

Store location in Kyiv

Test task

Lysenko

Maryna



Test Task problem:

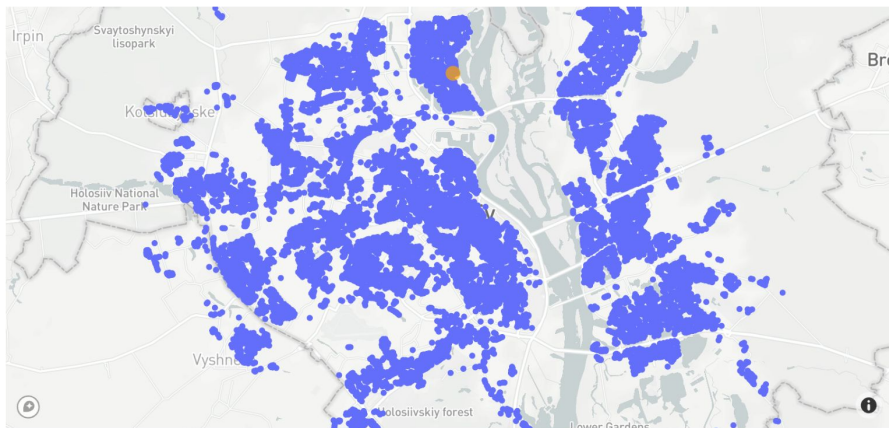
1. В Києві районів - 10, населення описане 12 752 геоточками
2. Магазинів Silpo - 92

Задача:

- аналіз кореляції характеристики магазину та населення
- візуалізація результатів дослідження,
- побудува геокарти розподілу характеристик
- ідеї для найкращих локацій для відкриття нового store

Вхідні дані - Population

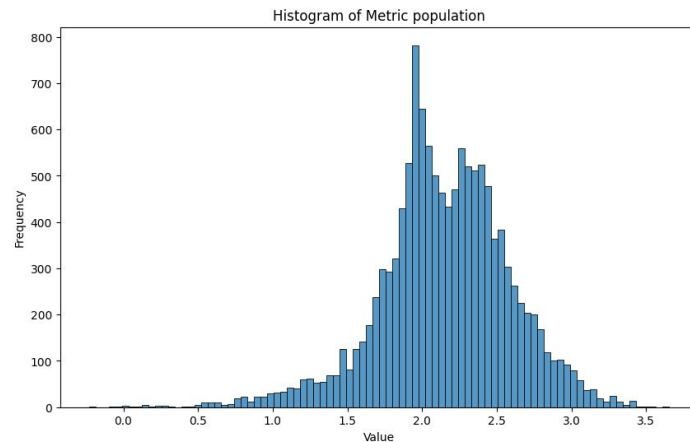
lat	lon	metric population
50,4360	30,6256	1,96
50,4352	30,6261	1,95
50,4364	30,6260	1,70
50,4362	30,6268	2,09
50,4354	30,6266	2,13



12 752 геоточок

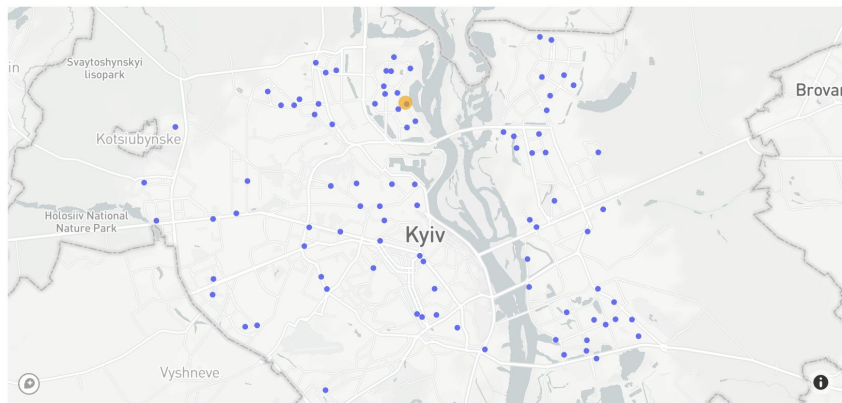


Середня метрика 2.149327
Мінімальна: -0.229571 (???)
Максимальна: 3.655685



Вхідні дані - store

Store	lat	long	Metric Store
Store 1	50,42	30,52	2,18
Store 2	50,50	30,51	2,64
Store 3	50,41	30,65	2,50
Store 4	50,47	30,48	2,38
Store 5	50,53	30,61	2,34

