

# Duże zbiory danych na potrzeby monitorowania miejskich rynków mieszkaniowych



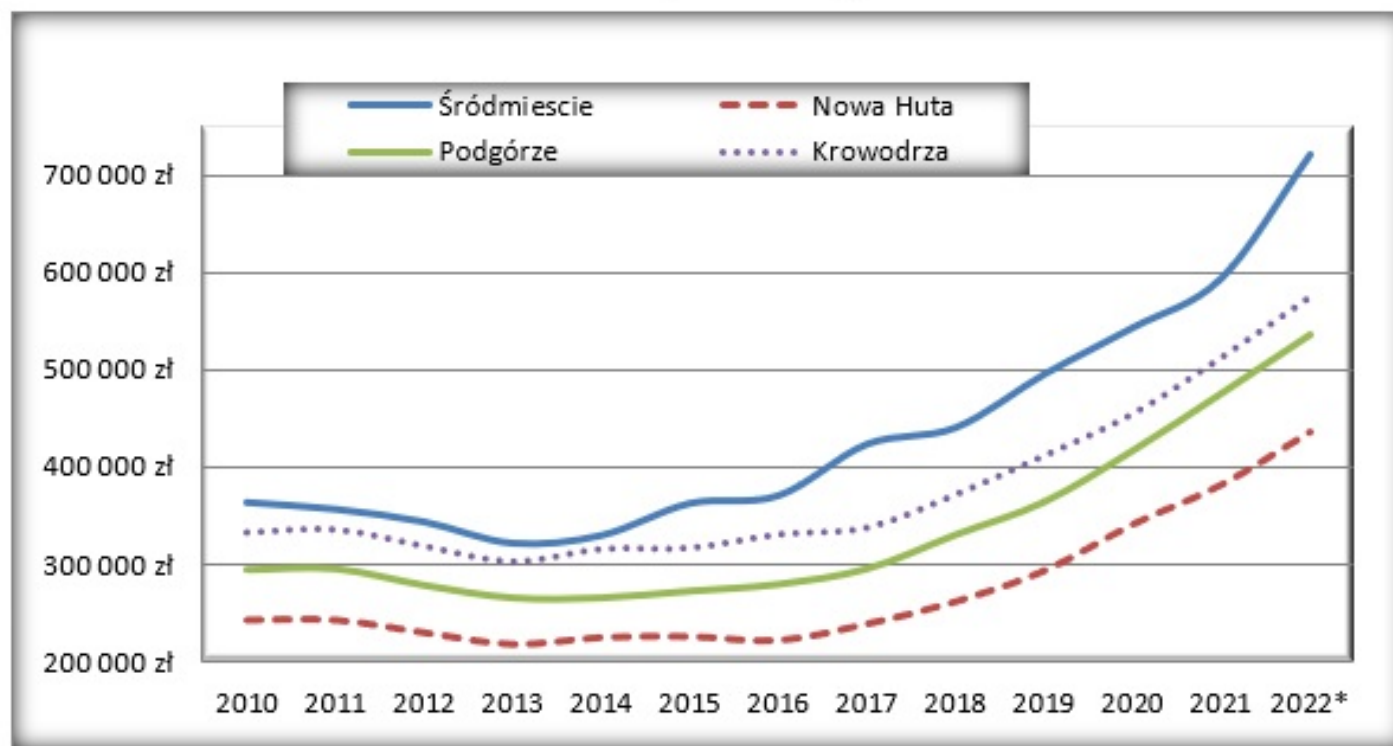
## Plan prezentacji:

- 1) Przesłanki wyboru tematu
- 2) Rodzaje dostępnych dużych zbiorów danych i scenariuszy ich zastosowania w badaniach rynków mieszkaniowych (wieloźródłowe duże zbiory danych według morfologii danych)
- 3) Możliwości i wyzwania związane z wykorzystaniem dużych zbiorów danych w badaniach rynków mieszkaniowych
- 4) Przyszłe kierunki i metody lepszego wykorzystania dużych zbiorów danych na rynku nieruchomości, w tym w wycenie.

## 1. Przesłanki wyboru tematu

- koszty pozyskania danych statystycznych lub ankietowych (powtarzalność badań ankietowych, agregacja roczna, opóźnienie w dostępie do danych, wąski zakres analizowanych cech)
- wady/zalety stosowanych dotychczas modeli
- specyfika rynku nieruchomości (cechy nieruchomości) – rynek mieszkaniowy jako barometr rozwoju gospodarczego
- ilość czy jakość danych ?

Średnia cena mieszkania na rynku wtórnym



Źródło: Instytut Analiz Monitor Rynku Nieruchomości – mrn.pl  
2022\* - wielkości doszacowane z uwagi na niepełne dane

## 2. Rodzaje dostępnych dużych zbiorów danych i scenariuszy ich zastosowania w badaniach rynków mieszkaniowych (wielozródłowe duże zbiory danych wg morfologii danych)

Wraz z powszechnym wykorzystaniem Internetu, technik komunikacji informacyjnej i Internetu Rzeczy ilość danych rośnie wykładniczo (Provost i Fawcett, 2013; Rathore i in., 2016).

W latach 80. pojawiło się pojęcie „ogromnych mas danych” (Toffler, 1980).

Termin „Big data” pojawił się w 1997 roku (Cox i Ellsworth, 1997), a następnie zyskał szerokie zainteresowanie społeczeństwa, w tym badaczy rynku nieruchomości.

Big data odnosi się do ogromnych, różnorodnych i złożonych zestawów danych, które trudno w **tradycyjny** sposób przechowywać, analizować lub wizualizować w celu dalszego przetwarzania (Sagiroglu i Sinanc, 2013).

Big data zapewnia „**pełne pokrycie**” danymi oznacza rewolucję w procesach badawczych (Pei i in. 2020).

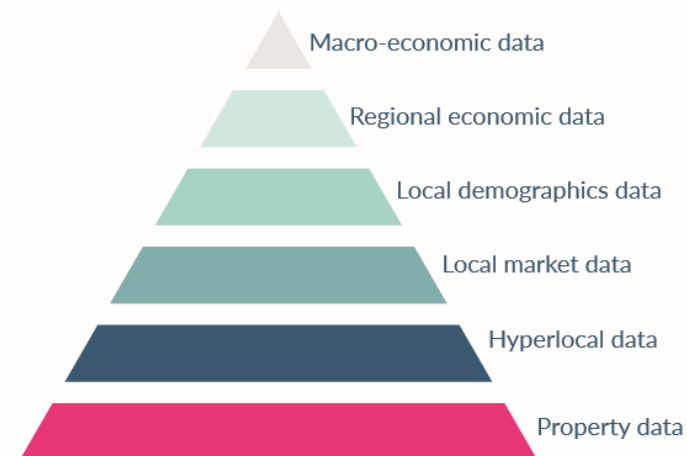
***7 V: volume, velocity, validity, veracity, variety, volatility, and value (Khan et al., 2014)***

**W literaturze wskazuje się na trzy główne scenariusze zastosowania Big Data w badaniach rynków mieszkaniowych:**

- monitorowanie dynamiki transakcji,
- wycena nieruchomości,
- zrozumienie zachowań uczestników rynku (głównie mieszkańców),
- wykorzystanie nowatorskich metod do gromadzenia danych i badań uwarunkowań rynkowych w oparciu o duże zbiory danych

