



Konferencja
Naukowo-Techniczna
Miasto i Transport

2025

METODY WERYFIKACJI I OCENY DANYCH BIGDATA Z KART SIM Z PRZYKŁADAMI

Karolina Jesionkiewicz-Niedzińska



**Wydział
Inżynierii Lądowej**

POLITECHNIKA WARSZAWSKA

Warszawa 8-9.05.2025

PLAN PREZENTACJI:

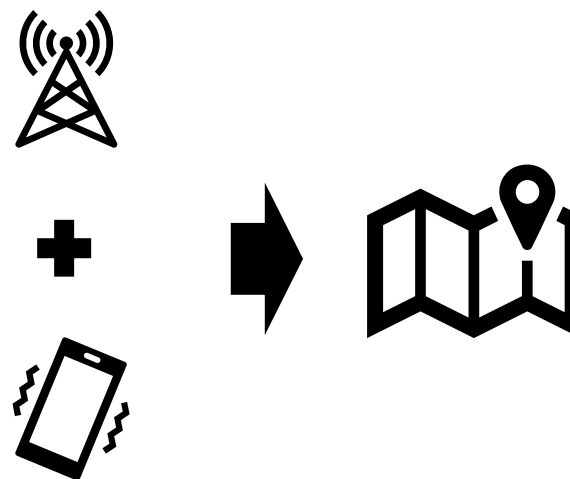
- 1.Charakterystyka danych z kart SIM
- 2.Metody weryfikacji
- 3.Niewykorzystany potencjał

Zrozumienie czym są dane z kart SIM może ustrzec przed błędami w ich wykorzystaniu

Rejestracja danych z kart SIM następuje przy:

- **każdej interakcji** (rozmowa, SMS, korzystanie z internetu- dane aktywne).
- **przełączaniu się między BTSami** (włączenie telefonu, poszukiwanie lepszego sygnału- gdy następują przemieszczenia- dane pasywne)

Są to dane: **punktowe, zdarzeniowe i częściowe.**

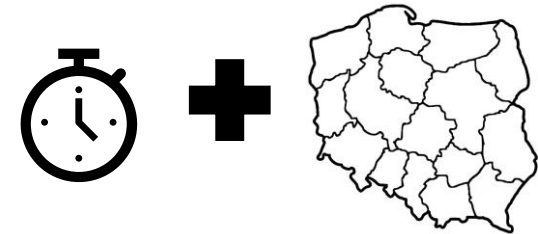


SIM: są mniej dokładne ale zbierane masowo

GPS: bardziej dokładne ale dotyczą tylko użytkowników z włączonym GPS

Charakterystyka danych z kart SIM

- Dane przestrzenno-czasowe: lokalizacja + czas
- Próbkowanie (rok, miesiąc, doba, godzina...)
- Agregacja (kraj, województwo, powiat, gmina, miasto, rejon...)



Metody weryfikacji danych:

- Sprawdzenie zbieżności danych w bazach danych z kart SIM
- Wykrywanie anomalii
- Walidacja semantyczna
- Porównanie z danymi referencyjnymi
- Porównanie danych od różnych operatorów

