



Dominik Krężolek

Uniwersytet Ekonomiczny w Katowicach
Wydział Informatyki i Komunikacji
Katedra Demografii i Statystyki Ekonomicznej
dominik.krezolek@ue.katowice.pl

Anna Cibis

Uniwersytet Ekonomiczny w Katowicach
Value Miners – Koło Naukowe Data Science
value.miners@uekat.pl

Izabela Staszczyszyn

Uniwersytet Ekonomiczny w Katowicach
Value Miners – Koło Naukowe Data Science
value.miners@uekat.pl

Marcin Pastecki

Uniwersytet Ekonomiczny w Katowicach
Value Miners – Koło Naukowe Data Science
value.miners@uekat.pl

Grzegorz Dargiewicz

Uniwersytet Ekonomiczny w Katowicach
Value Miners – Koło Naukowe Data Science
value.miners@uekat.pl

Mateusz Skrzydło

Uniwersytet Ekonomiczny w Katowicach
Value Miners – Koło Naukowe Data Science
value.miners@uekat.pl

Robert Kiermasz

Uniwersytet Ekonomiczny w Katowicach
Value Miners – Koło Naukowe Data Science
value.miners@uekat.pl

ATRAKCYJNOŚĆ MIASTA KATOWICE – PODEJŚCIE WIZUALNE

Streszczenie: Aglomeracje miejskie cechują się istotnym zróżnicowaniem z punktu widzenia postrzegania ich atrakcyjności przez mieszkańców. Termin „atrakcyjność” jest pojęciem subiektywnym, niemierzalnym i nieposiadającym jednoznacznej, uniwersalnej definicji. Głównym celem badania jest wykrycie i identyfikacja czynników istotnie wpływających na postrzeganie atrakcyjności miasta przez jego mieszkańców, jak również osoby przyjezdne. Założono, że nie tylko cena jest istotną determinantą skłaniającą do zakupu nieruchomości w danej lokalizacji, ale także szereg innych czynników. W pracy zaproponowano klasyfikację nieruchomości przy wykorzystaniu analizy skupień oraz przedstawiono graficzną prezentację wybranych czynników mogących w istotny sposób wpływać na postrzeganie atrakcyjności miasta Katowice przy wykorzystaniu autorskich aplikacji i kodów programistycznych.

Słowa kluczowe: analiza skupień, wizualizacja, atrakcyjność miasta, Data Science.

JEL Classification: C38, C63, R30, R31.

Wprowadzenie

Sposób postrzegania atrakcyjności miasta jest kwestią czysto indywidualną, zależną w głównej mierze od subiektywnych preferencji osoby oceniającej. Niemniej można wskazać szereg czynników, które w sposób naturalny wskazywałyby przewagę konkurencyjną jednej aglomeracji nad inną. Jednakże w większości przypadków determinowane jest to celem oceny. W prezentowanej pracy skoncentrowano się na propozycji budowy modelu wyceny nieruchomości na podstawie zbioru hipotetycznych czynników, takich jak: lokalizacja, powierzchnia użytkowa nieruchomości, odległość nieruchomości od centrum czy też pewnych charakterystycznych punktów, które mogą zachęcać inwestora do zakupu.

Szerokie zastosowanie w tego rodzaju analizach ma geostatystyka. W porównaniu z klasycznymi narzędziami pomiarowymi metody statystyki pozwalają uwzględniać lokalizację jedynie w sposób uproszczony, natomiast geostatystyka dostarcza metod opisujących ciągłość przestrzenną, która jest zasadniczą cechą wielu zjawisk obserwowanych na rynku nieruchomości. Jest ona zatem naturalnym uzupełnieniem dla analizy i modelowania statystycznego w badaniach cen transakcyjnych i czynników nań wpływających [Cellmer, 2014].

Z punktu widzenia metodologicznego w geostatystyce wizualnym narzędziem wykorzystywanym w analizach jest mapa definiowana jako pewien uogólniony obraz powierzchni Ziemi lub jej części wykonywany na płaszczyźnie, w określonej skali, według zasad odwzorowania kartograficznego, przy użyciu umownych znaków graficznych. Stanowi jednocześnie model wybranego zjawiska lub aspektu przestrzeni tworzącej geosystem. Mapa może być wykorzystana jako narzędzie służące do rozpoznawania i charakteryzowania układów przestrzennych, relacji i anomalii w przestrzennym zróżnicowaniu zjawisk oraz przekazywania uzyskanych informacji [Żyszkowska, Spallek, Borowicz, 2012].

Mapa atrakcyjności inwestycyjnej, która jest docelowym przedmiotem prowadzonego badania, może być źródłem użytecznych informacji, optymalizującym procesy decyzyjne związane z nieruchomościami dla różnych celów społeczno-gospodarczych. Do odbiorców takich opracowań należą m.in. jednostki administracji państwowej i samorządowej, urzędy skarbowe, banki, firmy ubezpieczeniowe, rzeczoznawcy majątkowi, zarządcy nieruchomości, pośrednicy w obrocie nieruchomościami, a przede wszystkim deweloperzy, fundusze inwestycyjne oraz osoby fizyczne. Wśród wymienionych uczestników rynku nieruchomości – potencjalnych beneficjentów wizualnej informacji o wybranej przestrzeni na rynku nieruchomości – istotną i niewątpliwie najliczniejszą grupę

stanowią osoby fizyczne. Dla nich inwestowanie w nieruchomości jest alternatywną formą lokaty kapitału przy dużym poziomie ryzyka inwestycyjnego w proponowane m.in. przez banki instrumenty finansowe oraz staje się popularną metodą indywidualnego zabezpieczenia finansowego w dynamicznie zmieniającym się otoczeniu prawnym i ekonomicznym.

Podstawą opracowywania takich map są najczęściej szczegółowe analizy sporządzane dla obszarów pojedynczych miast [Colwell, 2003; Bryan, Sarte, 2009]. Należy podkreślić, że rynek nieruchomości jest kształtowany przez skomplikowane procesy przestrzenne, jak też nieprzestrzenne, które łącznie oddziałują na wartość rynkową takiej inwestycji [Munroe, 2007]. Ceny i wartość nieruchomości określa nie tylko lokalizacja, ale także inne czynniki, które należy uwzględnić w analizie rynku [Earnhart i in., 2005; Palmquist, 2006].