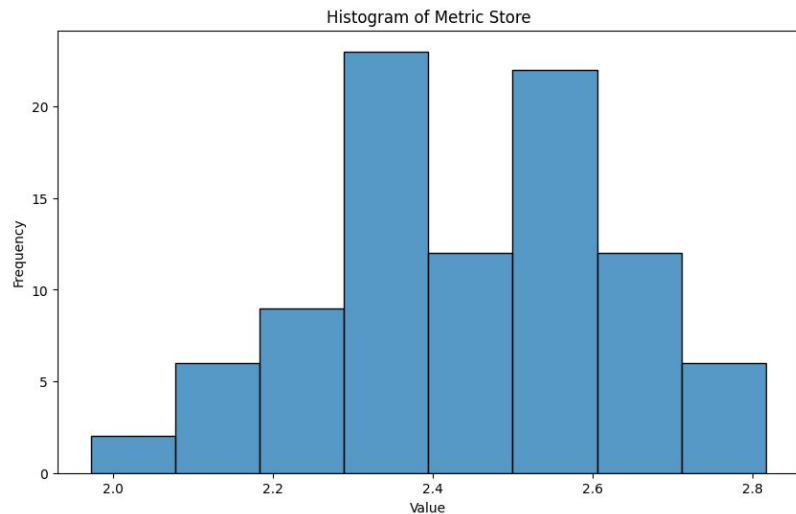


92 геоточок

Середня метрика: 2.438800

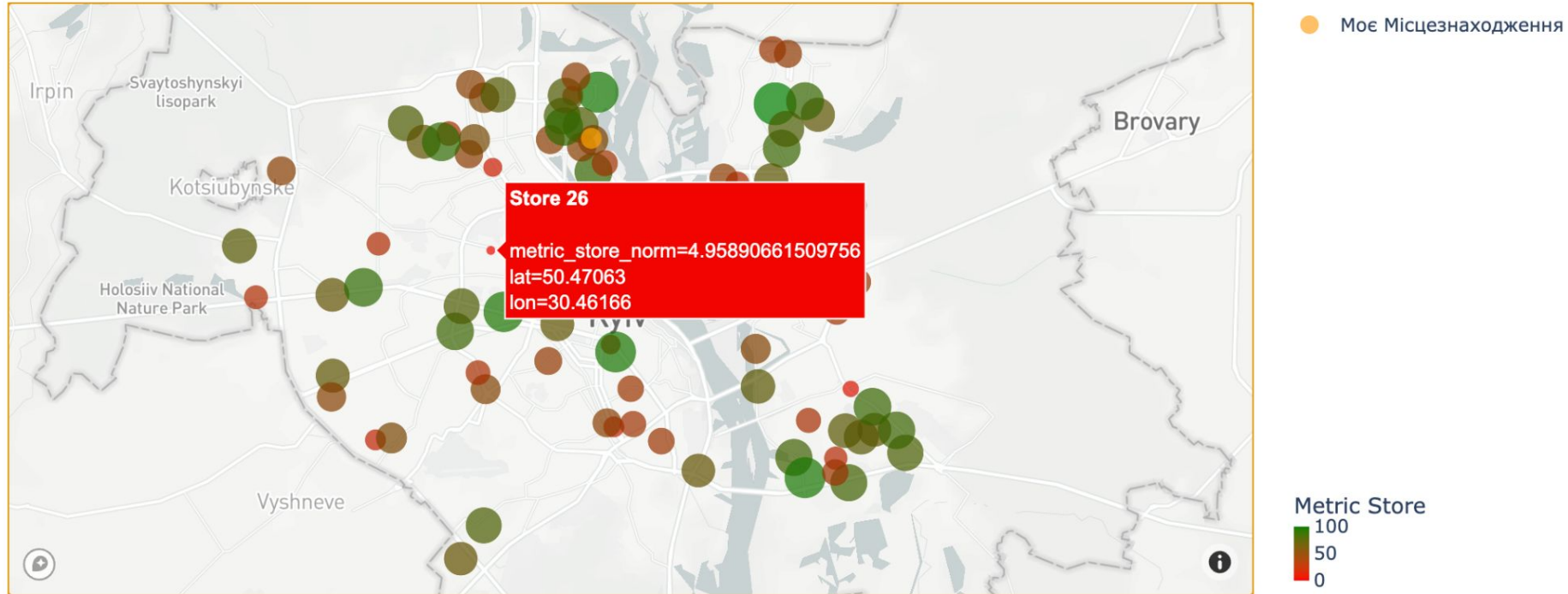
Мінімальна: 1.973128

Максимальна: 2.816308



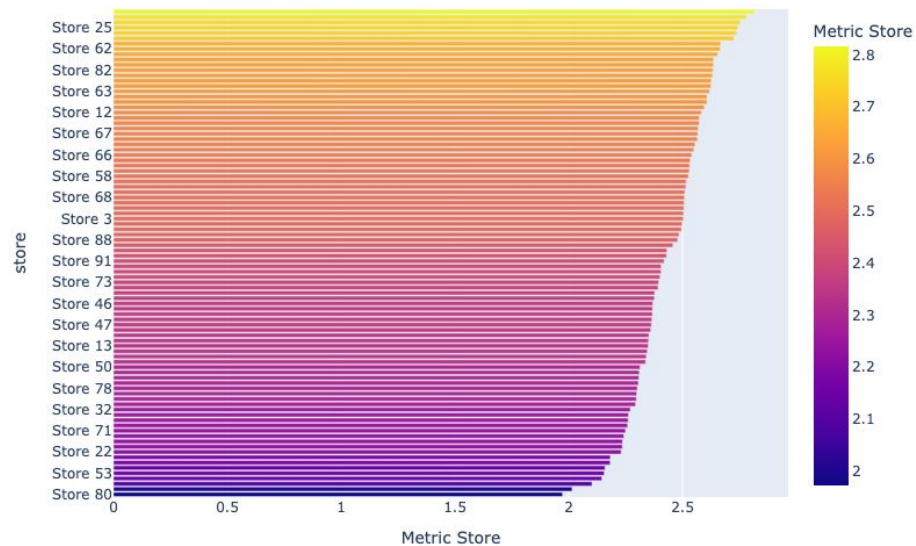
В подальшому нормалізую до 1-100

Геопросторовий розподіл



Топ-точки

Metric Store by Store



Етапи рішення

New data collection

Райони
Населення
Аренда
Місця поруч

FOURSQUARE



Population & Store joining

Distance
matrix

GeoPy



Population & Districts & Store feature



Виведення аналітичного правила



Data collection

Мета даного завдання полягає в визначенні оптимального району для відкриття магазину з урахуванням різних факторів:

- Тип району: Ідеальним варіантом є житловий район, де проживає багато людей. [[Райони Києва](#) + Foursquare API]
- Кількість мешканців: Чим більше населення в даному районі, тим вищий попит на продукти харчування [[Райони Києва](#) + аналіз вхідних даних]
- Середній дохід мешканців: Віддаємо перевагу районам з представниками робочого класу, оскільки вони є основною цільовою аудиторією [[Ціни на оренду квартири в Києві - середня вартість житла на DIM.RIA](#)].

