

POLITECHNIKA WARSZAWSKA

Wydział Elektroniki i Technik Informacyjnych

Instytut Automatyki i Informatyki Stosowanej

Praca dyplomowa magisterska

Konstrukcja i weryfikacja wskaźnika atrakcyjności inwestycji mieszkaniowych

Autor: inż. Patryk Czecko

Opiekun pracy: dr inż. Mariusz Kamola

Streszczenie

Konstrukcja i weryfikacja wskaźnika atrakcyjności inwestycji mieszkaniowych

Celem niniejszej pracy magisterskiej było zapoznanie się z rynkiem mieszkań ze zwróceniem szczególnej uwagi na m. st. Warszawę, przegląd istniejącej literatury i rozwiązań dotyczących modelowania cen nieruchomości, określenie i analiza cech wpływających na atrakcyjność mieszkań, która ma bezpośrednie odzwierciedlenie w ich cenie, a następnie stworzenie i weryfikacja modelu matematycznego zależności cen mieszkań od najistotniejszych parametrów je opisujących.

Podczas realizacji pracy udało się zgromadzić, przetworzyć i przeanalizować dane transakcyjne z ostatnich 2 lat dla ponad 7 tysięcy nieruchomości znajdujących się w 5 dzielnicach Warszawy. Za pomocą regresji hedonicznej wyznaczono model cen nieruchomości uwzględniający wybrane cechy mieszkań, takie jak metraż, kondygnacja oraz aspekty związane z ich lokalizacją: odległość od centrum miasta oraz położenie względem istotnych punktów użyteczności publicznej.

Słowa kluczowe: modelowanie, ceny, nieruchomości, regresja hedoniczna

Abstract

Construction and verification of real estate attractiveness index

The purpose of the thesis was to introduce the real estate market and price modeling topic, review related literature and solutions, define and analyze attributes determining apartments' attractiveness which is reflected in their prices and to form and verify a property pricing model that uses mentioned characteristics of houses.

Author succeeded to process and analyze data regarding over 7 thousands transactions that had taken place around 5 districts of Capital City of Warsaw during past 2 years. Using hedonic regression a pricing model has been prepared. Following attributes of properties and their location have been used: flat area, floor where the apartment is located, distance from city center and location related to points of interest.

Keywords: modelling, real estate, pricing, hedonic regression

1.	Wstęp	5
1.1.	Rynek mieszkań w Polsce	5
1.2.	Sformułowanie problemu.....	7
1.2.1.	Zagadnienie wyceny nieruchomości	7
1.2.2.	Czynniki wpływające na wartość nieruchomości	8
1.3.	Cel pracy	9
1.3.1.	Zakres pracy.....	10
2.	Przegląd obecnego stanu wiedzy.....	11
2.1.	Istniejące opracowania i literatura.....	11
2.1.1.	Publikacje dotyczące rynków zagranicznych.....	11
2.1.2.	Publikacje dotyczące rynku polskiego.....	12
2.2.	Rozwiązania automatyzujące proces oceny wartości nieruchomości.....	13
2.2.1.	Walk Score.....	14
2.2.2.	Szybko.pl.....	16
2.3.	Podsumowanie przedstawionych opracowań i rozwiązań.....	17
2.3.1.	Wskazanie nieopracowanej strony zagadnienia	18
2.3.2.	Doprecyzowanie celu pracy	19
3.	Analiza zagadnienia.....	20
3.1.	Szczegółowy opis sposobu modelowania i wyznaczania cen mieszkań.....	20
3.1.1.	Regresja hedoniczna.....	20
3.1.2.	AVM – automatyczny model wyceny.....	22
3.2.	Określenie składowych atrakcyjności mieszkania	23
3.2.1.	Parametryzacja opisu nieruchomości	24
3.2.2.	Lokalizacja nieruchomości.....	26
3.2.3.	Ogólny model cen mieszkań.....	27
3.2.4.	Założenie niezależności składowych	28
3.3.	Wykorzystane źródła danych.....	29
3.3.1.	Rejestr Cen i Wartości Nieruchomości.....	29
3.3.2.	Mapy Google	31
3.3.3.	Punkty użyteczności publicznej.....	33
3.3.4.	Czynniki opisujące mieszkania a dostępne dane	35
3.4.	Część techniczna analizy zagadnienia.....	35
3.4.1.	Pozyskiwanie danych.....	36

3.4.2.	Agregacja danych	37
3.5.	Sformułowanie hipotez dotyczących zależności ceny mieszkań od czynników	38
3.5.1.	Parametry mieszkań	38
3.5.2.	Cechy lokalizacji.....	41
3.6.	Weryfikacja postawionych hipotez	44
3.6.1.	Filtrowanie danych	45
3.6.2.	Zależność od numeru piętra.....	45
3.6.3.	Zależność od metrażu.....	47
3.6.4.	Zależność od piętra i metrażu	48
3.6.5.	Zależność od odległości od centrum	49
3.6.6.	Zależność od odległości od POI	52
3.6.7.	Regresja dla wszystkich mieszkań	53
3.6.8.	Próba korekty modelu.....	55
3.7.	Wnioski z analizy.....	57
4.	Podsumowanie.....	58
4.1.	Podsumowanie przeprowadzonych prac	58
4.2.	Konkluzje.....	58
4.3.	Możliwości dalszej analizy tematu	59

1. Wstęp

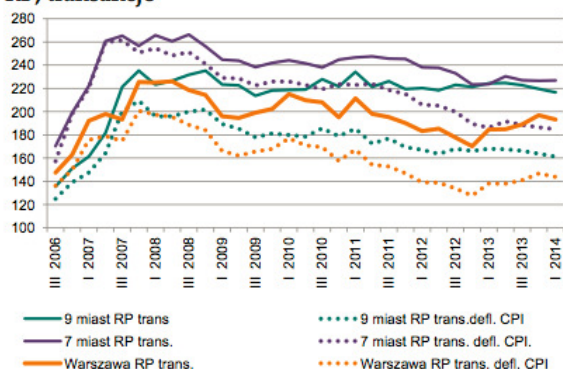
Niniejsza praca dyplomowa porusza zagadnienie wyceny nieruchomości, skupiając się w szczególności na lokalach użytkowych z rynku pierwotnego w m. st. Warszawie. Począwszy od zapoznania z rynkiem mieszkań zarówno w całym kraju jak i w stolicy i określenia, co de facto czyni mieszkanie atrakcyjnym dla potencjalnego nabywcy, poprzez przegląd istniejących rozwiązań i opracowań dotyczących problemu oraz wyboru metodyk i narzędzi odpowiednich do analizy zagadnienia, aż do postawienia i uzasadnienia autorskich hipotez dotyczących zależności pomiędzy wartością nieruchomości a czynnikami wpływającymi na komfort i jakość życia mieszkańców oraz ich weryfikacji za pomocą aparatu matematycznego z wykorzystaniem ogólnie dostępnych danych, praca przedstawia znalezienia odpowiedzi na pytanie, od jakich czynników i w jakim stopniu uzależnione są ceny mieszkań.

1.1. Rynek mieszkań w Polsce

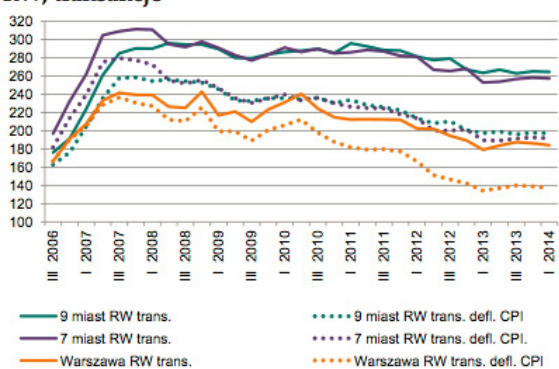
Ceny mieszkań w Polsce ulegają nieustannym, choć nieznacznym, zmianom. Wszelkie analizy rynku mieszkań w kraju opierają się na badaniu cen lokali w największych polskich miastach, którymi są miasta wojewódzkie.

Średnie ceny jednego metra kwadratowego po osiągnięciu szczytu w pierwszym kwartale 2008 roku spadały do roku 2013, po czym odnotowano kolejny, minimalny wzrost cen [1]. Jednak po urealnieniu cen z wykorzystaniem poziomu inflacji na przestrzeni ostatnich lat można wywnioskować, że sięgnęły one obecnie poziomów z roku 2006, gdyż spadły od 28% do 40%. Dodatkowo, uwzględniając fakt, że realna inflacja jest wyższa niż indeks CPI (wskaźnik cen towarów i usług konsumpcyjnych, najpopularniejsza miara inflacji) można stwierdzić, że poziom spadków jest jeszcze większy [1], co zostało przedstawione na rysunku 1.

Indeks średniej ważonej ceny m kw. mieszkania oraz realnej względem CPI (IV kw. 2002 r. = 100) – RP, transakcje



Indeks średniej ważonej ceny m kw. mieszkania oraz realnej względem CPI (IV kw. 2002 r. = 100) – RW, transakcje



Rys. 1: Wykres zmian średnich cen mieszkań oraz cen realnych względem CPI na przestrzeni lat. Źródło [1], na podstawie danych z NBP, GUS i AMRON

Jednak, mimo spadku cen, zauważyć można słabnący ruch na rynku mieszkań. Czas potrzebny na sprzedaż lokalu zwiększył się trzykrotnie w ciągu ostatnich 8 lat. W 2006 roku przeciętne mieszkanie znajdowało nabywcę w ciągu niecałych 2 miesięcy od wystawienia oferty, natomiast obecnie od pojawienia się ogłoszenia do sfinalizowania transakcji mija przynajmniej pół roku [1].

Często wykorzystywanym wskaźnikiem mówiącym o zamożności obywateli kraju jest liczba metrów kwadratowych, którą można nabyć za przeciętne wynagrodzenie (rys. 2). Jednak indeks ten potrafi się różnić nawet o 30% w zależności od źródła publikacji i wykorzystanej metody estymacji. Dane z systemu AMRON bazujące na dochodzie netto wskazują, że siła nabywcza wzrosła do roku 2014 o prawie 60% w porównaniu z rokiem 2007 [2].

Średnia cena 1 m² mieszkania w relacji do przeciętnego wynagrodzenia i wskaźnika dostępności mieszkania w latach 2004–2013



* średnia cena transakcyjna z 7 największych miast Polski (Warszawa, Wrocław, Kraków, Gdańsk, Poznań, Łódź, Białystok)

Rys. 2: Średnie ceny mieszkań w odniesieniu do przeciętnego wynagrodzenia. Źródło [1]

Przez rok 2015 znacznie wzrosła cena mieszkań na rynku wtórnym. W niektórych miastach ceny używanych mieszkań przewyższają ceny nowych, od deweloperów. Najwyżej cenione są mieszkania o powierzchni 38-60 metrów kwadratowych, które w Warszawie kosztują średnio 7778 zł/mkw w przypadku rynku pierwotnego oraz 8060 zł/mkw w przypadku rynku wtórnego [3]. Z przeanalizowanych raportów [3], [4] i [5] wynika, że ceny nowych mieszkań ulegały zmianom w ciągu ostatnich 2 lat, jednak nie tak drastycznym jak w przypadku nieruchomości z rynku wtórnego. Przykładowo, w ostatnim (drugim) kwartale 2016 roku, średnia cena sprzedanych mieszkań wzrosła o 2,2% w porównaniu do I kwartału i o 3% w porównaniu do analogicznego okresu w roku 2015. Dla porównania, zdecydowana większość mieszkań sprzedanych w przeciągu ostatnich 2 lat w Warszawie kosztowała od 6 do 10 tysięcy złotych za metr kwadratowy.

Podsumowując, ceny mieszkań ulegają nieustannym, lecz nieznacznym wahaniom, dotyczy to zwłaszcza nieruchomości pochodzących z rynku pierwotnego. Można więc założyć, że te zmiany w czasie nie miały istotnego wpływu na średnie ceny mieszkań w ciągu ostatnich 2 lat.

1.2. Sformułowanie problemu

Niniejszy podrozdział pracy stanowi sformułowanie zagadnienia będącego tematem pracy, a także opis próby zebrania i analizy czynników, które poprzez wpływ na komfort życia mieszkańców decydują o atrakcyjności mieszkania dla potencjalnego nabywcy, a w rezultacie kształtują ceny nieruchomości.

1.2.1. Zagadnienie wyceny nieruchomości

Z punktu widzenia polskiego prawa, wg Ustawy o gospodarce nieruchomościami z 21 sierpnia 1997 roku, wycena nieruchomości to proces określenia wartości rynkowej, odtworzeniowej i katastralnej danej nieruchomości. Z punktu widzenia niniejszej pracy istotna jest oczywiście wartość rynkowa, która jest uzależniona zarówno od popytu jak i podaży mieszkań. Zgodnie z prawem popytu i podaży, w warunkach niezmienności

pozostałych zjawisk rynkowych, nadwyżka popytu nad podażą powoduje wzrost ceny, natomiast nadwyżka podaży nad popytem jest przyczyną spadku ceny.

Podejście do zagadnienia przyjęte przez autora jest zgoła odmienne od klasycznej wyceny nieruchomości dokonywanej przez rzeczoznawców. W kontekście niniejszej pracy wyceną nieruchomości jest próba określenia ścisłych, matematycznych zależności pomiędzy cechami charakteryzującymi mieszkanie, a jego wartością rynkową, czyli de facto modelowanie cen nieruchomości w zależności od opisujących je czynników. Tak więc badana jest tu de facto strona konsumencka, czyli popyt na mieszkania. Pominięte zostały czynniki związane z podażą, czyli liczbą mieszkań na rynku oferowanych przez deweloperów.

Podsumowując, niniejsza praca bada konsumencką stronę zagadnienia wyceny nieruchomości z rynku pierwotnego, analizując czynniki wpływające na to, ile potencjalny nabywca jest skłonny zapłacić za konkretne mieszkanie.

1.2.2. Czynniki wpływające na wartość nieruchomości

Aby móc podjąć próbę tworzenia matematycznego modelu zależności cen mieszkań od cech charakteryzujących poszczególne lokale, należy określić zbiór tych własności oraz możliwość ich dyskretyzacji, tj. przypisania wartości liczbowej ze skończonego zbioru, określającej natężenie danej cechy. Przede wszystkim, należy wyróżnić cechy odnoszące się do danej inwestycji mieszkaniowej lub ogólniej mówiąc pewnego zbioru lokali, np. położonych w tej samej okolicy. Będą to pewne własności wspólne dla wielu mieszkań, jak choćby możliwość dojazdu windą – charakterystyczna dla wszystkich lokali w tym samym budynku, a przynajmniej w tej samej klatce schodowej.

Inne takie cechy to np. nasłonecznienie, rozumiane jako okres w ciągu dnia, podczas którego światło słoneczne wpada bezpośrednio do pomieszczeń, a zależne głównie od strony świata, z której są okna oraz, w mniejszym stopniu, piętra, na którym znajduje się rozważane mieszkanie. Niektóre własności, jak np. rozmieszczenie pokoi w lokalu, wiążą się z wygodą mieszkania, jednak są trudno mierzalne – można mówić jedynie o dobrym/dogodnym lub złym/niewygodnym rozkładzie pokoi.

Cechy takie jak nasłonecznienie lub układ mieszkania są charakterystyczne (posiadają identyczne lub bardzo podobne natężenie/wartość) dla pewnego zbioru lokali, mniejszego niż zbiór mieszkań w całej inwestycji. Jednak przy określaniu cech wspólnych pewnych grup lokali, dla uproszczenia rozważań należy ograniczyć się wyłącznie do cech charakterystycznych dla całych budynków/inwestycji mieszkaniowych, a cechy takie, jak nasłonecznienie czy wielkość balkonu potraktować jako cechy indywidualne lokalu.

Można również wskazać własności charakterystyczne dla konkretnej inwestycji mieszkaniowej (a co za tym idzie wspólne dla wszystkich lokali w jej obrębie), takie jak np. jakość materiałów użytych do budowy budynku, standard wykonania części wspólnych (klatek schodowych, placów zabaw itp.), a także cechy wspólne dla inwestycji różnych deweloperów, a znajdujących się blisko siebie, np. dostępność lokali usługowych, komunikacji zbiorowej, wskaźnik przestępczości na danym osiedlu itp.

Ograniczenie rozważań do wyłącznie nowych mieszkań w stanie surowym ułatwia analizę ze względu na identyczny standard wykończenia wszystkich mieszkań, a właściwie jego brak – zróżnicowanie jakości materiałów użytych do wykończenia mieszkania (tynki, podłogi, glazury itp.) oraz wyposażenia i mebli w lokalu mogłoby w znacznym stopniu wpływać na atrakcyjność i wartość poszczególnych mieszkań. Zdefiniowanie, kategoryzacja oraz dyskretyzacja cech mieszkań są tematem dalszych rozdziałów pracy – szczegółowej analizy zagadnienia.

1.3. Cel pracy

Podsumowując rozważania przedstawione we wstępie, niniejsza praca porusza zagadnienie wyceny nieruchomości od strony konsumenckiej. Celem pracy było zapoznanie się z rynkiem mieszkań, określenie i analiza cech wpływających na atrakcyjność mieszkań, która ma bezpośrednie odzwierciedlenie w ich jednostkowej cenie (za każdy metr kwadratowy), a następnie stworzenie i weryfikacja modelu matematycznego zależności cen mieszkań od najistotniejszych parametrów je opisujących. Zostały przedstawione, teoretycznie uzasadnione i sprawdzone pewne hipotezy dotyczące omawianej zależności. W niniejszej pracy przyjęto, że cena jednostkowa mieszkania jest miarą jego atrakcyjności, a więc wartości atrakcyjności są równe wartościom cen za metr kwadratowy.

1.3.1. Zakres pracy

Ceny mieszkań w obrębie kraju wykazują znaczne zróżnicowanie. W samych największych miastach ceny za metr kwadratowy mogą różnić się o ponad połowę – przykładowo średnia wartość nowego mieszkania o metrażu 38-60 mkw w Katowicach w styczniu 2016 roku wynosiła w 5022 zł, natomiast za podobne mieszkanie w Warszawie trzeba było zapłacić 7814 zł za każdy metr kwadratowy. W związku z powyższym należało ograniczyć rozważania, przynajmniej początkowo, do jednego miasta. Oczywistym wyborem była Warszawa, w której notuje się najwięcej transakcji kupna/sprzedaży mieszkań. Ponadto dane transakcyjne są udostępniane (pod pewnymi warunkami) przez Biuro Geodezji i Katastru Urzędu Miejskiego.

Fakt, że nowe mieszkania od deweloperów są zazwyczaj sprzedawane w stanie surowym, upraszcza analizę rynku nowych mieszkań, ponieważ na cenę lokalu nie wpływa czynnik związany z jakością indywidualnego wykończenia – można założyć, że wszystkie mieszkania są pod tym względem jednakowe, tj. nie posiadają żadnego wyposażenia.

Tak jak zostało to powiedziane we wstępie, kilkuprocentowe wahania cen na przestrzeni ostatnich kilku lat nie mają istotnego wpływu na ceny mieszkań, które dla Warszawy wahają się od 6 do 10 tys. zł za metr kwadratowy. Podsumowując powyższe fakty, za zakres niniejszej pracy przyjęto rynek nowych mieszkań oferowanych przez deweloperów na terenie stolicy. Analizie poddano transakcje, których dokonano w okresie ostatnich 2 lat.