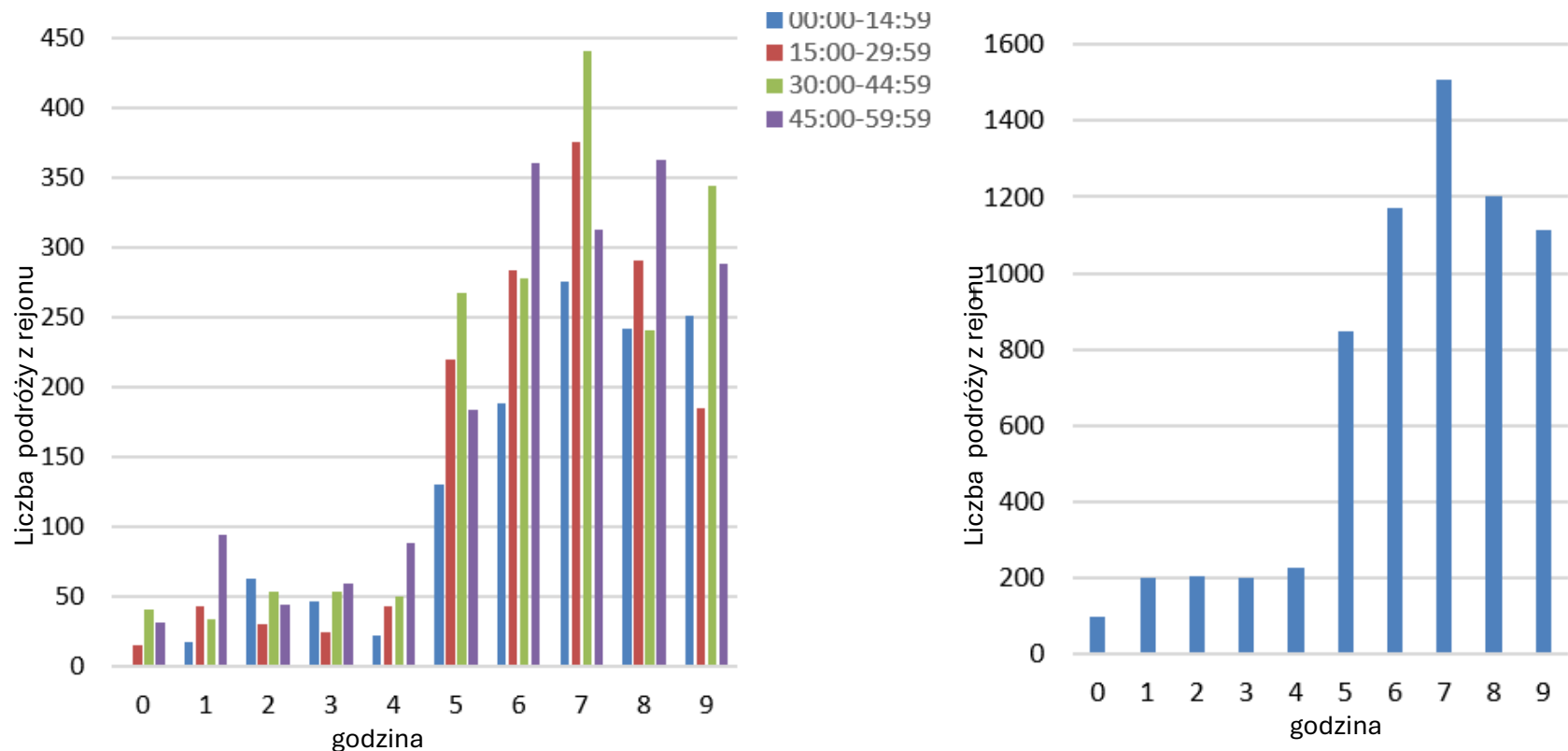


- Porównanie potencjału wybranych rejonów z liczbą podróży

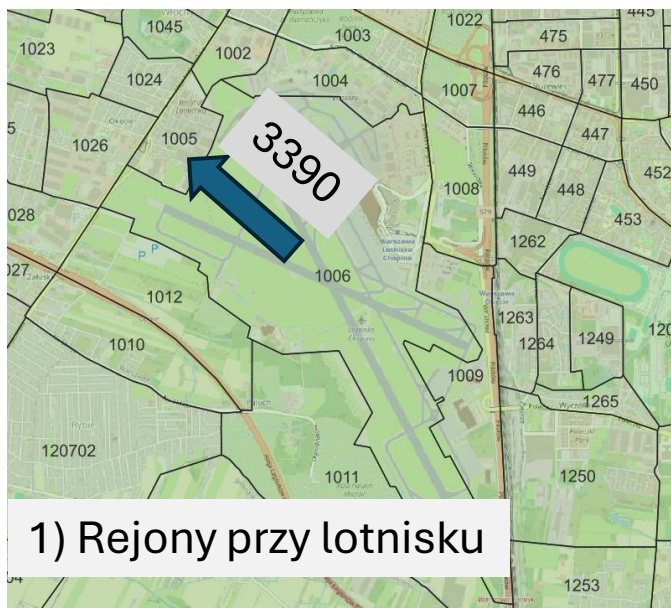
Map of the administrative district of Białka, showing the location of the village of Białka (yellow) and the village of Łańsk (red). The map includes surrounding villages and the number 1692.

Liczba wyjazdów z rejonu na podstawie potencjału i na podstawie podróży.

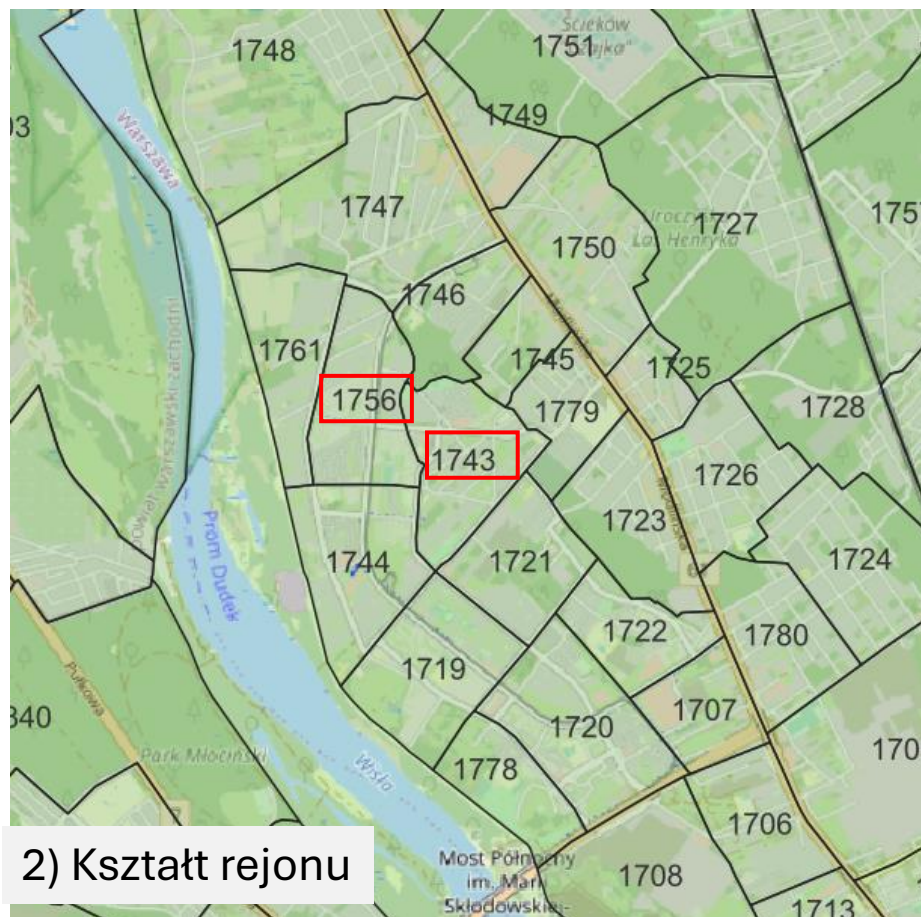
- Porównanie tych samych danych w bazach o różnym próbkowaniu (wyjazdy z rejonu w próbkowaniu 15 min i godzinowym)



