

Wstęp

Location Intelligence (LI) to dziedzina analityki, która integruje dane przestrzenne z danymi biznesowymi, wspomagając organizacje w podejmowaniu decyzji opartych na kontekście geograficznym. Rozwój infrastruktury rowerowej w miastach stał się kluczowym elementem planowania miejskiego, sprzyjającym zrównoważonemu rozwojowi oraz poprawie jakości życia mieszkańców.

Cel projektu

Głównym zadaniem projektu jest stworzenie modelu przewidującego długość ścieżek rowerowych w danym heksie na podstawie różnych argumentów dla miasta Amsterdam. Ostatecznym celem jest zastosowanie modelu na danych dla Krakowa i porównanie wyników z faktyczną długością ścieżek rowerowych w Krakowie.

Etapy:

- Stworzenie siatek H3 dla Krakowa i Amsterdamu:
System geoprzestrzennego indeksowania H3 dzieli powierzchnię Ziemi na heksagonalne komórki. W Pythonie siatki te można wygenerować przy użyciu biblioteki h3.
- Obliczenie długości ścieżek rowerowych wewnątrz każdego heksagonu:

Ten etap obejmuje stworzenie podstawowego zestawu danych przy użyciu biblioteki geopandas.

- Dodanie kolejnych argumentów:
 - Odległość od centrum miasta
 - Powierzchnia terenów zielonych wewnątrz heksa
 - Ilość budynków mieszkalnych czy punktów usługowych
 - Gęstość zaludnienia
 - Długość dróg o różnej klasyfikacji
- Trenowanie modeli

Należy przetestować różne typy modeli, korzystając z MLFlow do tworzenia eksperymentów i zapisu wyników. Dane powinny być podzielone na zestawy: treningowy, walidacyjny i testowy.

- Optymalizacja hiperparametrów

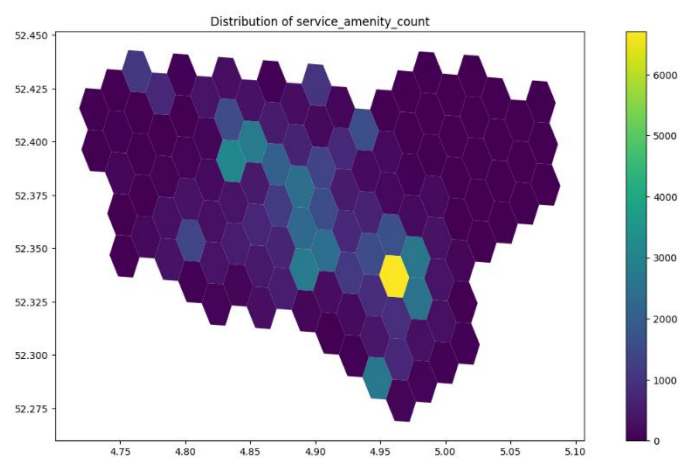
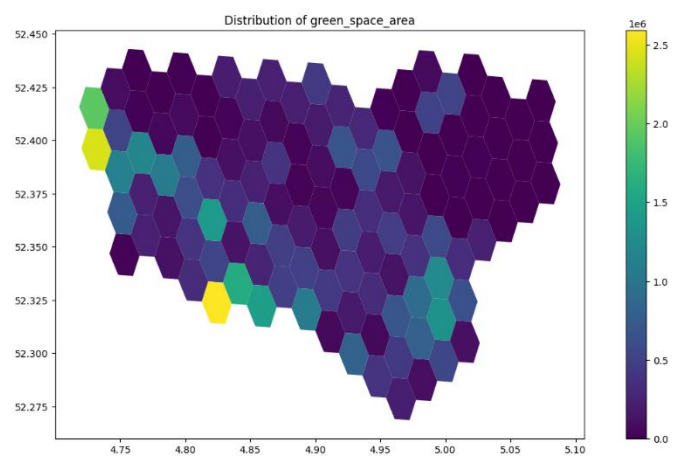
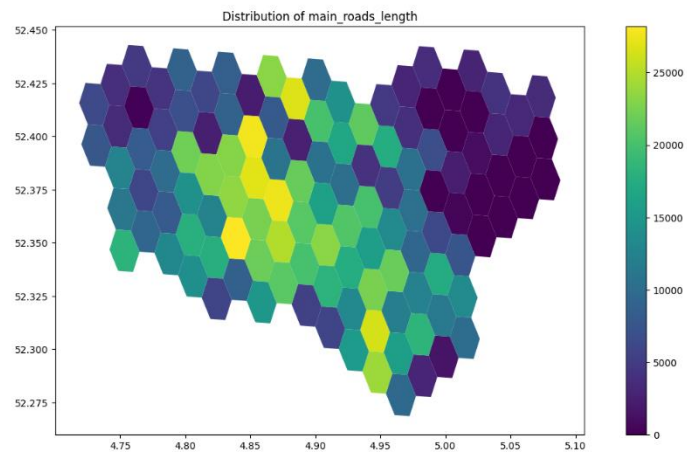
Możliwe metody optymalizacji obejmują np. Grid Search.