2. Przegląd obecnego stanu wiedzy

Niniejszy rozdział zawiera podsumowanie przeprowadzonych do tej pory badań i analiz dotyczących zagadnienia modelowania cen nieruchomości w kontekście zarówno polskiego rynku mieszkań, jak i rynków zagranicznych. Przedstawione opracowania poruszają różne strony problemu – zależność wartości mieszkań od ich otoczenia, np. bliskości terenów zielonych, akwenów, terenów przemysłowych, a także od zmian klimatu.

Ponadto, dokonano przeglądu dwu interesujących rozwiązań związanych z omawianym zagadnieniem, które wspomagają proces oceny wartości nieruchomości poprzez jego częściową automatyzację przy użyciu pewnych modeli. Pierwsze z nich to wskaźnik Walk Score używany za granicą (głównie w Stanach Zjednoczonych) do określania atrakcyjności konkretnych lokalizacji jak i całych miast pod względem wygody w poruszaniu się komunikacją zbiorową, bliskości punktów użyteczności publicznej itp. Drugie rozwiązanie to portal Szybko.pl, który oferuje bezpłatną usługę wyceny mieszkania po podaniu jego parametrów i lokalizacji. Algorytm wyceny opiera się na dotychczas zrealizowanych transakcjach kupna/sprzedaży mieszkań w najbliższym sąsiedztwie.

2.1. Istniejące opracowania i literatura

Zagadnienie wyceny nieruchomości oraz modelowania cen mieszkań było tematem wielu prac, które przedstawiały różne metody analizy problemu.

2.1.1. Publikacje dotyczące rynków zagranicznych

Praca [6] przedstawia analizę wpływu otoczenia nieruchomości na ceny jednostkowe mieszkań przy użyciu regresji hedonicznej. Badanie dotyczy miasta Kolonii w Niemczech. Autorzy stwierdzają, że zwiększenie powierzchni zieleni miejskiej w promieniu 500 metrów o 1% powoduje wzrost wartości mieszkania o 0,1%, natomiast zwiększenie powierzchni zajętej przez akweny o 1% w identycznym obszarze (500 m) powoduje wzrost ceny o 0,16%. Wynika z tego, że obecność zbiorników wodnych jest ceniona bardziej niż zieleń miejska i w większym stopniu wpływa na wartość nieruchomości. Dodatkowo, z każdym rokiem wieku

mieszkania, jego cena spada o 0,22%. Analiza dotyczyła 85046 transakcji dokonanych na przestrzeni lat 1995-2012.

Autorzy w pracy [7] wyróżniają następujące czynniki warunkujące ceny mieszkań: powierzchnia mieszkania/domu, liczba pięter, powierzchnia działki, standard i stan otoczenia mieszkania (rozumianego jako podwórko, ogród itp.) (dobry/słaby), liczba pokoi, liczba łazienek, obecność garażu. Przedstawiony został model wielomianowy opisujący cenę nieruchomości w zależności od liczbowej wartości powyższych parametrów, przy przyjęciu możliwych wartości 0 i 1 dla zmiennych takich, jak obecność garażu i stan otoczenia mieszkania. Próba użyta do badania to 81 mieszkań.

Publikacja [8] dotyczy wykorzystania Systemu Informacji Geograficznej (GIS) do stworzenia hedonicznego modelu cen nieruchomości. Badanie dotyczyło mieszkań w najbliższej okolicy Glasgow w Wielkiej Brytanii. Autorzy opisali metodę określania cech charakteryzujących nieruchomości, dzięki której udało się zdefiniować ponad 320 takich własności. Po analizie zostały odrzucone cechy o wysokiej wzajemnej korelacji, jak i te o znikomym wpływie na ceny mieszkań. Próba liczyła 3456 transakcji.

W pracy [9] przedstawione zostały wyniki analizy hedonicznego modelu cen mieszkań w hrabstwie Marion w stanie Indiana w Stanach Zjednoczonych, obejmującej 8772 transakcji zakupu/sprzedaży domów. Wg autorów czynnikami kształtującymi ceny domów w największym stopniu są cechy samych nieruchomości, takie jak liczba łazienek, wielkość ogrodu, obecność piwnicy i klimatyzacji. Lokalizacja domu i aspekty z nią związane, tj. czas dojazdu do centrum, bliskość szkół, w mniejszym stopniu determinują jego wartość.

2.1.2. Publikacje dotyczące rynku polskiego

Opracowanie [10] przedstawia analizę cen mieszkań w wybranych dzielnicach Warszawy, na podstawie próby 1485 cen z danych ofertowych z lat 2008-2010. Autor zweryfikował wpływ następujących czynników na wartość nieruchomości: obecność ochrony w budynku, obciążenie mieszkania kredytem hipotecznym, budowa budynku przez II wojnę światową, powierzchnia mieszkania, wiek mieszkania, średnia cena mieszkania w danej dzielnicy w minionym okresie. Przedstawiony przez autora model charakteryzuje się dopasowaniem wartości teoretycznych do empirycznych na poziomie 76,3%. Warto

podkreślić, że dane ofertowe, które zostały użyte podczas tworzenia modelu, mogą się nieznacznie różnić od rzeczywistych cen transakcyjnych.

Publikacja [11] opisuje metody wyznaczania hedonicznych indeksów cen mieszkań przy użyciu danych z Bazy Rynku Nieruchomości prowadzonej przez NBP. Autorzy posługując się wiedzą ekspercką określili 19 zmiennych wpływających bezpośrednio na atrakcyjność i ceny lokali, z czego 5 zmiennych odnosiło się do lokalizacji. Jednak ze względu na braki danych zbiór zmiennych wykorzystywanych w modelowaniu ograniczono do 12, z czego tylko 2 odnosiły się do lokalizacji. Pozostałe zmienne objaśniające określały m.in. standard wykończenia mieszkania, technologię budowy budynku, typ kuchni przyjmując wartość ze skończonego zbioru ustalonego przez autora. Ponadto, omawiane zmienne określały także liczbę pokoi czy obecność windy. Wykorzystano dane z 1494 transakcji dokonanych w latach 2006-2009 w obrębie stolicy. Zgodnie z wynikami analizy, usytuowanie mieszkania w dobrej dzielnicy podnosi cenę metra kwadratowego o około 29%, natomiast cena metra kwadratowego mieszkania jedno- lub dwupokojowego jest o około 8% wyższa niż cena metra kwadratowego większych mieszkań. Ponadto, mieszkania o wysokim standardzie wyposażenia są droższe od mieszkań o przeciętnym standardzie średnio o 9%.

2.2. Rozwiązania automatyzujące proces oceny wartości nieruchomości

Proces tworzenia modelu matematycznego cen mieszkań, składający się z określenia zmiennych objaśniających, doboru klas funkcji i wyznaczaniu współczynników jest dość pracochłonny i czasochłonny. Ponadto zagadnienie wymaga analizy znacznej liczby danych dotyczących cen i charakterystyk mieszkań (rzędu tysięcy) w celu uzyskania wyników w odpowiednim stopniu zgodnych z rzeczywistością.

W związku z powyższym niezbędne jest wykorzystanie narzędzi automatyzujących przynajmniej część procesu, potrafiących dokonać szybkiej analizy dużej ilości danych przy określonych warunkach. Spośród istniejących narzędzi wybrano dwa, które, choć działają w zgoła odmienny sposób i na różnych danych, wyznaczając inne rodzaje wartości, to automatyzują pewne czynności związane z zagadnieniem będącym tematem pracy.

Pierwsze z narzędzi to wskaźnik Walk Score używany do określania atrakcyjności konkretnych lokalizacji jak i całych miast pod względem wygody życia mieszkańców. Drugie rozwiązanie to portal Szybko.pl, oferujący bezpłatną usługę wyceny mieszkań w oparciu o dotychczas zrealizowane transakcje kupna/sprzedaży mieszkań w najbliższym sąsiedztwie.

2.2.1. Walk Score

Walk Score to firma oferująca usługi *walkability*¹ oraz wyszukiwania mieszkań poprzez stronę www (rys. 3) i aplikację mobilną [13]. Głównym i najbardziej znanym produktem jest indeks Walk Score, który przypisuje liczbowy wynik określający *walkability* dowolnego miejsca w Stanach Zjednoczonych, Kanadzie i Australii. Walk Score jest używany przez wiele agencji nieruchomości jako wiarygodny wskaźnik wygody codziennego życia oferowanej przez daną lokalizację [12]. Walk Score wyznacza również indeks dla polskich miast, jednak w tym przypadku funkcjonalność serwisu jest ograniczona.

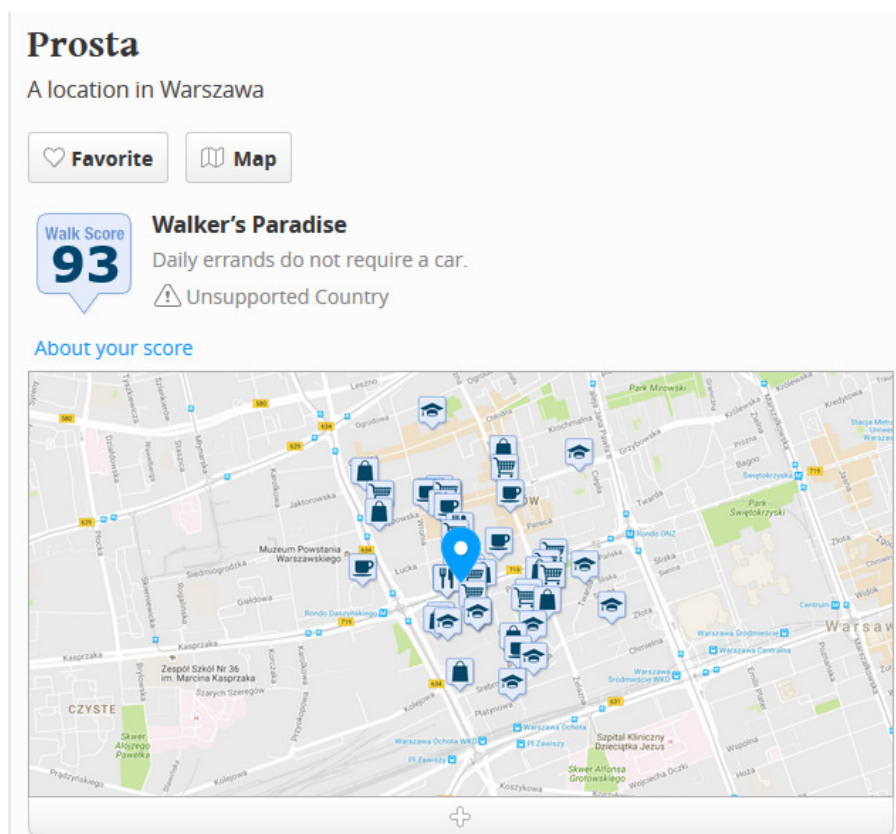


Rys. 3: Główna strona portalu WalkScore.com. Źródło [12]

¹ Słowo *walkability* nie ma polskiego odpowiednika. *Walkable* oznacza możliwy do przebycia pieszo, osiągalny na piechotę. W kontekście pracy *walkable* oznacza, że dana lokalizacja jest przyjazna dla poruszania się po mieście pieszo, ew. rowerem czy komunikacją zbiorową; wiele punktów użyteczności publicznej jest dostępnych w zasięgu krótkiego spaceru, bez potrzeby używania samochodu itp. *Walkability* jest więc miarą wygody, jaką oferuje dana lokalizacja dla osoby preferującej poruszanie się pieszo.

Wskaźnik Walk Score danej lokalizacji jest wyznaczany poprzez określenie liczby punktów użyteczności publicznej (POI) w najbliższym jej otoczeniu, tj. w zasięgu kilkuminutowego spaceru (rys. 4). Im więcej POI, takich jak np. poczta, restauracje, sklepy w okolicy, tym więcej codziennych obowiązków można załatwić nie używając żadnych środków komunikacji.

Walk Score jest zatem pewną miarą atrakcyjności lokalizacji mieszkań, może zatem być czynnikiem częściowo determinującym cenę lokali – im wyższy wskaźnik cechuje daną lokalizację, tym więcej będą kosztować mieszkania znajdujące się w okolicy. Naturalnym jest, że indeks rośnie w kierunku centrów miast, gdzie znajduje się dużo punktów użyteczności publicznej i mieszkania są odpowiednio droższe niż np. na dzielnicach obrzeżnych.



Rys. 4: Wskaźnik Walk Score wyznaczony dla ulicy Prostej w Warszawie, wraz z zaznaczonymi na mapie najbliższymi punktami użyteczności publicznej. Źródło [12]

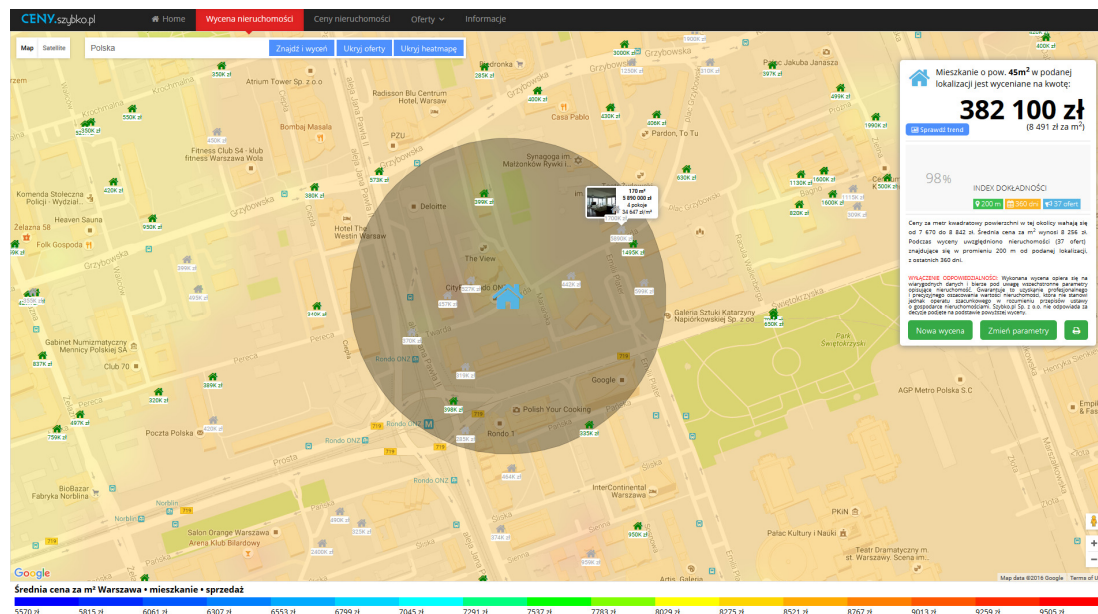
Portal Walk Score oferuje również wskaźniki Bike Score oraz Transit Score. Pierwszy – Bike Score – określa atrakcyjność lokalizacji dla rowerzystów, uwzględniając obecność ścieżek rowerowych i połączenia między nimi oraz pagórkowatość terenu, co przekłada się na dostępność i bliskość typowych celów podróży. Transit Score mówi natomiast o tym, jak

dobrze okolica jest obsługiwana przez komunikację zbiorową, co jest zależne od liczby linii i rodzaju środków transportu (autobus, metro) dostępnych w danym miejscu.

2.2.2. Szybko.pl

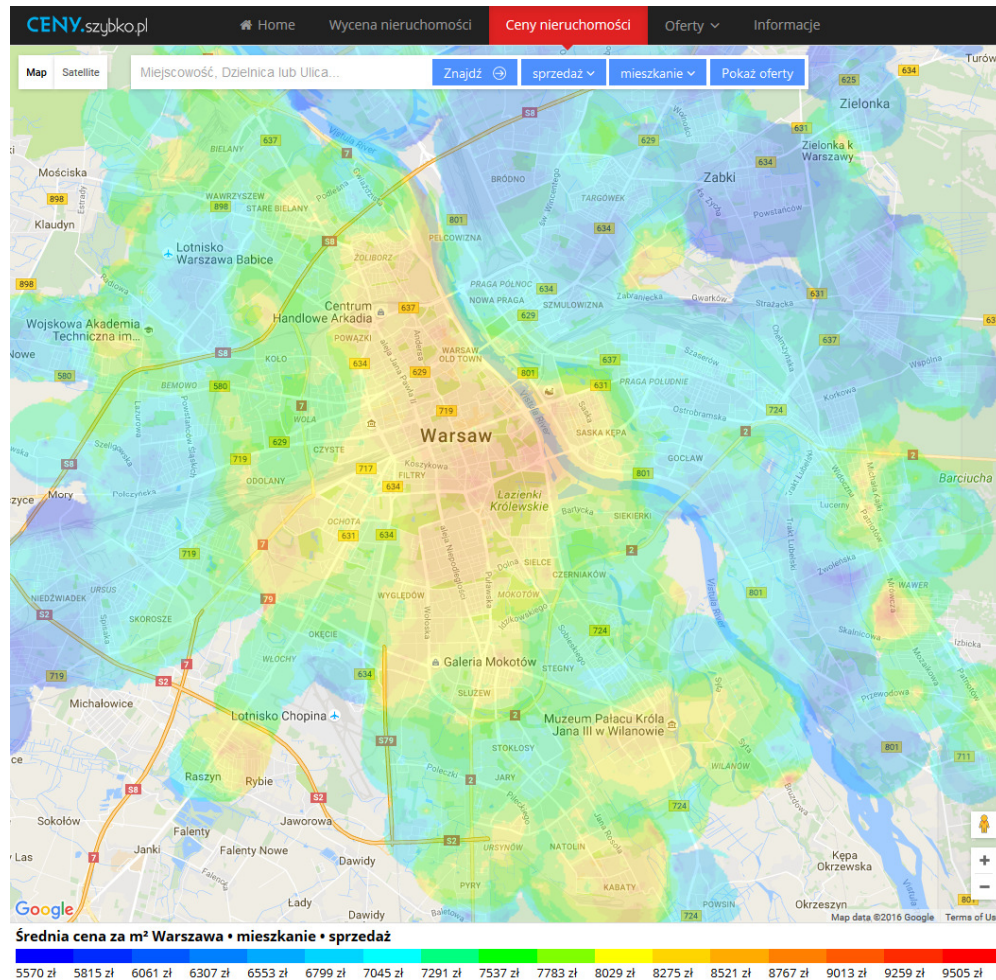
Platforma Szybko.pl jest grupą kilkudziesięciu serwisów ogłoszeniowych gromadzących różnego rodzaju oferty, w szczególności z rynku nieruchomości [14]. Oprócz publikacji ogłoszeń o sprzedaży/wynajmie nieruchomości, portal oferuje bezpłatną usługę wyceny mieszkania oraz informacje o cenach ofertowych mieszkań w różnych miastach w Polsce [15].

Aby serwis wyznaczył najbardziej prawdopodobną wartość nieruchomości, należy podać jej najważniejsze parametry, takie jak: dokładny adres, rodzaj nieruchomości, powierzchnia użytkowa, liczba pokoi, łazienek, piętro, obecność windy, parkingu, rok budowy budynku oraz standard wykończenia. Na podstawie tych danych oraz informacjach z ofert dotyczących lokali w najbliższym otoczeniu wybranej lokalizacji, ustalana jest najbardziej prawdopodobna cena mieszkania.