Odkrywanie i szacowanie odpowiednich mierników oraz opracowywanie map wiąże się z koniecznością pozyskiwania dużej ilości różnorodnych danych z zasobów publicznych oraz prywatnych. Dostęp do informacji (m.in. program otwierania danych publicznych Ministerstwa Cyfryzacji, geoportale, narzędzia GIS) oraz dynamiczny rozwój technologii informatycznych (m.in. usługi IaaS, PaaS w chmurze) pozwalają na efektywne pozyskiwanie i przetwarzanie danych.

1. Metodologia badawcza – analiza skupień

Analiza skupień polega na łączeniu badanych jednostek w grupy. Zakłada się, że każda badana obserwacja posiada specyficzne dla siebie cechy, które można zapisać w postaci wektora. Każdy wektor odpowiada jednej jednostce, a każda kolejna składowa w wektorze odpowiada wartości określonej cechy obserwacji. Przyjmijmy, że wszystkich badanych obserwacji jest n i każda opisana jest przez p cech, to i – tą jednostkę reprezentuje p –wymiarowy wektor x_i . Grupowanie opiera się na analizie zespołu wektorów $x_i, i = 1, 2, \dots, n$. W procesie tym zwykle zakłada się, że liczba grup g nie jest z góry znana, ale jest mniejsza niż liczba grupowanych jednostek, tzn.: $g < n$ [Błazejczyk-Majka, Kala, 2005].

Podstawowymi miarami grupowania są miary odległości lub podobieństwa. Miary podobieństwa, jak i miary odległości, są przedstawione za pomocą skalarów. Jest to istotny fakt, który pozwala na porównywanie obserwacji ze sobą, co w finalnym efekcie umożliwia proces grupowania. Im większa jest wartość miary podobieństwa, tym bardziej obiekty są do siebie podobne, co skutkuje przypisaniem ich do tej samej grupy. Przeciwnie interpretowana jest miara odległości,

która w praktyce jest wykorzystywana częściej niż podobieństwo. W odniesieniu do obserwacji opisanych cechami ilościowymi najczęściej stosowana jest odległość euklidesowa [Błażejczyk-Majka, Kala, 2005]. W niniejszej pracy zdecydowano się na wykorzystanie kwadratu odległości euklidesowej.

W analizie skupień wykorzystano również metodę Warda. Metoda Warda (porządkowanie nieliniowe) wykorzystująca analizę wariancji służy do oszacowania odległości między skupieniami. Odległość między skupieniami definiowana jest jako moduł różnicy między sumami kwadratów odległości punktów od środków grup, do których te punkty należą [Łogwiniuk, 2011].

Wykorzystanych w artykule danych nie standaryzowano przed rozpoczęciem procesu grupowania w celu zachowania rzeczywistych parametrów dla badanych lokali.

2. Analiza empiryczna

W empirycznej części badania przeprowadzono wizualną ocenę atrakcyjności miasta Katowice ze względu na zestaw różnych atrybutów powiązanych z analizowanymi nieruchomościami. W analizie wykorzystano takie zmienne, jak:

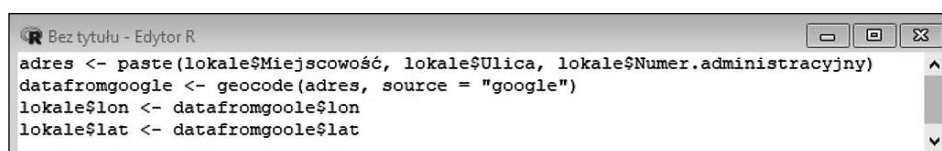
- Cena jednostkowa netto (PLN),
- Powierzchnia użytkowa lokalu (m^2),
- Liczba pokoi,
- Liczba szkół (w promieniu 1 km),
- Odległość do najbliższej szkoły (m),
- Liczba przystanków (w promieniu 1 km),
- Odległość do najbliższego przystanku (m),
- Liczba placówek medycznych (w promieniu 1 km),
- Odległość do najbliższej placówki medycznej (m),
- Liczba marketów (w promieniu 1 km),
- Odległość do najbliższego marketu (m),
- Odległość od centrum miasta (m).

Badanie podzielono na dwa zasadnicze etapy: wizualizację charakterystyk nieruchomości oraz analizę skupień.

2.1. Analiza wizualna wybranych charakterystyk nieruchomości

Metodyka tworzenia map wartości nieruchomości jako element tworzenia mapy atrakcyjności inwestycyjnej opiera się przede wszystkim na znajomości cen transakcyjnych, a także na relacji między wartością a odległością od pewnych charakterystycznych miejsc. Według niektórych badaczy tego tematu [Liu i in., 2007] kształtowanie się cen i wartości nieruchomości zależy od odległości od tzw. CBD (*Central Business District*), czyli elementów infrastruktury społecznej, obiektów służby zdrowia, kultury, szkół etc. Podobne podejście proponowali inni badacze [Bugs, 2007], wskazując, aby podstawą opracowania mapy wartości były strefy wyznaczone na podstawie m.in. odległości od centrum miasta, głównych ulic, miejsc szczególnych, wpływających na wartość, a także terenów zagrożonych powodzią. W wyniku analiz przestrzennych (z wykorzystaniem narzędzi informatycznych dedykowanych do tego celu) powstają mapy waloryzacji przestrzeni.

Wizualizacja charakterystyk poszczególnych lokali mieszkalnych przebiegała wieloetapowo. Niewątpliwie najbardziej pracochłonnym był etap przygotowania danych, polegający przede wszystkim na ich czyszczeniu, usuwaniu zbędnych informacji oraz doprowadzeniu do pełnej spójności analizowanego zbioru. W tym celu wykorzystano oprogramowanie Excel oraz RStudio. Następnie dla każdego lokalu mieszkalnego należało zdefiniować dane geolokalizacyjne, tj. długość i szerokość geograficzną. Wykorzystano język programistyczny R przy zastosowaniu skryptu danego kodem zamieszczonym na rys. 1.



```
adres <- paste(lokalne$Miejscowość, lokalne$Ulica, lokalne$Numer.administracyjny)
datafromgoogle <- geocode(adres, source = "google")
lokalne$lon <- datafromgoogle$lon
lokalne$lat <- datafromgoogle$lat
```

Rys. 1. Kod R definiujący geolokalizację

Źródło: Obliczenia własne.