próbkowanie	0	1	2	3	4	5	6	7	8	9
15 min	98	199	206	198	226	849	1173	1506	1200	1115
1h	90	189	193	186	206	803	1112	1406	1137	1069

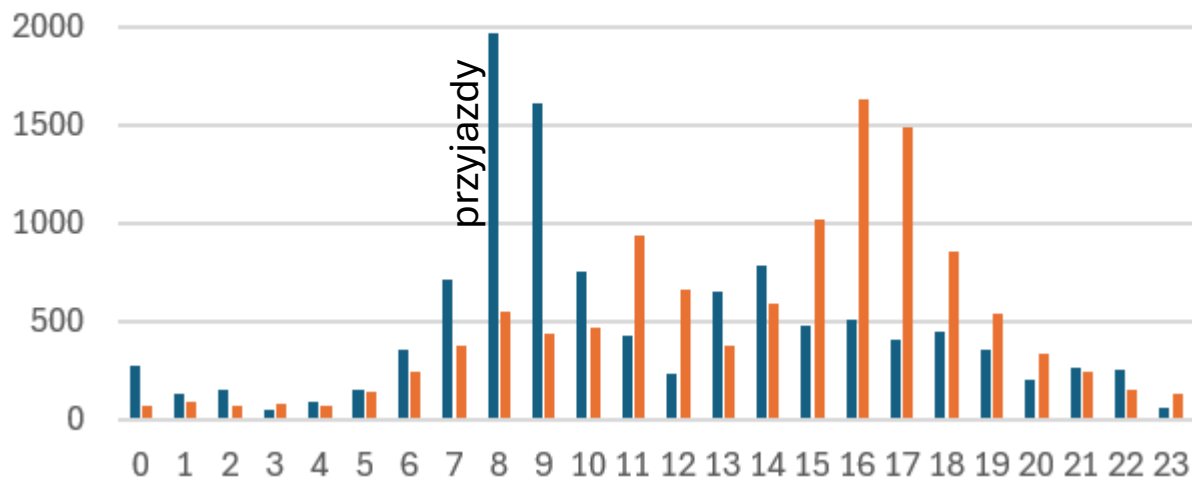
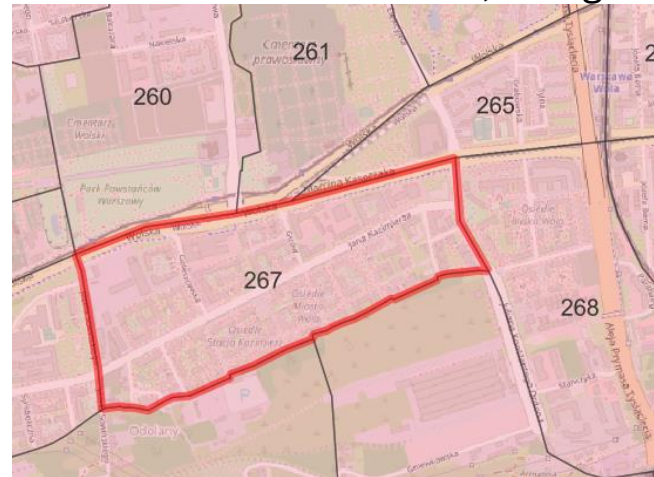
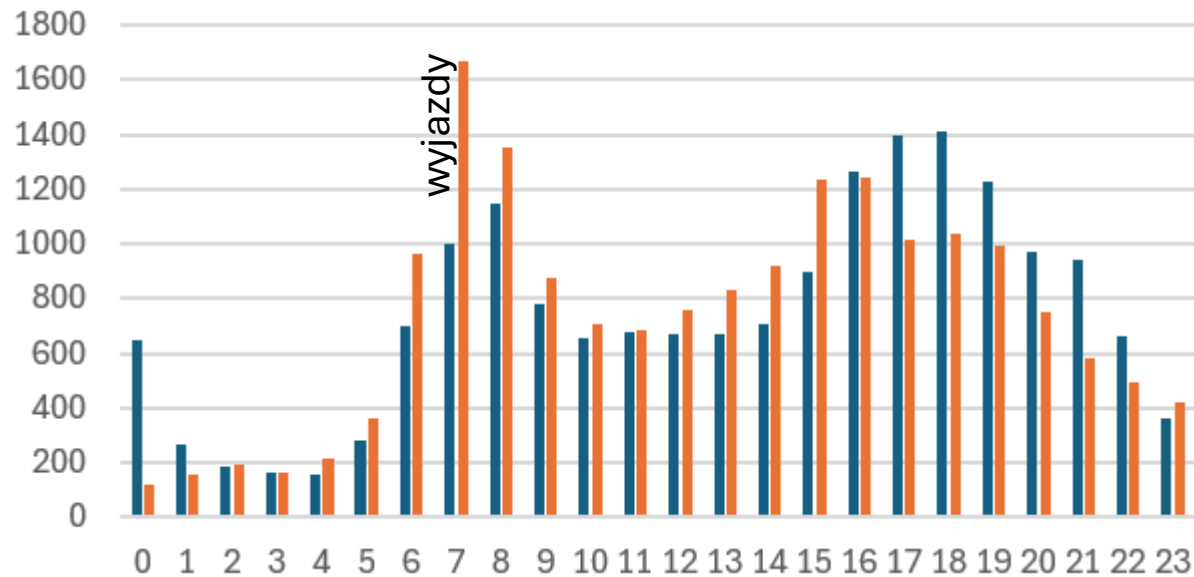