Rys. 5: Przykładowa wycena dwupokojowego mieszkania w centrum Warszawy. Na mapie zaznaczone są mieszkania z najbliższej okolicy wraz z ich cenami ofertowymi. Źródło: [15]

Portal oferuje ponadto informacje o cenach ofertowych mieszkań przedstawione na mapie (rys. 5) jako pojedyncze punkty odpowiadające ogłoszeniom oraz kartogram ukazujący średnie ceny mieszkań w mieście (kolor i odcień odpowiadają średnim cenom, zgodnie z legendą).



Rys. 6: Mapa ukazująca średnie ceny mieszkań. Źródło: [13]

2.3. Podsumowanie przedstawionych opracowań i rozwiązań

Omówione prace ukazują różne podejścia do modelowania cen mieszkań, zarówno na rynku polskim, jak i na rynkach zagranicznych. Rozpatrywano wiele czynników wpływających na atrakcyjność i wartość nieruchomości, takich jak parametry samego lokalu/domu, jego otoczenie oraz lokalizację. Zależnie od typu nieruchomości (mieszkanie w bloku albo dom wolnostojący) oraz od miasta, dla którego wybrano próbę transakcji lub ofert kupna/sprzedaży, można zauważyć, że różne czynniki w niejednakowym stopniu wpływają na wartość nieruchomości.

W przypadku domów jednorodzinnych można wywnioskować, że dla potencjalnych kupców istotne jest najbliższe otoczenie posesji, tj. ogródek, ulica, osiedle, a w mniejszym stopniu czynniki związane z jej lokalizacją, np. czas dojazdu do centrum. W przypadku mieszkań w blokach nieco większą rolę odgrywa lokalizacja, a mniejszą wagę przywiązuje się do czynników typowych dla domu jednorodzinnego. Wynika to prawdopodobnie z typu posesji oraz być może z preferencji nabywców.

Można wysnuć hipotezę, że nowe mieszkania w blokach kupują zazwyczaj młodzi ludzie, niekiedy nieposiadający jeszcze dzieci, aktywnie spędzający wolny czas i ceniący sobie bliskość miejsca pracy i centrum rozrywkowego miasta. Natomiast posesje jednorodzinne mogą znajdować kupców głównie pośród rodzin z dziećmi, które preferują mieszkanie we własnym domu z ogródkiem, w cichszej i spokojniejszej okolicy oddalonej od miejskiego zgiełku.

Z drugiej strony można by próbować wyciągać wnioski dotyczące preferencji mieszkańców miast o różnej wielkości, zaludnieniu czy też w różnych krajach. Jednak aby próbować wyciągać wnioski na ten temat, należałoby przeanalizować znacznie więcej opracowań dotyczących różnych miast w różnych krajach świata. Taka analiza może zostać pominięta, gdyż nie jest bezpośrednio związana z tematem pracy.

Przedstawione opracowania dotyczące Warszawy zawierają analizę i próby modelowania cen mieszkań z uwzględnieniem czynników zależnych głównie od wieku budynku, jakości wykończenia zarówno lokalu jak i całego budynku, liczby i charakteru pomieszczeń oraz w mniejszym stopniu lokalizacji nieruchomości. Nie rozpatrywano np. odległości od typowych celów podróży (centrum miasta, miejsca pracy, punkty użyteczności publicznej), która może być mierzona zarówno w jednostkach długości jak i orientacyjnym czasie podróży komunikacją zbiorową, własnym pojazdem bądź pieszo.

2.3.1. Wskazanie nieopracowanej strony zagadnienia

Tak jak wspomniano powyżej, przeprowadzone dotychczas analizy cen mieszkań tylko w niewielkim stopniu skupiały się na aspektach związanych z lokalizacją nieruchomości. W rzeczywistości położenie nieruchomości na mapie miasta przekłada się na dostępność usług, punktów użyteczności publicznej, sieci dróg oraz komunikacji zbiorowej, co bezpośrednio

wpływa na komfort życia mieszkańców. W związku z tym ceny gruntów, nieruchomości i lokali są odpowiednio wyższe w lepiej skomunikowanych częściach miasta – występuje naturalna tendencja do wzrostu cen w kierunku centrum, natomiast potencjalni nabywcy są skłonni zapłacić więcej za mieszkanie gwarantujące wyższy komfort życia.

Istniejące na polskim rynku rozwiązanie oferowane przez portal Szybko.pl potrafi określać przybliżone ceny mieszkań na podstawie wielu czynników, jednak głównym i decydującym wydają się być ceny transakcji dokonanych dla mieszkań z najbliższej okolicy. Ponadto szczegółowe parametry modelu używanego do wyznaczania cen nie są podane.

Rysunek 6 obrazuje średnie ceny mieszkań w Warszawie wg Szybko.pl. Zauważyć można wyższe średnie ceny nieruchomości w centrum oraz kilku innych miejscach. O ile mieszkania zlokalizowane w Śródmieściu są atrakcyjniejsze i droższe niż na obrzeżach, to ceny nieruchomości w podwarszawskim Raszynie czy obrzeżnych dzielnicach (Wawer, Targówek) wydają się być zawyżone. Można podejrzewać, że model cen mieszkań wykorzystywany przez portal Szybko.pl nie do końca poprawnie opisuje ich wartości w pewnych obszarach miasta. Stwarza to dodatkową motywację do opracowania alternatywnego sposobu określania wartości nieruchomości.

2.3.2. Doprecyzowanie celu pracy

Za cel pracy przyjęto zapoznanie się z rynkiem mieszkań, określenie i analizę czynników wpływających na atrakcyjność mieszkań, która ma bezpośrednie odzwierciedlenie w ich cenie, a następnie stworzenie i weryfikacja modelu matematycznego zależności cen mieszkań od najistotniejszych parametrów je opisujących.

Uwzględniając przeprowadzone dotychczas analizy oraz ich wyniki jak i pomysły zaczerpnięte z przedstawionych istniejących rozwiązań postanowiono w niniejszej pracy skupić się na kształtujących ceny nieruchomości cechach wynikających z dokładnego położenia na mapie miasta – w odniesieniu do istniejących punktów użyteczności publicznej, sieci dróg oraz sieci komunikacji zbiorowej – a także opisujących same lokale jak i całe inwestycje mieszkaniowe w sposób ilościowy.

3. Analiza zagadnienia

Niniejszy rozdział opisuje przeprowadzoną analizę problemu modelowania cen nieruchomości. Przedstawione zostały następujące zagadnienia:

- Metody wykorzystywane przy modelowaniu – regresja hedoniczna i automatyczny model wyceny (AVM),
- Cechy składające się na atrakcyjność mieszkania – analiza i wybór najistotniejszych i mających potencjalnie największy wpływ na wartość nieruchomości cech związanych z lokalem, budynkiem oraz lokalizacją,
- Wykorzystane źródła danych – przegląd i opis wykorzystania źródeł danych o atrybutach i cenach mieszkań oraz charakteryzujących lokalizacje pod kątem dostępności usług, komunikacji zbiorowej i sieci dróg.

Dalsza część przedstawia opis zbierania, analizy i interpretacji danych:

- Akwizycja i agregacja danych – opis sposobów pozyskiwania, łączenia i analizowania danych,
- Sformułowanie i weryfikacja hipotez dotyczących zależności cen mieszkań od omówionych czynników,
- Wnioski wyływające z wykonanej analizy – potwierdzone i odrzucone hipotezy.

3.1. Szczegółowy opis sposobu modelowania i wyznaczania cen mieszkań

Do badania zależności cen nieruchomości od cech je opisujących wykorzystano metodę regresji hedonicznej. Natomiast do wyznaczania najbardziej prawdopodobnej wartości mieszkania użyto automatycznego modelu wyceny (AVM).

3.1.1. Regresja hedoniczna

Regresja jest metodą statystyczną, która umożliwia badanie zależności występujących pomiędzy danymi oraz dzięki temu przewidywanie wartości pewnych wielkości z wykorzystaniem dostępnych i znanych wartości innych wielkości [16].

Użycie regresji w praktyce sprowadza się do dwóch etapów:

- konstruowanie modelu – polega na stworzeniu modelu regresyjnego, czyli funkcji opisującej zależności pomiędzy zmienną objaśnianą a zmiennymi objaśniającymi. Taka funkcja może przybrać postać zarówno prostego wzoru matematycznego, jak również skomplikowanego algorytmu w postaci drzewa regresyjnego czy też sieci neuronowej. Model jest konstruowany w taki sposób, aby jak najlepiej pasował do danych ze zbioru uczącego (zawierającego zmienne objaśniające i zmienne objaśniane).
- stosowanie modelu (scoring) – wykorzystanie skonstruowanego wcześniej modelu na danych zawierających wyłącznie zmienne objaśniające w celu określenia na ich podstawie wartości oczekiwanej zmiennej objaśnianej.

Koncepcja hedonicznego modelu cen opiera się na następującym założeniu: dobra heterogeniczne (czyli takie, które różnią się między sobą pewnymi cechami, np. nieruchomości) mogą zostać przedstawione jako zespół pewnych cech je charakteryzujących [11]. Z punktu widzenia konsumenta każdej z tych cech przypisana jest pewna użyteczność, natomiast cena produktu jest sumą (nieobserwowalnych) użyteczności poszczególnych cech, gdyż same cechy nie są przedmiotem transakcji i nie mają określonych wartości pieniężnych. Modele hedoniczne przybierają postać modeli ekonometrycznych, w których zmienną objaśnianą jest cena – wartość produktu, a zmiennymi objaśniającymi – charakterystyki produktu mające według założenia istotny wpływ na ostateczną cenę [11]. Cena heterogenicznego dobra jest w związku z tym sumą wartości poszczególnych cech je charakteryzujących opisanych za pomocą zmiennych objaśniających oraz innych czynników odzwierciedlonych w składniku losowym.

Zgodnie z powyższym, ogólny model wartości dobra heterogenicznego przybiera następującą postać:

$$W = \sum_{i=1}^n w_i(c_i) + \varepsilon \quad (1)$$

gdzie:

W – wartość dobra heterogenicznego (np. cena mieszkania),

c_i – natężenie pewnej cechy dobra (np. powierzchnia mieszkania),

w_i - funkcja przyporządkowująca natężeniu pewnej cechy c_i pewną wartość,

$w_i(c_i)$ –wartość, jaką pewna cecha c_i dobra stanowi dla konsumenta,

ε – składnik losowy modelu,

$i \in \{1,2,3, \dots, n\}$,

n – liczba cech opisujących mieszkanie.

W związku z charakterem powyższej zależności, do wyznaczania modelu wartości nieruchomości została wykorzystana regresja liniowa, która zakłada liniową zależność pomiędzy zmienną objaśnianą, a zmiennymi objaśniającymi [17]. Jeżeli przyjąć, że p_i jest wartością atrakcyjności wynikającej z natężenia pewnej cechy i jest proporcjonalne do $w_i(c_i)$, zagadnienie sprowadza się do regresji liniowej wartości nieruchomości W względem parametrów p_i , a więc:

$$W = \sum_{i=1}^n a_i p_i + \gamma + \varepsilon \quad (2)$$

gdzie a_i jest współczynnikiem liniowej zależności dla $i \in \{1,2,3, \dots, n\}$, natomiast γ to pewna stała wartość.

Podsumowując, metoda regresji hedonicznej zakłada, że wartość pewnego dobra (w tym przypadku mieszkania) jest sumą wartości poszczególnych, obserwowalnych jego cech, takich jak np. jakość wykończenia, atrakcyjność lokalizacji itp. Powyższe cechy są traktowane jako zmienne objaśniające, natomiast cena jest zmienną przez nie objaśnianą. Celem jest wyznaczenie takiego modelu, który mógłby z powodzeniem przewidywać w przyszłości ceny nowych mieszkań.

3.1.2. AVM – automatyczny model wyceny

AVM (Automated Valuation Model - Automatyczny Model Wyceny) jest usługą umożliwiającą wyznaczanie wartości nieruchomości przy użyciu modelu matematycznego oraz bazy danych zawierającej informacje o innych nieruchomościach na rynku. Algorytm analizuje duże ilości danych oraz cechy konkretnej nieruchomości, aby następnie oszacować wartość rynkową danego obiektu [18].

Technologia AVM wykorzystywana jest podczas wyceny nieruchomości przez rzeczoznawców, instytucje kredytowe oraz inwestorów. Wykonywana przez AVM wycena powstaje w wyniku połączenia badania i przetwarzania dostępnych danych oraz zależności

przyjętych w modelu. Wykorzystywane są dwa rodzaje danych: informacje o cechach związanych z nieruchomością oraz dane transakcyjne [14].

AVM zazwyczaj obejmuje swym badaniem [18]:

- opinie rzeczoznawców majątkowych i analityków rynku,
- informacje na temat nieruchomości,
- ceny transakcyjne,
- analizę porównawczą podobnych nieruchomości.

Podsumowując, AVM służy do szybkiego przewidywania cen mieszkań przy użyciu przeanalizowanych danych o innych nieruchomościach i dotychczas zrealizowanych transakcjach oraz cechach opisujących wyceniane mieszkania.

W niniejszej pracy założenia AVM zostały wykorzystane przy tworzeniu i weryfikacji modelu cen nieruchomości, natomiast docelowo AVM będzie wykorzystywany do przewidywania wartości nowych mieszkań z wykorzystaniem opracowanego modelu oraz historycznych danych transakcyjnych. Regresja hedoniczna była przeprowadzana przy wykorzystaniu automatycznego modelu wyceny do określania przewidywanych wartości nieruchomości na podstawie informacji o cechach je charakteryzujących.

3.2. Określenie składowych atrakcyjności mieszkania

W celu stworzenia matematycznego modelu opisującego wartość nieruchomości, należy przede wszystkim określić cechy je opisujące. Zgodnie z hedonicznym modelem cen, który opiera się na założeniu, że dobra (w tym wypadku mieszkania) można przedstawić jako agregat ich cech istotnych z punktu widzenia konsumenta, należało wyróżnić atrybuty je charakteryzujące i tworzące omawiany hedoniczny koszyk atrybutów:

- mieszkanie, takie jak np. liczba pokoi, piętro;
- całą inwestycję (budynek lub grupę budynków), np. jakość materiałów użytych przy budowie, obecność garażu podziemnego;
- lokalizację nieruchomości, np. odległość od centrum.

Powyższe cechy połączono w 2 grupy:

1. odnoszące się do standardu mieszkania (połączenie atrybutów samego mieszkania oraz inwestycji, w której się znajduje) oraz
2. dotyczące lokalizacji nieruchomości.

Spośród własności charakteryzujących mieszkania należało wybrać te, którym można przypisać wartości liczbowe z pewnego zbioru tak, aby było możliwe ilościowe określenie natężenia danych cech, co jest warunkiem koniecznym, aby móc uwzględnić je w procesie modelowania.

3.2.1. Parametryzacja opisu nieruchomości

Lokal mieszkalny może być scharakteryzowany przez szereg cech wpływających w mniejszym lub większym stopniu na komfort mieszkańców. Poza podstawowymi, takimi jak metraż i liczba pomieszczeń, istotne są również wysokość mieszkania, charakter pomieszczeń – liczba łazienek, sypialni, rodzaj kuchni, obecność i powierzchnia balkonu. Niektóre cechy są trudne, a czasami wręcz niemożliwe do dokładnego opisanie ilościowego.

Poniżej przedstawiono cechy charakterystyczne dla lokali mieszkalnych oraz inwestycji mieszkaniowych wraz z propozycjami zbiorów wartości liczbowych określających natężenie danej cechy. Atrybuty i zbiory wartości zostały częściowo zaczerpnięte z literatury przedstawionej w rozdziale 2.1.

Tabela 1: Cechy mieszkań wpływające na komfort i atrakcyjność nieruchomości

Cecha	Typ zbioru wartości cechy (ciągły/dyskretny)	Zbiór wartości opisujących natężenie cechy
Metraż	Ciągły	Wartość powierzchni w metrach kwadratowych
Liczba pokoi	Dyskretny	Liczba pokoi (= 1, 2, 3, 4...)
Liczba łazienek	Dyskretny	Liczba łazienek (= 1, 2...)
Liczba sypialni	Dyskretny	Liczba sypialni (= 1, 2...)
Miejsce postojowe przypisane do lokalu	Dyskretny	0 – brak miejsca 1 – miejsce pod blokiem 2 – miejsce w garażu
Wysokość lokalu	Ciągły	Wartość wysokości lokalu w metrach albo 0 – przeciętna wysokość (ok. 2,5 m) 1 – lokal wysoki (ok. 3 m)
Obecność/powierzchnia balkonu/loggii	Ciągły/ Dyskretny	0 – brak balkonu/loggii 1 – balkon/loggia w lokalu albo

		Wartość powierzchni balkonu/loggii w metrach kwadratowych
Rodzaj kuchni (zamknięta/otwarta)	Dyskretny	0 – zamknięta 1 – otwarta (W dużej mierze zależne od preferencji użytkowników lokalu)
Rozkład pokoi	Możliwość arbitralnej oceny (dobry/przeciętny/słaby)	2 – dobry 1 – przeciętny 0 - słaby
Naświetlenie mieszkania	Dyskretny	0 – okna tylko od północy 1 – okna z innych stron
Widok z okna	Możliwość arbitralnej oceny (ładny, przeciętny, słaby)	2 – ładny 1 – przeciętny 0 - słaby
Rodzaj umowy sprzedaży	Użytkowanie wieczyste/prawo własności	0 – użytkowanie wieczyste 1 – prawo własności

Tabela 2: Cechy inwestycji mieszkaniowych wpływające na komfort i atrakcyjność nieruchomości

Cecha	Typ zbioru wartości cechy (ciągły/dyskretny)	Zbiór wartości opisujących natężenie cechy
Obecność windy	Dyskretny	0 – brak windy 1 – winda w budynku
Obecność ochrony	Dyskretny	0 – brak ochrony 1 – ochrona w budynku/na osiedlu
Dziedziniec z placem zabaw	Dyskretny	0 – brak dziedzińca 1 – dziedziniec przy budynku
Jakość materiałów użytych przy budowie budynku	Możliwość arbitralnej oceny (dobra, przeciętna, słaba)	2 – dobra 1 – przeciętna 0 - słaba
Recepcja i portier	Dyskretny	0 – brak recepcji 1 – recepcja w budynku
Jakość wykończenia przestrzeni wspólnych	Możliwość arbitralnej oceny (dobra, przeciętna, słaba)	2 – dobra 1 – przeciętna 0 - słaba
Liczba kondygnacji budynku	Dyskretny	Liczba kondygnacji (= 1, 2, 3...)
(Dodatkowy) parking pod budynkiem dla gości	Dyskretny	0 – brak parkingu 1 – parking pod budynkiem

Wszystkie powyższe cechy wpływają w pewnym stopniu na komfort życia mieszkańców, zazwyczaj w oczywisty sposób – jak np. obecność windy lub miejsca parkingowego), a czasami zależnie od indywidualnych preferencji – np. rodzaj kuchni czy rozkład pokoi. Do analizy warto wykorzystać cechy wpływające w sposób oczywisty na wygodę mieszkańców.