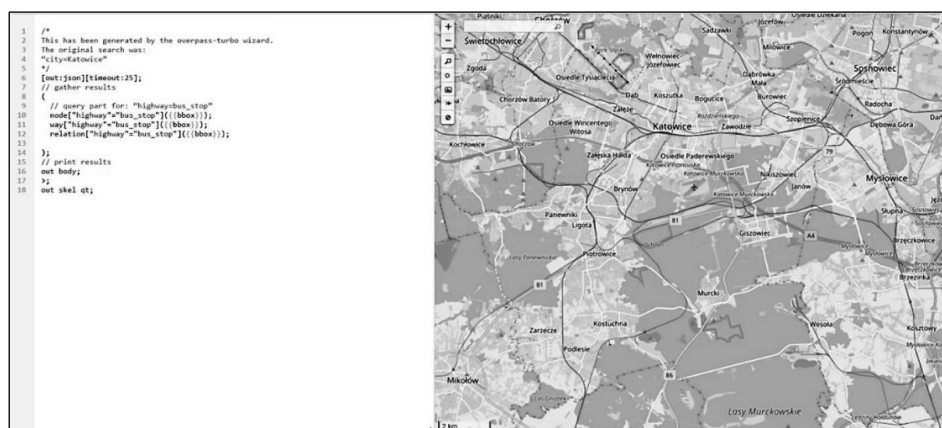
Kolejno przy pomocy funkcji `geocode()` wywoływana jest usługa geocoder interfejsu programistycznego API¹ witryny Google Maps², która dla dowolnego

¹ API/Interfejs programistyczny (ang. *Application Programming Interface*) to „ściśle określony zestaw reguł i ich opisów, w jaki programy komputerowe komunikują się między sobą” [www 1].

² Google Maps – serwis internetowy pozwalający na szukanie na mapach obiektów, oglądanie powierzchni Ziemi, sprawdzanie natężenia ruchu ulicznego w czasie rzeczywistym oraz obli-

przekazanego jej adresu zwraca jego współrzędne geograficzne. Do funkcji `geo-code()` jako parametr przekazywana jest zmienna *adres*, a wynik zapytania, dwie zmienne liczbowe, zachowywany jest w zmiennej *datafromgoogle*.

Celem wygenerowania informacji o obiektach typu przychodnie, szpitale, apteki, obiekty sportowe, sklepy wielkopowierzchniowe oraz przystanki komunikacji miejskiej wykorzystano aplikację internetową OpenStreetMap, a także internetową aplikację do filtrowania danych z OpenStreetMap [www 2], tworzone przez społeczność internetową na licencji typu OpenSource³, dostarczające darmowe dane geograficzne. W celu uzyskania danych, korzystając z przeglądarki internetowej, należy ograniczyć obszar wyszukiwania widoczny w postaci mapy (rys. 2), a następnie uruchomić skrypt (rys. 3) języka Overpass QL [www 4], który będzie pracował na zadanym ograniczonym obszarze (tu: wyszukiwał obiekty podanego typu).



Rys. 2. Aplikacja internetowa do filtrowania danych OpenStreetMap

Źródło: [www 4].

czanie odległości między punktami na mapie w zależności od wybranego środka transportu (podróż samochodem, samolotem, transportem publicznym, rowerem, pieszo).

³ OpenSource – zapewnia swoim użytkownikom prawo do legalnego oraz darmowego uruchamiania, kopiowania, rozpowszechniania, analizowania, modyfikacji oraz ulepszania i rozbudowy istniejących produktów [www 3].

```

1  /*
2  This has been generated by the overpass-turbo wizard.
3  The original search was:
4  "city=Katowice"
5  */
6  [out:json][timeout:25];
7  // gather results
8  (
9    // query part for: "highway=bus_stop"
10   node["highway"]="bus_stop"({{bbox}});
11   way["highway"]="bus_stop"({{bbox}});
12   relation["highway"]="bus_stop"({{bbox}});
13
14  );
15  // print results
16  out body;
17  >;
18  out skel qt;

```

Rys. 3. Skrypt języka OverpassQL

Źródło: Obliczenia własne.

Obiekty, które były przedmiotem analizy, zostały opisane w języku Overpass QL następującymi kluczami przedstawionymi w tabeli 1.

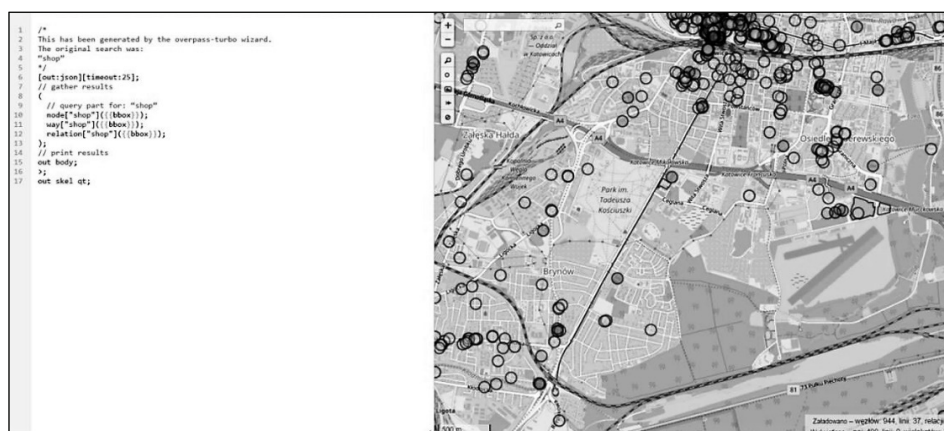
Tabela 1. Klucze w języku Overpass QL

Obiekt	Kategoria	Tag
Przystanki transportu publicznego	highway	bus_stop
Obiekty sportowe	leisure	sports_centre, pitch, fitness_centre, gym, swimming_pool
Obiekty powiązane ze zdrowiem	healthcare, amenity	dentist, doctors, hospital, pharmacy, clinic
Sklepy wielkopowierzchniowe	shop	mall, supermarket

Źródło: Obliczenia własne.

Po wykonaniu skryptu uzyskano wyniki w wersji graficznej (rys. 4) oraz w formie tekstowej, dostępnej w formacie JSON⁴ (rys. 5).

⁴ JSON – tekstowy format wymiany danych komputerowych bazujący na języku JavaScript. Komunikat JSON jest tablicą asocjacyjną, wszystkie dane są zmiennymi, natomiast nazwy obiektów muszą być otoczone znakami cudzysłowu. Może przyjmować wartości typu *string* (ciąg znaków), *number* (liczba typu *double*), stałe typu *null* bądź *false*. Komunikat jest kodowany w unidzie, domyślnie UTF-8.