Districts data

	District	Population	Area (km2)	district_eng	latlng	1-bedroom	2-bedroom	3-bedroom	4-bedroom	lat	lon
0	Голосіївський	247600	156.0	Holosiivskyi District	[50.3268905, 30.5672193]	700	1000	1300	4880	50.326890	30.567219
1	Дарницький	314700	134.0	Darnytskyi District	[50.4153964, 30.69359919999999]	420	750	1000	1377	50.415396	30.693599
2	Деснянський	358300	148.0	Desnianskyi District	[50.521161, 30.6830731]	210	263	289	1925	50.521161	30.683073
3	Дніпровський	354700	67.0	Dniproviskyi District	[50.4535873, 30.601521]	407	650	1099	1386	50.453587	30.601521
4	Оболонський	319000	108.6	Obolonskyi District	[50.54672979999999, 30.4372576]	508	743	1000	3418	50.546730	30.437258
5	Печерський	152000	27.0	Pecherskyi District	[50.41980059999999, 30.560858]	1000	1200	1999	4610	50.419801	30.560858
6	Подільський	198100	34.0	Podilskyi District	[50.4938086, 30.4428556]	700	1000	1250	4080	50.493809	30.442856
7	Святошинський	340700	110.0	Sviatoshynskyi District	[50.339191, 30.2487279]	381	604	1000	1020	50.339191	30.248728
8	Солом'янський	383259	40.0	Solomianskyi District	[50.4282552, 30.4529175]	526	900	1000	3180	50.428255	30.452918
9	Шевченківський	218900	26.6	Shevchenkivskyi District	[50.4642048, 30.4664647]	850	947	1500	2117	50.464205	30.466465

Вручну з
офіційного
портала Києва

Вручну з domria

geocoder.google

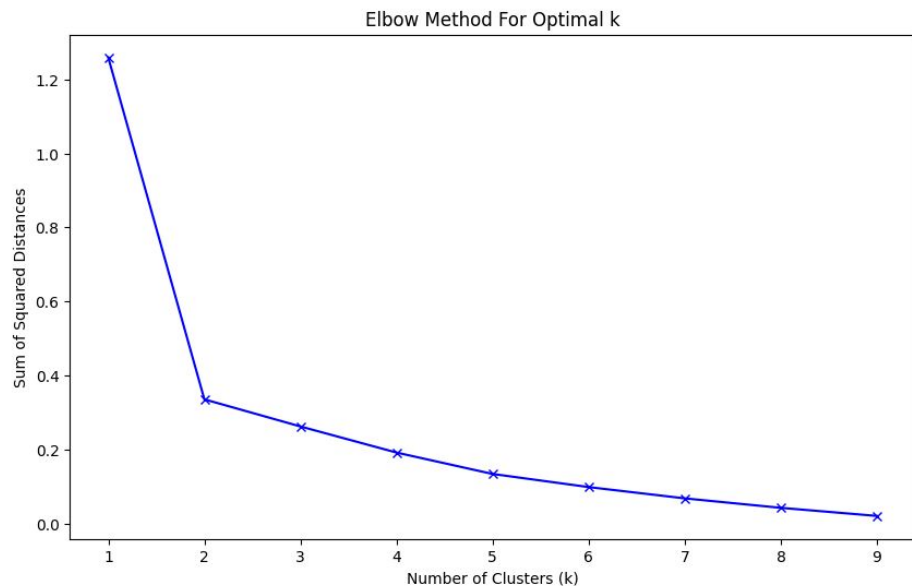


Google Geocode API

Most Common Venue

District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Darnytskyi District	Drugstore	Grocery Store	Post Office	Convenience Store	Cosmetics Store	Food and Beverage Retail	Breakfast Spot	Cafeteria	Café	Church
Desnianskyi District	BBQ Joint	Arcade	Art Gallery	Art Museum	Arts and Crafts Store	Asian Restaurant	Bakery	Beach	Beer Store	Bistro
Dniprovskyi District	Candy Store	Coffee Shop	Cosmetics Store	Medical Center	Bakery	Clothing Store	Fast Food Restaurant	Grocery Store	Pet Supplies Store	Tea Room
Holosiivskyi District	Shopping Mall	Children's Clothing Store	Cosmetics Store	Electronics Store	Grocery Store	Arcade	Beach	Beer Store	Bookstore	Boutique
Obolonskyi District	Coffee Shop	Blood Bank	Cosmetics Store	Dining and Drinking	Gastropub	Grocery Store	Pizzeria	Pub	Veterinarian	Arcade
Pecherskyi District	Coffee Shop	Bakery	Escape Room	Art Gallery	Bookstore	Boutique	Café	Grocery Store	History Museum	Scenic Lookout
Podilskyi District	Post Office	Coffee Shop	Grocery Store	Caucasian Restaurant	Eastern European Restaurant	Toy Store	Bakery	Beer Store	Blood Bank	Boutique
Shevchenkovskyi District	Coffee Shop	Café	Grocery Store	Bookstore	Dairy Store	Electronics Store	Pizzeria	Seafood Restaurant	Arts and Crafts Store	Bistro
Solomianskyi District	Coffee Shop	Italian Restaurant	Beer Store	Caucasian Restaurant	Bakery	Drugstore	Grocery Store	Medical Center	Pet Supplies Store	Sporting Goods Retail
Sviatoshynskyi District	Coffee Shop	Miscellaneous Store	Beer Store	Convenience Store	Cosmetics Store	Education	Farmers Market	Night Club	Post Office	Restaurant

Спроба кластеризації на основі топ-місць і опису районів



При застосуванні кластерів один виділився в групу 2
Деснянський
Та всі інші в групу 1

Поєднання даних

Для подальших досліджень важливо вдало приписати магазин потенційному клієнту

Спосіб 1.