3.2.2. Lokalizacja nieruchomości

Osobną kategorię stanowią cechy nieruchomości związane z jej położeniem na mapie miasta, a nieodnoszące się w żaden sposób do samego mieszkania/budynku. Można tu wyróżnić cechy związane z komfortem mieszkania i spędzania czasu wolnego, takie jak natężenie hałasu czy bliskość terenów zielonych, jak również aspekty odnoszące się do tkanki miejskiej, np. czas dojazdu do pracy czy dostępność komunikacji zbiorowej. Tak jak w przypadku atrybutów związanych ze standardem nieruchomości, niektóre cechy są trudne, a czasami wręcz niemożliwe do dokładnego opisanie ilościowego.

Poniżej przedstawiono cechy związane z lokalizacją nieruchomości wraz z propozycjami zbiorów wartości liczbowych określających natężenie danej cechy. Atrybuty i zbiory wartości zostały częściowo zaczerpnięte z literatury przedstawionej w rozdziale 2.1.

Tabela 3: Cechy lokalizacji wpływające na komfort i atrakcyjność nieruchomości

Cecha	Typ zbioru wartości cechy (ciągły/dyskretny)	Zbiór wartości opisujących natężenie cechy
Bliskość centrum	Ciągły	Odległość od centrum miasta albo czas dojazdu do centrum miasta (samochodem/rowerem/komunikacją zbiorową)
Dostępność komunikacji miejskiej	Ciągły/Dyskretny	Czas dojazdu do centrum miasta komunikacją miejską albo odległość od najbliższego przystanku albo liczba środków komunikacji miejskiej odjeżdżających z najbliższego przystanku w ciągu godziny
Natężenie hałasu	Cecha trudna do zmierzenia, możliwość arbitralnej oceny (wysokie, średnie, niskie)	0 – wysokie natężenie hałasu 1 – średnie natężenie hałasu 2 – niskie natężenie hałasu
Wskaźnik przestępczości	Dyskretny (na podstawie danych statystycznych)	0 – wysoki 1 – średni 2 – niski albo liczba przestępstw popełnionych w danej dzielnicy/osiedlu w ciągu ostatniego roku
Bliskość terenów zielonych	Ciągły/dyskretny	Odległość do najbliższego parku albo liczba terenów zielonych w promieniu np. 2 kilometrów
Dostępność usług	Ciągły/dyskretny	Odległość do najbliższego punktu usługowego wybranej kategorii (np. sklep, poczta, bank) albo liczba punktów usługowych w promieniu np. 1 kilometra

Warto zauważyć, że każdy potencjalny nabywca nieruchomości może w inny sposób priorytetyzować powyższe cechy, np. w zależności od tego czy podróżuje komunikacją zbiorową czy własnym samochodem, a także cenić sobie bardziej bliskość własnego miejsca pracy niż bliskość ogólnie pojętego centrum miasta.

3.2.3. Ogólny model cen mieszkań

Podsumowując powyższe rozważania, atrakcyjność nieruchomości jest pewną funkcją standardu mieszkania/budynku oraz atrakcyjności lokalizacji. Zakładając niezależność tych czynników, atrakcyjność (wartość) mieszkania można przedstawić jako sumę tych dwóch podstawowych składowych:

$$A = A_N + A_L + \varepsilon + \gamma = F_N(W_N, \alpha_N) + F_L(W_L, \alpha_L) + \gamma + \varepsilon \quad (3)$$

gdzie:

A – atrakcyjność mieszkania,

A_N – atrakcyjność nieruchomości (mieszkania/budynku),

A_L – atrakcyjność lokalizacji nieruchomości,

F_N – funkcja atrakcyjności nieruchomości, przypisująca zestawowi cech W_N związanych z mieszkaniem/budynkiem pewną wartość (atrakcyjność),

W_N – wektor cech związanych z mieszkaniem/budynkiem,

α_N – wektor parametrów przypisanych do cech związanych z mieszkaniem/budynkiem,

F_L – funkcja atrakcyjności lokalizacji, przypisująca zestawowi cech W_L związanych z lokalizacją nieruchomości pewną wartość (atrakcyjność),

W_L – wektor cech związanych z lokalizacją nieruchomości,

α_L – wektor współczynników przypisanych do cech związanych z lokalizacją nieruchomości,

ε – składnik losowy modelu,

γ – stała wartość (hipotetyczna cena mieszkania o zerowej atrakcyjności).

W opracowywanym modelu przyjęto założenie, że powyższe główne składowe, czyli atrakcyjność samej nieruchomości oraz atrakcyjność lokalizacji, można przedstawić jako sumy funkcji poszczególnych atrybutów. Jest to zgodne z teorią modeli hedonicznych, która zakłada, że każde mieszkanie jest charakteryzowane przez zbiór istotnych z punktu widzenia konsumenta cech.

W związku z tym atrakcyjność mieszkania/budynku będzie opisana jako suma pewnych funkcji atrybutów ją charakteryzujących:

$$A_N = f_{N1}(C_{N1}, \alpha_{N1}) + f_{N2}(C_{N2}, \alpha_{N2}) + \dots + f_{Nn}(C_{Nn}, \alpha_{Nn}) + \varepsilon \quad (4)$$

gdzie:

A_N – atrakcyjność nieruchomości (mieszkania/budynku),

f_{Ni} – i -ta składowa atrakcyjności nieruchomości, wynikająca z pewnej cechy C_i ,

C_{Ni} – i -ta cecha związana z atrakcyjnością nieruchomości, taka jak np. obecność windy,

α_{Ni} – współczynnik przypisany do i -tej cechy związanej z atrakcyjnością nieruchomości,

ε – składnik losowy modelu,

$i \in \{1, 2, 3, \dots, n\}, C_{Ni} \in W_N, \alpha_{Ni} \in \alpha_N$

Analogicznie, atrakcyjność lokalizacji będzie opisana jako suma pewnych funkcji atrybutów ją charakteryzujących:

$$A_L = f_{L1}(C_{L1}, \alpha_{L1}) + f_{L2}(C_{L2}, \alpha_{L2}) + \dots + f_{Ln}(C_{Ln}, \alpha_{Ln}) + \varepsilon \quad (5)$$

gdzie:

A_L – atrakcyjność lokalizacji nieruchomości,

f_{Li} – i -ta składowa atrakcyjności lokalizacji, wynikająca z pewnej cechy C_i ,

C_{Li} – i -ta cecha związana z atrakcyjnością lokalizacji, taka jak np. odległość od centrum,

α_{Li} – współczynnik przypisany do i -tej cechy związanej z atrakcyjnością lokalizacji,

ε – składnik losowy modelu,

$i \in \{1, 2, 3, \dots, n\}, C_{Li} \in W_L, \alpha_{Li} \in \alpha_L$

3.2.4. Założenie niezależności składowych

Tak jak wspomniano powyżej, zostało przyjęte ważne założenie o niezależności standardu inwestycji mieszkaniowych od ich lokalizacji. Jednak w rzeczywistości można dostrzec pewną korelację pomiędzy tymi czynnikami. W pewnych lokalizacjach powstaje więcej budynków w wysokim standardzie (np. apartamentowce) – dotyczy to bardziej prestiżowych lokalizacji, takich jak np. ścisłe centrum miasta czy chociażby sąsiedztwo ambasad i innych ważnych instytucji. Z drugiej strony, istnieją w miastach miejsca, gdzie powstaje dużo typowych bloków w średnim standardzie wykonania – dotyczy to zazwyczaj dużych, rozwijających się osiedli.

3.3. Wykorzystane źródła danych

Niniejszy podrozdział przedstawia dostępne i wykorzystane źródła danych. Zasadniczo wykorzystane źródła danych można podzielić na dwie kategorie: informacje o cenach i standardzie mieszkań oraz dane dotyczące aspektów związanych z ich lokalizacją w mieście. Natomiast z punktu widzenia tworzonego modelu atrakcyjności mieszkań, dane można podzielić na dane wejściowe (zmienne objaśniające – parametry modelu) oraz wartości objaśniane, czyli ceny transakcyjne mieszkań.

3.3.1. Rejestr Cen i Wartości Nieruchomości

Rejestr Cen i Wartości Nieruchomości (RCiWN) dla obszaru Warszawy prowadzi Prezydent m. st. Warszawy za pośrednictwem Biura Geodezji i Katastru Urzędu m. st. Warszawy [19].

Na informacje zawarte w RCiWN składają się dane pochodzące z:

- odpisów aktów notarialnych pochodzących z kancelarii notarialnych,
- szacunków przekazywanych przez rzeczoznawców majątkowych.

Baza RCiWN zawiera usystematyzowane dane w postaci kart informacyjnych nieruchomości: gruntowych, budynkowych i lokalowych. Karty te zawierają następujące informacje:

- data transakcji,
- nr repertorium aktu notarialnego,
- położenie nieruchomości,
- dane ewidencyjne działki, budynku, lokalu
- strony transakcji,
- rodzaj transakcji,
- informacje o nieruchomości,
- cena transakcji.

O dostęp do danych zawartych w Rejestrze Cen i Wartości Nieruchomości mogą wnioskować w szczególności [19]:

- rzeczoznawcy majątkowi – na podstawie ustawy z dnia 21 sierpnia 1997r. o gospodarce nieruchomościami (tj. Dz. U. z 2010r. Nr 102, poz. 651 ze zm.),
- uczelnie dla celów dydaktycznych – na podstawie ustawy z dnia 27 lipca 2005r. Prawo o szkolnictwie wyższym (Dz. U. z 2005r. Nr 164, poz. 1365 ze zm.).

Dane, które udało się pozyskać do celów niniejszej pracy, dotyczą transakcji zrealizowanych w okresie styczeń 2014 – luty 2016 na terenie pięciu dzielnic Warszawy: Śródmieścia, Woli, Bemowa, Żoliborza i Bielan. Transakcje dotyczą nieruchomości zarówno z rynku pierwotnego, jak i wtórnego, przy czym wykorzystane zostaną dane o transakcjach związanych z nowymi mieszkaniami. Pozyskana baza zawiera następujące dane:

- data transakcji,
- cena transakcyjna,
- powierzchnia lokalu,
- cena jednostkowa lokalu (za 1 metr kwadratowy),
- numer działki i obrębu,
- dokładny adres,
- liczba pokoi,
- numer piętra,
- dane notariusza oraz informacja o kupującym i sprzedającym
- uwagi, zawierające nieusystematyzowane dane dotyczące umowy oraz:
 - obecności garażu lub miejsca postojowego,
 - obecności i powierzchni balkonu/loggii,
 - udziału w prawach do gruntu,
 - ewentualnym wyposażeniu mieszkania,
 - rozmieszczeniu pokoi,
 - liczbie sypialni, łazienek.

Niestety, nie wszystkie transakcje są opisane kompletem danych, co uniemożliwia użycie ww. informacji (zawartych w uwagach odnośnie lokalu) do tworzenia modelu cen mieszkań. W związku z tym jedyne informacje o mieszkaniu, które zostały użyte podczas modelowania to:

- cena transakcyjna,
- powierzchnia lokalu,
- cena jednostkowa lokalu (za 1 metr kwadratowy),
- dokładny adres,
- liczba pokoi,
- numer piętra.

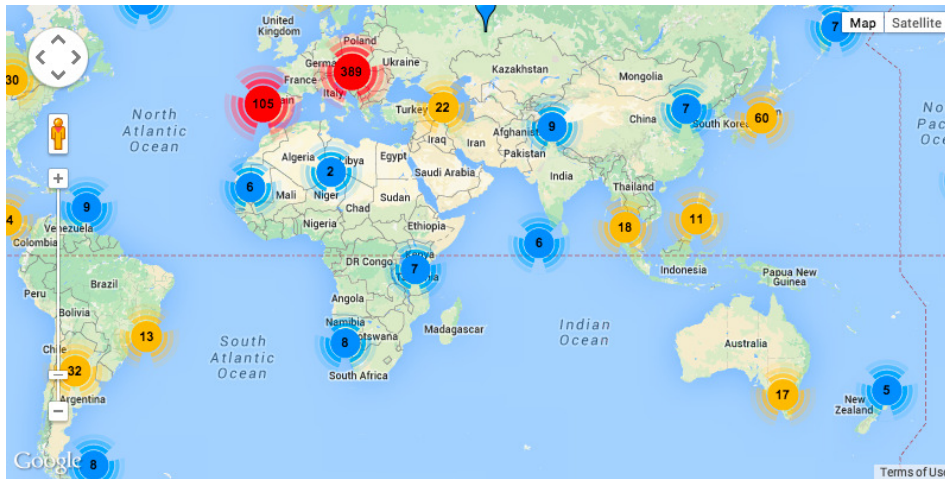
Cena jednostkowa lokalu jest zmienną objaśnianą modelem. Zostaną sformułowane i przedstawione pewne hipotezy dotyczące zależności ceny za metr kwadratowy od piętra, na którym znajduje się mieszkanie, liczby pokoi, powierzchni oraz lokalizacji. Należy przypomnieć, że lokalizacja wiąże się z wieloma czynnikami mogącymi kształtować ceny nieruchomości, np. bliskość centrum, dostępność komunikacji zbiorowej oraz punktów użyteczności publicznej itd.

3.3.2. Mapy Google

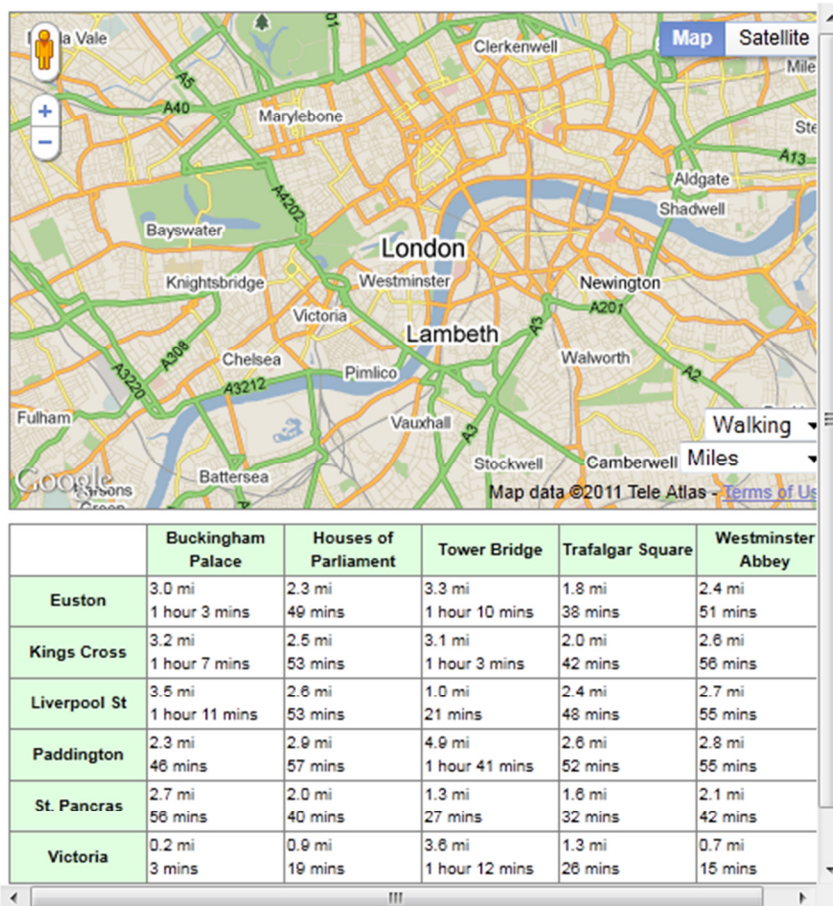
Aplikacja Mapy Google jest potężnym narzędziem oferującym wiele możliwości. Poza dostępnymi dla zwykłych użytkowników mapami świata, informacjami o komunikacji zbiorowej i ruchu drogowym oraz nawigacją, twórcy narzędzia przygotowali interfejs programistyczny (API) oferujący dodatkowe możliwości, takie jak geokodowanie (ustalanie współrzędnych geograficznych na podstawie innych danych geograficznych, np. kodu pocztowego, ulicy i miasta) za pomocą Geocoding API, zautomatyzowane obliczanie tras i czasów podróży pomiędzy dowolnymi lokalizacjami na świecie przy użyciu Distance Matrix API (rys. 8) czy wizualizacja dowolnych danych na mapie za pomocą Google Maps JavaScript API (rys. 7). Długość podróży może być określana dla ruchu pieszo, samochodem, rowerem czy z wykorzystaniem komunikacji miejskiej.

Odwołania do interfejsu Map Google w celu uzyskania danych geolokalizacyjnych (współrzędnych GPS) przypisanych dla danych adresów oraz wyznaczanie czasów podróży między lokalizacjami odbywa się za pomocą żądania HTTP GET, na które serwer Google odpowiada danymi w formacie JSON. Natomiast wizualizacja danych na mapie realizowana jest za pomocą JavaScript w przeglądarce klienta (rys. 7), przy pobraniu pewnych danych z serwerów Google.

Funkcjonalność oferowana przez Mapy Google zostanie wykorzystana do wyznaczania odległości lokalizacji od centrum miasta, czasów dojazdu do centrum miasta oraz do wizualizacji danych i wyników w celu ich prezentacji oraz ewentualnej dalszej analizy.



Rys. 7: Przykład zastosowania API Map Google do wizualizacji danych. Źródło [20]



Rys. 8: Przykład zastosowania API Map Google do obliczania czasu podróży (Distance Matrix API). Źródło [21]

Podsumowując, dzięki interfejsowi programistycznemu Map Google istnieje możliwość określenia istotnych czynników związanych z lokalizacją nieruchomości, takich jak odległość i czas dojazdu do dowolnie określonego centrum miasta czy bliskość innych istotnych miejsc znajdujących się w bliższej bądź dalszej okolicy.