Rys. 4. Aplikacja internetowa do filtrowania danych OpenStreetMap – lokalizacja obiektów

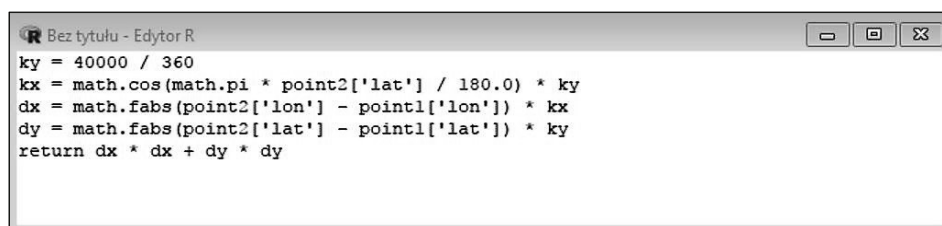
Źródło: [www 4].

```
8 "elements": [  
9   {  
10     "type": "node",  
11     "id": 26304526,  
12     "lat": 50.2210272,  
13     "lon": 18.9896572,  
14     "tags": {  
15       "bus": "yes",  
16       "highway": "bus_stop",  
17       "name": "Ochojec Wapienna",  
18       "public_transport": "stop_position"  
19     }  
20   }  
21 ]
```

Rys. 5. Dane w formacie JSON

Źródło: Obliczenia własne.

Celem zdefiniowania dla każdego adresu mieszkalnego obiektów zainteresowania znajdujących się w promieniu jednego kilometra od danego lokalu mieszkalnego zastosowano algorytm, który dla każdego lokalu mieszkalnego oblicza odległość od każdego obiektu zainteresowania, a następnie weryfikuje, czy odległość ta jest mniejsza od 1000 metrów. Algorytm ten korzysta z następujących funkcji zaprezentowanej w skrypcie R na rys. 6.



Rys. 6. Algorytm wyznaczania odległości

Źródło: Obliczenia własne.

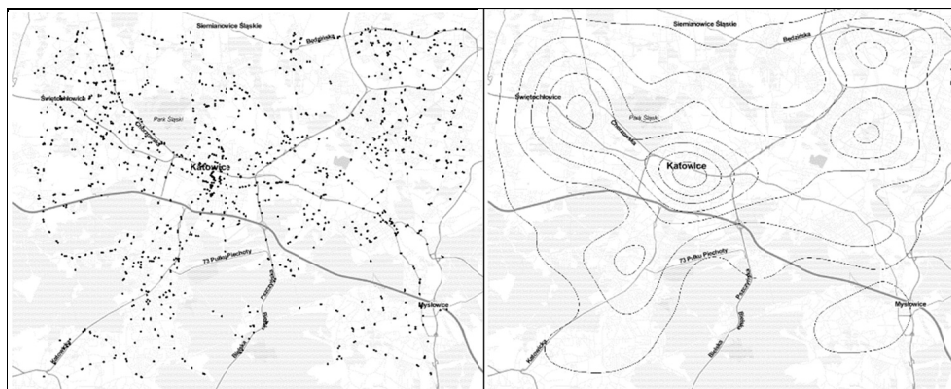
W przedstawionym algorytmie point1 to współrzędne lokalu mieszkalnego, point2 – współrzędne obiektu zainteresowania. Ponieważ odległość jednego kilometra uznano za stosunkowo niewielką, zastosowano pewne uproszczenie polegające na zignorowaniu krzywizny Ziemi, przekształceniu stopni na kilometry i obliczeniu odległości, korzystając z twierdzenia Pitagorasa.

Do wyznaczenia pieszej odległości pomiędzy dwiema współrzędnymi wykorzystano bibliotekę pyroulib2, udostępnioną przez OpenStreetMap [www 5]. Biblioteka ta pozwala na wyznaczanie odległości dla danego środka transportu przy wykorzystaniu publicznych danych udostępnionych przez OpenStreetMap.

Operacja wyznaczenia odległości jest najbardziej złożona obliczeniowo z dotychczas opisanych procedur. Ilość tych operacji jest też zdecydowanie największa – dla każdego lokalu należy wyznaczyć trasę dla co najmniej kilkudziesięciu obiektów. Aby zwiększyć efektywność kodu, wykorzystano klasę Pool biblioteki multiprocessing dostępnej w środowisku Python, a następnie podzielono działanie programu na liczbę podprocesów wynikającą z ilości rdzeni procesora komputera, na którym kod był wykonywany. Pozwoliło to znacząco przyspieszyć obliczenia.

Znając odległości piesze obiektów znajdujących się w obrębie jednego kilometra (liczone w linii prostej) od każdego z lokali, można było wyznaczyć obiekt najbliższy lokalowi w danej grupie, co polegało na posortowaniu listy według odległości od najmniejszej do największej i wybraniu pierwszego elementu z posortowanej listy. W celu obliczenia odległości pieszej lokalu mieszkalnego od centrum miasta (jako centrum miasta przyjęto punkt o współrzędnych: szerokość geograficzna 50.259592, długość geograficzna 19.022224, odpowiadający punktowi o adresie Katowice, Rynek) napisano z użyciem języka Python program, który korzystając z funkcji gmaps.directions() z interfejsu programistycznego (API) Google Maps, oblicza pieszą odległość między dwoma punktami na mapie.

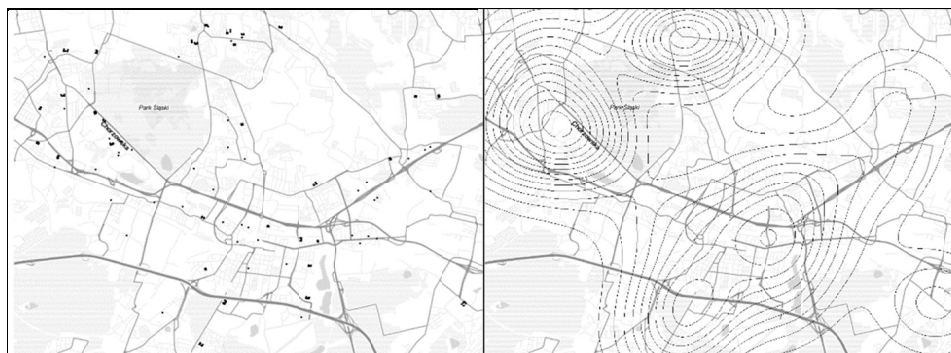
Ostateczne wyniki wizualizacji obiektów mogących potencjalnie determinować cenę nieruchomości w Katowicach przedstawiono na rys. 7-11.