z rejonu	do rejonu	godz	l. podróży
1756	1743	1	170
1720	1743	1	11
1721	1743	1	11
1744	1743	1	11
1746	1743	1	11
1707	1743	1	5
1728	1743	1	5
1745	1743	1	5
1748	1743	1	5



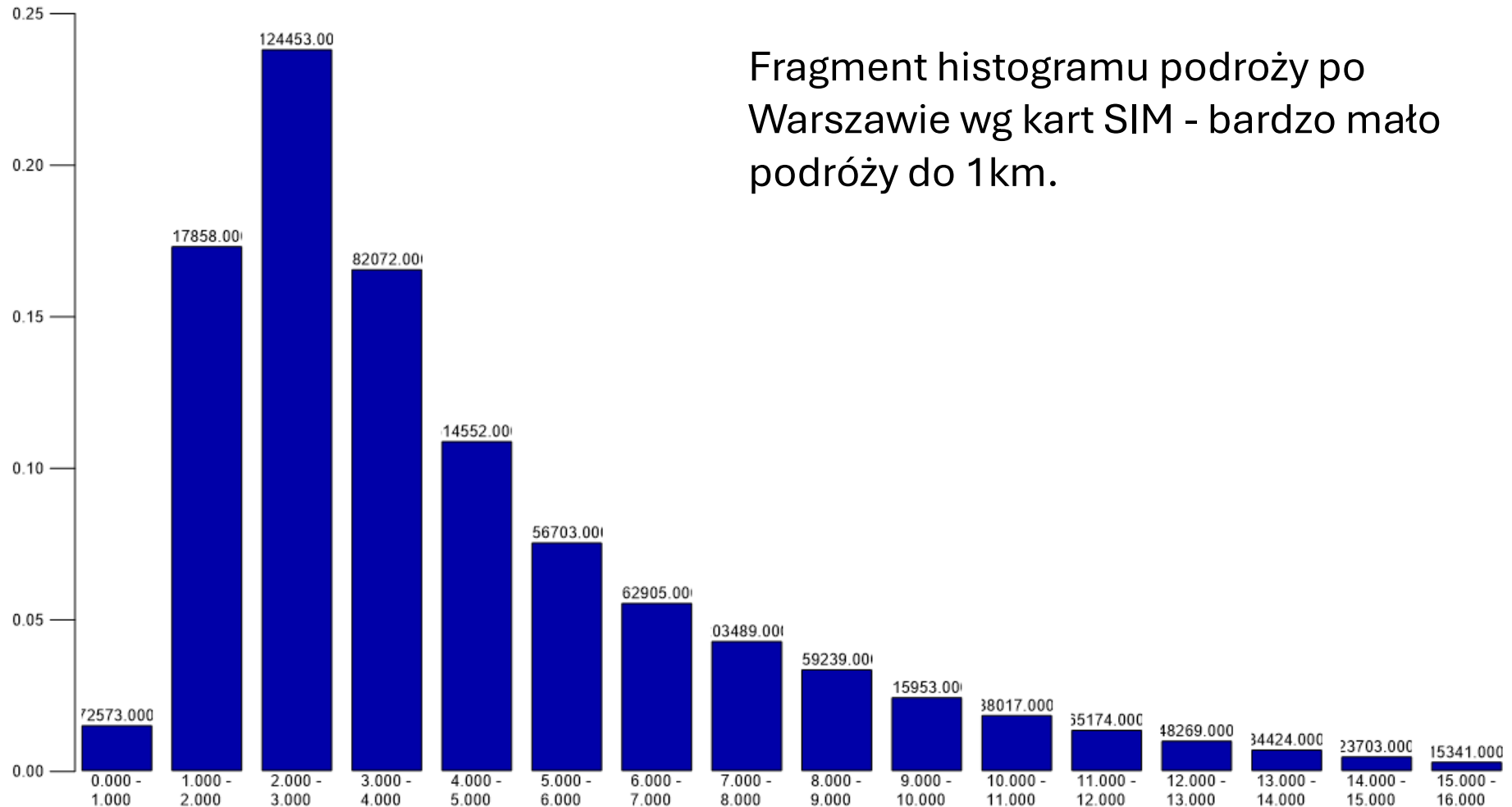
Najwięcej podróży w ciągu 1 godziny do rejonu 1743 ma miejsce między 23:00-24:00 (267 podróży) i występuje z rejonu 1756.