**Przesłanki wykorzystania technologii Big Data na rynku mieszkaniowym:**

- decyzje podejmowane ze wsparciem dużych zbiorów danych są dalej naukowe i efektywne w porównaniu z tradycyjnymi metodami empirycznymi
- dynamiczne monitorowanie cen nieruchomości w czasie rzeczywistym (na wzór notowań giełdowych)
- wydajność wyceny nieruchomości



## Dowody naukowe wykorzystania dużych zbiorów danych na rynku nieruchomości:

- Zastosowanie big data znacznie **poprawia dokładność wyceny** nieruchomości (Guo i in., 2020; Soltani i in., 2022).
- np. Bourassa i in. (2021) porównali efektywność modelu HPI na dużych zbiorach danych i bez nich i stwierdzili **niewielki wkład** dużych zbiorów danych w moc przewidywania cen mieszkań.
- Problem **niespójnego wymiaru czasowego danych**, ponieważ duże zbiory danych są tworzone przez praktyków, a nie specjalnie na potrzeby badań naukowych, co powoduje dylemat wnioskowania statystycznego (Anagnostopoulos i in., 2016).

## Wsparcie praktyki:

- **Zasoby stron Internetowych** (np. Craigslist w Stanach Zjednoczonych, Idealista w Hiszpanii) spontaniczne zasilanie w informacje o nieruchomościach stron Internetowych przez Internautów, w tym informacje niestrukturalne, takie jak adresy, zdjęcia i komentarze geotekstowe. Informacje te były trudne do uzyskania w poprzednich badaniach mieszkaniowych.
- Technologia **web crawler** przeglądania zasobów stron Internetowych rozwiązuje problem źródła danych i ręcznego wyszukiwania (wydajne, niedrogi, szybkie narzędzie zbierania danych w Internecie)

## **Dane można również podzielić w trzy inne grupy (źródło pozyskania oraz charakter danych)**

- 1) Dane numeryczne (statystyka publiczna, instytucje, agencje nieruchomości, portale nieruchomościowe, liczba wejść, liczba kliknięć)
- 2) Dane wyszukiwane w Internecie
- 3) Dokumentacja administracyjna

### **Dane numeryczne:**

- statystyka publiczna (GUS)
- agencje nieruchomości (MLS)
- portale nieruchomościowe (np. gratka.pl)
- instytucje (NBP, Związek Banków Polskich)
- prywatne portale (np. Infoludek)

Tabela 2 Dynamika cen transakcyjnych m<sup>2</sup> mieszkania w analizowanych miastach w IV kw. 2022 r.

Miasto	Dyn. cen nominalnych kw./kw		Dyn. cen nominalnych r/r		Dyn. cen realnych wzgl. CPI */ kw./kw.		Dyn. cen realnych wzgl. CPI */ r/r	
	RPT	RWT	RPT	RWT	RPT	RWT	RPT	RWT
Gdańsk	-7,7%	0,9%	-2,3%	12,2%	-10,9%	-2,6%	-16,7%	-4,4%
Gdynia	13,9%	-4,0%	23,6%	8,7%	-10,0%	-7,3%	5,4%	-7,4%
<b>Kraków</b>	<b>-3,6%</b>	<b>0,1%</b>	<b>20,0%</b>	<b>10,5%</b>	<b>-2,6%</b>	<b>-11,9%</b>	<b>-1,2%</b>	<b>-14,1%</b>
Łódź	0,5%	-7,3%	8,5%	3,7%	-3,0%	-10,5%	-3,5%	-11,6%
Poznań	-0,2%	-5,2%	13,6%	3,3%	-3,7%	-8,5%	-3,2%	-11,9%
Wrocław	-0,6%	-0,6%	10,0%	13,0%	-4,1%	-4,1%	-6,3%	-3,7%
Warszawa	0,8%	1,1%	10,4%	1,6%	-2,7%	-2,4%	-5,9%	-13,4%

Źródło: NBP

### Kraków - ceny ofertowe nieruchomości mieszkaniowych na rynku wtórnym

Jednostka: pln/m<sup>2</sup>

Kategoria: Ceny ofertowe na rynku wtórnym

Częstotliwość: Kwartalna

Branża: Premium



#### LEGENDA:

— Kraków - ceny ofertowe nieruchomości mieszkaniowych na rynku wtórnym



## Dane wyszukiwania w Internecie

- Badanie opinii (np. portale społecznościowe, komentarze, wyszukiwarki wg haseł) – monitorowanie opinii publicznej, co zapewnia natychmiastową reakcję odzwierciedlającą potrzeby i zainteresowania uczestników rynku

Kwantyfikując liczbę wyszukiwań użytkowników Internetu, badacze wygenerowali różne wskaźniki zainteresowania opinii publicznej, takie jak np.:

- indeks nastrojów na rynku nieruchomości (Dietzel, 2016),
- zainteresowanie opinii publicznej polityką dotyczącą nieruchomości (Zhou i in., 2022),
- obawy (Li i in., 2022).

Wykorzystali te wskaźniki do zbadania:

- wpływu przyczynowego na poszczególne fazy cyklu koniunkturalnego
- wolumenu wyszukiwania do przewidywania wolumenu transakcji mieszkaniowych (Lee i in., 2019)
- zmian cen mieszkań (Rizun i Baj-Rogowska, 2021).

## Dokumentacja administracyjna

Rejestry administracyjne zawierają zestawy cech:

- demograficznych takich jak płeć, wiek, poziom wykształcenia i pochodzenie etniczne właściciela domu (Ambrus i in., 2020; Bakens i de Graaff, 2020; Francke i Korevaar, 2021),
- zapisy hipotek mieszkaniowych (Li i Kuang, 2012; Groot i Lejour, 2018)
- zużycie energii w budynkach mieszkalnych w gospodarstwach domowych do wykrywania wolnych mieszkań (Zhou i in., 2022).



