukiwanie przestrzeni hiperparametrów, testując różne kombinacje wartości w celu znalezienia tych, które dają najlepsze wyniki. Grid Search wykorzyztało następujące hiperparametry do optymalizacji modelu:

- **n_estimators** - liczba drzew w lesie. Testowane wartości to 100 i 200.
- **max_depth** - maksymalna głębokość drzewa. Testowane wartości to 10, 20 oraz brak ograniczenia (None).
- **min_samples_split** - minimalna liczba próbek wymagana do podziału węzła. Testowane wartości to 2, 5 oraz 10.
- **min_samples_leaf** - minimalna liczba próbek wymagana do utworzenia liścia. Testowane wartości to 1, 2 oraz 4.

Procedura Grid Search została przeprowadzona z wykorzystaniem kroswalidacji (cross-validation) z trzema (cv=3) podziałami danych. Wykorzystanie kroswalidacji pozwala na ocenę wydajności modelu na różnych podzbiorach danych, co zwiększa wiarygodność wyników i pomaga uniknąć przeuczenia modelu (overfitting).

5.4 Testowanie modelu

5.4.1 Strategia podziału danych

Pierwszym krokiem w testowaniu modelu był podział zbioru danych na zestawy treningowe i testowe. Biorąc pod uwagę geograficzną różnorodność w zbiorze danych, zapewniliśmy, że każde miasto było odpowiednio reprezentowane w obu zestawach. Początkowo przeznaczono 5% danych na testowanie; jednak obserwowany błąd średniej bezwzględnej (MAE) był zbyt wysoki. W związku z tym rozmiar zestawu testowego został zwiększony do 15%, co znacząco poprawiło wyniki modelu.

5.4.2 Optymalizacja hiperparametrów

W celu zoptymalizowania wydajności modelu, przeprowadzono optymalizację hiperparametrów za pomocą metody Grid Search. Grid Search został przetestowany na 3 różnych zestawach danych, aby zapewnić, że wybrane hiperparametry są optymalne dla wszystkich miast. Ostatecznie wybrano zestaw hiperparametrów 5.3, który zapewniał najniższy błąd średniej bezwzględnej dla większości miast.

Zestaw 1:

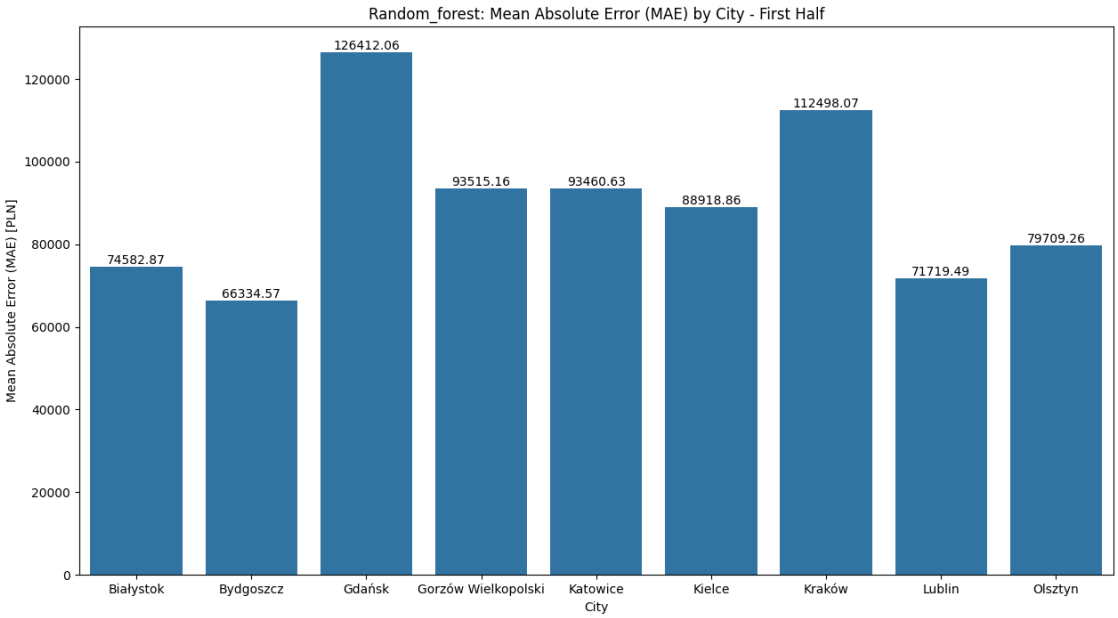
- **n_estimators** - 50
- **max_depth** - 5, 15
- **min_samples_split** - 5, 15
- **min_samples_leaf** - 2, 6