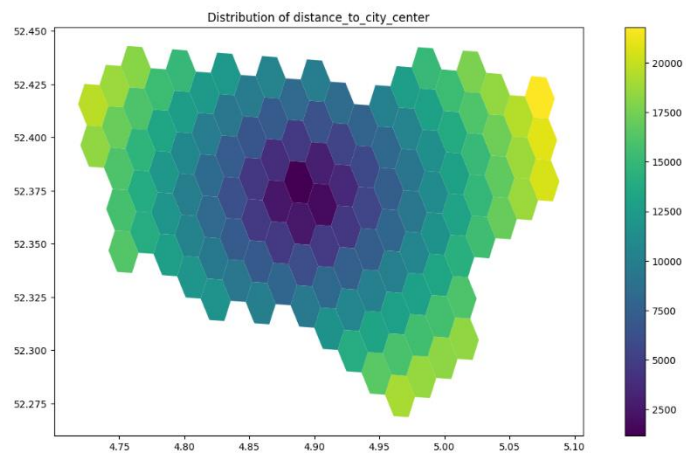
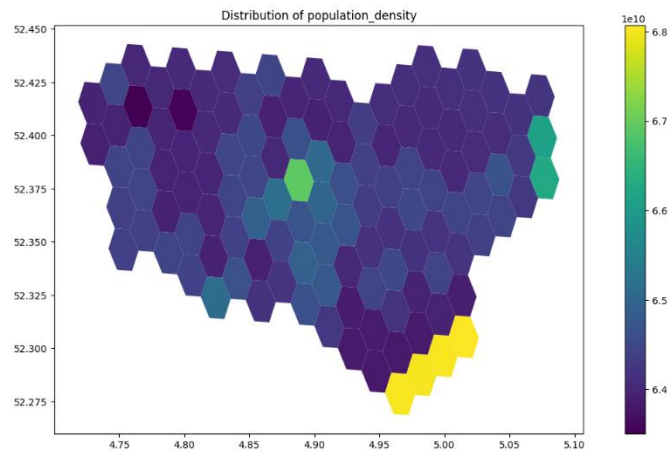
- Wykorzystanie najlepszego modelu na siatce dla Krakowa

Otrzymany model należy zastosować na siatce H3 dla Krakowa i porównać wyniki z faktyczną długością ścieżek rowerowych w Krakowie.