Walidacja semantyczna- spodziewany rozkład godzinowy potencjałów

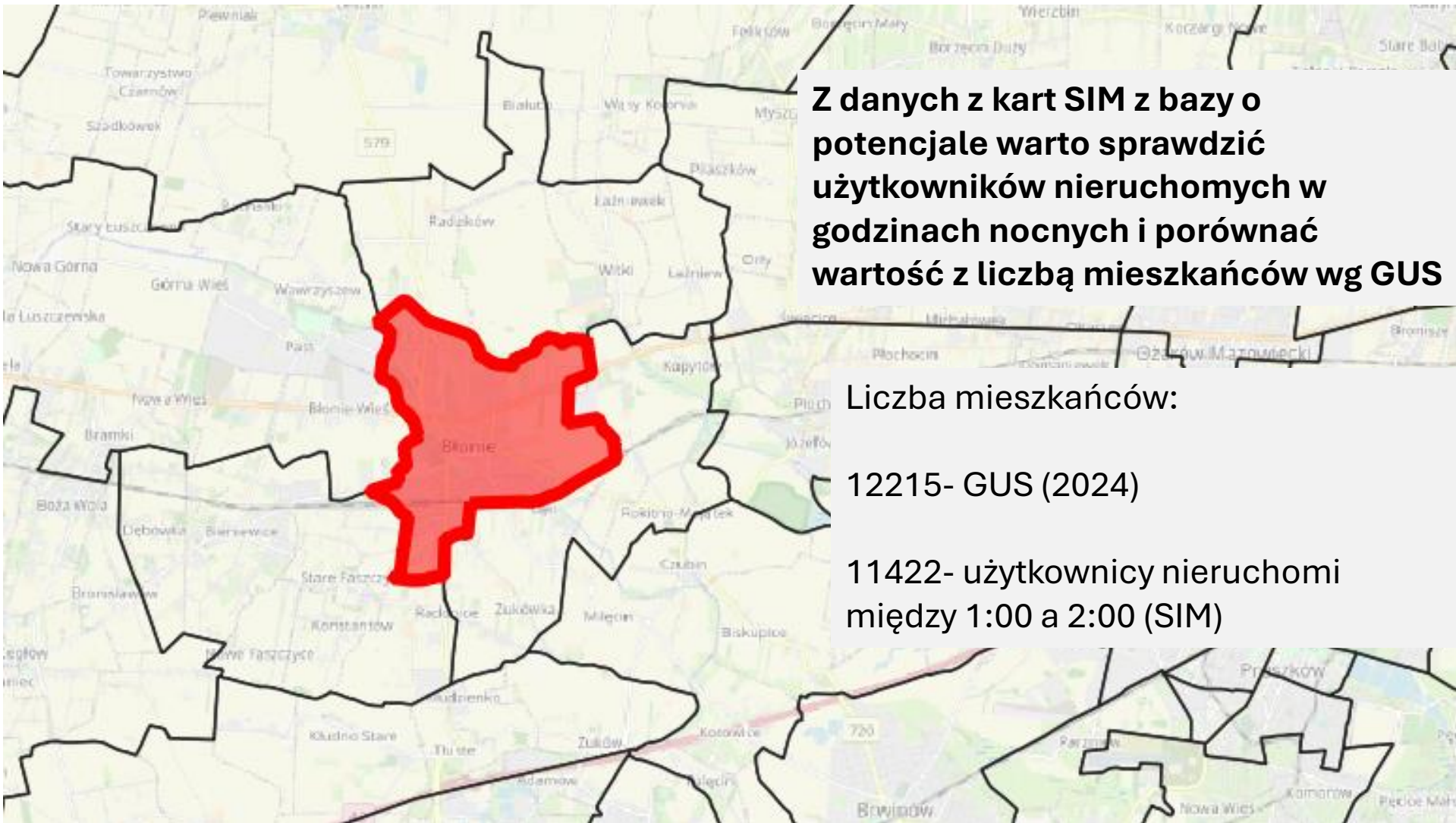


Zabudowa biurowa

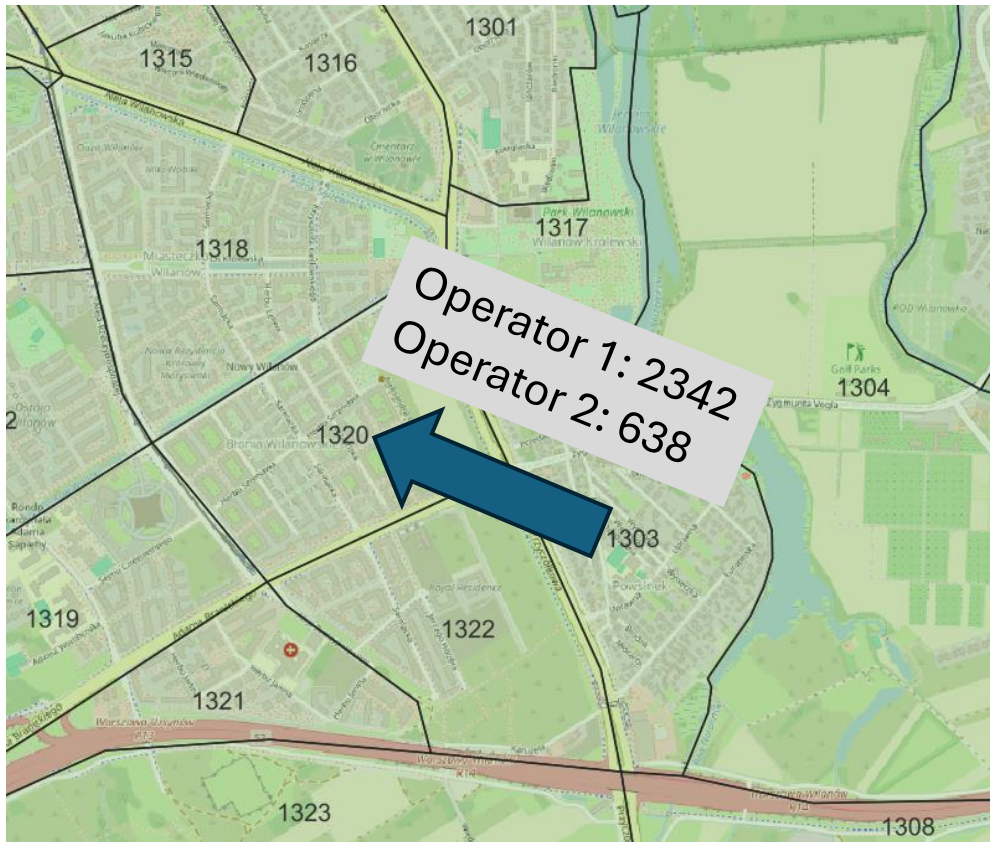




Porównanie z danymi referencyjnymi- dane demograficzne, liczba mieszkańców