Zestaw 2:

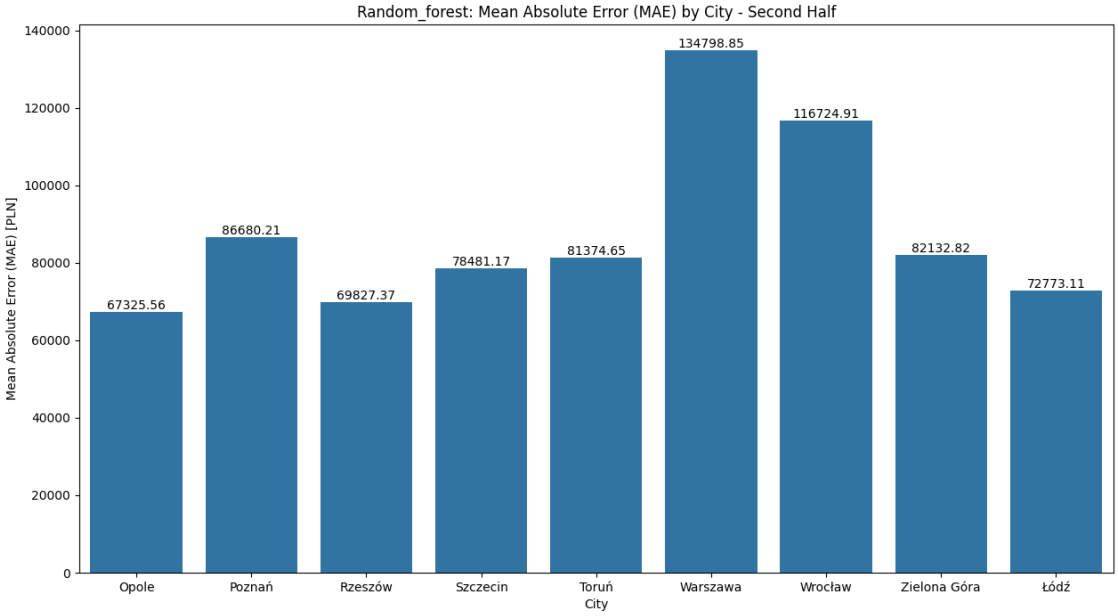
- **n_estimators** - 100
- **max_depth** - 10, 20
- **min_samples_split** - 3, 8
- **min_samples_leaf** - 1, 3

Zestaw 3: Domyślny zestaw hiperparametrów przyjęty w 5.3.

Wyniki zestawu 1:

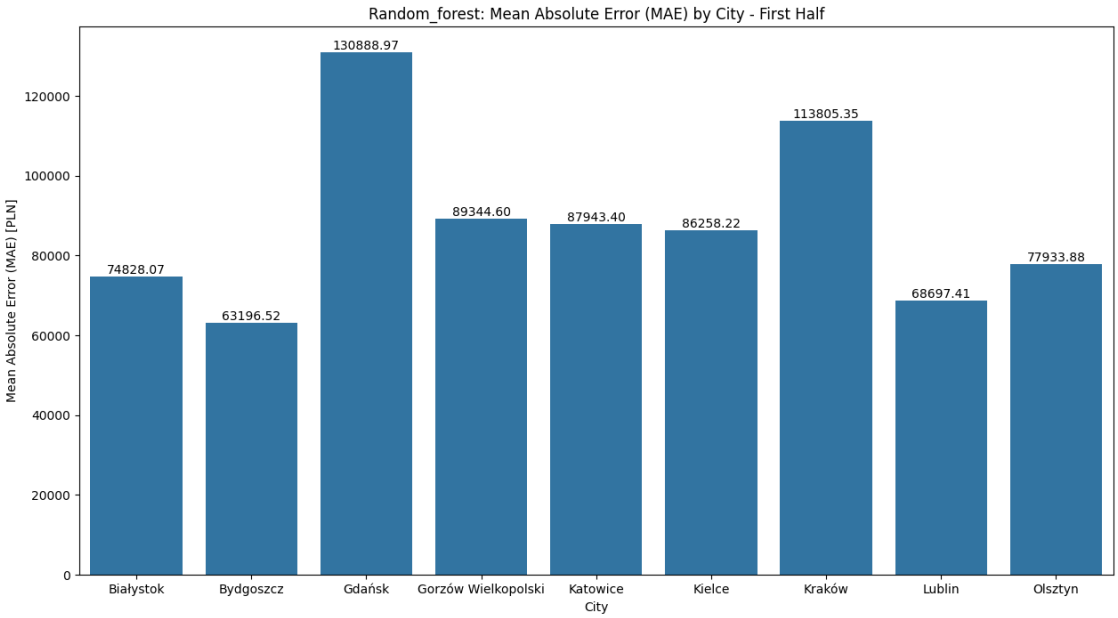


Rysunek 37: Średni Błąd Absolutny dla modelu drzewa decyzyjnego dla pierwszego zestawu parametrów (pierwsza połowa miast)

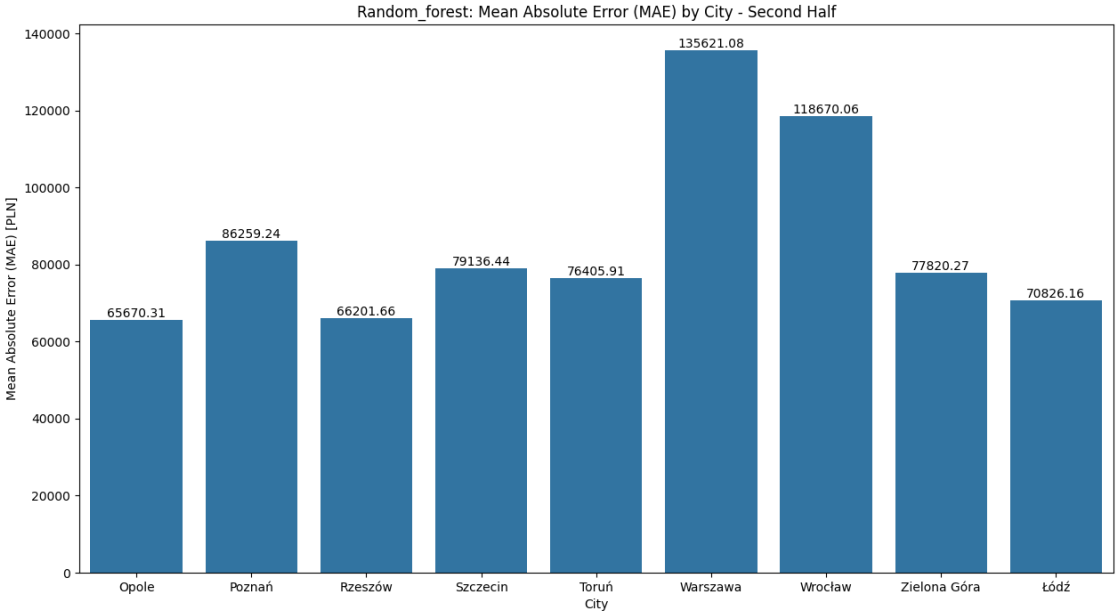


Rysunek 38: Średni Błąd Absolutny dla modelu drzewa decyzyjnego dla pierwszego zestawu parametrów (druga połowa miast)

Wyniki zestawu 2:

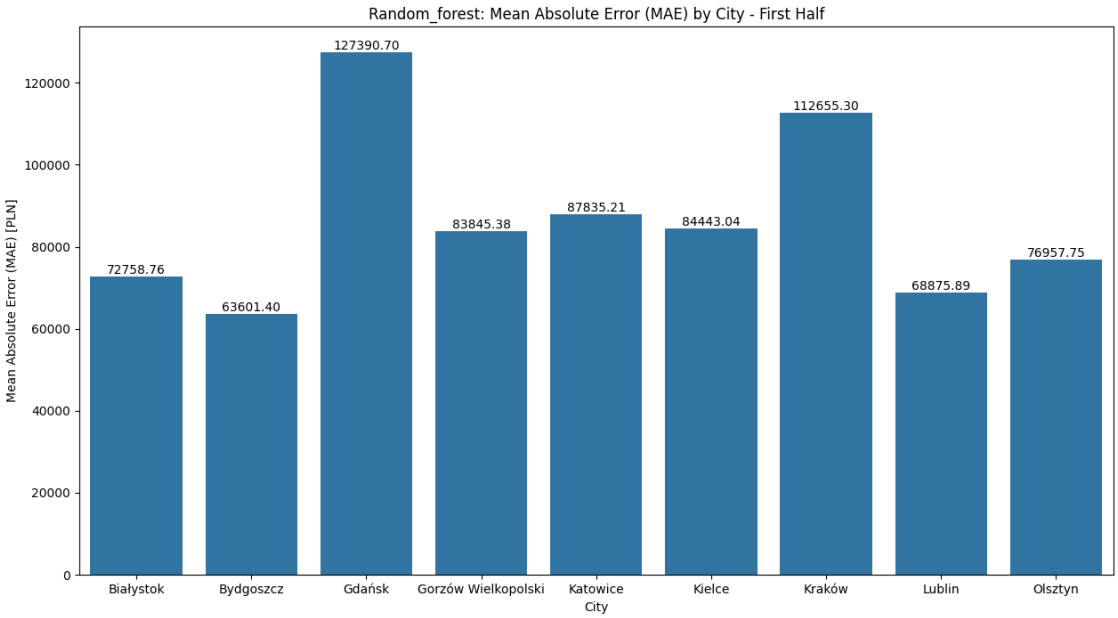


Rysunek 39: Średni Błąd Absolutny dla modelu drzewa decyzyjnego dla drugiego zestawu parametrów (pierwsza połowa miast)

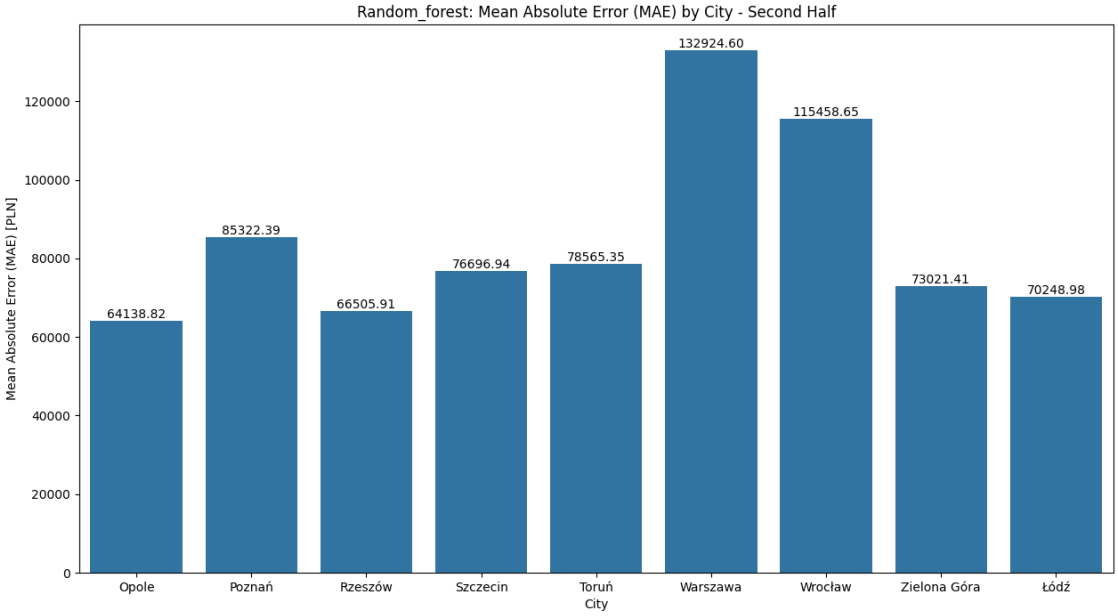


Rysunek 40: Średni Błąd Absolutny dla modelu drzewa decyzyjnego dla drugiego zestawu parametrów (druga połowa miast)

Wyniki zestawu 3:



Rysunek 41: Średni Błąd Absolutny dla modelu drzewa decyzyjnego dla trzeciego zestawu parametrów (pierwsza połowa miast)