## Trzy typy danych (zgodnie z ich metodami pozyskiwania i źródłami):

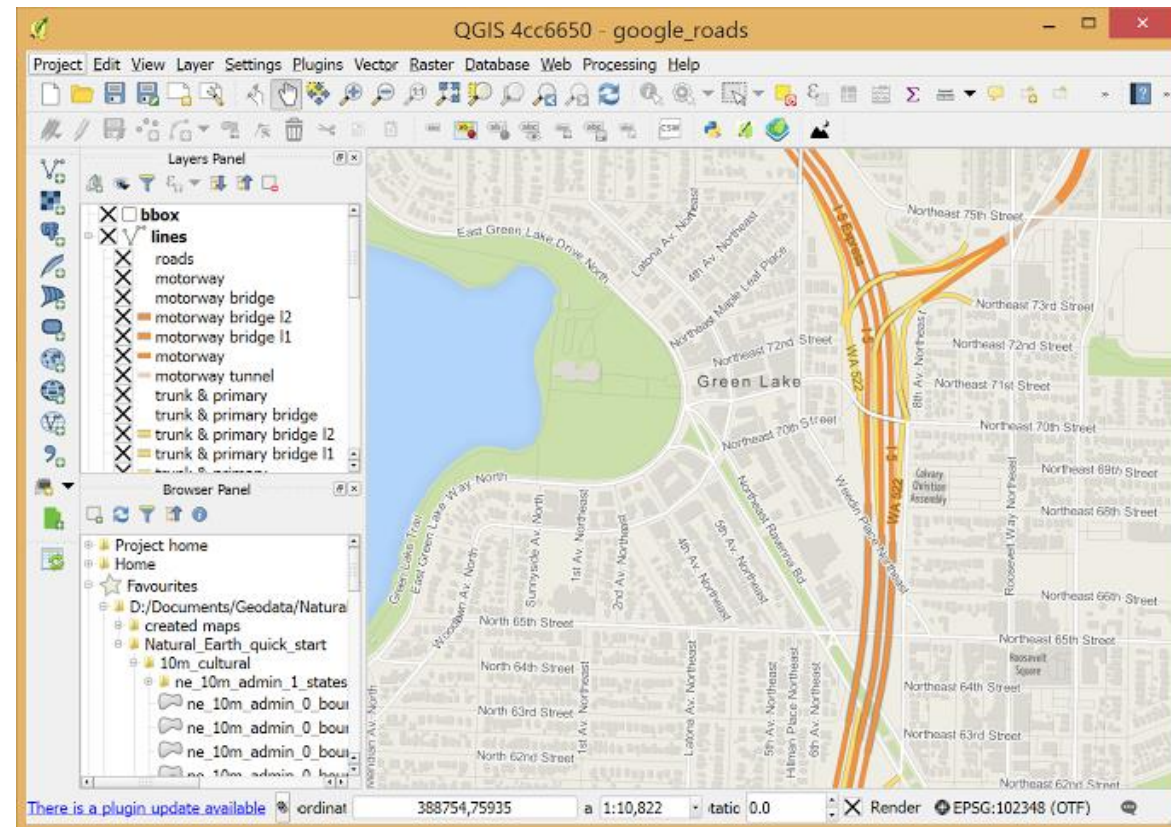
- 1) Internet (platformy biur pośrednictwa, dane z map internetowych, dane o zachowaniach związanych z wyszukiwaniem w Internecie, dane o obrazach i dane tekstowe)
- 2) Teledetekcja (obrazy wielospektralne, światło nocne, radar laserowy)
- 3) IoT Internet rzeczy