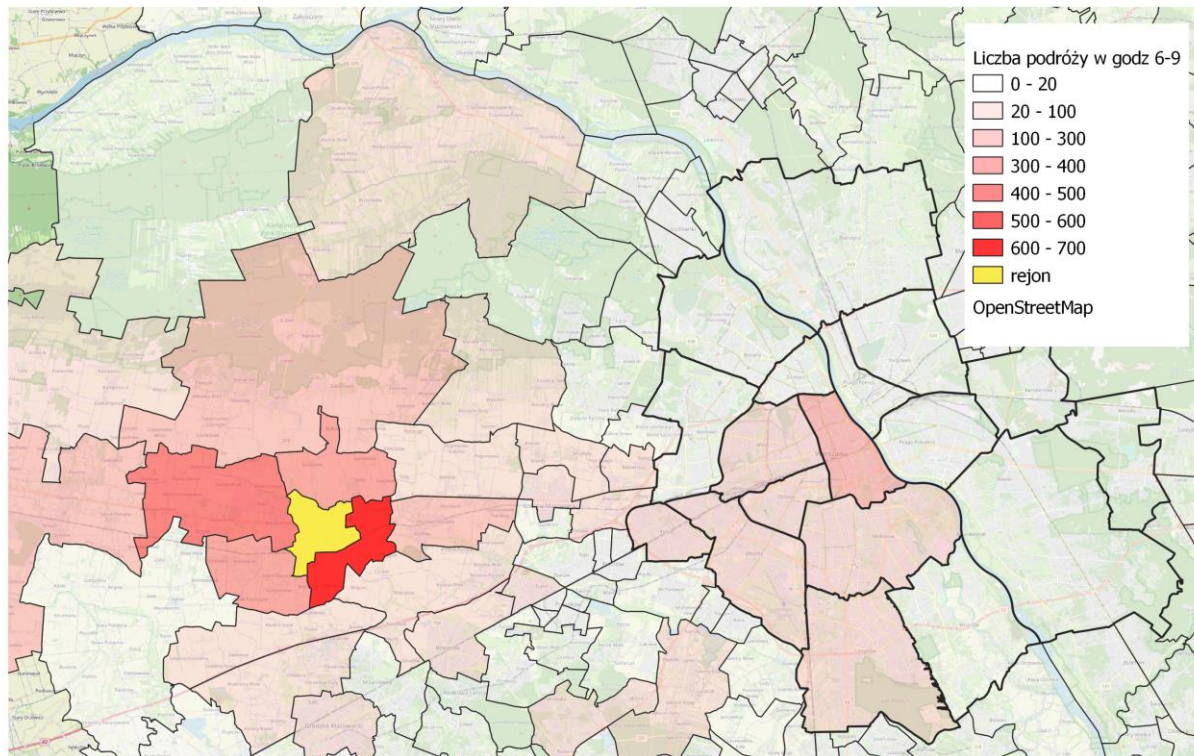
Porównywanie danych od różnych operatorów



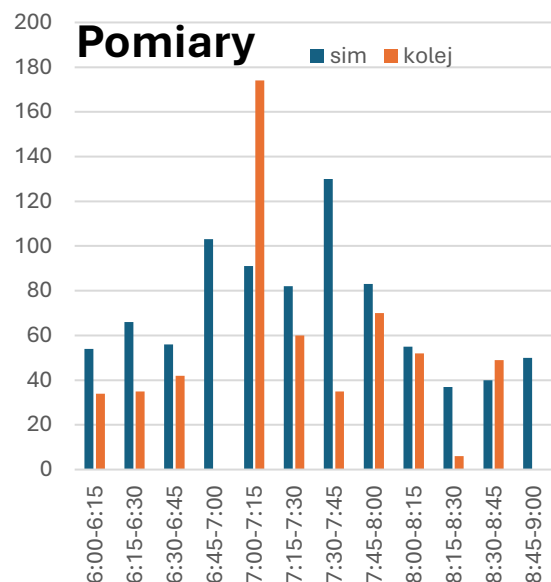
**Przykład podróży w dobie
pomiędzy dwoma rejonami u
dwóch różnych operatorów**

Jak pracować na bazach danych z kart SIM?