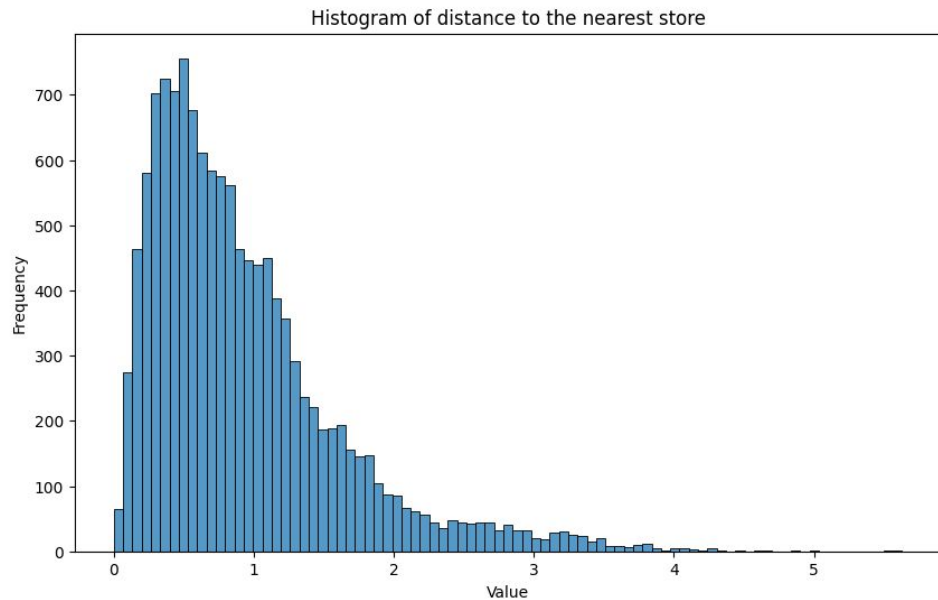
Спрощений розрахунок геоцентричної відстані
(geopy.distance geodesic)

Спосіб 2.

З урахуванням дорог - GoogleMaps API
(mode="walking" / "driving")



Відстань до найближчого магазину



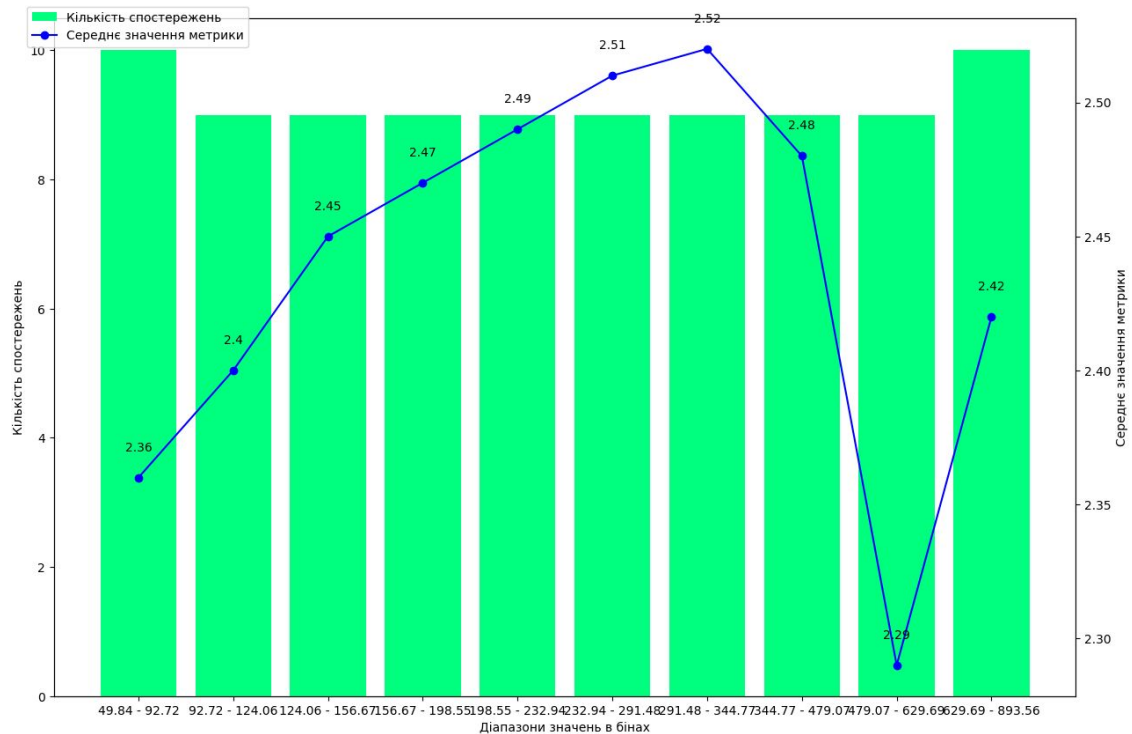
В середньому відстань до магазину
200-1200м

Feature engineering

- К-сть / сума / частка популяції в радіусі [0.2, 0.5, 1, 2, 3] та більше 3 км
- Загальна к-сть популяції
- Відстань для популяції - середня, мін, макс
- Кількість та частка популяції в бінах за відстанню