3.3.3. Punkty użyteczności publicznej

Punkty użyteczności publicznej (ang. Points Of Interest - POI) są to punkty znajdujące się najczęściej na powierzchni Ziemi, które są w pewien sposób użyteczne dla ludzi. Do POI zalicza się takie miejsca jak:

- lokale gastronomiczne (bary, restauracje),
- punkty handlowo-usługowe (sklepy, centra handlowe, banki),
- placówki administracji publicznej (urzędy, komisariaty, ambasady),
- placówki oświaty (przedszkola, szkoły, uczelnie wyższe),
- atrakcje kulturalne i rozrywkowe (kino, teatr, centrum kultury),
- węzły komunikacyjne (przystanki, dworce, stacje metra),
- atrakcje sportowe, turystyczne, rekreacyjne

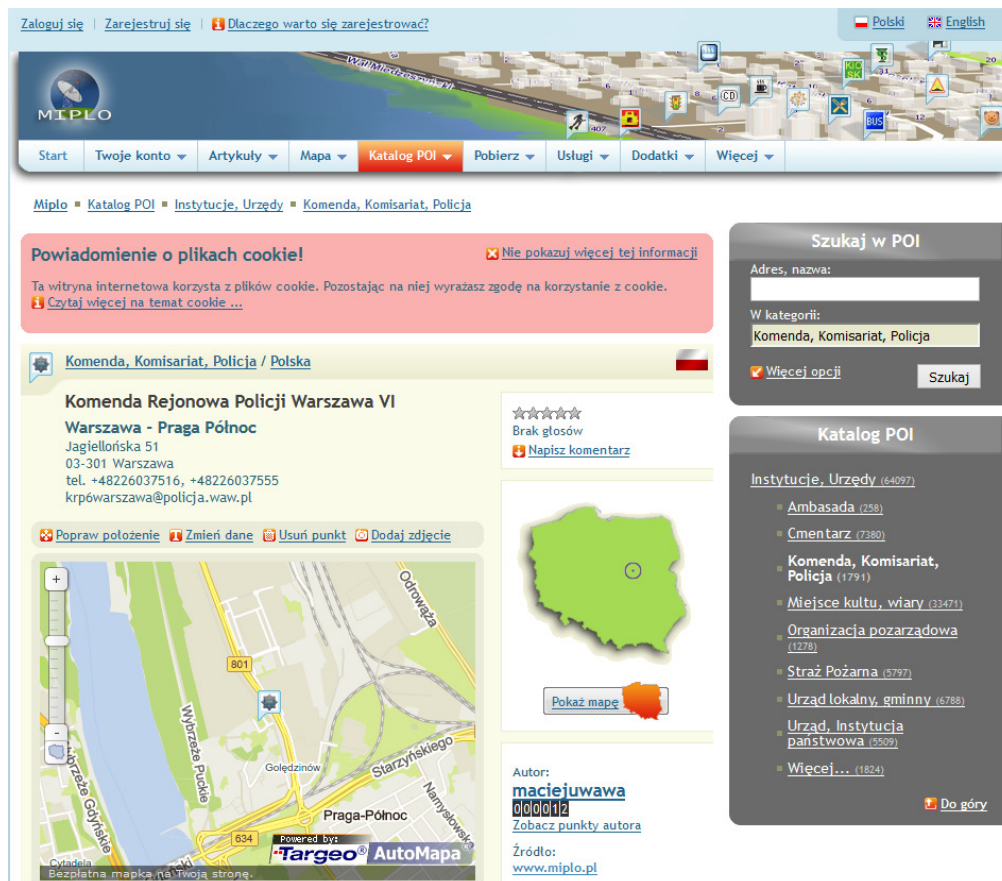
oraz wiele innych.

Punkty użyteczności publicznej są stosowane w szczególności w aplikacjach nawigacyjnych odbiorników GPS lub systemach GIS (ang. Geographic Information System, System Informacji Geograficznej). Podstawowe informacje opisujące POI to współrzędne geograficzne, kategoria, nazwa i adres.

Dane o punktach użyteczności publicznej w danym mieście mogą stanowić cenną informację o atrakcyjności lokalizacji nieruchomości, poprzez możliwość dokładnego określenia jej położenia względem miejsc istotnych z punktu widzenia komfortu życia mieszkańców, takich jak punkty handlowo-usługowe, placówki oświaty, węzły komunikacyjne, atrakcje rozrywkowe itp.

Bazy danych POI są zazwyczaj dostarczane i aktualizowane przez producentów urządzeń do nawigacji. Niestety, takie dane nie są z reguły dostępne bezpłatnie. Jednak producenci oprogramowania AutoMapa [22] prowadzą serwis zawierający codziennie aktualizowaną przez użytkowników systemu bazę punktów użyteczności publicznej –

miplo.pl [23]. Strona zawiera katalog ponad 700 tysięcy POI, spośród których ponad 53 tysiące znajduje się w Warszawie. Dane nie są możliwe do bezpłatnego pobrania ze strony, oferowane są jedynie dla posiadaczy płatnej licencji programu AutoMapa. Jednak wszystkie punkty użyteczności publicznej skatalogowane w serwisie miplo.pl można przeglądać na stronie (rys. 9). Zatem można w pewien sposób użyć tych danych do określania atrakcyjności nieruchomości.



Rys. 9: Informacja o przykładowym POI w serwisie miplo.pl.

Podsumowując, katalog punktów użyteczności publicznej dostępny w serwisie miplo.pl został wykorzystany do określenia czynników wpływających na atrakcyjność i wartość nieruchomości związanych z takimi aspektami jak dostępność usług (bliskość sklepów, lokali usługowych, placówek publicznych), bliskość węzłów komunikacyjnych oraz innych miejsc istotnych z punktu widzenia komfortu życia mieszkańców jak np. tereny zielone.

3.3.4. Czynniki opisujące mieszkania a dostępne dane

Podsumowując powyższy podrozdział należy skonfrontować składowe atrakcyjności nieruchomości opisane w podrozdziale 3.2. z dostępnymi danymi. Spośród wielu aspektów, które w mniejszym lub większym stopniu wpływają na ceny mieszkań do analizy należy wybrać te, które są dokładnie opisane lub można je jednoznacznie określić dla danego lokalu. Warto podkreślić, że czynniki związane z samym mieszkaniem lub budynkiem/inwestycją powinny być opisane dokładnie i identycznie dla wszystkich mieszkań. Niestety, dane transakcyjne z Rejestru Cen i Wartości Nieruchomości zawierają wiele nieuporządkowanych i niekompletnych danych, co uniemożliwia ich użycie w procesie modelowania cen nieruchomości.

Do analizy wykorzystane zostaną zatem następujące parametry opisujące mieszkania przedstawione w poniższej tabeli.

Tabela 4: Parametry nieruchomości wykorzystane podczas tworzenia modelu

Atrybuty związane z mieszkaniem i inwestycją	Atrybuty związane z lokalizacją
Cena lokali	Odległość od centrum miasta
Cena jednostkowa (za metr kwadratowy)	Czas dojazdu do centrum
Metraż	Odległość od wybranych punktów użyteczności publicznej
Liczba pokoi	Dostępność komunikacji zbiorowej
Piętro	

3.4. Część techniczna analizy zagadnienia

Niniejszy podrozdział dotyczy technicznej strony modelowania, czyli sposobu pozyskiwania danych w formie elektronicznej oraz ich agregacji i wykorzystania do obliczeń. Różne dane są udostępniane na różne sposoby i wymagają innych sposobów ich pozyskania i przetworzenia. Wszystkie zgromadzone informacje należało przechowywać w spójny sposób w jednym miejscu, aby były łatwo dostępne podczas wykonywania obliczeń związanych z tworzeniem i weryfikacją modelu.

3.4.1. Pozyskiwanie danych

Dane z Rejestru Cen i Wartości Nieruchomości zostały pozyskane na podstawie ustawy z dnia 27 lipca 2005r. Prawo o szkolnictwie wyższym, która umożliwia wnioskowanie o dane transakcyjne przez studentów uczelni wyższych. Tak jak wspomniano wcześniej, wykorzystane dane pochodzą z RCIWN i dotyczą transakcji zrealizowanych w okresie styczeń 2014 – luty 2016 na terenie pięciu dzielnic Warszawy: Śródmieścia, Woli, Bemowa, Żoliborza i Bielan.

Dane dotyczące punktów użyteczności publicznej ze strony miplo.pl nie są dostępne bezpłatnie w przystępnej formie, jednak można je przeglądać na stronie dostawcy – umożliwiło to użycie techniki web scraping do ich pozyskania. Metoda polega na zautomatyzowanym wysyłaniu zapytań HTTP do serwera www w celu pobrania wszystkich potrzebnych informacji. W ramach pracy został przygotowany skrypt w języku Python, który automatycznie przeanalizował ponad 50 000 stron www i wyodrębnił z kodu HTML informacje o identyfikatorze, nazwie oraz kategorii POI. Unikatowy identyfikator POI został wykorzystany do pobrania współrzędnych geograficznych odpowiadających punktowi za pomocą żądania HTTP GET wysłanego do serwera miplo.pl:

```
http://www.miplo.pl/c/point.php?p=xxxxxxxx
```

gdzie xxxxxxxx to unikatowy ID punktu. Serwer za każdym razem zwracał dokładne współrzędne GPS w formacie JSON, które zostały przypisane danemu punktowi i wykorzystane w dalszym etapie do określania parametrów związanych z lokalizacją mieszkań.

Interfejs Map Google został wykorzystany do określenia dokładnych lokalizacji wszystkich adresów inwestycji, dla których pozyskano informacje o transakcjach. Spośród wszystkich danych transakcyjnych wyodrębniono unikatowe adresy. Użycie API polegało na wysłaniu dwóch rodzajów żądań HTTP GET do serwera Google:

- https://maps.googleapis.com/maps/api/geocode/json?address=ADRES_INWESTYCJI&key=KLUCZ_API – zwracającego współrzędne GPS (długość i szerokość geograficzną) w formacie JSON,
- https://maps.googleapis.com/maps/api/distancematrix/json?origins=ADRES_INWESTYCJI&destinations=ADRES_CENTRUM&key=KLUCZ_API –

zwracającego odległość i czas podróży między lokalizacją dane inwestycji a arbitralnie określonym centrum miasta w formacie JSON.

Pozyskane dane zostały później wykorzystane przy określaniu atrakcyjności lokalizacji nieruchomości.

3.4.2. Agregacja danych

Zgromadzone dane były przechowywane w tablicy słownikowej (dictionary) dostępnych w języku Python. Tablica składała się z elementów przypisanych kluczom – adresom inwestycji. Każdy element tablicy składał się z następujących danych:

- lista mieszkań w danej inwestycji,
- współrzędne geograficzne danej inwestycji,
- odległość i czas podróży do centrum miasta,
- liczba punktów użyteczności publicznej w określonej odległości od inwestycji.

Natomiast każde mieszkanie było opisane przez zestaw zmiennych:

- cena całkowita,
- cena za metr kwadratowy,
- metraż,
- liczba pokoi,
- piętro.

W celu szybszego dostępu do danych podczas wykonywania regresji informacje dotyczące wszystkich mieszkań były dodatkowo przechowywane w formie listy, która zapewnia szybszy dostęp do danych.

Dane o punktach użyteczności publicznej przechowywane były w liście. Każdy element zawierał następujące informacje o POI:

- identyfikator,
- nazwa,
- nazwa kategorii,
- numer kategorii,
- współrzędne geograficzne.

Odległości pomiędzy lokalizacjami mieszkań a wybranymi punktami użyteczności publicznej były wyznaczane za pomocą metody Vincentego, która polega na iteracyjnym obliczaniu przybliżonej odległości dwu punktów znajdujących się na powierzchni sferoidy [24], z wykorzystaniem pozyskanych danych o współrzędnych geograficznych inwestycji i obiektów.

3.5. Sformułowanie hipotez dotyczących zależności ceny mieszkań od czynników

Niniejszy podrozdział przedstawia hipotezy dotyczące zależności cen mieszkań od wybranych czynników. Hipotezy zostały sformułowane na podstawie pewnych wniosków wyciągniętych z analizy rynku mieszkań oraz opracowań zarówno bezpośrednio związanych z zagadnieniem modelowania cen (omówionych w rozdziale drugim) jak i dotyczących uwarunkowań przestrzennych i komunikacyjnych m. st. Warszawy. Zgodnie z ogólnym modelem cen mieszkań przedstawionym w rozdziale 3.2.3 wartość mieszkania jest sumą funkcji atrybutów je opisujących.

3.5.1. Parametry mieszkań

Tabela 4 przedstawia cechy opisujące mieszkania, które dostępne w pozyskanych danych i zostały ujęte w analizie. Cena za metr kwadratowy jest zmienną objaśnianą w modelu, natomiast cena całkowita wynika bezpośrednio z metrażu mieszkania. Zatem czynnikami mogącymi wpływać na wartości nieruchomości są: metraż, liczba pokoi oraz piętro.

Przeglądając i analizując oferty deweloperów na rynku można zauważyć, że ceny identycznych mieszkań w jednym pionie bardzo często rosną wraz z numerem piętra, na którym lokal się znajduje. Sugeruje to, że potencjalni nabywcy rzeczywiście bardziej cenią sobie mieszkania na wyższych piętrach. Może być to spowodowane ładniejszym widokiem z okien takiego lokalu, większym nasłonecznieniem czy mniejszym natężeniem hałasu dochodzącego z dziedzińca czy ulicy. Obecnie, gdy właścicielom wszystkie budynki wyposażone są w windy, mieszkanie na wyższym piętrze nie wiąże się z żadnym dyskomfortem, który mógłby być spowodowany koniecznością pokonywania wielu stopni pieszo.

Poniżej przedstawiono ceny wybranych lokali w kilku realizowanych obecnie inwestycjach mieszkaniowych. Wybrane zostały mieszkania w tych samych pionach, lecz na różnych piętrach.

Tabela 5: Porównanie cen identycznych mieszkań na różnych piętrach.

Adres inwestycji	Kasprzaka		Lirowa		Instalatorów	
Piętro i cena	Mieszkanie 84,05 m ²		Mieszkanie 55,34 m ²		Mieszkanie 35,80 m ²	
	II	690 709 zł	I	453 788 zł	I	243 451 zł
	III	713 748 zł	II	453 788 zł	II	247 519 zł
	IV	718 662 zł	III	459 322 zł	III	253 122 zł
	V	737 914 zł	IV	459 322 zł	IV	260 193 zł

Można zauważyć, że ceny mieszkań rosną wraz z piętrem – niekoniecznie każda kondygnacja oznacza wzrost ceny, czasami mieszkania na kilku sąsiadujących piętrach kosztują tyle samo. Jednak widać ogólną tendencję wzrostu cen wraz z wysokością, na jakiej lokal się znajduje.

Kolejnym czynnikiem, który może wpływać na ceny jednostkowe mieszkań, jest ich powierzchnia. Rzeczywiście, przyglądając się ofertom deweloperów na rynku, można zauważyć pewny trend spadkowy ceny lokalu wraz ze wzrostem jego powierzchni. Można to intuicyjnie uzasadnić w następujący sposób: częścią wartości mieszkania jest sam komfort jego posiadania, niezależny od dostępnej powierzchni użytkowej.

Można więc założyć, że mieszkanie kosztuje $A + Bm$ zł, gdzie A to wartość hipotetycznego mieszkania o zerowej powierzchni (czyli cena „dachu nad głową”), B to cena za każdy metr kwadratowy lokalu a m to jego metraż. W praktyce cena mieszkania jest określana jako iloczyn metrażu i pewnej ceny jednostkowej C przedstawionej w ofercie i obecnej w danych transakcyjnych. Zakładając dalej, że wartości A i B są stałe, tzn. że B jest pewną rzeczywistą wartością każdego metra kwadratowego mieszkania, można porównać obie ceny. Tak więc $A + Bm = Cm$, gdzie $A, B = const$, w związku z tym $C = \frac{A}{m} + B$. Wynika z tego, że w pewnym uproszczeniu cena jednostkowa mieszkania jest proporcjonalna od odwrotności jego metrażu, a więc większe mieszkania kosztują mniej, co jest zgodne z obserwowalnym na rynku trendem.

Tabela 6: Porównanie cen mieszkań o różnym metrażu na tych samych piętrach.

Adres inwestycji	Kasprzaka		Garażowa		Sikorskiego	
Metraż i cena jednostkowa	Mieszkania I p.		Mieszkania I p.		Mieszkania I p.	
	42,57 m ²	8350 zł	41,24 m ²	9100 zł	39,78 m ²	8712 zł
	59,26 m ²	8200 zł	48,06 m ²	8800 zł	46,13 m ²	8650 zł
	80,78 m ²	8100 zł	55,34 m ²	8600 zł	59,37 m ²	8594 zł
	91,26 m ²	8040 zł	70,70 m ²	8100 zł	74,92 m ²	8446 zł

Zaobserwowane na rynku mieszkaniowym trendy nie do końca wpisują się w teoretyczną zależność, aczkolwiek można przyjąć, że odwrotność metrażu jest jednym z czynników modelującym wartości nieruchomości.

Liczba pokoi jest cechą bezpośrednio wynikającą z metrażu mieszkania, choć można trafić na lokale (w tym samym budynku) o podobnej powierzchni różniące się liczbą pokoi, maksymalnie o jeden. Jednak na rynku nie występuje wiele takich mieszkań, co utrudnia analizę i dokładne określenie zależności. W związku z tym powyższy czynnik został pominięty w modelowaniu.

Podsumowując, atrakcyjność wynikająca z parametrów mieszkań jest sumą funkcji atrybutów je charakteryzujących zgodnie ze wzorem (4):

$$A_N = f_{N1}(C_{N1}, \alpha_{N1}) + f_{N2}(C_{N2}, \alpha_{N2}) + \dots + f_{Nn}(C_{Nn}, \alpha_{Nn}) \quad (4)$$

Biorąc pod uwagę zmienne charakteryzujące mieszkanie (piętro i metraż) oraz przyjęte zależności, atrakcyjność mieszkania wynikającą z aspektów związanych wyłącznie z samym lokalem określona jest jako:

$$A_N = \alpha_{N1}p + \frac{\alpha_{N2}}{m} \quad (6)$$

gdzie:

A_N – atrakcyjność nieruchomości (mieszkania/budynku),

p – piętro, na którym znajduje się lokal,

m – metraż lokalu,

α_{N1}, α_{N2} – współczynniki liniowej zależności.

3.5.2. Cechy lokalizacji

Oczywistym jest, że w każdym większym mieście ceny mieszkań są wyższe w bliższej odległości od jego centrum. Jest to spowodowane bliskością strategicznych lokalizacji w mieście, lepszą dostępnością punktów użyteczności publicznej. Śródmieścia są reprezentacyjnymi, prestiżowymi i chętnie odwiedzanymi częściami miast. Ponadto, na wysoką cenę mieszkań zlokalizowanych w centrum wpływają ceny działek, które również wynikają z powyższych czynników. Dodatkowo, takich działek jest niewiele, ze względu na gęstą zabudowę centralnych części miast, co dodatkowo wpływa na ich cenę.

Jednak z punktu widzenia mieszkańców, istotna jest nie tylko odległość, ale również czas potrzebny na dostanie się do centrum miasta przy użyciu własnego samochodu lub komunikacji zbiorowej. Tak więc mieszkania znajdujące się blisko ważnych ciągów komunikacyjnych czy węzłów przesiadkowych są zazwyczaj droższe niż te zlokalizowane w gorzej skomunikowanych miejscach. Dobrze skomunikowana lokalizacja stanowi wartość samą w sobie, gdyż pozwala na sprawniejsze poruszanie się po całym mieście, nie tylko podróżowanie do centrum. Dlatego np. mieszkania znajdujące się na obrzeżach miast, ale położone w bliskim sąsiedztwie stacji metra są bardziej atrakcyjne niż te bliżej centrum, ale leżące w słabo skomunikowanej okolicy (np. oddzielone od reszty miasta trudnymi do pokonania przeszkodami terytorialnymi takimi jak tory kolejowe).

Ludzie kupujący mieszkania cenią sobie również bliskość punktów użyteczności publicznej takich jak sklepy, centra handlowe, lokale gastronomiczne, szkoły i przedszkola, przychodnie czy apteki. Wynika to z zakresu typowych codziennych aktywności mieszkańców, oczywiście poza pracą, takich jak wyprawienie dzieci do szkoły, zakupy czy wizyta u lekarza. Ma to również odzwierciedlenie w reklamach i broszurach deweloperów, którzy chętnie zachwalają dogodną lokalizację swoich inwestycji w odniesieniu do omawianych punktów użyteczności publicznej.