- Określić potrzebę i uszczegóławiać dane tylko tam gdzie jest to konieczne
- Przemyśleć dla jakich dni chcemy pobrać danę
- Ustalić jakie parametry pośrednie (budowa bazy danych) i o jakim próbkowaniu
- Łączyć z innymi źródłami danych (pomiar, GPS, rozkłady jazdy transportu zbiorowego, zagospodarowanie itd.)



Łączenie danych z kart SIM z innymi źródłami danych **znacząco zwiększa ich wartość analityczną**, pozwalając uzyskać pełniejszy, dokładniejszy i bardziej użyteczny obraz przemieszczeń ludzi.



Liczba podróży z Błonia na kierunku do Warszawy (SIM)

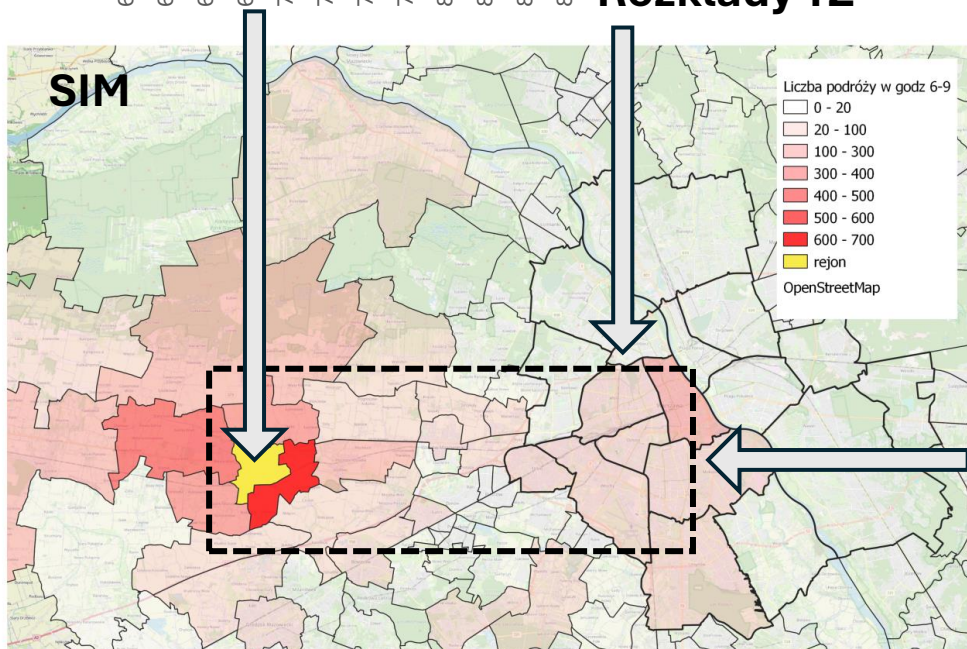
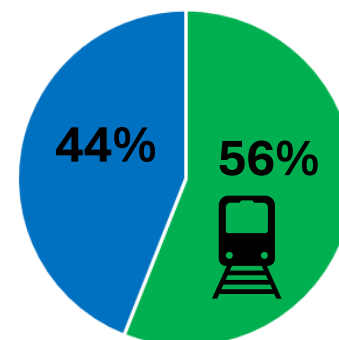


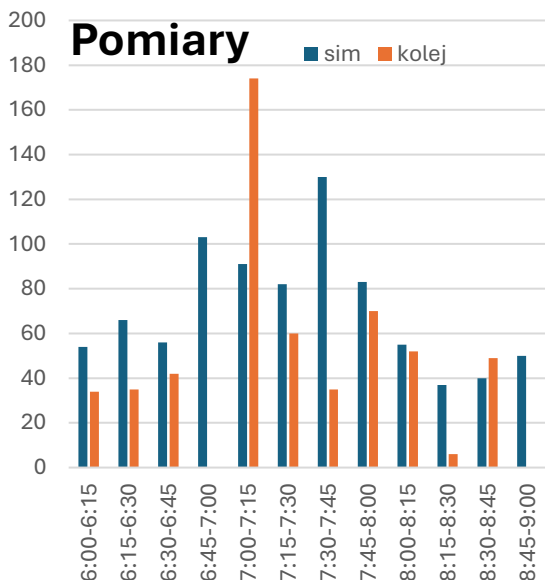
Pomiary i ankiety na stacjach kolejowych w Błoniu