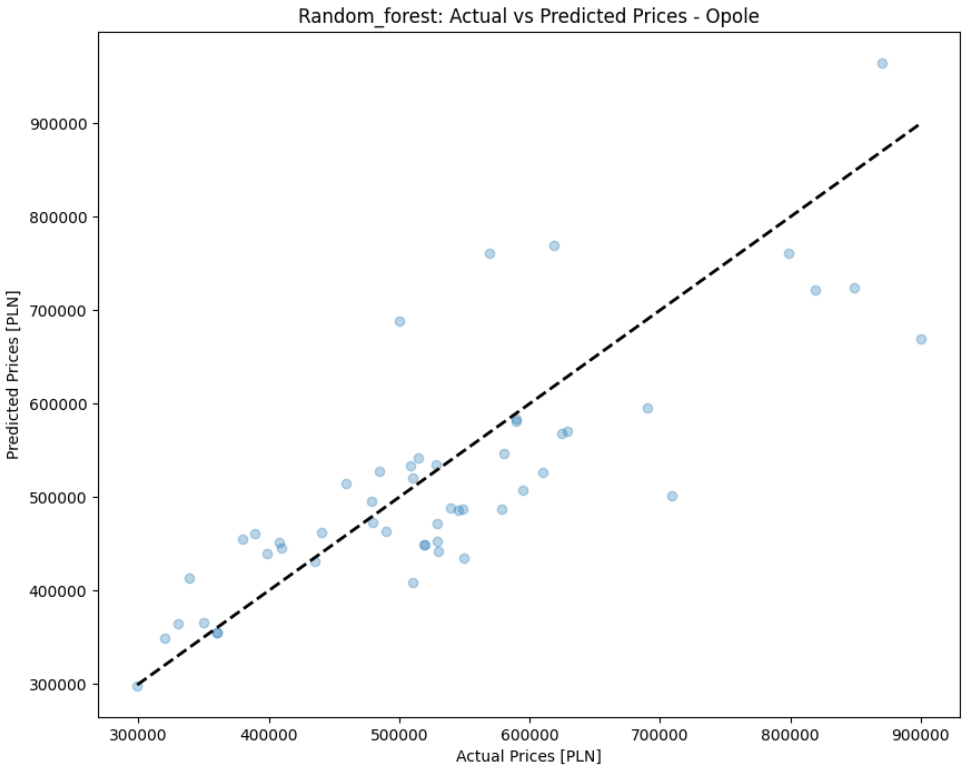


Rysunek 42: Średni Błąd Absolutny dla modelu drzewa decyzyjnego dla trzeciego zestawu parametrów (druga połowa miast)