Rys. 7. Rozmieszczenie (panel lewy) oraz zagęszczenie (panel prawy) przystanków w Katowicach

Źródło: Obliczenia własne.

Główna koncentracja przystanków autobusowych zlokalizowana jest w centralnej części Katowic oraz w okolicach osiedla Tysiąclecia. Dwa skupiska zaobserwowano także w obszarze osiedli Dąbrówka Mała oraz Zawodzie, gdzie zlokalizowana jest pętla autobusowo-tramwajowa.

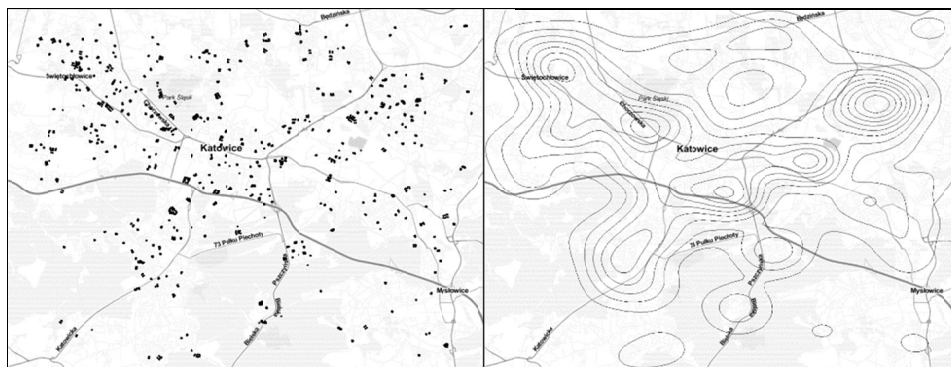


Rys. 8. Rozmieszczenie (panel lewy) oraz zagęszczenie (panel prawy) sklepów wielkopowierzchniowych w Katowicach

Źródło: Obliczenia własne.

Sklepy wielkopowierzchniowe (markety) zlokalizowane są głównie w obszarze osiedla Tysiąclecia oraz osiedli Wełnowiec-Józefowiec, Dąb i Koszutka.

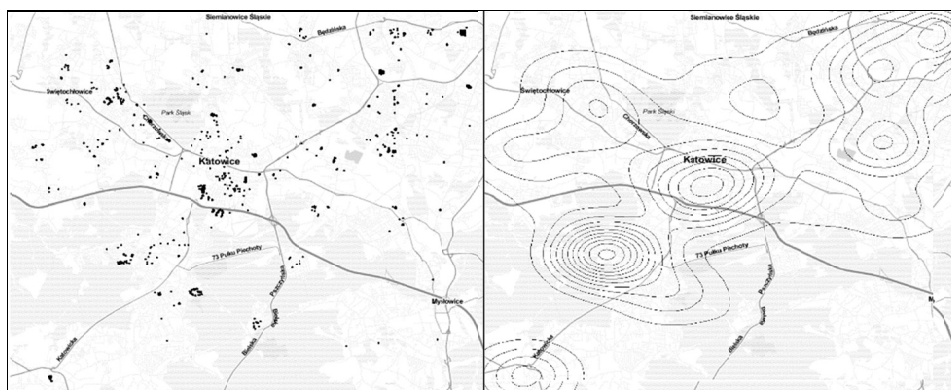
Można zaobserwować ich zagęszczenie także w Śródmieściu i ścisłym Centrum, jednakże wynika to przede wszystkim z lokalizacji dwóch głównych galerii handlowych: Silesia City Center oraz Galeria Katowicka.



Rys. 9. Rozmieszczenie (panel lewy) oraz zagęszczenie (panel prawy) obiektów sportowych w Katowicach

Źródło: Obliczenia własne.

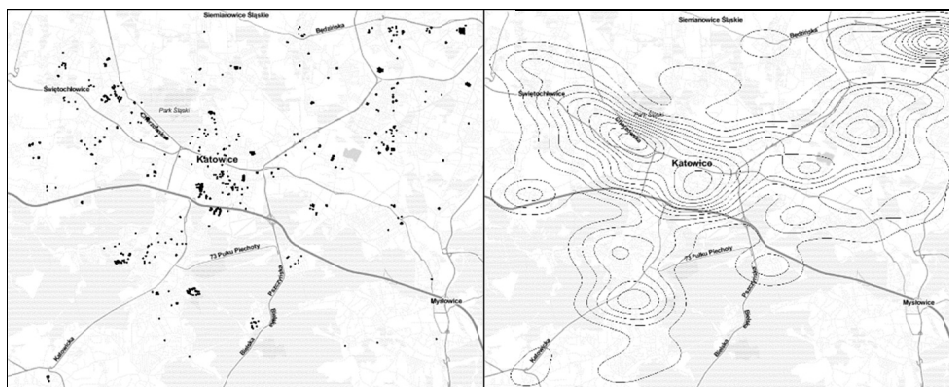
Rozmieszczenie obiektów sportowych na obszarze miasta Katowice jest dość równomierne. Najwięcej znajduje się w obszarze osiedli zlokalizowanych w sąsiedztwie parków: Śląskiego, Parku Kościuszki czy Doliny Trzech Stawów. Dużą gęstość tej klasy obiektów zaobserwowano także w dzielnicach Zawodzie, Szopienice i Dąbrówka Mała.



Rys. 10. Rozmieszczenie (panel lewy) oraz zagęszczenie (panel prawy) obiektów związanych z opieką zdrowotną w Katowicach

Źródło: Obliczenia własne.

Wszelkie obiekty związane z szeroko rozumianą opieką medyczną zlokalizowane są w sąsiedztwie największych katowickich centrów medycznych, tj. na Ochojcu oraz w Śródmieściu i Centrum. Dodatkowo obiekty takie dość gęsto występują na osiedlu Tysiąclecia, najbardziej zaludnionym osiedlu miasta Katowice.