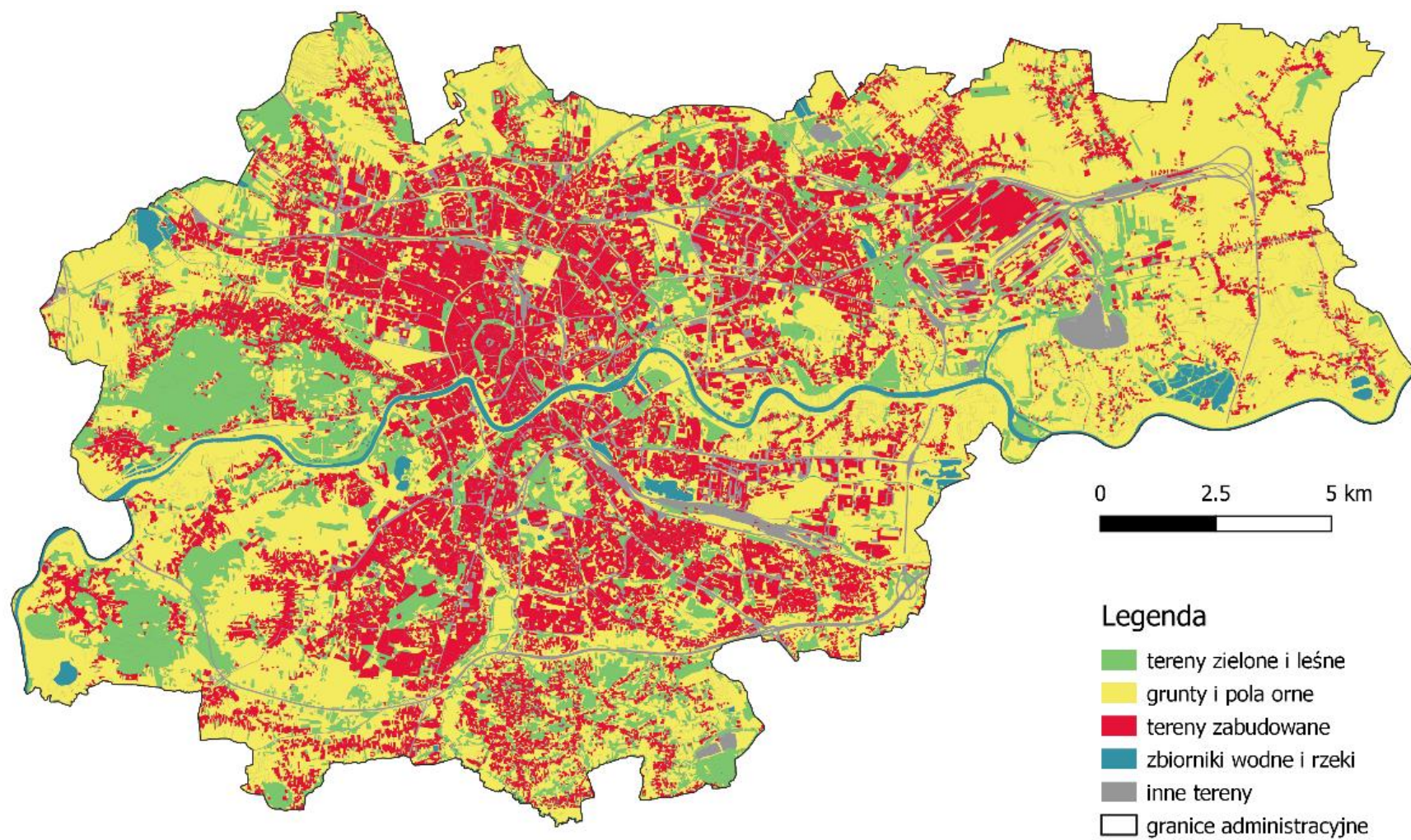
### 1. Internetowe duże zbiory danych

#### Platformy biur pośrednictwa

#### Internetowe dane mapowe

- Geoportal
- QGis
- Google Earth



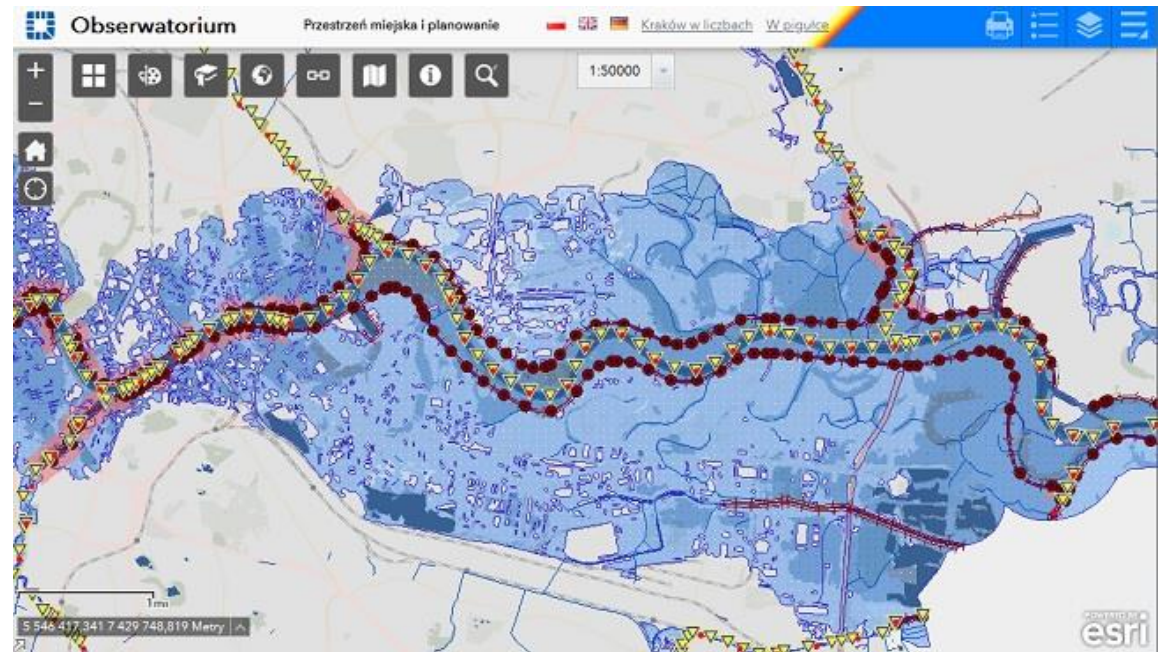


## Kraków z uwagi na sposób wykorzystania terenu

Źródło: UM Kraków

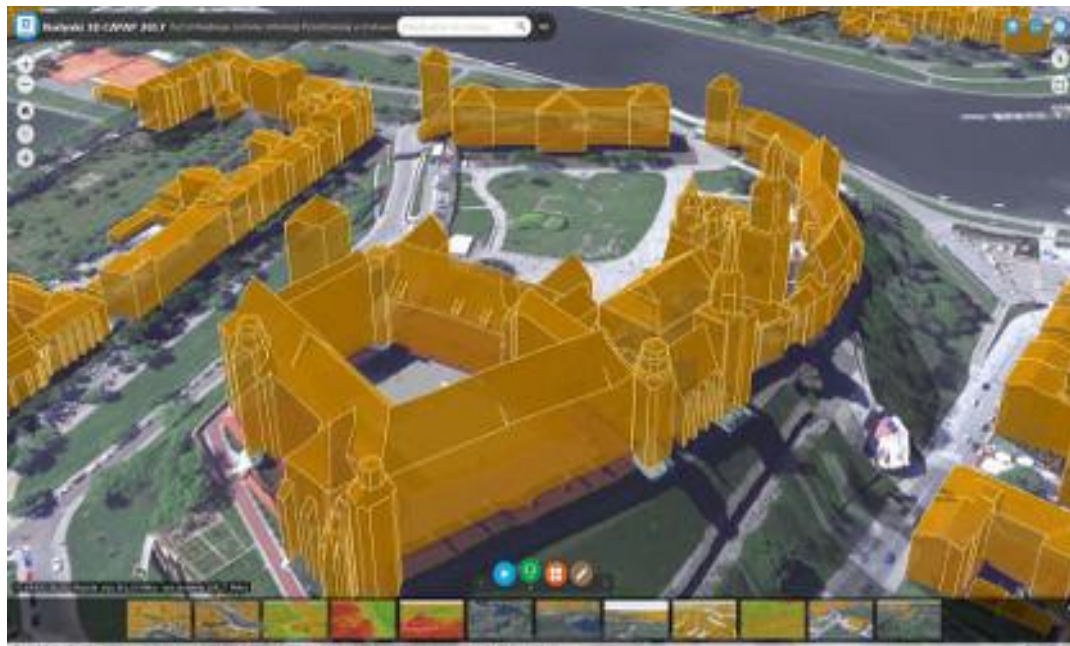


**Geoportal Kraków – Mapa hałasu 2022**



**Geoportal Kraków – Mapa powodziowa**

- obrysy 2D budynków z bazy BDOT10k,
- dane wysokościowe z lotniczego skanowania laserowego LIDAR o gęstości 12 punktów na m<sup>2</sup>, numeryczny model terenu (NMT) o 1 m oczku siatki , granice działek ewidencyjnych z numerami,



**Budynki 3D CAPAP 2017**

## Wyszukiwania Internetowe

- Badacze uważają, że dane dotyczące zachowań w wyszukiwarkach internetowych odzwierciedlają z wyprzedzeniem intencje użytkowników wyszukujących oraz zmiany popytu na rynku nieruchomości (przyszły popyt).
- Istniejące badania potwierdzają również, że czy to na dojrzałych rynkach o wysokiej przejrzystości, takich jak Stany Zjednoczone, czy na rynkach wschodzących o dużej asymetrii informacji, takich jak Indie i Chiny, zmiany wolumenu wyszukiwań słów kluczowych w Internecie odzwierciedlają z góry kilkumiesięczne wahania cen mieszkań.

*Publikowanym indeksom cen nieruchomości brakuje aktualności, mają długi cykl publikacji.*

## Dane wizualne

## Dane tekstowe

Kang i in., (2019) uważają, że **na ceny mieszkań mają wpływ postawy psychologiczne człowieka**. Dane tekstowe składają się ze słów edytowanych przez ludzi i zawierają bogate informacje sentymentalne do wyodrębnienia. Rodzaj danych wykorzystywanych przez naukowców obejmuje ogłoszenia dotyczące mieszkań online i wiadomości dotyczące nieruchomości.

## 2. Technologie teledetekcji

**Technologia teledetekcji** może dostarczać dane obrazu o wielu rozdzielczościach, wielu widmach i wielu czasach. Mogą pomóc w postrzeganiu zewnętrznych warunków środowiskowych, ponieważ zapewniają widok z lotu ptaka ocenianej nieruchomości i otaczającego ją środowiska społeczności. W szczególności ma trzy kierunki zastosowania, jak następuje:

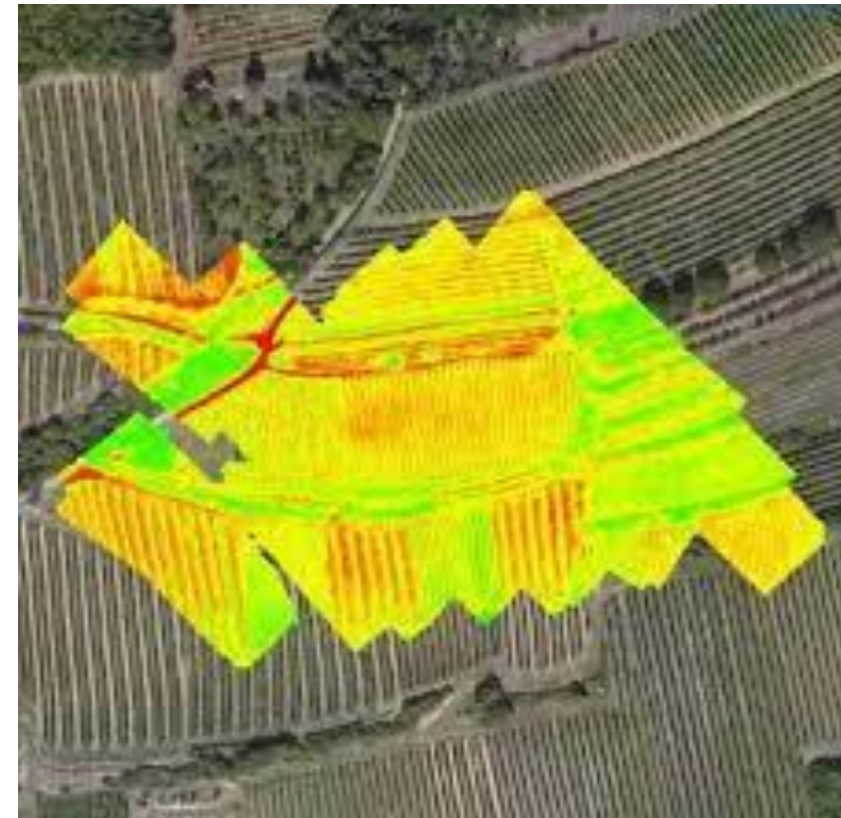
**Obrazy wielospektralne:** Informacje o środowisku powierzchni dostarczane przez obrazy wielospektralne są znacznie bogatsze niż obrazy w rzeczywistych kolorach. Zastosowanie multispektralnych obrazów teledetekcyjnych koncentruje się głównie na obliczaniu wskaźnika środowiskowego. To niewątpliwie oferuje bardziej przekonujące atrybuty związane z sąsiedztwem.