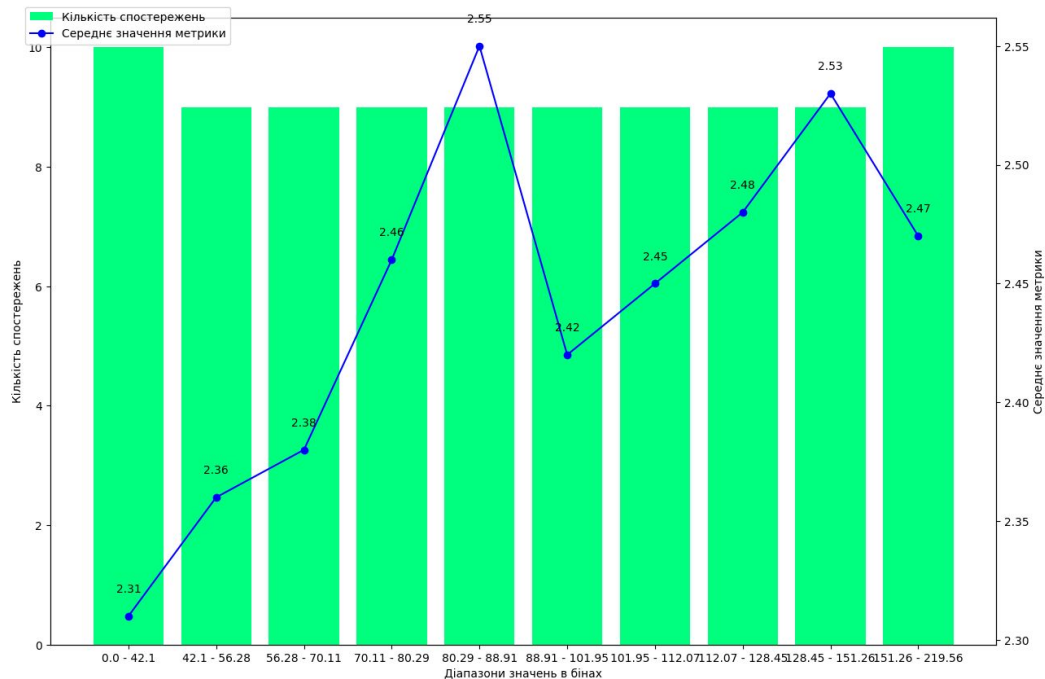
Всього: 38 фіч

Залежність metric_store від sum_metric_population



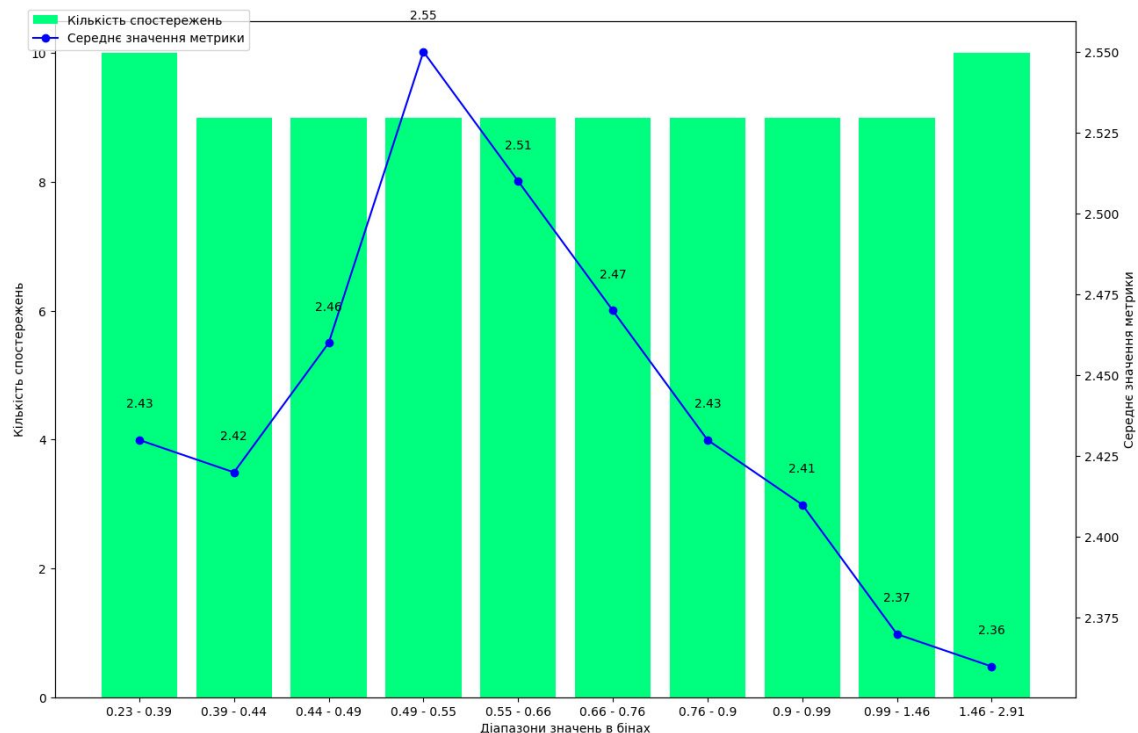
До 350 залежність зростання метрики зі зростом значення фічі
Найкраще значення метрик де 291.48 - 344.77