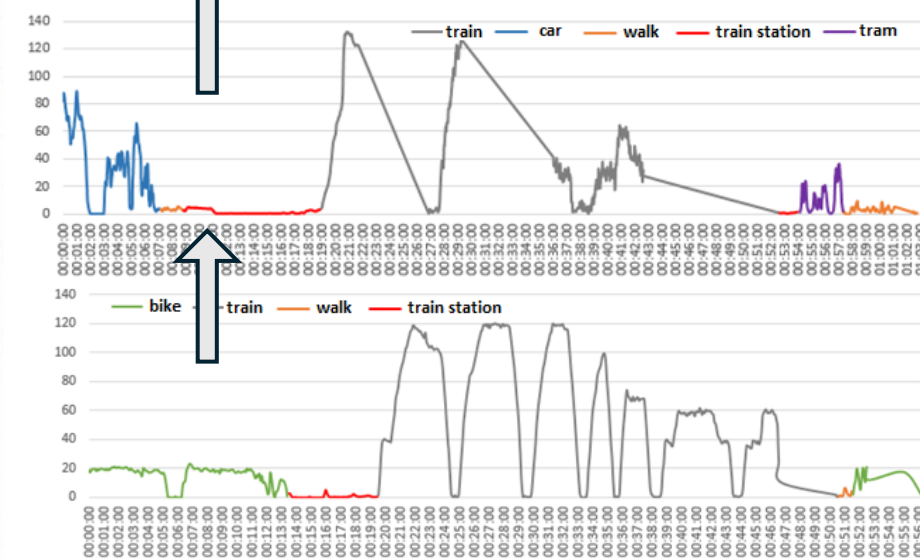
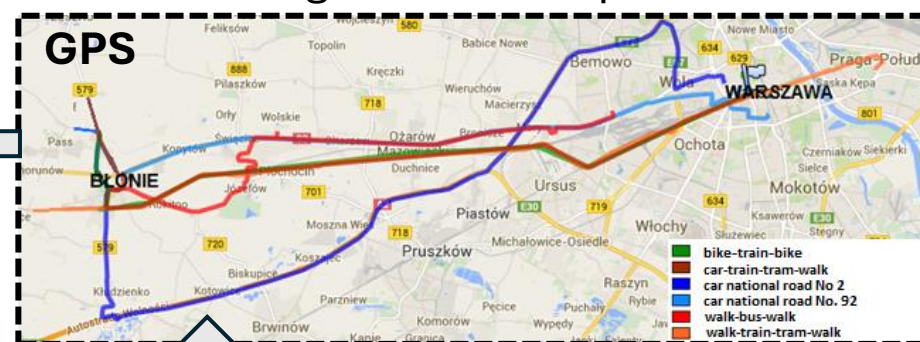
Podział zadań przewozowych
56% koleją
44% samochód/autobus

Rozkłady TZ

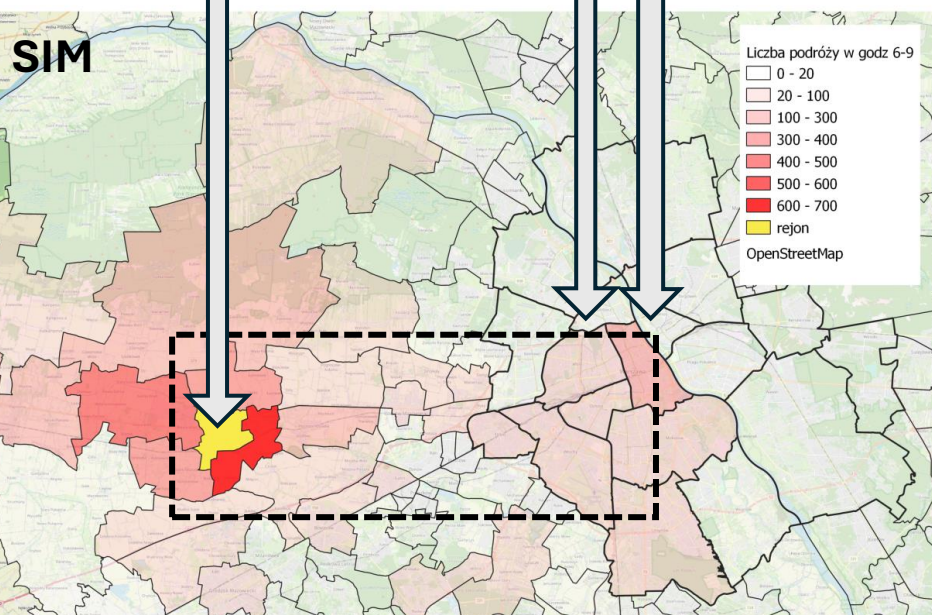




Same dane z kart SIM są cenne, ale mają ograniczenia, ich uzupełnienie o inne źródła danych niweluje tę słabości i zwiększa możliwości ich wykorzystania.



SIM



Podsumowanie

- Dane z kart SIM wymagają przetworzenia, by były użyteczne.
- Ogromny, wciąż niewykorzystany potencjał analityczny.



Konferencja
Naukowo-Techniczna
Miasto i Transport

2025

Dziękuję za uwagę

Karolina Jesionkiewicz-Niedzińska



**Wydział
Inżynierii Lądowej**

POLITECHNIKA WARSZAWSKA

Warszawa 8-9.05.2025