**Światło nocne:** nieruchomości mają silne powiązania z atrybutami społecznymi i ekonomicznymi, Np. stwierdzono, że światło nocne (NTL) jest wysoce przestrzennie skorelowane z (PKB) i gęstością zaludnienia. NTL skutecznie pokazuje dobrobyt miast i dlatego niektórzy badacze używali go również do modelowania cen mieszkań.



**Radar laserowy:** może dostarczyć **nowych trójwymiarowych parametrów przestrzennych dla modelu oceny**. Do tej pory odpowiednie badania obejmowały użycie go w powietrzu do pomiaru powierzchni użytkowej i kubatury domów, a także zbadały wpływ kąta patrzenia na morze na wartość domów na plaży za pomocą informacji geomorficznych o przybrzeżnych obszarach. Wada- zastosowanie na dużych obszarach jest nieefektywne kosztowo i niewydajne.

***Cena nieruchomości jest ściśle związana ze środowiskiem sąsiedztwa. Badanie porównawcze pokazuje, że połączenie obrazów teledetekcyjnych i innych danych geoprzestrzennych może znacznie poprawić dokładność wyceny nieruchomości.***





### 3. IoT Internet rzeczy

Internet rzeczy jest szeroko stosowany w różnych dziedzinach ze względu na jego **działanie w czasie rzeczywistym i interaktywne właściwości**. W wyniku tych zalet, IoT został również eksperymentalnie prowadził badania w zakresie wyceny nieruchomości.

#### Dane z kart inteligentnych (SCD)

anonimowe dane dotyczące podróży ludzi, są powszechnie używane do reprezentowania dostępności różnych zasobów publicznych w obszarach, w których mieszkają ludzie. Obliczając **wskaźniki dostępności**, można zbadać związek między nimi a cenami mieszkań.

#### Dane pozycjonowania GPS

Śledząc pozycję urządzenia wyposażonego w moduł GPS w czasie rzeczywistym, można ocenić pozycję lub trajektorię aktywności jednej lub więcej osób.

W dziedzinach związanych z nieruchomościami GPS samochodowy i GPS telefonu komórkowego są wykorzystywane do dostarczania informacji o działalności człowieka, które następnie można analizować w celu **określenia preferencji** ludzi dotyczących nieruchomości z perspektywy popytu.

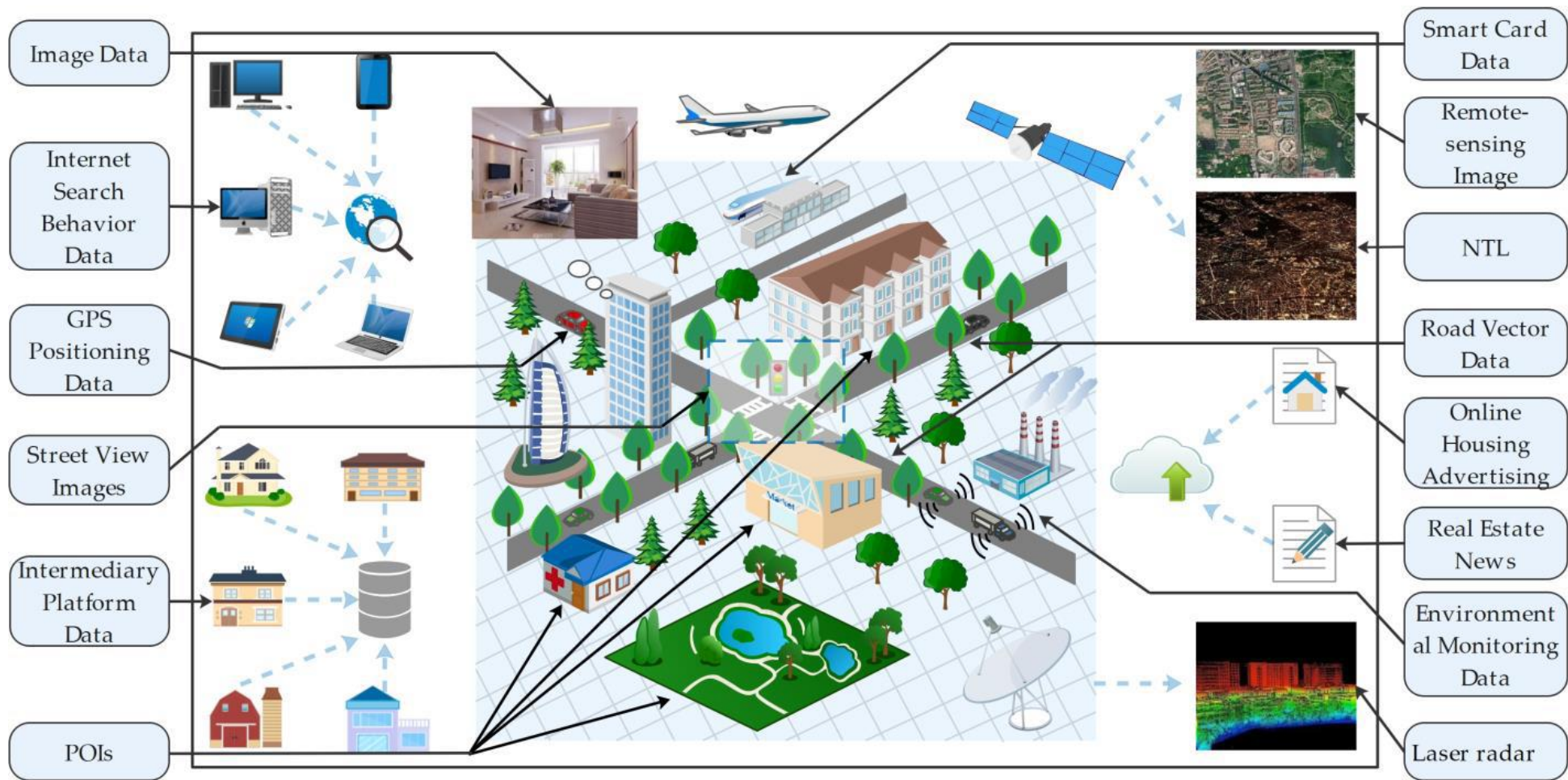
## Dane z monitoringu środowiska

są dostarczane głównie przez instytucje rządowe, samorządowe za pomocą czujników środowiskowych, w tym dotyczących **natężenia hałasu, jakości powietrza, zanieczyszczenia wody i innych informacji o środowisku.**

Na przykład Mei i in. (2020) stwierdzili oczywistą ujemną korelację między zanieczyszczeniem powietrza w miastach a cenami domów w Pekinie w Chinach.