Dodatkowo, o dogodności lokalizacji świadczy również liczba punktów usługowych w zasięgu krótkiego spaceru, czyli kilkuset metrów. Duża liczba przydatnych miejsc na osiedlu i w jego okolicy wpływa pozytywnie na wygodę mieszkańców. Ma to odzwierciedlenie np. w omawianym wcześniej wskaźniku WalkScore.

Podczas tworzenia modelu zostały wykorzystane zatem następujące cechy lokalizacji nieruchomości:

- odległość od centrum
- czas dojazdu do centrum
- odległość od najbliższego lokalu gastronomicznego,
- odległość od najbliższego centrum handlowego,
- odległość od najbliższej szkoły/przedszkola,
- odległość od najbliższej apteki,
- odległość od najbliższej przychodni,
- odległość od najbliższej placówki pocztowej,
- odległość od najbliższego przystanku komunikacji zbiorowej,
- liczba POI powyższych kategorii w promieniu 500 metrów.

Aby wykorzystać lokalizacje powyższych punktów w procesie modelowania, należy znać ich dokładne położenie. O ile współrzędne geograficzne punktów użyteczności publicznej są znane, o tyle centrum miasta jest dość szerokim pojęciem. Należy zatem precyzyjnie określić jedną lokalizację, która będzie jak najlepiej reprezentować centrum Warszawy.

Ważną cechą śródmieścia jest to, że skupia wiele miejsc pracy, lokali usługowych, obiektów związanych z kulturą i rozrywką, a także jest bardzo dobrze skomunikowane – leży przy ważnych ciągach komunikacyjnych, popularnych przystanków autobusowych, tramwajowych i stacji metra. W powszechnym rozumieniu, centrum Warszawy stanowi Rondo R. Dmowskiego znajdujące się u zbiegu Alei Jerozolimskich oraz ulicy Marszałkowskiej. Wiele map wskazuje ten punkt jako środek stolicy.

Jednak ściśle określając punkt, który zostanie przyjęty za centrum miasta w modelu, należy wziąć pod uwagę inne aspekty. W ciągu ostatnich lat zrealizowanych zostało wiele prestiżowych inwestycji mieszkaniowych i biurowych na północny zachód od wspomnianego Ronda Dmowskiego. Ponadto uruchomiona została II linia metra, która krzyżuje się z pierwszą również w pewnej odległości od Ronda. Analizując fachowe opracowania zagadnienia, takie jak Studium Komunikacyjne Śródmieścia Warszawy [25], przedstawiające wyniki pomiarów natężenia ruchu generowanego w dni robocze przez budynki biurowe w centrum stolicy, można stwierdzić, że centrum miasta, rozumiane jako typowy cel podróży

związanych z codziennymi dojazdami, do pracy jest rzeczywiście przesunięte w kierunku północno-zachodnim od Ronda Dmowskiego.

W związku z wymienionymi faktami i czynnikami w niniejszej pracy postanowiono za centrum m. st. Warszawy przyjąć okolice Ronda ONZ. Rondo ONZ leży przy ważnych ciągach komunikacyjnych (Al. Jana Pawła II, ul. Świętokrzyska/Prosta, II linia metra, skrzyżowanie linii tramwajowych), w jego bliskim sąsiedztwie powstało i nadal powstaje wiele prestiżowych budynków mieszkalnych i biurowych. Względem tego miejsca określano odległość poszczególnych nieruchomości rozumianą jako odległość od centrum oraz czas dojazdu do centrum.

Wszystkie powyżej omówione aspekty związane z lokalizacją mieszkania odnoszą się do odległości od szczególnych miejsc w mieście, takich jak centrum, punkty komunikacyjne czy punkty użyteczności publicznej. Jednak atrakcyjność lokalizacji jest związana de facto z bliskością omawianych miejsc. Tak więc należy określić klasę funkcji, która będzie modelować zależność wartości mieszkań od wyżej wymienionych czynników. Należy tu podkreślić, że czas podróży do centrum miasta może być traktowany jako odległość od centrum, wyrażona jednostkach czasu zamiast odległości.

Przed wszystkim, atrakcyjność lokalizacji rośnie wraz ze zmniejszaniem odległości od wybranych punktów, można więc założyć że jest do niej odwrotnie proporcjonalna. Jednak w przypadku prostej zależności $y \sim \frac{1}{x}$, atrakcyjność wynikająca z bliskości poszczególnych kategorii istotnych miejsc rosłaby w nieskończoność wraz ze zbliżaniem się nich. Dlatego funkcja modelująca powinna przyjąć formę $y \sim \frac{1}{x+a}$, gdzie a jest pewną stałą, którą można interpretować jako odwrotność hipotetycznej atrakcyjności wynikającej z bezpośredniego położenia nieruchomości przy wybranym istotnym punkcie, czyli maksymalnej atrakcyjności wynikającej z bezpośredniego sąsiedztwa tego punktu.

Tak więc zgodnie ze wzorem (5):

$$A_L = f_{L1}(C_{L1}, \alpha_{L1}) + f_{L2}(C_{L2}, \alpha_{L2}) + \dots + f_{Ln}(C_{Ln}, \alpha_{Ln}) \quad (5)$$

atrakcyjność lokalizacji jest sumą atrakcyjności wynikających z bliskości poszczególnych kategorii miejsc istotnych z punktu widzenia komfortu mieszkańców:

$$A_L = \sum_{i=1}^n \frac{\alpha_{Li}}{x_i + C_i} + \beta L_p \quad (7)$$

gdzie:

A_L – atrakcyjność lokalizacji nieruchomości,

x_i – odległość punktu i -tej kategorii,

C_i – odwrotność maksymalnej atrakcyjności związanej z punktem i -tej kategorii,

α_{Li}, β – współczynniki liniowej zależności,

n – liczba kategorii POI objaśniających wartość nieruchomości,

L_p – liczba POI wybranych kategorii w promieniu 500 metrów,

$i \in \{1, 2, 3, \dots, n\}$.

Podsumowując wszystkie powyższe rozważania, proponowany model wskaźnika atrakcyjności mieszkań wygląda następująco:

$$A = \alpha_{N1}p + \frac{\alpha_{N2}}{m} + \sum_{i=1}^n \frac{\alpha_{Li}}{x_i + C_i} + \beta L_p + \gamma + \varepsilon \quad (8)$$

gdzie A jest atrakcyjnością (a więc ceną) danego mieszkania, natomiast pozostałe parametry (stałe i zmienne) zostały wyjaśnione wcześniej.

3.6. Weryfikacja postawionych hipotez

Niniejszy podrozdział przedstawia wyniki analiz przeprowadzonych na zgromadzonych danych w celu weryfikacji postawionych wcześniej hipotez o zależności wartości mieszkań od wybranych czynników. Zostało sprawdzone, jak duży wpływ na cenę lokalu ma piętro, na którym się znajduje, metraż oraz odległość od centrum miasta (Rondo ONZ) i wybranych punktów użyteczności publicznej.

Miarą dopasowania modelu był współczynnik determinacji R^2 określony w następujący sposób:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

gdzie:

y_i – rzeczywista wartość zmiennej y dla i -tego pomiaru,

\hat{y}_i – wartość zmiennej objaśnianej na podstawie modelu dla i -tego pomiaru,

\bar{y} – średnia arytmetyczna rzeczywistych wartości zmiennej objaśnianej,

n – wielkość próby danych,

$i \in \{1,2,3, \dots, n\}$.

3.6.1. Filtrowanie danych

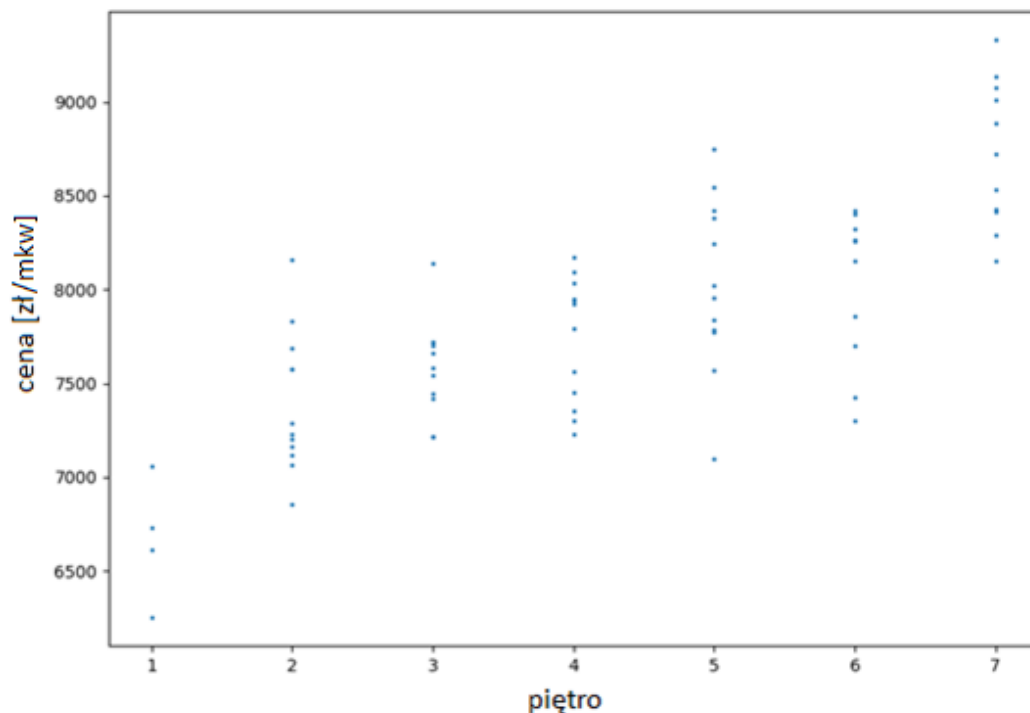
Analizy analogiczne do opisanych poniżej zostały przeprowadzone dla danych po wstępnym odfiltrowaniu. Wybrano mieszkania znajdujące się na piętrach 2-5, aby uzyskać zależność od kondygnacji jak najbardziej zbliżoną do liniowej (zakładając odchylenia na skrajnych piętrach). Następnie skupiono się na mieszkaniach o podobnym metrażu (dwu- i trzypokojowych), aby skupić się wyłącznie na zależności ceny od kondygnacji. Transakcje filtrowano również pod względem ich wartości – wstępnie odrzucono lokale o cenach za metr znacznie odbiegających od średniej dla danej inwestycji (pozostawiono nieruchomości o cenach różnych o mniej niż 2 odchylenia standardowe od średniej).

Filtrowanie danych nie wpłynęło jednak prawie w żadnym stopniu na wyniki analizy i dokładność modelowania. Przy uwzględnieniu wszystkich transakcji otrzymywano podobne wyniki i nie gorsze parametry modelu co przy pracy na filtrowanych danych. Tak więc, jeżeli nie zaznaczono inaczej, analizie poddawane były wszystkie pozyskane dane transakcyjne.

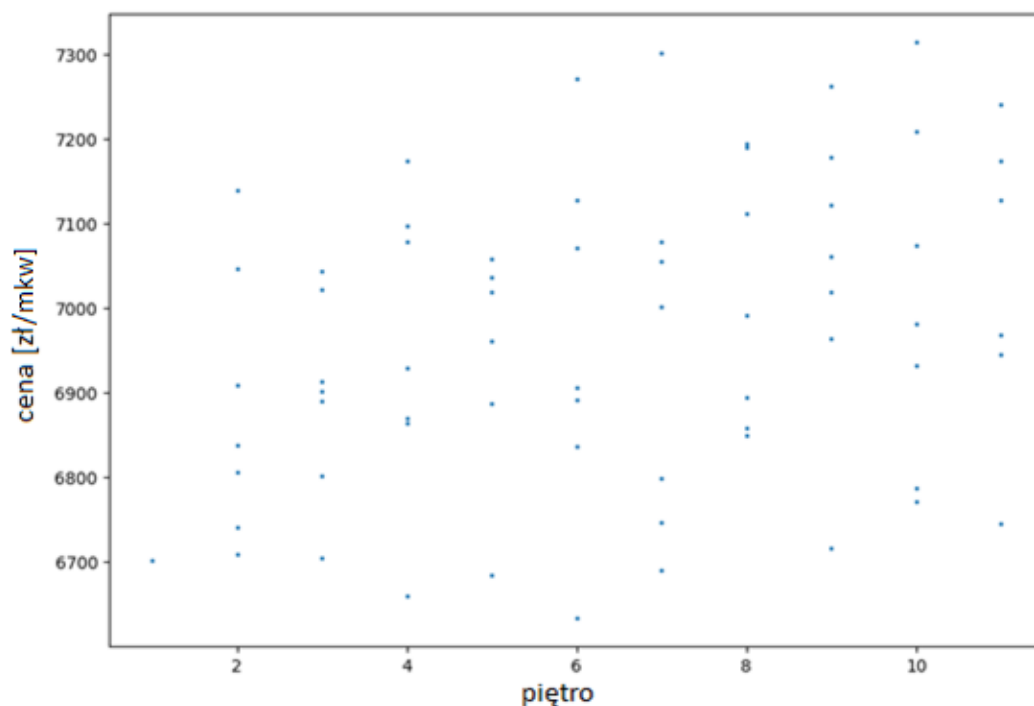
3.6.2. Zależność od numeru piętra

Pierwszą z postawionych hipotez był liniowy wzrost ceny mieszkań wraz z numerem piętra, na którym się ono znajduje. W celu weryfikacji hipotezy przeanalizowano wybrane inwestycje, dla których pozyskano informacje o dużej liczbie transakcji (większej niż 20) znajdujących się na różnych kondygnacjach.

Dla różnych budynków otrzymano różne wyniki. W przypadku niektórych widać wyraźny trend wzrostu ceny wraz z piętrem (rys. 10), natomiast w przypadku innych nie da się dostrzec tej zależności (rys. 11). Może to być spowodowane faktem, że mieszkania w obrębie tych samych pięter znacznie się różnią (np. rozkładem pokoi, naświetleniem, obecnością balkonu/loggii), co powoduje, że ich ceny są zróżnicowane w dużym stopniu, przez co trend związany z numerem kondygnacji jest niemożliwy do dostrzeżenia.



Rys. 10: Wykres zależności ceny jednostkowej mieszkań od piętra dla inwestycji przy ul. Bryłowskiej 2 (74 mieszkania). Zauważalny jest wyraźny wzrost cen za metr kwadratowy wraz z numerem piętra.



Rys. 11: Wykres zależności ceny jednostkowej mieszkań od piętra dla inwestycji przy ul. Sowińskiego 53E (69 mieszkań). Ceny mieszkań na poszczególnych piętrach różnią się znacznie niż pomiędzy kondygnacjami, trudno dostrzec jakąkolwiek zależność ceny od piętra, na którym znajduje się mieszkanie.

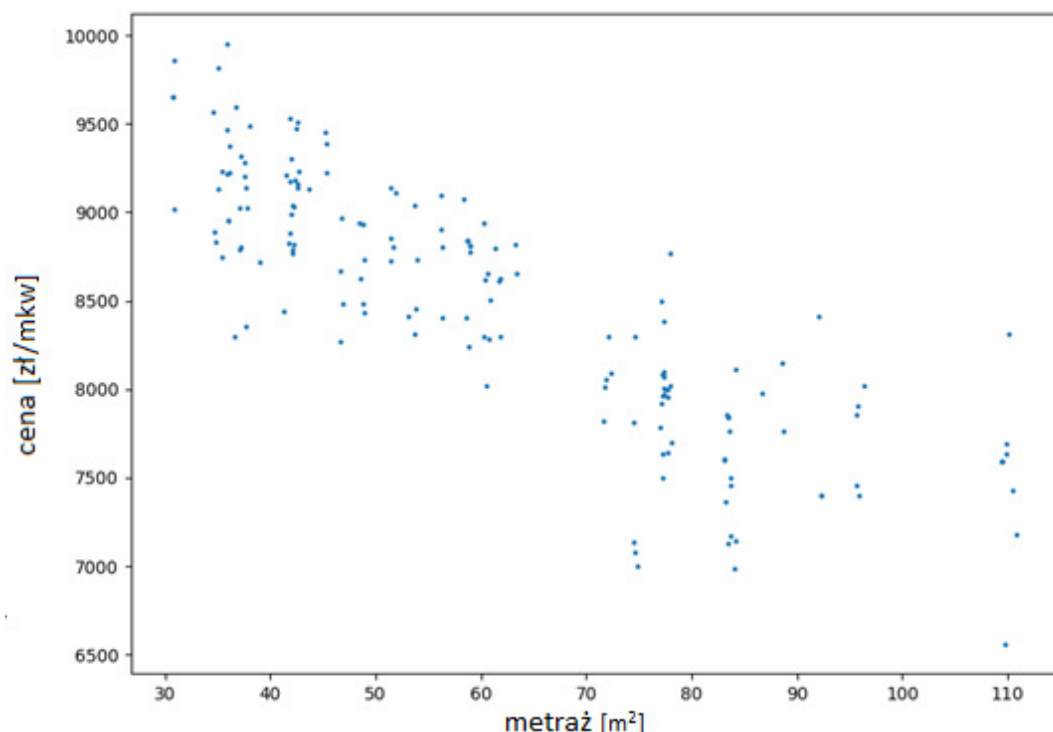
Po przeprowadzeniu regresji względem piętra dla poszczególnych inwestycji otrzymano współczynniki determinacji wahające się od 0,13 do 0,85. Średnia wartość

współczynnika R^2 ważonego liczbą mieszkań w inwestycji wyniosła 0,33. Podsumowując, zależność liniowa wartości mieszkania od kondygnacji jest wyraźna jedynie w przypadku niektórych inwestycji.

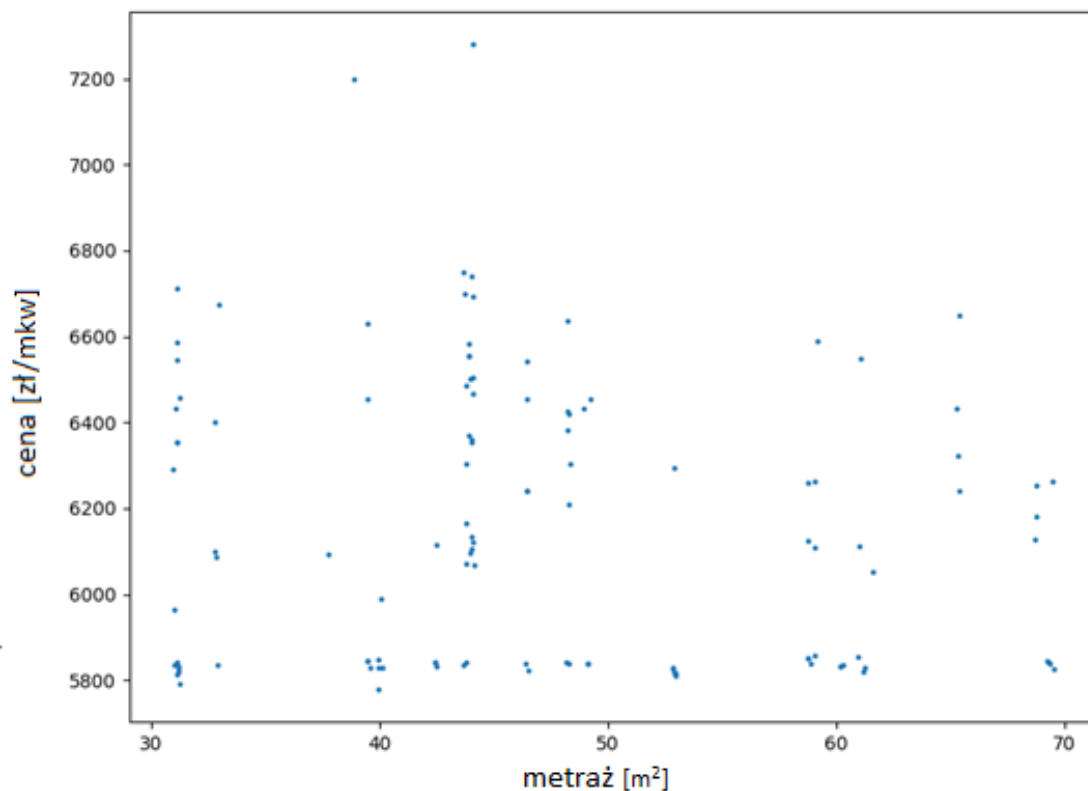
3.6.3. Zależność od metrażu

Kolejną hipotezą była odwrotna proporcjonalność wartości mieszkań wraz do metrażu. W celu weryfikacji hipotezy przeanalizowano wybrane inwestycje z dużą liczbą lokali o różnych powierzchniach użytkowych.