cnt_pop_within_0.5km



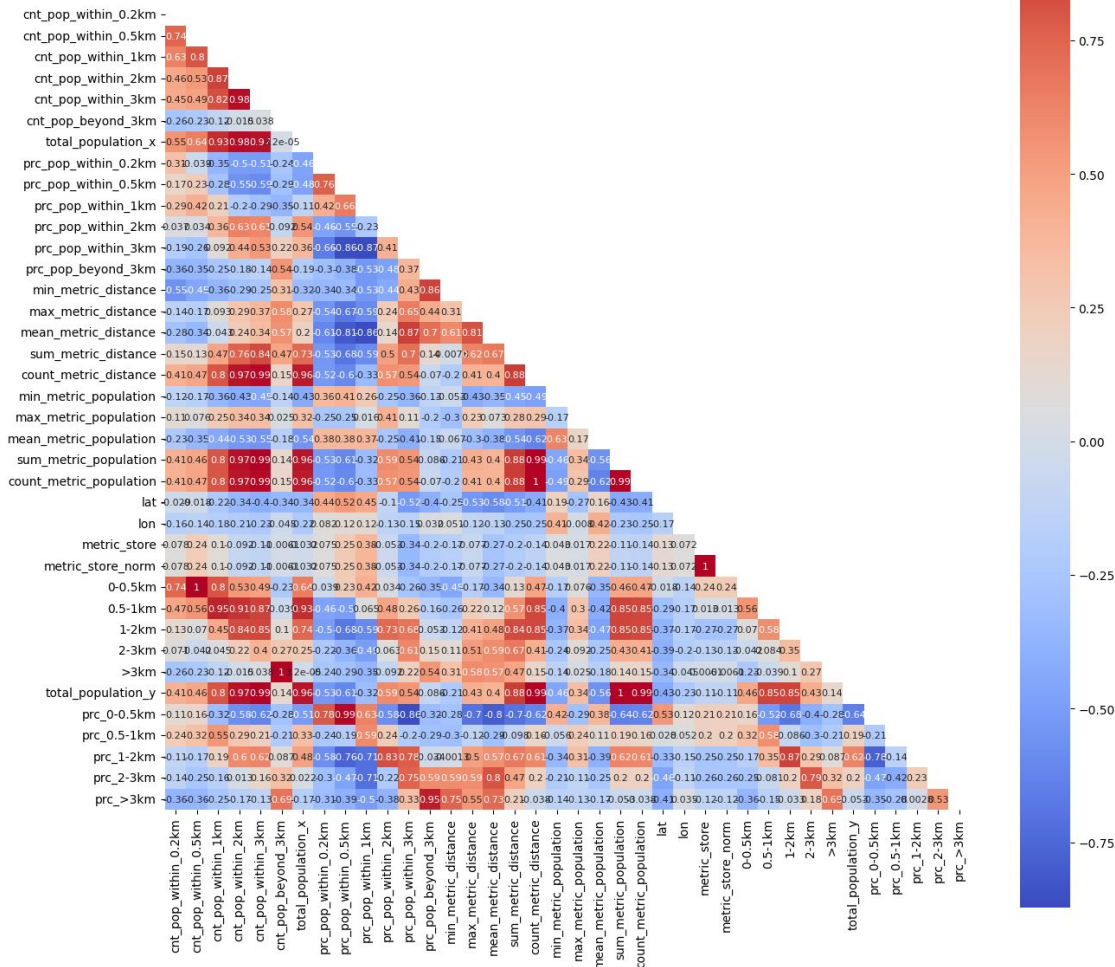
Чим більше популяції в маленькому радіусі - тим краще метрика