Badanie przeprowadzone przez Zambrano-Monserrate i Ruano (2019) w Machala City w Ekwadorze pokazuje, że każdy wzrost hałasu w środowisku o 1 decybel obniża ceny domów o 1,97%.





**Mapa powiązań Big Data i źródeł ich pozyskiwania**

Źródło: Wiewin, 2022.

#### 4. Możliwości i wyzwania związane z wykorzystaniem dużych zbiorów danych w badaniach rynków mieszkaniowych

Bez względu na to, jakie metody są stosowane, analiza przestrzenna GIS musi być uwzględniona w analizie na kilku etapach:

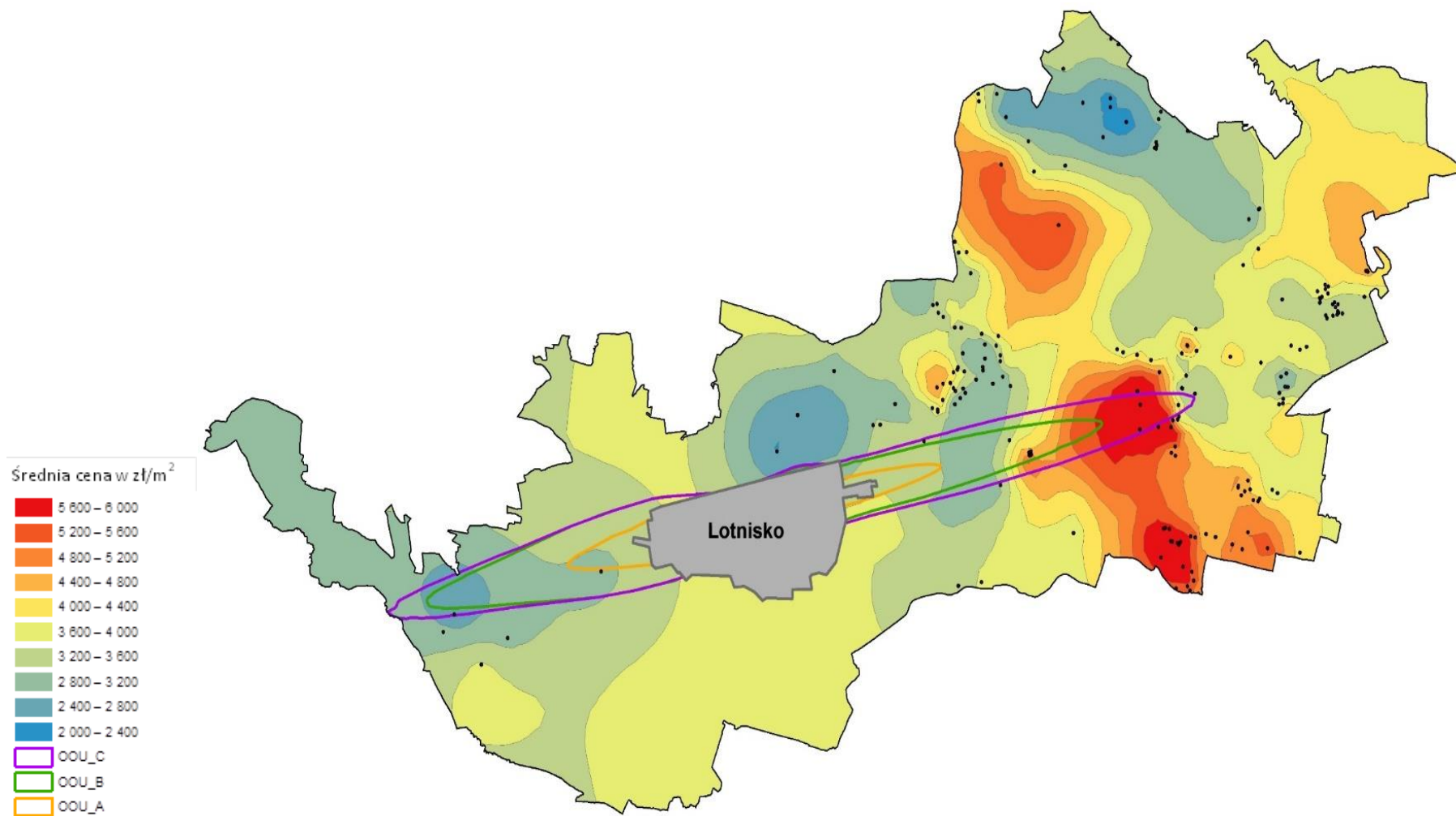
- Wstępne przetwarzanie danych
- Modelowanie przestrzenne
- Analiza przestrzenna GIS
- Nauczanie maszynowe