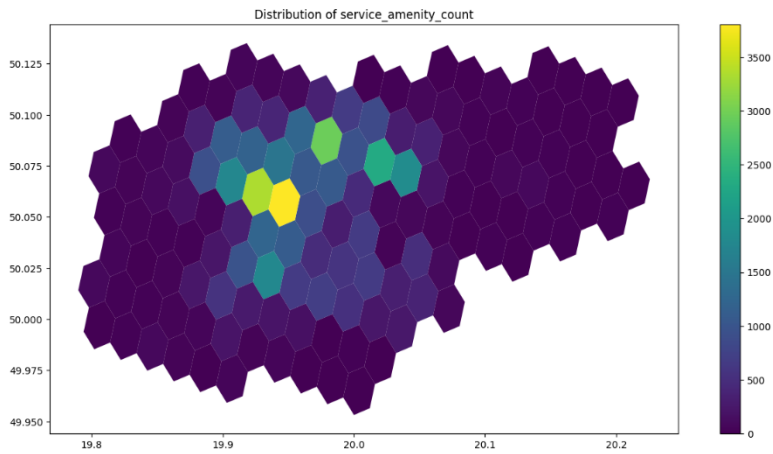
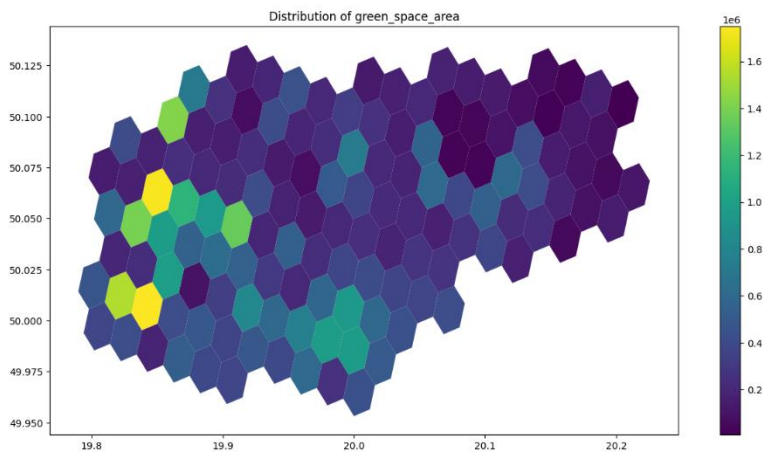
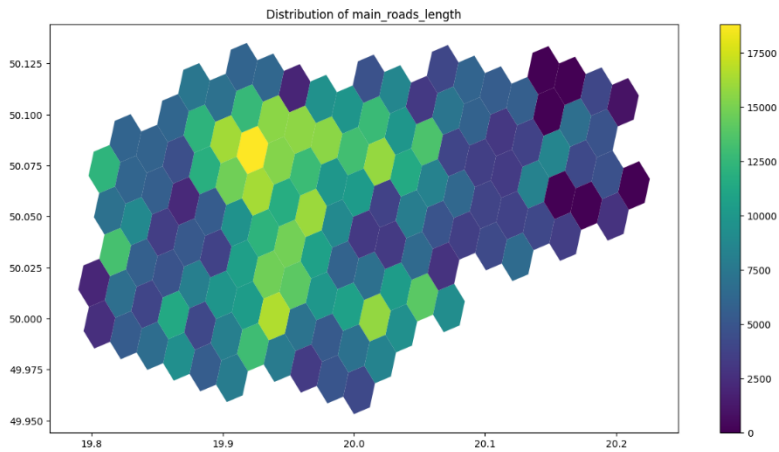
Poniżej zaprezentowano rozkład poszczególnych danych w ujęciu przestrzennym dla Krakowa i Amsterdamu w siatce heksagonalnej.