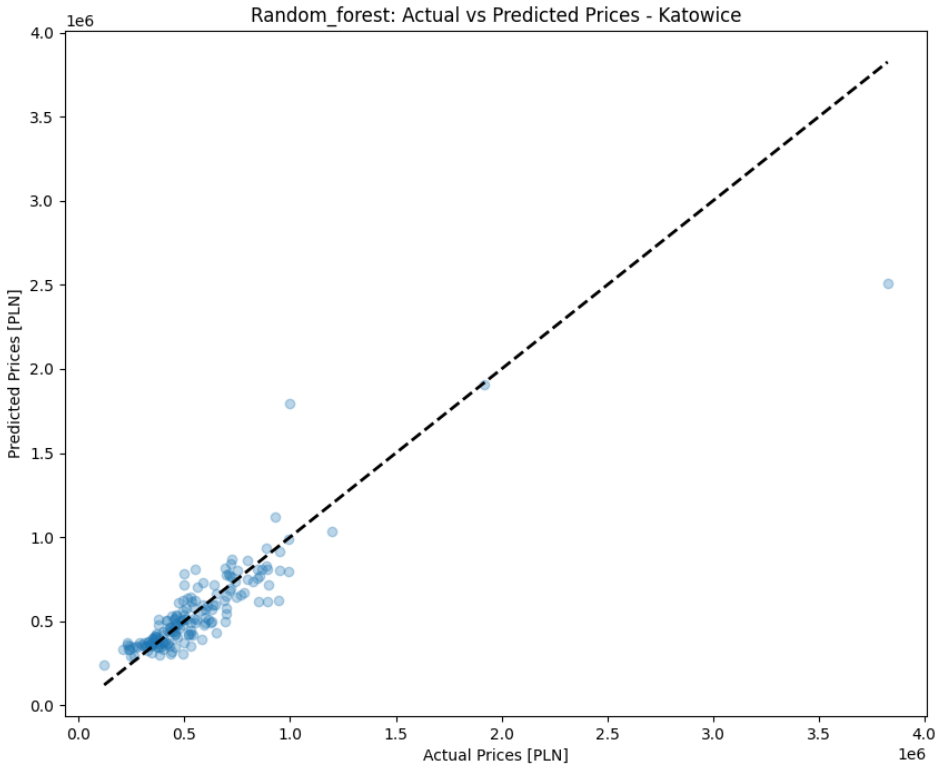
5.4.3 Wyniki modelu dla różnych miast

Poniżej przedstawiono wyniki modelu dla 3 z 18 miast wojewódzkich w Polsce. Miasta zostały wybrane pod kątem dostępności mieszkań na sprzedaż oraz różnorodności cen mieszkań w celu zapewnienia reprezentatywności wyników.

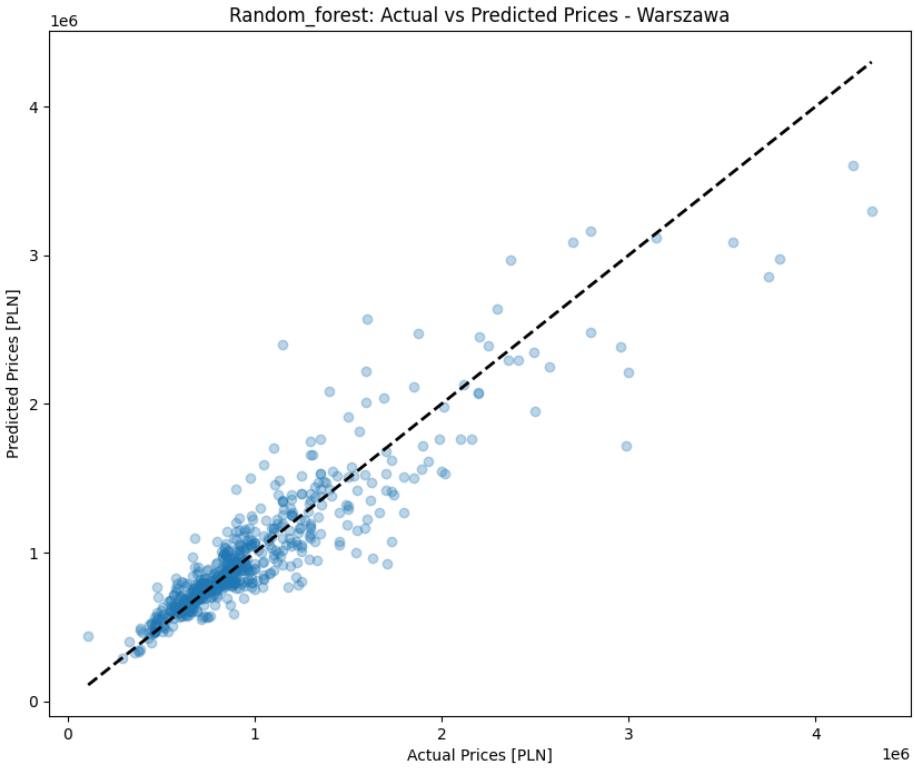
Rzeczywista cena mieszkania vs. przewidywana cena mieszkania



Rysunek 43: Rzeczywista cena mieszkania vs. przewidywana cena mieszkania dla modelu drzewa decyzyjnego w Opolu