##### **Wstępne przetwarzanie danych**

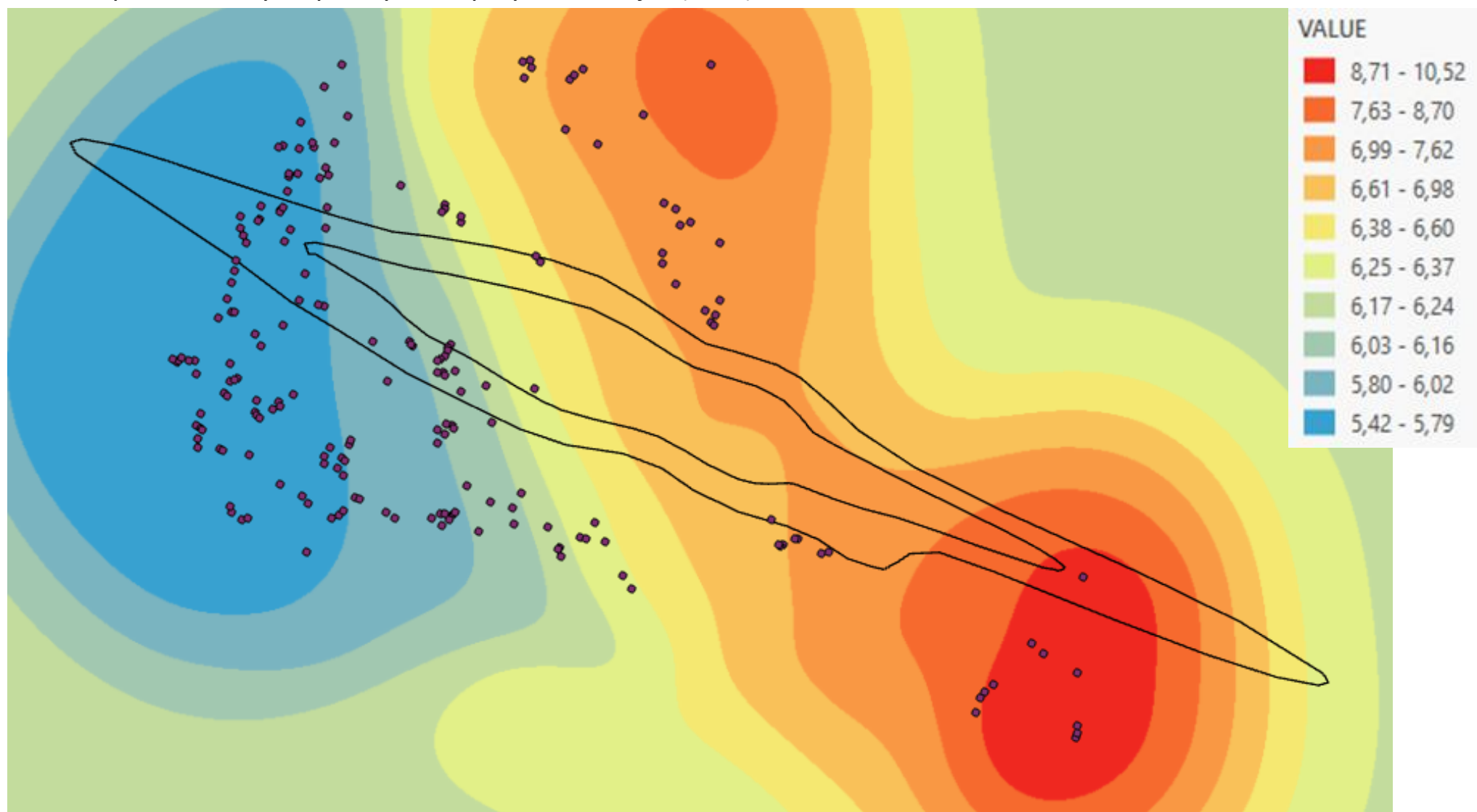
- duże zbiory danych pozwalają na pozyskanie wielu danych wielowymiarowych, ale mogą powodować poważne problemy w zakresie współliniowości i nadmiernego dopasowania, jeśli zostaną bezpośrednio zastosowane w modelach statystycznych (wybór cech, redukcja wymiarowości, normalizacja).
- dane przechwytywane przez Internet są **nieuchronnie pełne „brudnych” danych**, czyszczenie danych staje się warunkiem wstępnym analizy dużych zbiorów danych, stanowiącej ważną część wstępnego przetwarzania.

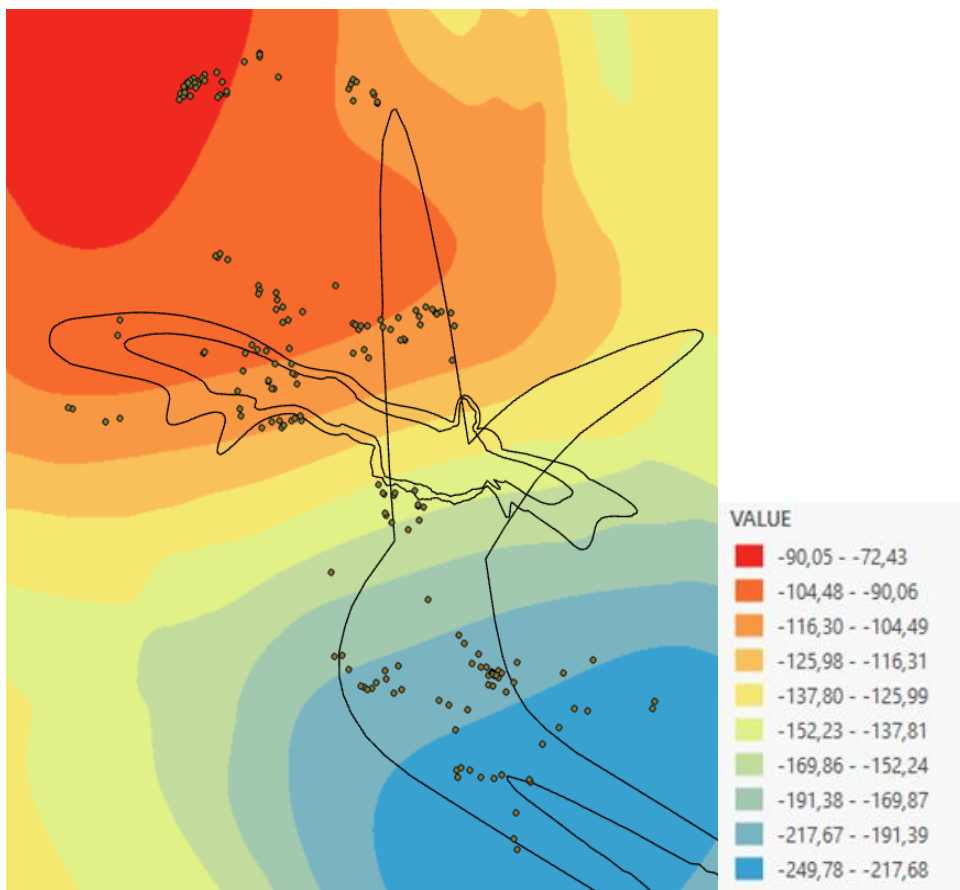
# Przestrzenne zróżnicowania średnich cen gruntów (interpolacja przestrzenna metodą krzygu zwykłego) PL

Kraków

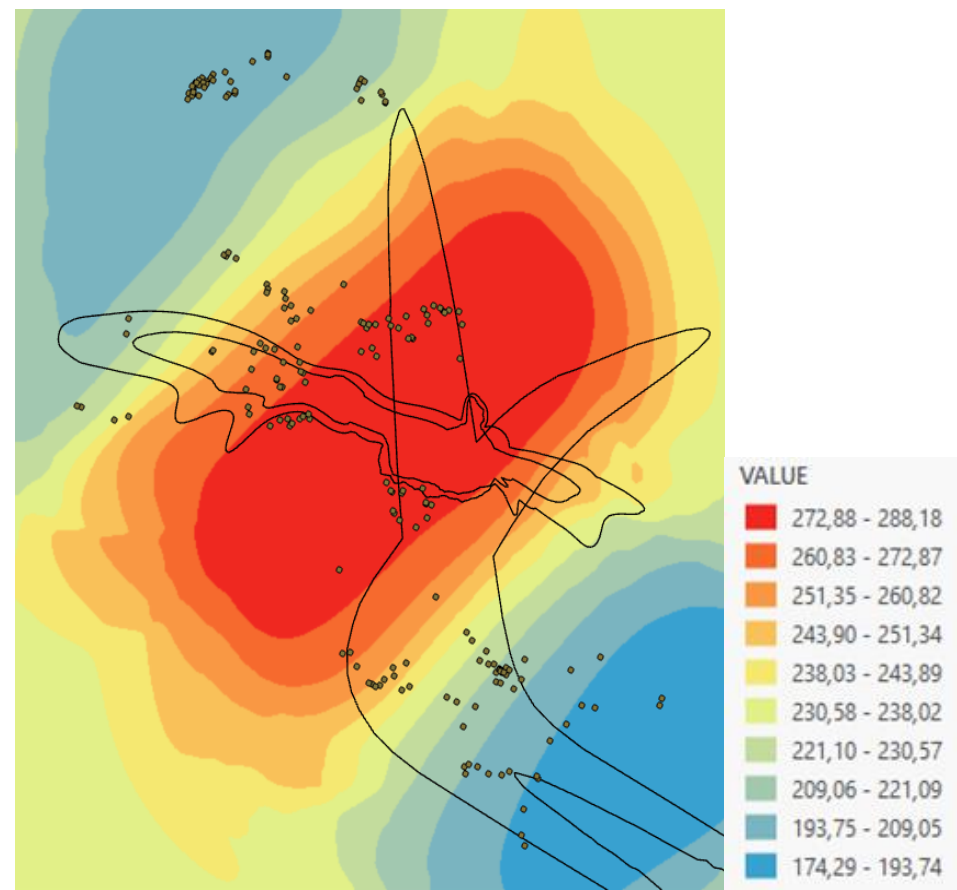


Rozkład przestrzenny współczynnika przy zmiennej D (data) modelu GWR dla cen w otoczeniu PL Gdańsk