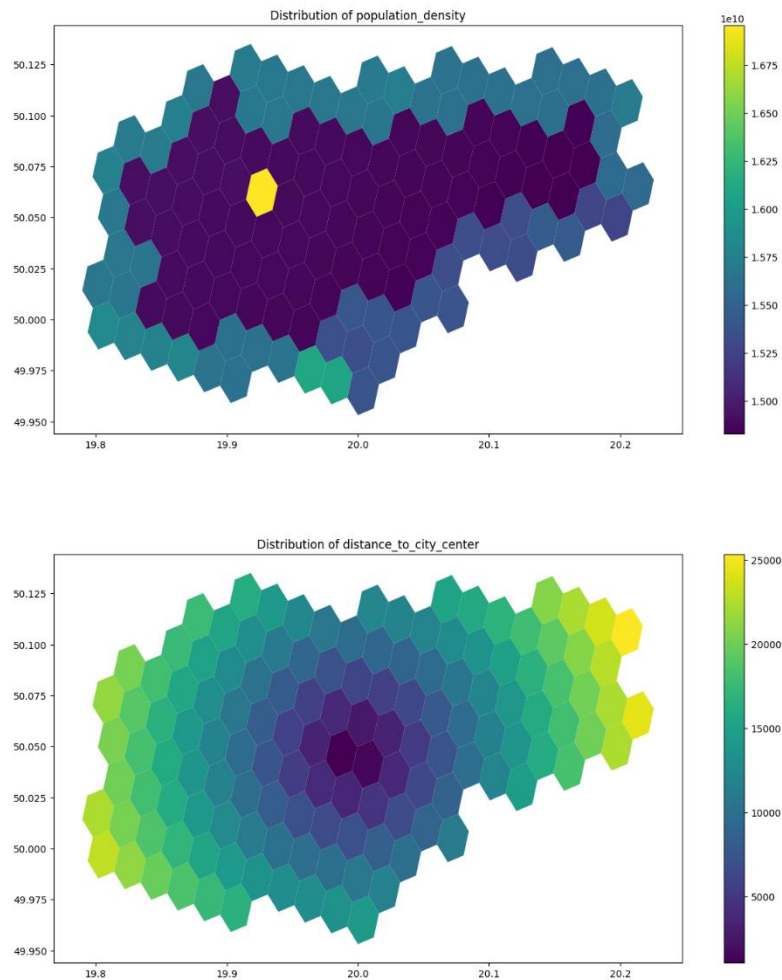
Amsterdam:





Kraków



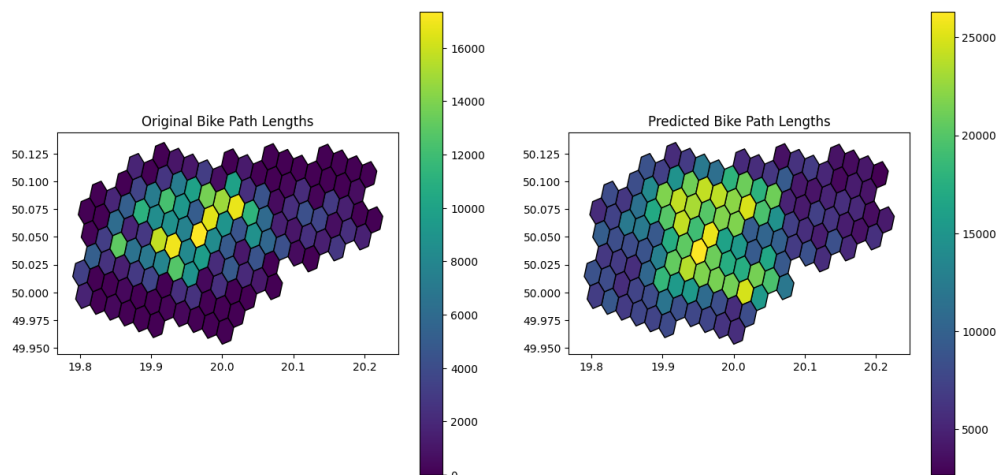


W projekcie został wykorzystany model Lasu losowego. Wynik ewaluacji modelu:

MAE: 4730.0379181633025, MSE: 36941670.135272406, RMSE: 6077.965953777004, R2: 0.6163335399364275

Współczynnik R-kwadrat, mówiący jak dobrze model wyjaśnia zmienność danych, jest na poziomie 0,62. Oznacza to, że model wyjaśnia około 62,15% zmienności danych.

Działanie modelu na przykładzie miasta Krakowa zostało przedstawione na obrazie poniżej:



Model trafnie przewidział większe zagęszczenie ścieżek rowerowych w centralnej części miasta oraz mniejsze jej zagęszczenie w obszarach oddalonych od centrum. Szczególnie dobrze widoczna jest uszczuplona infrastruktura rowerowa w obszarach Nowej Huty.