mean_metric_distance

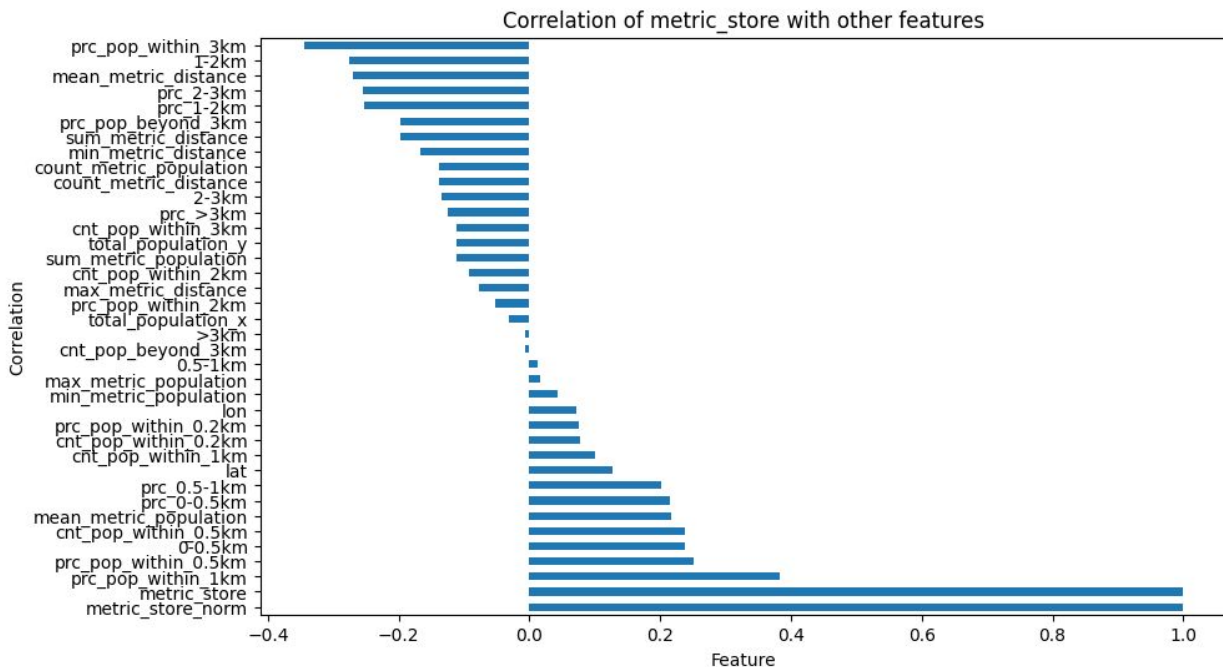


Чим вища середня дистанція до магазину- тим гірша метрика

Кореляція фічей

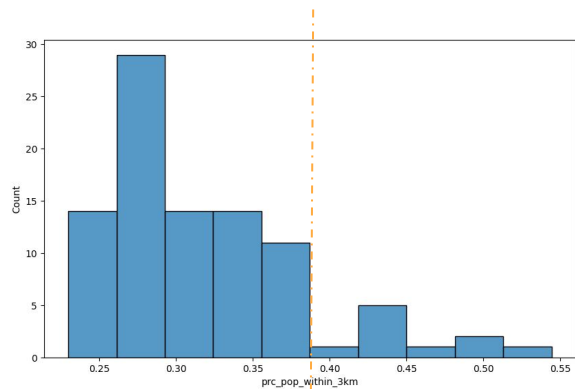
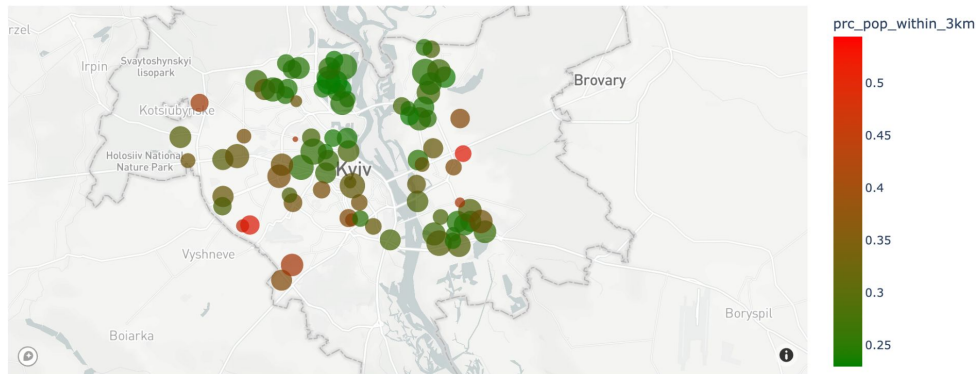


Знайдені фічі з середньою метрикою



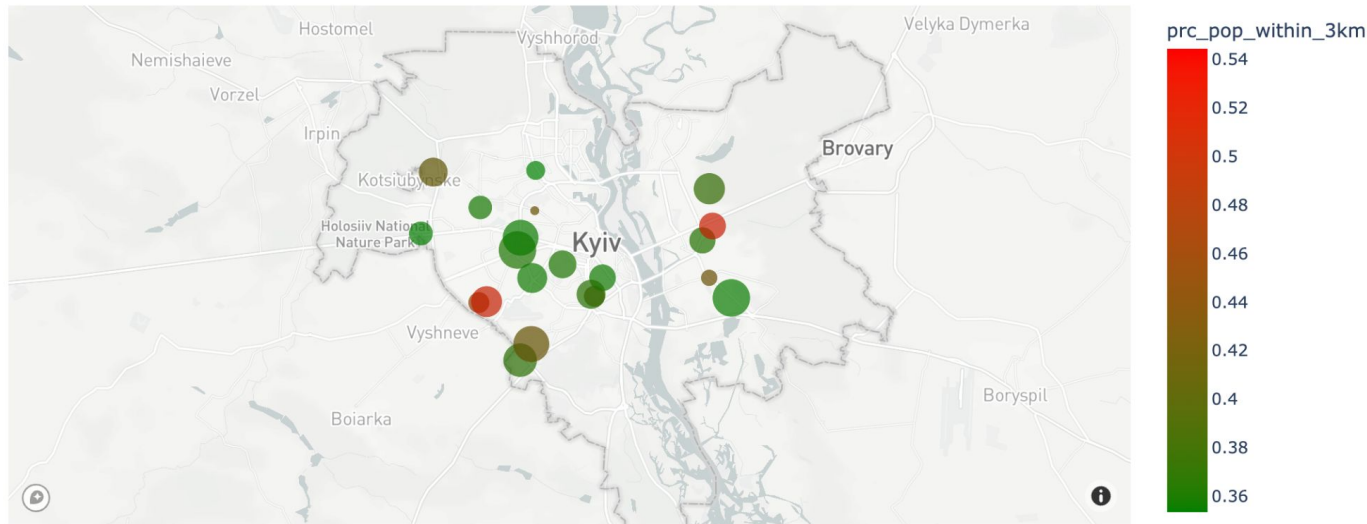
Найгірше впливають на метрику магазину **prc_pop_within_3km**

Розподіл метрики



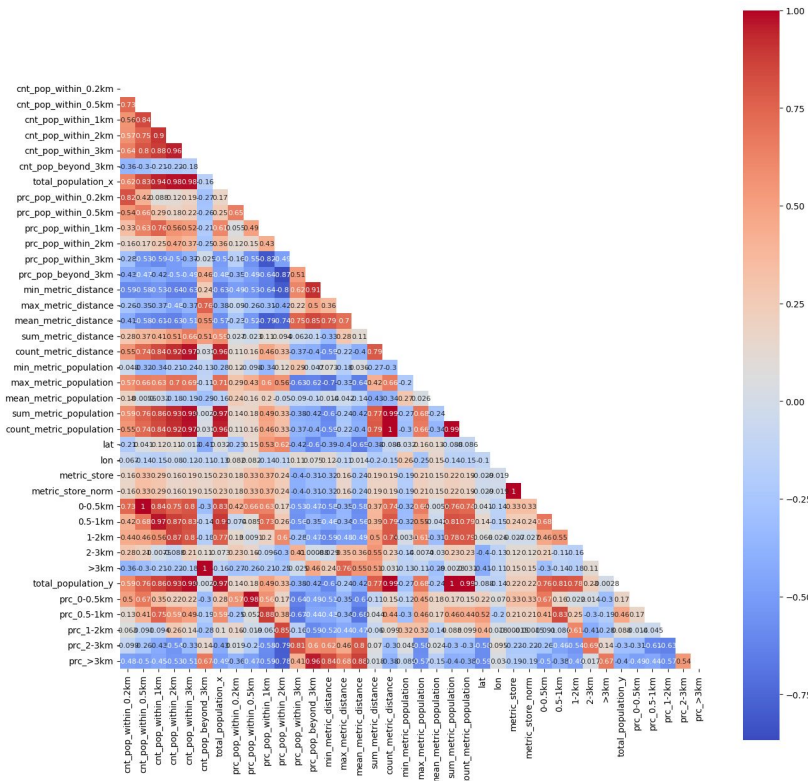
Обираю для подальшого аналізу магазини де велика частка популяції має велику відстань