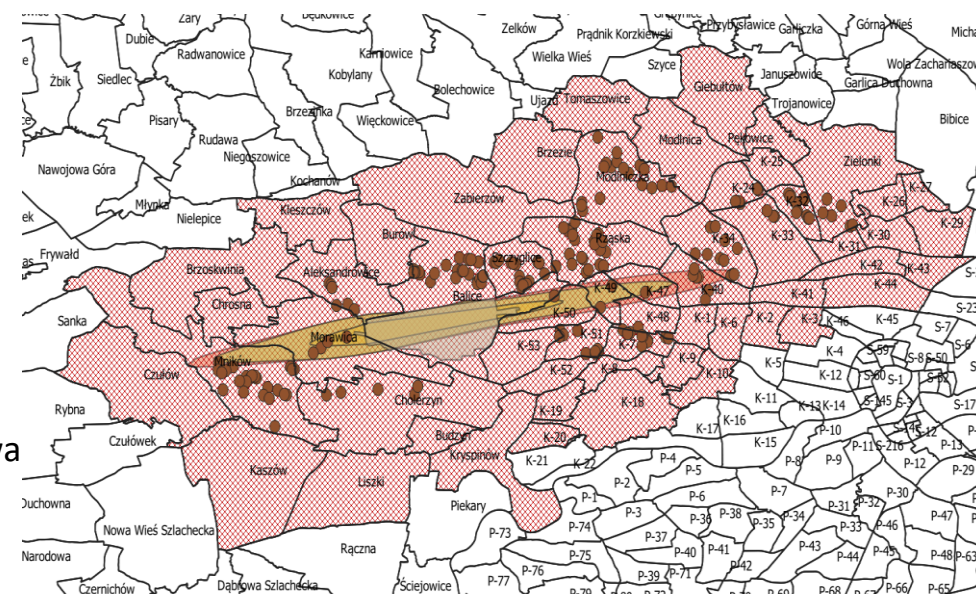
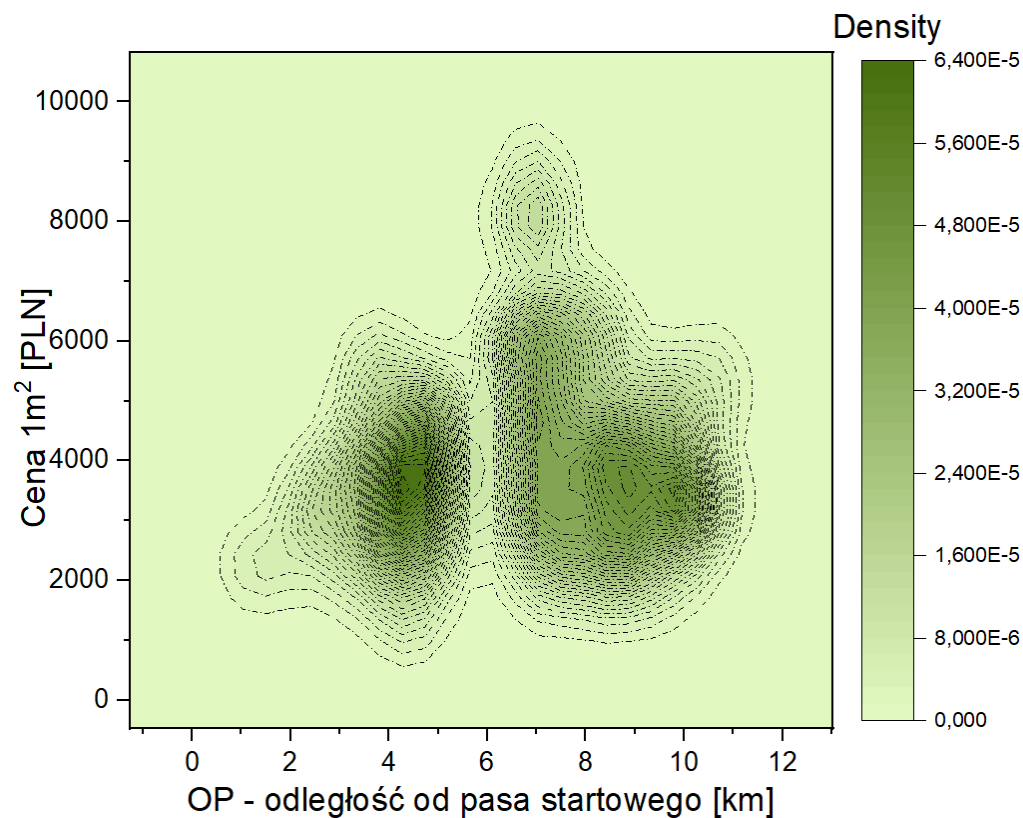




Rozkład przestrzenny współczynnika przy zmiennej BUS (odległość od przystanków autobusowych) modelu GWR dla cen w otoczeniu PL Poznań



Rozkład przestrzenny współczynnika przy zmiennej ATTRACTION (odległość od charakterystycznych miejsc) modelu GWR dla cen w otoczeniu PL Poznań



Wykres 2D nieparametrycznej funkcji gęstości prawdopodobieństwa opartej na estymacji jądrowej gęstości transakcji w zależności od ceny jedn. domu i odległości od pasa startowego

Szerzej o zrealizowanym projekcie Sowa 2020 i realizowanym projekcie Sowa 2023 czytaj na stronie:

[Projekty – Katedra Ekonomiki Nieruchomości i Procesu Inwestycyjnego \(uek.krakow.pl\)](http://uek.krakow.pl)



## Wnioski

1. Łatwa dostępność i niski koszt dużych zbiorów danych umożliwiają rozwój np. dziedziny wyceny nieruchomości w kierunku wycen masowych.
2. Powstają darmowe platformy oferujące zbiory danych oraz automatyczne systemy wyceny dla każdego użytkownika.
3. Duże zbiory danych online mogą być wrażliwe na zmiany nastrojów rynkowych co pozwala bezpośrednio mierzyć nastawienie opinii publicznej do rynku nieruchomości.
4. Internetowy indeks nieruchomości może służyć do przewidywania z wyprzedzeniem zmian na rynku.

### **Pozostają pytania metodyczne:**

*W jaki sposób efektywnie wykorzystać duże zbiory danych, wybrać zestawy danych ewaluacyjnych i jak ustalić kryteria zastosowania dużych zbiorów danych do ewaluacji?*

Dziękuję za uwagę ...