Rys. 11. Rozmieszczenie (panel lewy) oraz zagęszczenie (panel prawy) szkół i placówek edukacyjnych w Katowicach

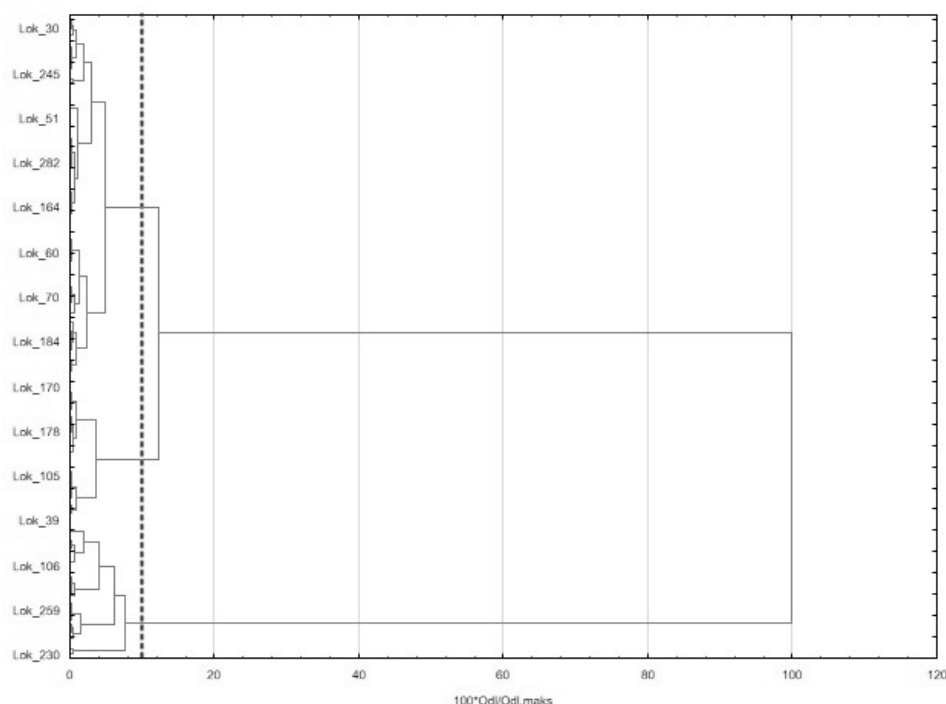
Źródło: Obliczenia własne.

Analizując rozmieszczenie ośrodków edukacyjnych na terenie miasta, można zauważyć, że w tym zakresie również dominuje osiedle Tysiąclecia, ale także znaczne wartości uzyskały również Śródmieście i Centrum. Istotna liczba szkół znajduje się w dzielnicach południowo-zachodnich oraz w sąsiedztwie miasta Sosnowiec (dzielnicie wschodnie). W obszarach tych zlokalizowane są przede wszystkim szkoły ponadgimnazjalne oraz uczelnie wyższe.

2.2. Analiza skupień

Kolejnym etapem badania była próba określenia pewnych jednorodnych obszarów miasta Katowice ze względu na zestaw zmiennych mogących istotnie determinować cenę za nieruchomość. Po wyeliminowaniu przypadków z niepełnymi informacjami, na podstawie opinii eksperta do spraw obrotu nieruchomościami zdecydowano o wyborze dla każdej z dzielnic maksymalnie 20 obserwacji, tak aby każda z badanych dzielnic cechowała się pewną charakterystyczną próbą reprezentatywną. Zmienne, które uwzględniono w tej części badania, stanowiły podstawę przedstawionych w poprzednim podrozdziale prezentacji graficznych.

Pierwszym podjętym krokiem w opisywanej analizie było stworzenie dla badanych zmiennych dendrogramu, tak aby wyznaczyć optymalną liczbę skupień do dalszej analizy. Otrzymany dendrogram zaprezentowano na rys. 12.



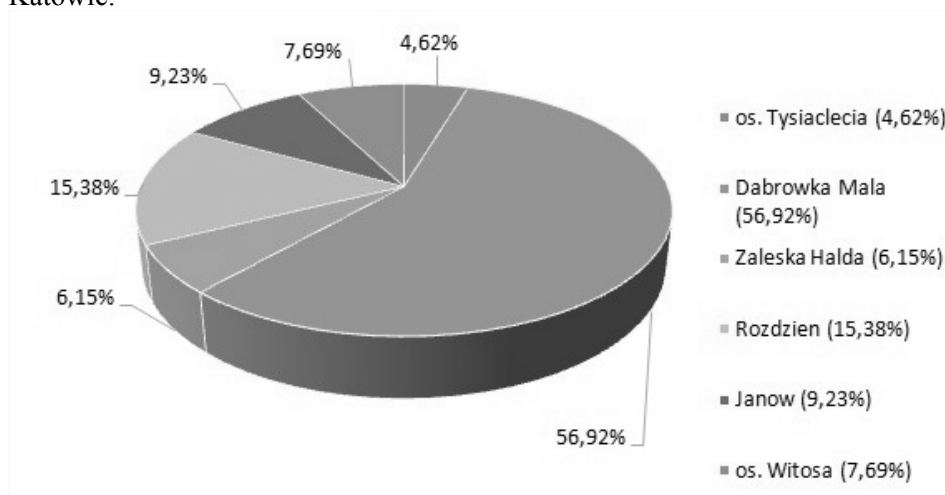
Rys. 12. Szacowanie liczby skupień

Źródło: Obliczenia własne.

W celach obliczeniowych przy tworzeniu dendrogramu zdecydowano o wykorzystaniu kwadratu odległości euklidesowej do obliczenia odległości między przypadkami oraz o wykorzystaniu Metody Warda. Metoda ta znacząco różni się od wszystkich pozostałych proponowanych rozwiązań przez oprogramowanie Statistica, ponieważ do obliczenia odległości między skupieniami wykorzystuje się podejście analizy wariancji. Głównym jej założeniem jest minimalizacja sumy kwadratów odchyleń dowolnych dwóch skupień, które mogą zostać uformowane na każdym etapie.

Na podstawie wiedzy eksperckiej podjęto decyzję o podziale dendrogramu w punkcie, który identyfikuje trzy skupienia. Niestety wbrew postawionym oczekiwaniom otrzymane skupienia nie opierały się na lokalizacji geograficznej nieruchomości. W pierwszym otrzymanym skupieniu zauważono największy

procent (ponad 50%) reprezentowany przez jedną dzielnicę. Jest to skupienie z nieruchomościami położonymi średnio najdalej od centrum miasta w porównaniu z pozostałymi grupami. Kompensuje to na pewno najniższa średnia cena netto oraz najmniejsze średnie odległości przystanków i szkół. Jak zaprezentowano na rys. 13, analizowane obiekty to głównie reprezentanci północnej części Katowic.