Dla różnych budynków otrzymano różne wyniki. W przypadku niektórych widać wyraźny trend spadku ceny wraz ze wzrostem powierzchni użytkowej (wzrost ceny przy wzroście odwrotności metrażu; rys. 12), natomiast w przypadku innych trudno zauważyć tę zależność (rys. 13). Może to być spowodowane faktem, tak jak w przypadku zależności ceny od kondygnacji, że mieszkania o podobnej powierzchni znacznie się różnią (np. rozkładem pokoi, naświetleniem, obecnością balkonu/loggii), co powoduje, że ich ceny są zróżnicowane w dużym stopniu, przez co trend związany z metrażem jest niemożliwy do dostrzeżenia.



Rys. 12: Wykres zależności ceny jednostkowej mieszkań od metrażu dla inwestycji przy ul. Siedmiogrodzkiej 1 (163 mieszkania). Zauważalny jest wyraźny spadek cen za metr kwadratowy wraz ze wzrostem powierzchni użytkowej mieszkania.



Rys. 13: Wykres zależności ceny jednostkowej mieszkań od metrażu dla inwestycji przy ul. Batalionów Chłopskich 87C (116 mieszkań). Ceny mieszkań o podobnych powierzchniach różnią się znacznie niż w zależności od metrażu, trudno dostrzec jakąkolwiek zależność ceny od powierzchni użytkowej lokalu.

Po przeprowadzeniu regresji względem odwrotności metrażu dla poszczególnych inwestycji otrzymano współczynniki determinacji wahające się od 0,09 do 0,74. Średnia wartość współczynnika R^2 ważonego liczbą mieszkań w inwestycji wyniosła jedynie 0,26. Podsumowując, zależność liniowa wartości mieszkania od kondygnacji jest wyraźna jedynie w przypadku niektórych inwestycji.

3.6.4. Zależność od piętra i metrażu

Następnym krokiem w analizie zależności było przeprowadzenie regresji z wykorzystaniem dwóch zmiennych objaśniających wartość (atrakcyjność) lokalu: kondygnacji oraz metrażu. Zostanie więc przyjęta zależność wynikająca ze wzorów (7) i (8):

$$A = A_N + A_L + \varepsilon + \gamma = \alpha_{N1}p + \frac{\alpha_{N2}}{m} + A_L + \gamma + \varepsilon \quad (10)$$

gdzie:

A – atrakcyjność mieszkania,

A_N – atrakcyjność nieruchomości (mieszkania/budynku),

A_L – atrakcyjność lokalizacji nieruchomości,

p – piętro, na którym znajduje się lokal,

m – metraż lokalu,

α_{N1}, α_{N2} – współczynniki liniowej zależności.

Wartości współczynników α_{N1}, α_{N2} oraz A_L (przyjęta w tym momencie jako stała) zostaną wyznaczone poprzez zastosowanie regresji liniowej.

Oczywiście, tak jak w poprzednich przypadkach, dla różnych budynków otrzymano różne wyniki. Jednak tym razem współczynniki determinacji przyjmowały wartości z zakresu od 0,16 do 0,91, natomiast średnia wartość R^2 ważona liczbą mieszkań w inwestycji wyniosła 0,58.

Zależność wartości mieszkania od piętra i metrażu postanowiono wykorzystać w modelu, przyjmując stałe współczynniki proporcjonalności ceny do kondygnacji i odwrotności powierzchni użytkowej α_{N1} i α_{N2} . Oba współczynniki zostały zdefiniowane jako średnie wartości takich współczynników wyznaczonych za pomocą regresji liniowej dla wszystkich inwestycji, ważone liczbą mieszkań w każdym budynku. Wartości omawianych współczynników wyniosły:

$$\alpha_{N1} = 136,74 \text{ [zł]},$$

$$\alpha_{N2} = 31421 \text{ [zł]}$$

i takie zostały przyjęte do dalszej analizy, czyli weryfikacji zależności cen lokali od parametrów związanych z lokalizacją.

3.6.5. Zależność od odległości od centrum

Kolejną postawioną hipotezą była zależność ceny od odwrotności odległości od centrum. W celu weryfikacji hipotezy przeanalizowano ustalone średnie ceny mieszkań dla wszystkich omawianych inwestycji.

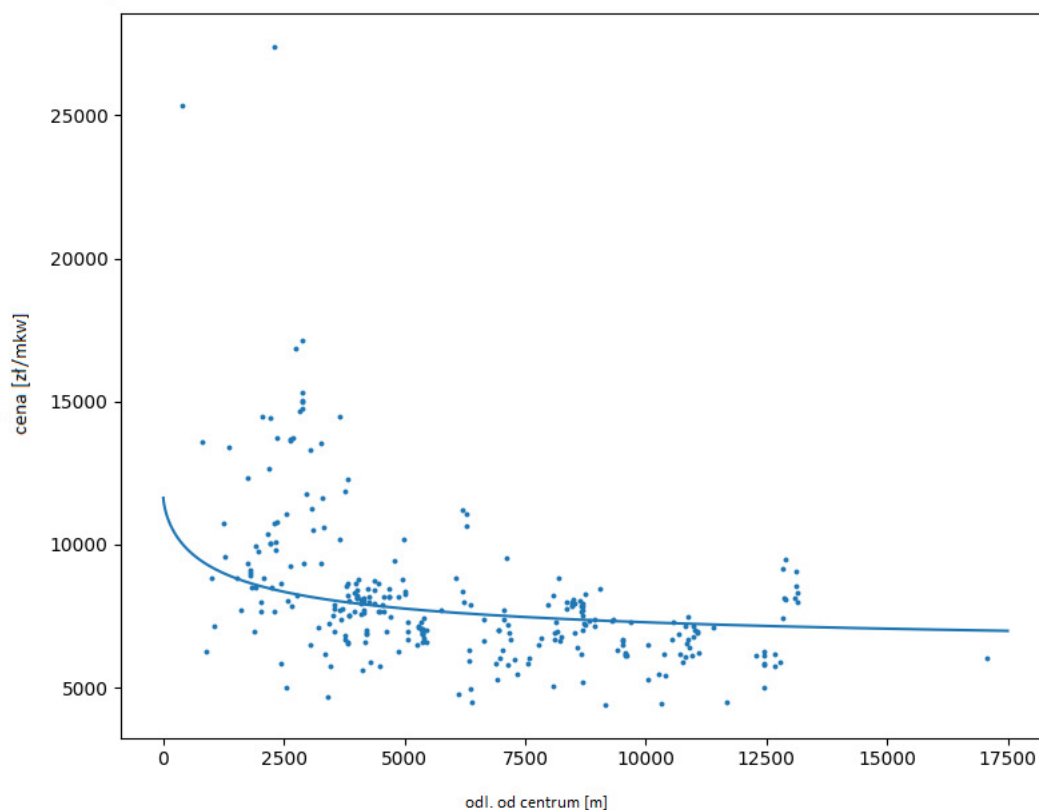
Pierwszym krokiem było ustalenie pewnej ceny bazowej mieszkania dla każdego budynku. Dokonano tego na trzy sposoby i wyznaczono trzy różne wartości:

1. średnia arytmetyczna cen wszystkich mieszkań w obrębie danej inwestycji (rys. 14),

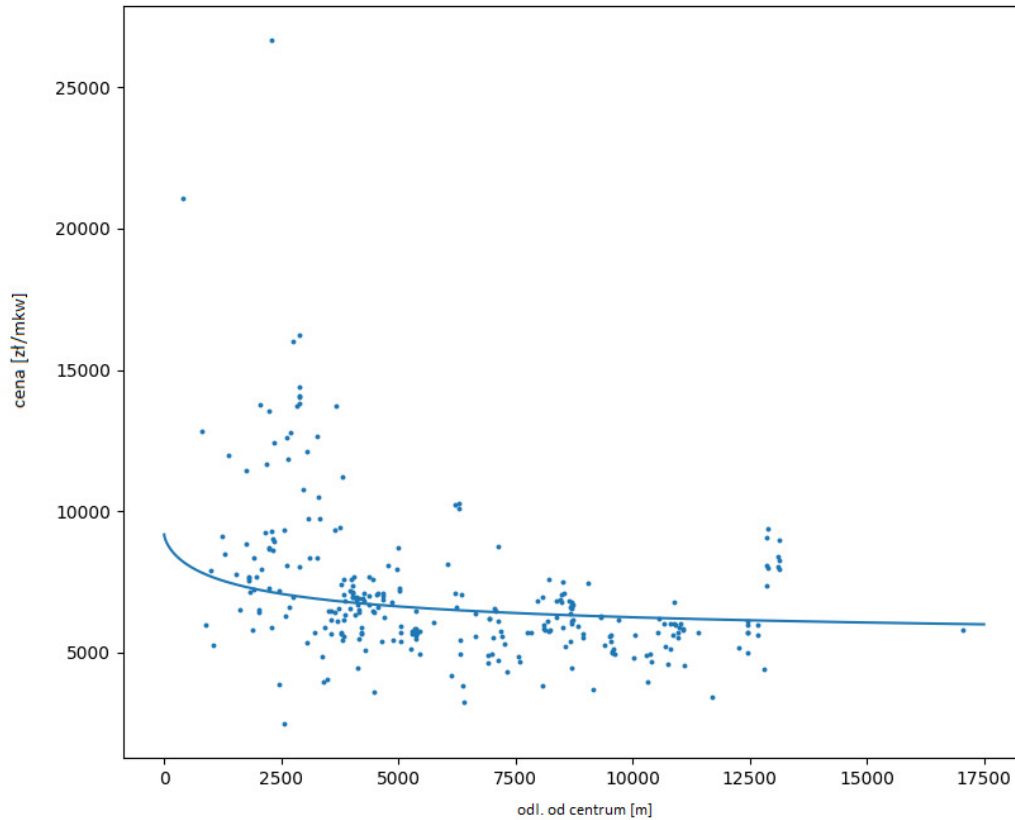
- średnia arytmetyczna cen wszystkich mieszkań z uwzględnieniem korekty wynikającej z założonej wcześniej zależności ceny od piętra i metrażu (rys. 15).
- stała będąca wynikiem przeprowadzonej regresji względem kondygnacji i metrażu dla każdej inwestycji z osobna (tylko dla wybranych inwestycji, dla których przeprowadzono regresję) (rys. 16).

Zgodnie z hipotezą, wartość (atrakcyjność) mieszkania wynosi $A = \frac{C}{x+D} + B$, gdzie C , D i B to pewne stałe, x to odległość lokalizacji od centrum (Ronda ONZ), natomiast A to atrakcyjność, (czyli cena jednostkowa) lokalu.

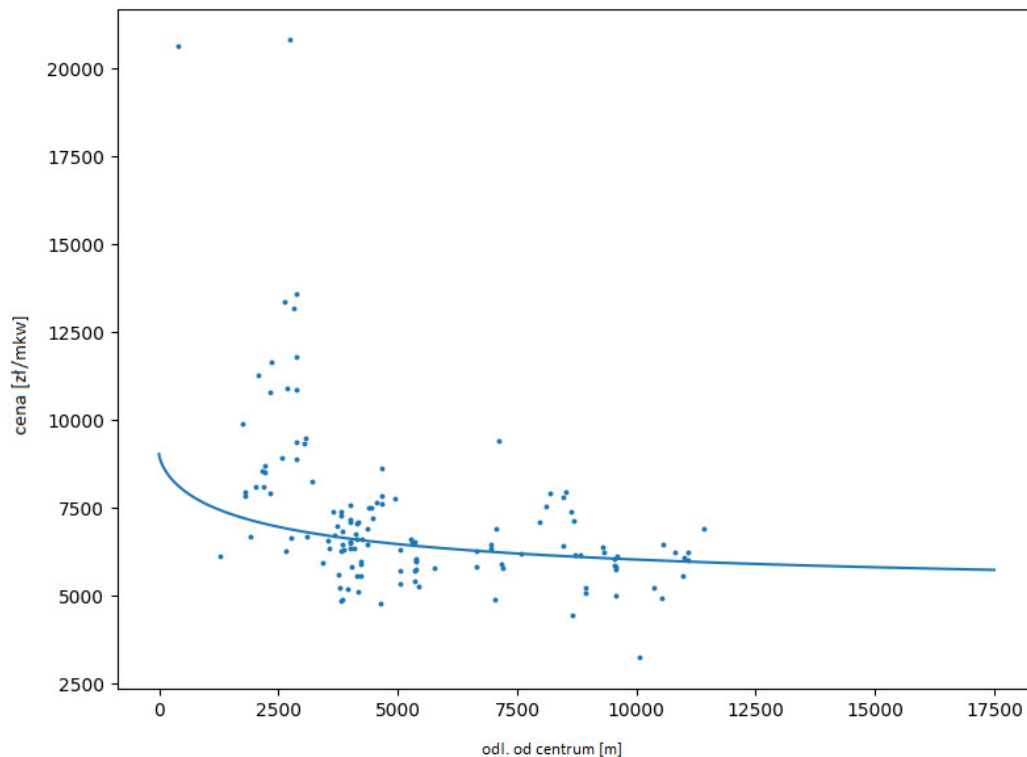
Dla wyznaczonych wartości bazowych cen mieszkań we wszystkich lokalizacjach dokonano regresji liniowej względem zmiennej objaśniającej z , gdzie $z = \frac{1}{x+D}$, nazywaną bliskością centrum, natomiast zmienną objaśnianą była oczywiście cena za metr kwadratowy. Aby dobrać wartość stałej D maksymalizującej współczynnik determinacji R^2 , wykorzystano narzędzie do optymalizacji (solver z pakietu SciPy dla języka python, minimalizacja metodą gradientu sprzężonego [28]).



Rys. 14: Wykres zależności średniej ceny jednostkowej mieszkań we wszystkich lokalizacjach (obliczonej wg sposobu 1) w zależności od odległości do centrum. Zauważalny jest spadek wartości lokali wraz z oddalaniem się od centrum.



Rys. 15: Wykres zależności średniej ceny jednostkowej mieszkań we wszystkich lokalizacjach (obliczonej wg sposobu 2) w zależności od odległości do centrum. Wykres różni się niewiele od poprzedniego



Rys. 16: Wykres zależności średniej ceny jednostkowej mieszkań w wybranych lokalizacjach (obliczonej wg sposobu 3) w zależności od odległości do centrum. Widoczna jest nieznaczna różnica względem poprzednich wykresów.

Po przeprowadzeniu regresji liniowej względem bliskości centrum, dla bazowych cen mieszkań w inwestycjach wyznaczonych na 3 sposoby, wyznaczonych otrzymano wyniki, które naniesiono na powyższe wykresy w postaci krzywych (rys. 14, 15 i 16):

$$1. A = 6207 + \frac{846500}{x+156}, R^2 = 0,41$$

$$2. A = 5317 + \frac{773500}{x+201}, R^2 = 0,35$$

$$3. A = 4894 + \frac{988300}{x+239}, R^2 = 0,49$$

Można zauważyć, że najlepszy wynik daje uwzględnienie cen bazowych mieszkań wynikających z regresji względem kondygnacji i metrażu.

Po przeprowadzeniu analogicznej analizy z uwzględnieniem czasu podróży samochodem do centrum zamiast odległości od centrum otrzymano następujące wyniki:

$$1. A = 6265 + \frac{1092000}{x-156}, R^2 = 0,35$$

$$2. A = 5278 + \frac{1076000}{x-201}, R^2 = 0,29$$

$$3. A = 4669 + \frac{1506000}{x-239}, R^2 = 0,44$$

Warto zauważyć, że otrzymano ujemne wartości współczynników w mianowniku, co jest niezgodne z postawioną wcześniej hipotezą. Jednak w związku z faktem, że błąd modelu jest większy w przypadku uwzględnienia czasu podróży do centrum, postanowiono przyjąć do dalszej analizy założenie o zależności ceny od bliskości centrum.

3.6.6. Zależność od odległości od POI

Kolejna postawiona hipoteza mówi, że odległość nieruchomości od wybranych kategorii punktów użyteczności publicznej wpływa w podobny sposób na jej atrakcyjność. Sprawdzono, która kategoria POI po dołączeniu do modelu (razem z odległością od centrum) (zgodnie ze wzorem (7) - $A_L = \sum_{i=1}^n \frac{\alpha_{Li}}{x_i + C_i} + \beta L_p$) pozwala osiągnąć większy współczynnik determinacji modelu. Przy dołączaniu do modelu wartości mieszkań wybranych kategorii punktów użyteczności publicznej otrzymano następujące wartości wskaźników R^2 :

- najbliższy lokal gastronomiczny - $R^2 = 0,51$,
- najbliższe centrum handlowe - $R^2 = 0,54$,
- najbliższa szkoła/przedszkole - $R^2 = 0,56$,

- najbliższa apteka - $R^2 = 0,53$,
- najbliższa przychodnia - $R^2 = 0,51$,
- najbliższa placówka pocztowa - $R^2 = 0,53$,
- najbliższy przystanek komunikacji zbiorowej - $R^2 = 0,58$.

Wynika z tego, że w największym stopniu na wygodę mieszkańców wpływa bliskość przystanków komunikacji miejskiej. Ujęcie w modelu odległości od najbliższego przystanku oraz najbliższej szkoły/przedszkola pozwoliło uzyskać współczynnik determinacji $R^2 = 0,61$.

Następnie przeanalizowano dodatkowy wpływ liczby punktów użyteczności publicznej wyżej wymienionych kategorii w promieniu 500 metrów na atrakcyjność mieszkań. Włączenie do modelu omawianej miary atrakcyjności lokalizacji pozwoliło nieznacznie zwiększyć dokładność modelu – wskaźnik R^2 osiągnął wartość 0,65.

Podsumowując, wyznaczony model atrakcyjności lokalizacji ma następującą postać:

$$A = 4384 + \frac{5858000}{x_1 + 295} + \frac{153700}{x_2 + 117} + 18.9 * L_p \quad (11)$$

gdzie:

A – atrakcyjność (cena) lokalu,

x_1 – odległość centrum w metrach,

x_2 – odległość najbliższego przystanku komunikacji miejskiej w metrach,

L_p – liczba POI wszystkich wybranych kategorii w promieniu 500 metrów.