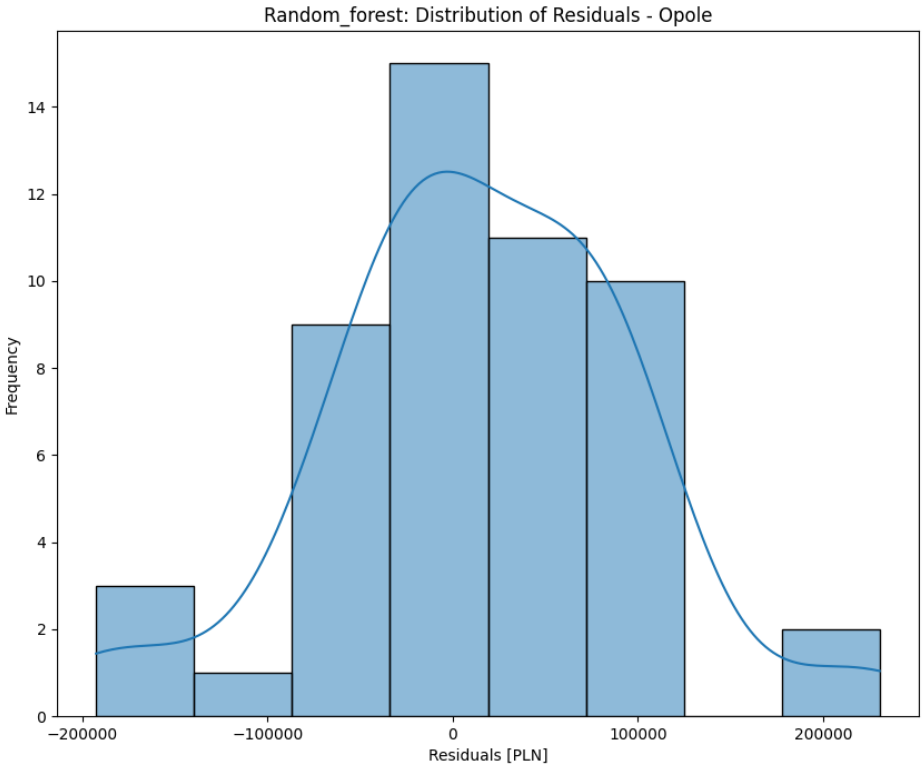


Rysunek 44: Rzeczywista cena mieszkania vs. przewidywana cena mieszkania dla modelu drzewa decyzyjnego w Katowicach

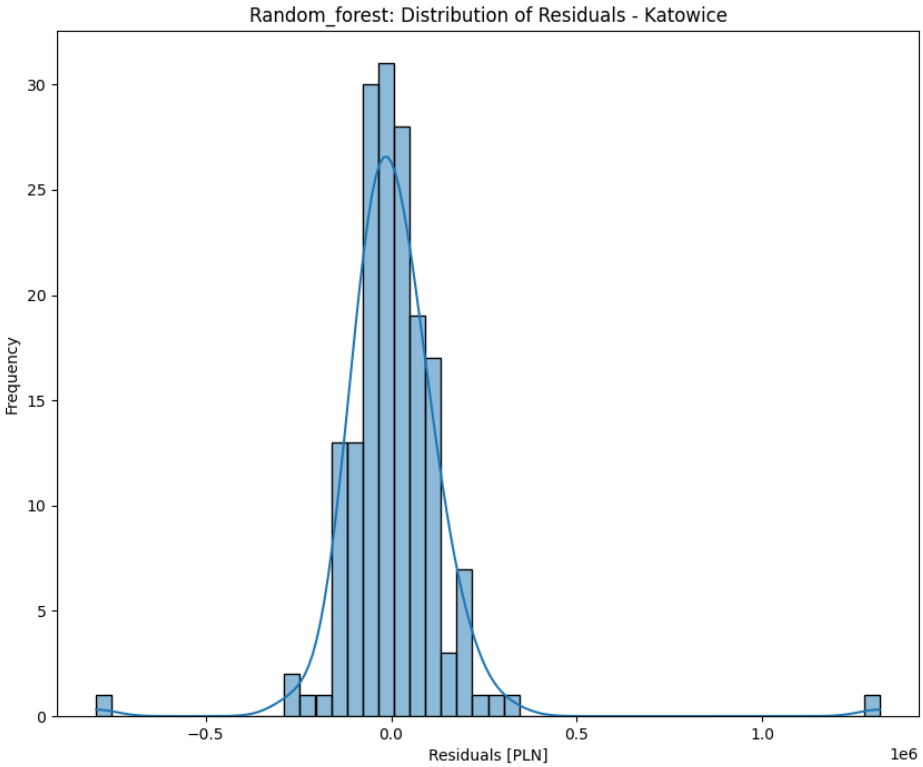


Rysunek 45: Rzeczywista cena mieszkania vs. przewidywana cena mieszka w Warszawie dla modelu drzewa decyzyjnego