Точки для подальшого аналізу



Візуально якщо не брати до уваги околиці для яких нормально більш велика відстань до магазину дуже вибивається магазин 26. Аналіз того, що знаходиться поруч по карті вказує на те що фактично населення ще не заїхало в ці новобудови і недостатньо покриття навіть цієї точки тому не буду пропонувати нові точки поруч

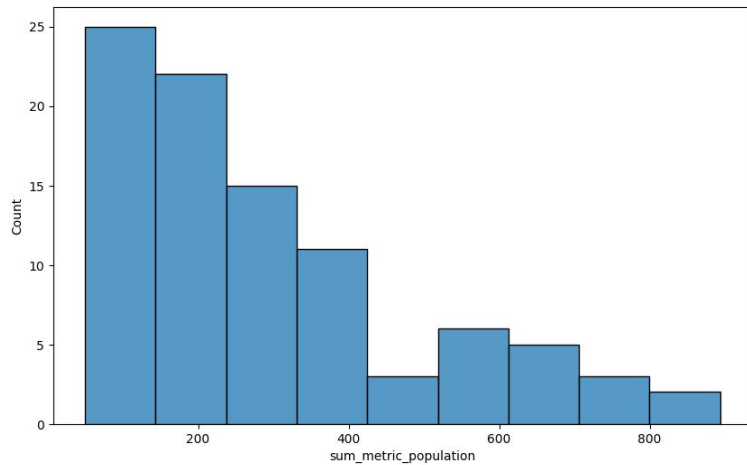
Що впливає на метрики цих 22 магазинів



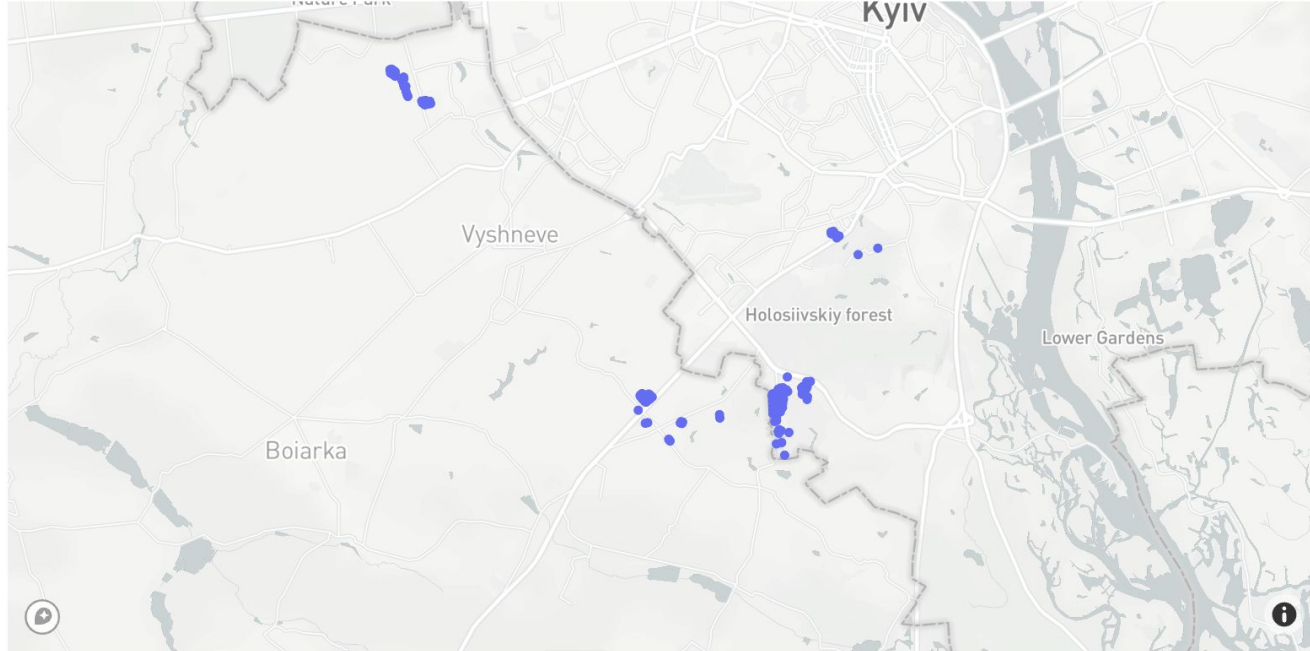
```
- 0.404355 з
   prc pop within 3km
```

Звуження точок

- # Визначення порогових значень
- `distance_threshold = 3.0` # в км, мінімальна відстань до найближчого Сільпо
-
- # Фільтрація точок, що задовольняють обрані критерії
- `potential_points = population_full_data[`
- `(population_full_data["distance"] > distance_threshold)`
- `& (population_full_data["sum_metric_population"] > 350)`
- `]`



Точки з потребою в магазині ближче



Центри точок
[50.35256878,
30.4820822]
[50.43007764,
30.32619834]
[50.34768888,
30.42619204]

На основі отриманих даних є велика завантаженість поблизу околиць і при цьому хороші метрики у магазинів, відкрити додаткові точки поблизу топ-найвіддаленніших (виведені на останній карті) потенційно може покращити метрики.

Питання



Дякую за увагу