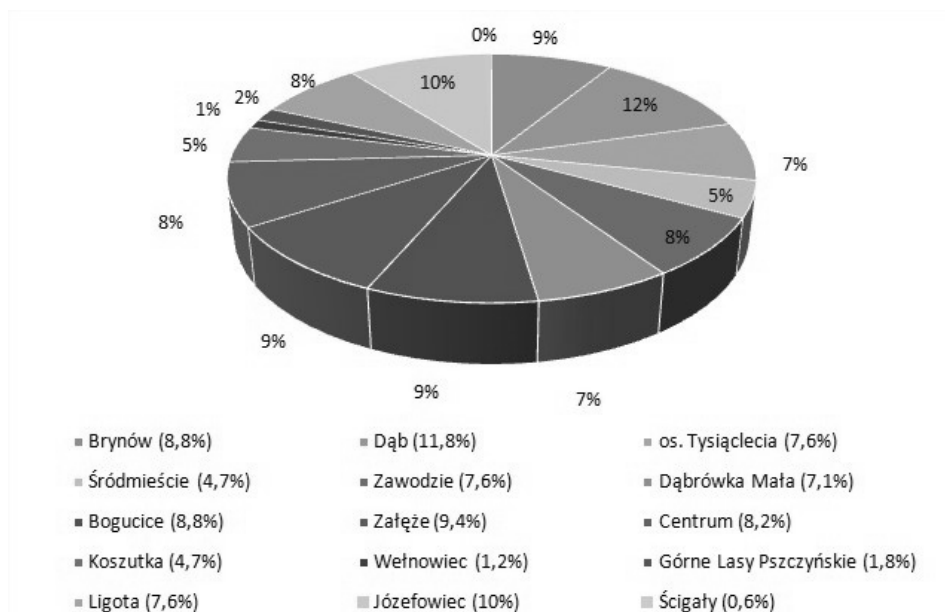


Rys. 13. Rozkład nieruchomości według dzielnic – skupienie 1

Źródło: Obliczenia własne.

Drugie skupienie pod względem udziału procentowego dzielnic prezentuje się inaczej niż skupienie pierwsze. Nie ma tam zdecydowanego lidera udziałowego, lokale rozkładają się równomiernie w kontekście liczebności w poszczególnych dzielnicach stanowiących skupienie. Liczba reprezentowanych dzielnic jest ponad dwukrotnie większa, co skutkuje brakiem możliwości przypisania ich do jakiegoś konkretnego obszaru Katowic (rys. 14).

Dane zaprezentowane w tabeli 2 pokazują uśrednione charakterystyki nieruchomości w każdym ze skupień. Skupienie drugie charakteryzuje się największą ceną netto. Wynik jest niewątpliwie powiązany z odległością od centrum dla nieruchomości należących do tego skupienia. Zaobserwowano ponadto najwyższe przeciętne wartości dla punktów zainteresowania publicznego. Można zatem stwierdzić, że w tej grupie należałoby szukać najbardziej opłacalnego mieszkania.



Rys. 14. Rozkład nieruchomości według dzielnic – skupienie 2

Źródło: Obliczenia własne.

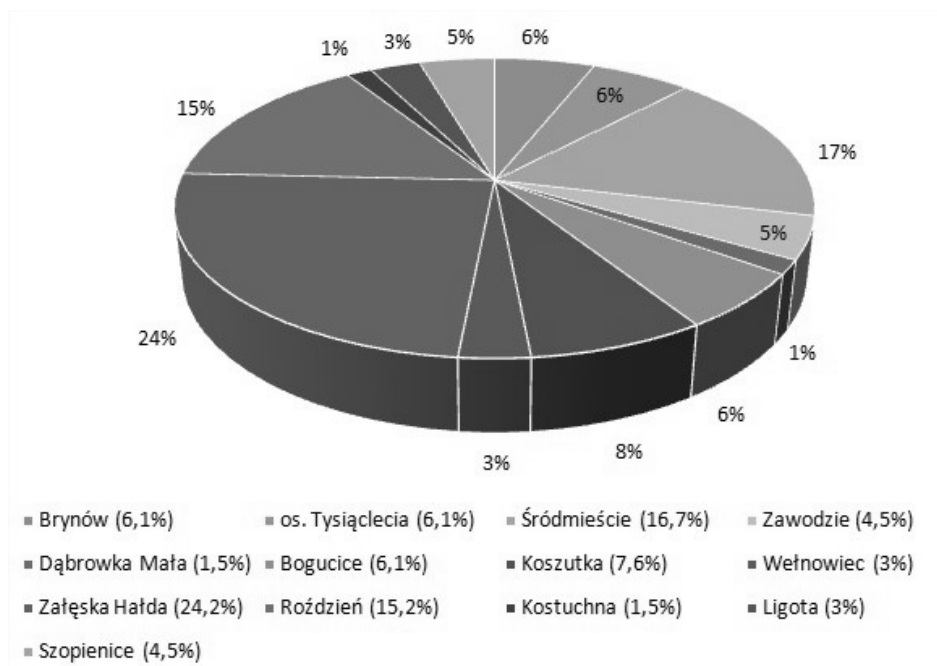
Tabela 2. Zestawienie średnich cech dla otrzymanych skupień

Zmienne	Skupienie 1	Skupienie 2	Skupienie 3
cena_j_netto	2543,3	3656,9	3195,0
pow_użytk	51,3	54,2	55,0
pokoje	2,2	2,2	2,2
szkoły	13,8	19,3	16,1
szkoły-odl	496,5	748,3	1074,0
przystanki	24,1	35,9	26,7
przystanki_odl	327,4	411,9	349,4
medyczne	10,2	14,8	5,0
medyczne_odl	1353,5	762,2	6285,8
markety	10,2	14,8	5,0
markety_odl	1353,5	762,2	6285,8
centrum_odl	4979,7	2587,3	3753,8

Źródło: Opracowanie własne.

Skupienie 3, podobnie jak skupienie 1, reprezentują nieruchomości zlokalizowane w północnej części Katowic, jednakże warto nadmienić, że istotny odsetek to także lokale z dzielnicy Śródmieście (rys. 15). Nieruchomości należące do skupienia 3 zlokalizowane są relatywnie dalej od centrum miasta, skupienie 3 średnio nie ma najbliżej do centrum miasta. Są to przede wszystkim dzielnice połu-

dniowe, np. Kostuchna, które zlokalizowane są na obrzeżach miasta. Analizując atrybuty mieszkań, można dostrzec, że to lokale należące do tego skupienia posiadają charakterystyki o wartościach pośrednich pomiędzy skupieniami 1 oraz 3. Jednakże otoczenie tych nieruchomości wypada zdecydowanie najmniej korzystnie.



Rys. 15. Rozkład nieruchomości według dzielnic – skupienie 3

Źródło: Obliczenia własne.