Należy przypomnieć, że powyższy model powstał na skutek analizy cen bazowych mieszkań w wybranych inwestycjach (z liczbą lokali większą niż 20 – wg dostępnych danych transakcyjnych). Ceny bazowe zostały wyznaczone poprzez regresję względem numeru kondygnacji i metrażu wszystkich mieszkań w obrębie wskazanych inwestycji.

3.6.7. Regresja dla wszystkich mieszkań

Następnym etapem analizy było uwzględnienie wszystkich dostępnych danych transakcyjnych w regresji. Przyjęto hipotezy o zależności cen lokali od:

- piętra,
- metrażu,
- odległości od centrum,

- odległości od najbliższego przystanku komunikacji miejskiej,
- liczby punktów użyteczności publicznej w okolicy 500 metrów.

Przyjęto następujący model cen mieszkań:

$$A = \alpha_{N1}p + \frac{\alpha_{N2}}{m} + \frac{\alpha_{L1}}{x_1+C_1} + \frac{\alpha_{L2}}{x_2+C_2} + \beta L_p + \gamma + \varepsilon \quad (12)$$

gdzie:

A – atrakcyjność (cena) lokalu,

p – piętro, na którym znajduje się lokal,

m – metraż mieszkania,

x_1 – odległość inwestycji od centrum w metrach,

x_2 – odległość budynku od najbliższego przystanku komunikacji miejskiej w metrach,

L_p – liczba POI wszystkich wybranych kategorii w promieniu 500 metrów od budynku,

$\alpha_{N1}, \alpha_{N2}, \alpha_{L1}, \alpha_{L2}, C_1, C_2, \beta, \gamma$ – stałe,

ε – błąd modelu.

Współczynniki modelu związane ze zmianą ceny w zależności od kondygnacji i metrażu zostały wyznaczone wcześniej:

$$C_{N1} = 136,74 \text{ [zł]},$$

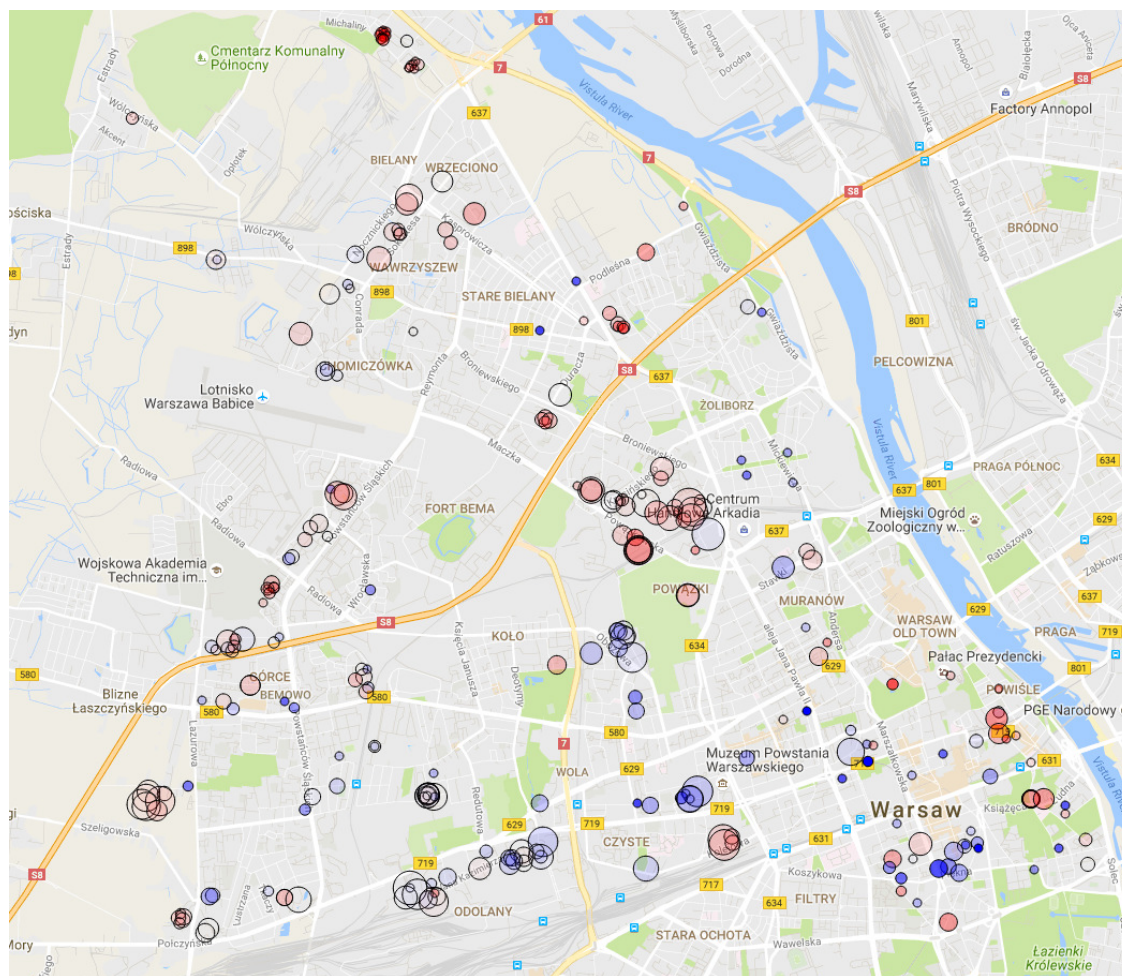
$$C_{N2} = 31421 \text{ [zł]}.$$

Wyznaczone wartości przyjęto jako stałe dla wszystkich mieszkań i inwestycji. Natomiast współczynniki związane z lokalizacją (odległością od centrum i położeniem względem punktów użyteczności publicznej) zostały ponownie wyznaczone za pomocą regresji liniowej i solvera.

W wyniku analizy wszystkich mieszkań z próby, która liczyła 7158 transakcji, otrzymano następujący model atrakcyjności (wartości) nieruchomości:

$$A = 4354 + 136,74 p + \frac{31421}{m} + \frac{10700000}{x_1+261} + \frac{359100}{x_2+105} + 15,1 * L_p \quad (13)$$

Współczynnik determinacji powyższego modelu osiągnął wartość $R^2 = 0,76$.



Rys. 17: Wizualizacja błędów modelowania dla poszczególnych inwestycji na mapie. Wielkość punktu jest proporcjonalna do pierwiastka 4 stopnia z liczby transakcji zarejestrowanych dla danej lokalizacji. Kolor czerwony oznacza niedoszacowanie, a niebieski – przeszacowanie wartości lokalu przez model. Natężenie barwy oznacza natomiast skalę niedoszacowania/przeszacowania wyznaczonej jako średni względny błąd modelu dla wszystkich mieszkańców z danej inwestycji.

3.6.8. Próba korekty modelu

Przedstawiając średni względny błąd modelowania dla wszystkich inwestycji (rys. 17) można dostrzec pewne zbiorowiska inwestycji niedoszacowanych i przeszacowanych, przy czym tych pierwszych jest więcej. Rozbieżność cen modelowanych i rzeczywistych można próbować wyjaśniać na wiele sposobów. Niedoszacowanie mieszkań w dzielnicy Żoliborz może być spowodowana prestiżem tej dzielnicy, natomiast przeszacowanie lokali w okolicach Odolan (część dzielnicy Wola) może wynikać z niższych cen działek znajdujących się na terenach pokolejowych. Niedoskonałość przedstawionego sposobu szacowania cen może

być spowodowana różnicami w jakości budynków budowanych przez różnych deweloperów, ponieważ z powodu braku odpowiednich danych niemożliwe było określenie czynników związanych ze standardem poszczególnych inwestycji mieszkaniowych.

Jednak obserwując wizualizację można zauważyć, że nieruchomości osiągają wyższe średnie ceny, gdy są zlokalizowane w większych skupiskach. Może to być spowodowane tym, że pojedynczy nowy blok postawiony w wolnym miejscu pomiędzy starszymi budynkami nie jest tak atrakcyjny jak nieruchomość znajdująca się na nowym osiedlu, otoczona nowoczesnymi, ładnymi budynkami. Dobrym przykładem jest ulica Powstańców Śląskich, gdzie znajdują się zarówno nowe osiedla o zwartej zabudowie jak i pojedyncze nowe budynki umieszczone na terenie starych osiedli. Z rysunku 17 wynika, że model niedoszacował cen mieszkań znajdujących się w budynkach na terenie nowego osiedla, natomiast lokale z inwestycji umiejscowionych pośród starszych budynków mają w rzeczywistości ceny niższe niż wynika to z estymacji.

Postanowiono zweryfikować powyższą hipotezę poprzez określenie miary opisującej najbliższe otoczenie nieruchomości i włączenie jej do modelu cen. Założono, że budynek znajduje się w obrębie nowego osiedla, jeżeli liczba nowych bloków w promieniu 200 metrów jest większa niż jeden. Za nowe budynki przyjęto wszystkie nieruchomości, których dotyczyły dane transakcyjne pozyskane i analizowane w niniejszej pracy. Zmienną logiczną określającą otoczenie nieruchomości (nowoczesne/stare) zdefiniowano jak powyżej i włączono do opracowanego wcześniej modelu (wzór 8). W celu wyznaczenia parametrów i dokładności modelu przeprowadzono analogiczną procedurę jak poprzednio. Tak więc założono, że cena (atrakcyjność) mieszkania jest opisana wzorem:

$$A = \alpha_{N1}p + \frac{\alpha_{N2}}{m} + \frac{\alpha_{L1}}{x_1 + C_1} + \frac{\alpha_{L2}}{x_2 + C_2} + \beta L_p + \delta N + \gamma + \varepsilon \quad (14)$$

gdzie N to zmienna logiczna opisująca sąsiedztwo nieruchomości, a δ to stała.

Współczynnik determinacji zaktualizowanego modelu osiągnął wartość $R^2 = 0,77$, tak więc włączenie kolejnego parametru nie przyniosło widocznej poprawy. W związku z tym uznano opisany parametr związany z otoczeniem budynku za nieistotny i za ostateczną wersję zależności opisującej ceny mieszkań przyjęto wzór (10) z wyznaczonymi współczynnikami (11).

3.7. Wnioski z analizy

Udało się zgromadzić, przetworzyć i przeanalizować dane transakcyjne dla ponad 7 tysięcy nieruchomości znajdujących się w 5 dzielnicach Warszawy. Przeprowadzona analiza pozwoliła zweryfikować pewne hipotezy związane z zależnością cen mieszkań na rynku pierwotnym. Na zgromadzonych danych udało się potwierdzić hipotezy przedstawione w rozdziale 3.5. Udowodniono, że na atrakcyjność i cenę jednostkową mieszkania ma wpływ piętro, na którym się znajduje, metraż, odległość od centrum miasta, odległość najbliższego przystanku komunikacji miejskiej oraz liczba punktów użyteczności publicznej w najbliższej okolicy. Hipoteza o wpływie sąsiedztwa nieruchomości (nowe/stare osiedle) została odrzucona.

4. Podsumowanie

Niniejszy rozdział zawiera podsumowanie przeprowadzonych prac. Przedstawione zostały konkluzje wynikające z wykonanego badania zagadnienia modelowania cen nieruchomości. Zaproponowano również dalsze możliwe ścieżki pracy nad zagadnieniem.

4.1. Podsumowanie przeprowadzonych prac

Celem pracy było zapoznanie się z rynkiem mieszkań, określenie i analiza cech wpływających na atrakcyjność mieszkań oraz stworzenie i weryfikacja modelu matematycznego opisującego ceny mieszkań uwzględniającego wybrane parametry nieruchomości.

Przeanalizowano sytuację na rynku nieruchomości w Polsce i dokonano przeglądu literatury i opracowań dotyczących zagadnienia modelowania cen mieszkań zarówno na rynku polskim jak i na rynkach zagranicznych. Zwrócono uwagę na istniejące rozwiązania automatyzujące proces analizy danych, z których zostały zaczerpnięte pewne inspiracje.

Następnie przedstawione zostały teoretyczne podstawy hedonicznych indeksów cen mieszkań oraz modelowania za pomocą regresji. Przeanalizowane zostały czynniki wpływające na komfort życia mieszkańców, a co za tym idzie na atrakcyjność nieruchomości, związane z samym mieszkaniem oraz jego lokalizacją. Opisano sposoby pozyskiwania i gromadzenia danych z różnych źródeł oraz skonfrontowano określone czynniki determinujące ceny mieszkań z informacjami dostępnymi w zebranych danych. Poprzez analizę danych udało się potwierdzić pewne hipotezy dotyczące cech nieruchomości wpływających na ich cenę. Cel pracy został osiągnięty.

4.2. Konkluzje

W rozdziale dotyczącym przeglądu istniejącego stanu wiedzy określono nieopracowaną stronę zagadnienia. W niniejszej pracy, w przeciwieństwie do innych opracowań tematu, zwrócono większą uwagę na aspekty związane z lokalizacją mieszkań i

informacjami, które są możliwe do wydobycia przy użyciu istniejących technologii. Udało się potwierdzić wpływ położenia nieruchomości na mapie miasta na ich ceny rynkowe. Zwrócono również uwagę na główne czynniki związane z samymi lokalami, tj. wielkość mieszkania oraz kondygnację, na którym się znajduje. Wyznaczony model wartości nieruchomości cechował się zadowalającym współczynnikiem determinacji.

4.3. Możliwości dalszej analizy tematu

Zagadnienie modelowania cen mieszkań nie zostało wyczerpane w niniejszej pracy, również biorąc pod uwagę określony zakres badania – rynek nowych mieszkań w Warszawie.

Pierwszym – i najważniejszym – powodem jest fakt, że udało się pozyskać dane transakcyjne wyłącznie dla części miasta. Informacje o transakcjach z całej stolicy dałyby szerszy pogląd na sytuację na warszawskim rynku oraz z pewnością dokładniejsze wyniki analizy. Prawdopodobnie udałoby się zauważyć więcej lokalizacji, w których mieszkania są droższe bądź tańsze niż wynikałoby to z modelu oraz wyciągnąć pewne wnioski, np. związane z położeniem nieruchomości po obu stronach Wisły. Gdyby dane z Rejestru Cen i Wartości Nieruchomości posiadały więcej uporządkowanych informacji na temat mieszkań, np. obecności miejsca parkingowego czy balkonu, być może udałoby się stworzyć dokładniejszy model uwzględniający czynniki nie wzięte pod uwagę przy wykonanej analizie.

Ponadto, można by próbować określać inne czynniki związane z lokalizacją nieruchomości determinujące ich cenę, takie jak chociażby położenie względem terenów zielonych czy innych typowych celów podróży mieszkańców Warszawy, jak np. dzielnica biurowa na Mokotowskim Służewcu, gdzie pracuje obecnie niemal 100 tysięcy osób [26]. Należy jednak podkreślić, że bardzo istotnym dla otrzymania jak najbardziej rzetelnych wyników analizy jest możliwość analizy jak największej ilości danych dotyczących transakcji z całej Warszawy.

Bibliografia:

- [1] Trader21 (2014), „Rzut okiem na rynek nieruchomości”, <http://independenttrader.pl> (Dostęp: 13 września 2016)
- [2] A. Prajsnar (2014), „Ceny mieszkań wczoraj i dziś”, <http://nf.pl> (Dostęp: 13 września 2016)
- [3] M. Gawin (2016), „Raport z rynku mieszkań – styczeń 2016”, <http://www.bankier.pl> (Dostęp: 13 września 2016)
- [4] M. Gawin (2016), „Raport z rynku mieszkań – luty 2016”, <http://www.bankier.pl> (Dostęp: 13 września 2016)
- [5] REAS (2016), „Rynek mieszkaniowy w Polsce, II kwartał 2016”, <http://www.reas.pl/> (Dostęp: 13 września 2016)
- [6] J. Kolbe, H. Wuestemann (2015), „Estimating the Value of Urban Green Space: A hedonic Pricing Analysis of the Housing Market in Cologne, Germany”, SFB 649: Economic Risk Discussion Paper 2015-002
- [7] G. Hu, J. Wang, W. Feng (2013), „Multivariate Regression Modeling for Home Value Estimates with Evaluation using Maximum Information Coefficient”, *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, s. 69-81
- [8] Iain R. Lake, Andrew A. Lovett, Ian J. Bateman, Brett Day (2000), „Using GIS and large-scale digital data to implement hedonic pricing studies”, *International Journal of Geographical Information Science*, s. 521-541
- [9] J. Ottensmann, S. Payton, J. Man (2008), „Urban Location and Housing Prices within a Hedonic Model”, *Journal of Regional Analysis and Policy* 38, s. 19-35
- [10] R. Zbyrowski (2010), „Szacowanie wartości nieruchomości na podstawie modeli ekonometrycznych”, *Equilibrium* t. 4 nr 1 s. 241-252
- [11] E. Tomczyk, M. Widłak (2010), „Konstrukcja i własności hedonicznego indeksu cen mieszkań dla Warszawy”, *Bank i Kredyt* 41 (1), s. 99-128
- [12] <https://www.WalkScore.com> (Dostęp: 13 września 2016)
- [13] https://en.wikipedia.org/wiki/Walk_Score (Dostęp: 13 września 2016)

- [14] <https://www.szybko.pl/> (Dostęp: 13 września 2016)
- [15] <https://ceny.szybko.pl/wycena-nieruchomosci> (Dostęp: 13 września 2016)
- [16] [https://pl.wikipedia.org/wiki/Regresja_\(statystyka\)](https://pl.wikipedia.org/wiki/Regresja_(statystyka)) (Dostęp: 13 września 2016)
- [17] <https://pl.wikipedia.org/wiki/Regresja liniowa> (Dostęp: 13 września 2016)
- [18] https://en.wikipedia.org/wiki/Automated_valuation_model (Dostęp: 13 września 2016)
- [19] <http://www.mapa.um.warszawa.pl/BGIK/RCiWN.html> (Dostęp: 13 września 2016)
- [20] <https://github.com/googlemaps/js-marker-clusterer> (Dostęp: 13 września 2016)
- [21] <http://www.programmableweb.com/news/calculate-distances-intelligently-google-distance-matrix-api/2011/07/22> (Dostęp: 13 września 2016)
- [22] <http://www.automapa.pl/> (Dostęp: 13 września 2016)
- [23] <http://www.miplo.pl/> (Dostęp: 13 września 2016)
- [24] https://en.wikipedia.org/wiki/Vincenty's_formulae (Dostęp: 13 września 2016)
- [25] A. Brzeziński, M. Dobrosielski, T. Dybicz i in. (2009), „*Studium komunikacyjne Śródmieścia Warszawy w związku z planowanymi zmianami zagospodarowania przestrzennego*”
- [26] <http://natemat.pl/8081,sluzewiec-miejsce-w-ktorym-jest-najwiecej-samochodow-w-polsce-i-najmniej-miejsc-do-parkowania> (Dostęp: 13 września 2016)
- [27] https://pl.wikipedia.org/wiki/Wsp%C3%B3%C5%82czynnik_determinacji (Dostęp: 13 września 2016)
- [28] https://pl.wikipedia.org/wiki/Metoda_gradientu_sprężonego (Dostęp: 13 września 2016)