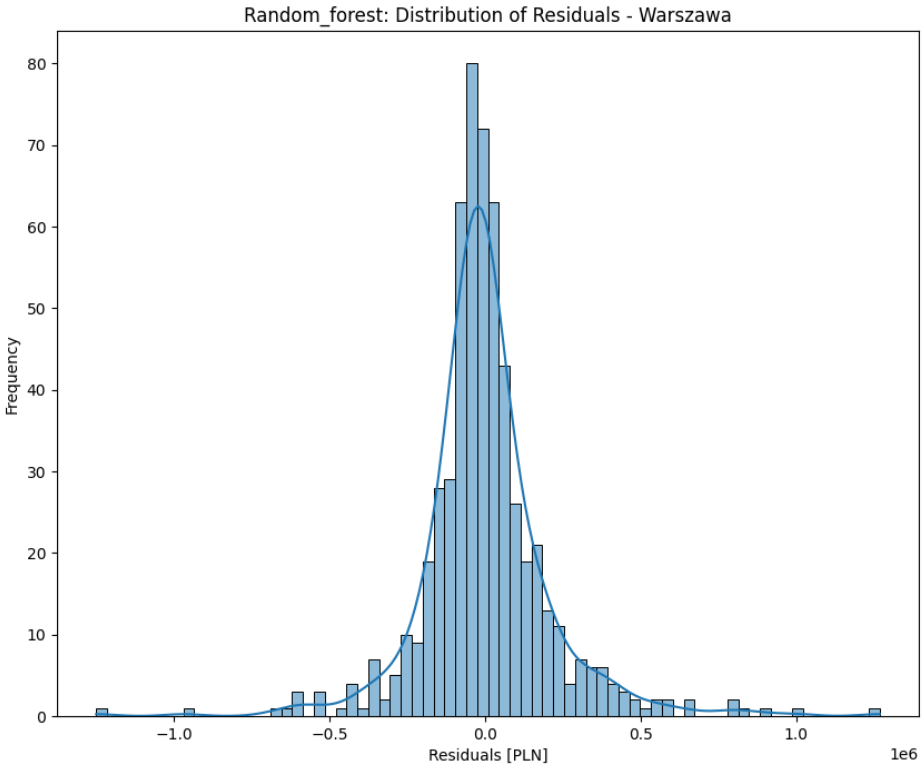
Reszty - różnica między rzeczywistą ceną mieszkania a przewidywaną ceną mieszkania



Rysunek 46: Reszty dla modelu drzewa decyzyjnego w Opolu



Rysunek 47: Reszty dla modelu drzewa decyzyjnego w Katowicach



Rysunek 48: Reszty dla modelu drzewa decyzyjnego w Warszawie

Jak możemy zauważyć na wykresach rzeczywistej ceny mieszkania vs. przewidywanej ceny mieszkania, model drzewa decyzyjnego dobrze radzi sobie z przewidywaniem cen mieszkań, jeśli posiada wystarczająco duże dane treningowe. Reszty dla modelu drzewa decyzyjnego są rozłożone równomiernie wokół zera, co sugeruje, że model jest dobrze dopasowany do danych.

6 Wnioski

Model drzewa decyzyjnego pokazał, że jest w stanie dobrze przewidywać ceny mieszkań w różnych miastach wojewódzkich, pod warunkiem, że dostępne są wystarczająco duże dane treningowe. W miastach takich jak Warszawa, Kraków czy Gdańsk, gdzie popyt na mieszkania jest wysoki, model dokładnie przewidywał ceny, co jest kluczowe dla inwestorów i potencjalnych nabywców. W miastach o mniejszym popycie, takich jak Bydgoszcz czy Gorzów Wielkopolski, model również radził sobie dobrze, choć dokładność przewidywań była nieco niższa ze względu na większą zmienność cen.

Analiza reszt wykazała, że model drzewa decyzyjnego generował reszty rozłożone równomiernie wokół zera, co sugeruje, że model jest dobrze dopasowany do danych. Równomierne rozłożenie reszt wskazuje na brak systematycznych błędów w przewidywaniach modelu, co jest ważne dla jego wiarygodności i stabilności.

Dokładność przewidywań modelu była silnie zależna od jakości i ilości dostępnych danych. Precyzyjne dane wejściowe, odpowiednio przetworzone i oczyszczone, pozwoliły na uzyskanie dokładniejszych wyników. W przyszłości dalsze wzbogacenie zbioru danych o dodatkowe zmienne oraz aktualizacja danych mogą dodatkowo poprawić skuteczność modelu.