Podsumowując wyniki analizy skupień, nie udało się jednoznacznie zdefiniować obszarów miasta Katowice cechujących się pewnymi unikatowymi charakterystykami, które mogłyby wpływać na cenę nieruchomości. Trudno tym samym odpowiedzieć na pytanie, dlaczego kupujący (inwestorzy indywidualni) preferują poszczególne nieruchomości w Katowicach i jak ich wybór powiązany jest z ceną takiej nieruchomości. Analiza pozwala przypuszczać, że ceny lokali mieszkalnych zależą głównie od rozmiaru mieszkania/domu, natomiast pozostałe czynniki mają mniejsze znaczenie dla zakupu nieruchomości.

Podsumowanie

Ocena atrakcyjności i konkurencyjności poszczególnych aglomeracji nie jest zadaniem prostym i wymaga niejednokrotnie zastosowania zaawansowanych narzędzi analitycznych. Sposób postrzegania obiektu przez jednostkę warunkowany jest wieloma czynnikami, a co za tym idzie, trudno takie zjawisko mierzyć. W artykule podjęto próbę oceny atrakcyjności miasta Katowice rozumianej w kontekście popytu inwestycyjnego na rynku nieruchomości. Analizowano charakterystyki mogące stanowić istotne determinanty przy procesie wyceny lokali mieszkalnych. Badanie podzielono na dwa etapy: wizualny i analityczny. W części wizualnej wykorzystano autorskie kody programistyczne, za pomocą których zaprezentowano graficznie wybrane charakterystyki nieruchomości w rozmieszczeniu geolokalizacyjnym na mapie miasta. Wyniki pozwalają ocenić gęstość rozmieszczenia poszczególnych obiektów w obszarze Katowic. Zaobserwowano, iż centrum miasta nie jest najbardziej zagęszczonym obszarem ze względu na badane atrybuty. W wielu przypadkach dominują dzielnice obrzebne. Można to tłumaczyć preferencją mieszkańców do zasiedlania obszarów oddalonych od centrum, bliżej terenów zielonych. W części analitycznej zaproponowano wykorzystanie analizy skupień przy hipotetycznym założeniu, że miasto podzielone jest na pewne quasi-rozłączne dystrykty ze względu na charakterystyki mogące stanowić istotne determinanty dla wyceny nieruchomości. Jednakże analiza nie pozwoliła jednoznacznie zweryfikować tego sądu. Wskazano pewne skupienia ze względu na badane atrybuty, jednakże nie są one powiązane z lokalizacją geograficzną lokali mieszkalnych. Należy podkreślić, iż przeprowadzona analiza jest badaniem wstępnym, którego główny cel to rozpoznanie struktury miasta z punktu widzenia wyceny nieruchomości i preferencji mieszkańców determinujących taką wycenę. Na kolejnym etapie weryfikacji poddany zostanie model hedoniczny wyceny nieruchomości przy wykorzystaniu danych klasy Big Data.

Literatura

- Błajejczyk-Majka L., Kala R. (2005), *Metody analizy skupień do charakterystyki użytków rolnych wybranych państw Unii Europejskiej*, „Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu”, t. 7, z. 5, s. 5-10.
- Bryan K.A., Sarte P. (2009), *Semiparametric Estimation of Land Price Gradients Using Large Data Sets*, “Federal Reserve Bank of Richmond Economic Quarterly”, Vol. 95(1), s. 53-74.

- Bugs G. (2007), *Urban Land Value Map: A Case Study in Eldorado do Sul- Brazil*, GIS applications, Instituto Superior de Estatística e Gestão da Informação, Universidade Nova de Lisboa – UNL, Master of Science on Geospatial Technologies.
- Cellmer R. (2014), *Modelowanie przestrzenne w procesie opracowywania map wartości gruntów*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn.
- Colwell P.F. (2003), *Tender Mercies: Efficient and Equitable Land Use Change*, "Real Estate Economics", Vol. 25, No. 4, s. 525-537.
- Earnhart D.H., Ebihara T., Glicksman R., Haider-Markel D. (2005), *Shaping Corporate Environmental Behavior and Performance: The Impact of Enforcement and Non-Enforcement Tools*, University of Kansas, https://cfpub.epa.gov/ncer_abstracts/index.cfm/fuseaction/display.abstractDetail/abstract/1026/report/F (dostęp: 27.03.2018).
- Liu Y. Zheng B., Huang L., Tang X. (2007), *Urban Residential Land Value Analysis: Case Danyang*, "Geospatial Information Science", Vol. 33(5), s. 228-234.
- Łogwiniuk A. (2011), *Zastosowanie metod taksonomicznych w analizie porównawczej dostępu do infrastruktury ICT przez młodzież szkolną w Polsce*, „Ekonomia i Zarządzanie”, vol. 3, nr 1, s. 7-23.
- Munroe D.K. (2007), *Exploring the Determinants of Spatial Pattern in Residential Land Markets: Amenities and Disamenities in Charlotte*, "Environment and Planning B: Planning and Design", Vol. 34(2), NC, USA, s. 336-354.
- Palmquist R.B. (2006), *Property Value Models*, "Handbook of Environmental Economics", Vol. 2, s. 763-819.
- Żyszkowska W., Spallek W., Borowicz D. (2012), *Kartografia tematyczna*, Wydawnictwo Naukowe PWN, Warszawa.
- [www 1] https://en.wikipedia.org/wiki/Application_programming_interface (dostęp: 24.05.2017).
- [www 2] http://wiki.openstreetmap.org/wiki/Pl:Map_Features (dostęp: 14.05.2017).
- [www 3] evolpe.pl (dostęp: 28.05.2017).
- [www 4] <https://overpass-turbo.eu> (dostęp: 14.05.2017).
- [www 5] <https://svn.openstreetmap.org/applications/routing/pyroutelib2/> (dostęp: 17.05.2017).

THE ATTRACTIVENESS OF THE CITY KATOWICE – VISUAL-BASED APPROACH

Summary: Urban agglomerations are described by different characteristics in terms of from their attractiveness. The term "attractiveness" is subjective, not measurable and without an universal definition. The main purpose of this study is to identify factors that significantly affect the perception of the city's attractiveness by its residents, as well as by the visitors. It has been assumed that not only price is an important factor to buy

a property in a given location, but also other factors. In this paper, we propose to classify real estate using cluster analysis and we present in a graphical way some selected factors that may significantly influence the perception of the attractiveness of the city of Katowice.

Keywords: cluster analysis, visualization, attractiveness of the cities, Data Science.