

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО”  
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

# Дослідження раціональності відкриття закладів відпочинку залежно від регіону методами просторової економетрики

**Виконав:**  
студент 4-го курсу  
групи КА-31 ІПСА «КПІ»

**Луца А.М.**

**Керівник:**

**к. ф.-м. н., доцент Каніовська І. Ю.**

# Вступ



- **Об'єкт дослідження:** дані з просторовими ефектами;
- **Предмет дослідження:** застосування методів просторової економетрики для дослідження даних з порталу Foursquare;
- **Мета дослідження:** проаналізувати предмет дослідження та дослідити ефективність застосування просторової економетрики у задачах кластеризації та рекомендації;
- **Методи дослідження:** методи виявлення просторової автокореляції та моделі просторової авторегресії.

# Постановка задачі

- З порталу Foursquare.com взяти дані про відвідування різних закладів відпочинку в місті Києві. На їх основі всю територію Києва розбити на 56 однакових районів, на основі чого дані агрегувати відповідно до цього розбиття.
- Дослідити наявність просторових ефектів в цілому по місту Києву, та у всіх районах зокрема.
- Запропонувати спосіб обчислення матриці ваг для врахування не тільки географічних характеристик районів.
- Дослідити 3 просторові авторегресійні моделі, і визначивши найкращу – зробити висновки.

# Актуальність

- Задача рекомендації і прийняття рішень є провідною задачею в сучасному бізнесі.
- Багато хто використовує сучасні методи у великих компаніях, проте малий і середній бізнес страждає від невикористання найновіших інструментів.
- Є необхідність у дослідженні коректних методів для дослідження раціональності створення бізнесу в різних місцях, залежно від розташування локації та категорії, в якій хоче працювати людина.

# Основні поняття просторової економетрики

Просторова автокореляція – це зв'язок між значеннями одного параметру, який виходить за межі певної географічної області, в якій він з'являється. Вона вимірює схожість об'єктів в межах області, ступінь, з якою просторове явище корелює з самим собою, тобто просторова автокореляція є оцінка кореляції параметру в посиленні на географічне розташування та вплив сусідів.

Просторова автокореляція існує тоді, коли існує систематична просторова зміна значень певного параметру. Ця зміна може існувати в двох формах: додатна, або від'ємна автокореляція

# Способи виявлення глобльних просторових ефектів

- Гамма індекс
- Статистика кількості приєднань
- Індекс Морана I
- Індекс Гірі С

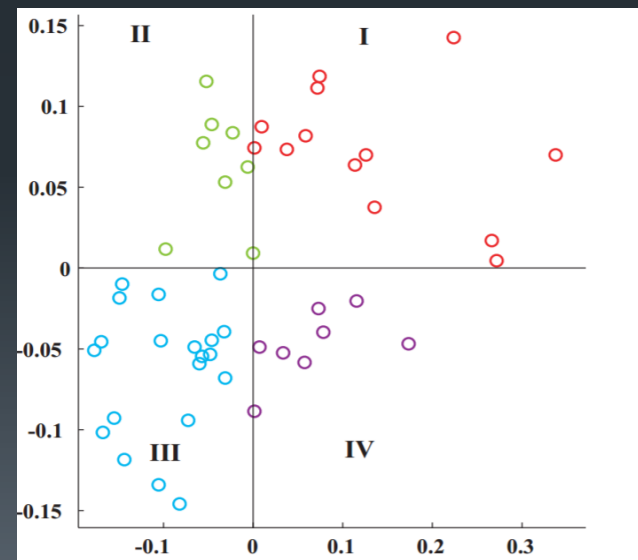
Перевіряється гіпотеза про те, що за деякого рівня значущості між локаціями всередині карти немає просторових ефектів.

# Локальний індекс Морана I

Локальний показник просторової асоціації – статистика, що задовільняє:

1. LISA для кожного спостереження дає уявлення про ступінь просторової залежності значущої змінної від схожих значень з локалей навколо вибраного регіону;
2. Сума індикаторів LISA для всіх спостережень пропорційна глобальному показнику.

Діаграма Морана – відстежує звязки між нев'язками для кожної локації та середніми нев'язками для сусідів



# Матриця вагових коефіцієнтів

- За адміністративно-територіальним устроєм:

$$w_{ij} = 1, (i, j) \in A$$

$$w_{ij} = 1, (i, j) \in A$$

- Метод фіксованого ядра

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2, & d_{ij} < b \\ 0, & d_{ij} \geq b \end{cases}$$

$b$  – ширина смуги, в якій ми вважаємо локації сусідніми,  $d_{ij}$  - відстань між локаціями, як відстань між 2 точками-центрами локацій.



# Матриця вагових коефіцієнтів

- **Метод модифікованого адаптивного ядра**

$$w_{ij} = \begin{cases} \left( 1 - \left( \frac{d_{ij} - \frac{1}{m} \sum_k^m \frac{c_{kj}}{c_k} * b}{b} \right)^p \right)^p, & d_{ij} < b \\ 0, & d_{ij} \geq b \end{cases}$$

$b$  – ширина смуги, в якій ми вважаємо локації сусідніми;

$d_{ij}$  - відстань між локаціями, як відстань між 2 точками-центрами локацій;

$m$  – кількість категорій закладів відпочинку;

$c_k$  - загальна кількість закладів в категорії  $k$  по всьому місту;

$c_{kj}$  - кількість закладів в категорії  $k$  в локації  $j$ .

# Модель просторового лагу (SAR)

$$\begin{aligned}y &= \rho W y + X \beta + \varepsilon, \\y &= (I_n - \rho W)^{-1} X \beta + (I_n - \rho W)^{-1} \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n)\end{aligned}$$

Обчислення відбувається за наступною схемою:

1. Застосовується МНК для моделі  $y = X \beta_0 + \varepsilon_0$ .
2. Застосовується МНК для моделі  $W y = X \beta_L + \varepsilon_L$ .
3. Обчислюється нев'язка  $e_0 = y - X \hat{\beta}_0$  та  $e_L = W y - X \hat{\beta}_L$ .
4. За допомогою обчисленої нев'язки знаходиться оцінка  $\rho$  за допомогою максимізації логарифмічної функції правдоподібності  $\ln(L)$
5. Маючи оцінку  $\hat{\rho}$ , що максимізує  $\ln(L)$  обчислюємо оцінки  $\hat{\beta} = \hat{\beta}_0 - \hat{\rho} \hat{\beta}_L$  та  $\hat{\sigma}^2 = \frac{(e_0 - \hat{\rho} e_L)^T (e_0 - \hat{\rho} e_L)}{n}$ .

# Модель просторових помилок (SEM)

$$\begin{aligned}y &= X\beta + u \\u &= \lambda W u + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n)\end{aligned}$$

Обчислення відбувається за наступною схемою:

1. Знаходимо  $\tilde{y} = y - \lambda W y$ .
2. Знаходимо  $\tilde{X} = X - \lambda W X$ .
3. Застосовується МНК для моделі  $\tilde{y} = \tilde{X}\beta + \varepsilon$ .
4. Обчислюється нев'язка  $e = \tilde{y} - \tilde{X}\hat{\beta}$ .
5. За допомогою обчисленої нев'язки знаходиться оцінка  $\lambda$  за допомогою максимізації логарифмічної функції правдоподібності
6. Маючи оцінку  $\hat{\rho}$ , що максимізує  $\ln(L)$  обчислюємо оцінку  $\hat{\sigma}^2 = \frac{e^T e}{n}$ .

# Просторова модель Дарбіна (SDM)

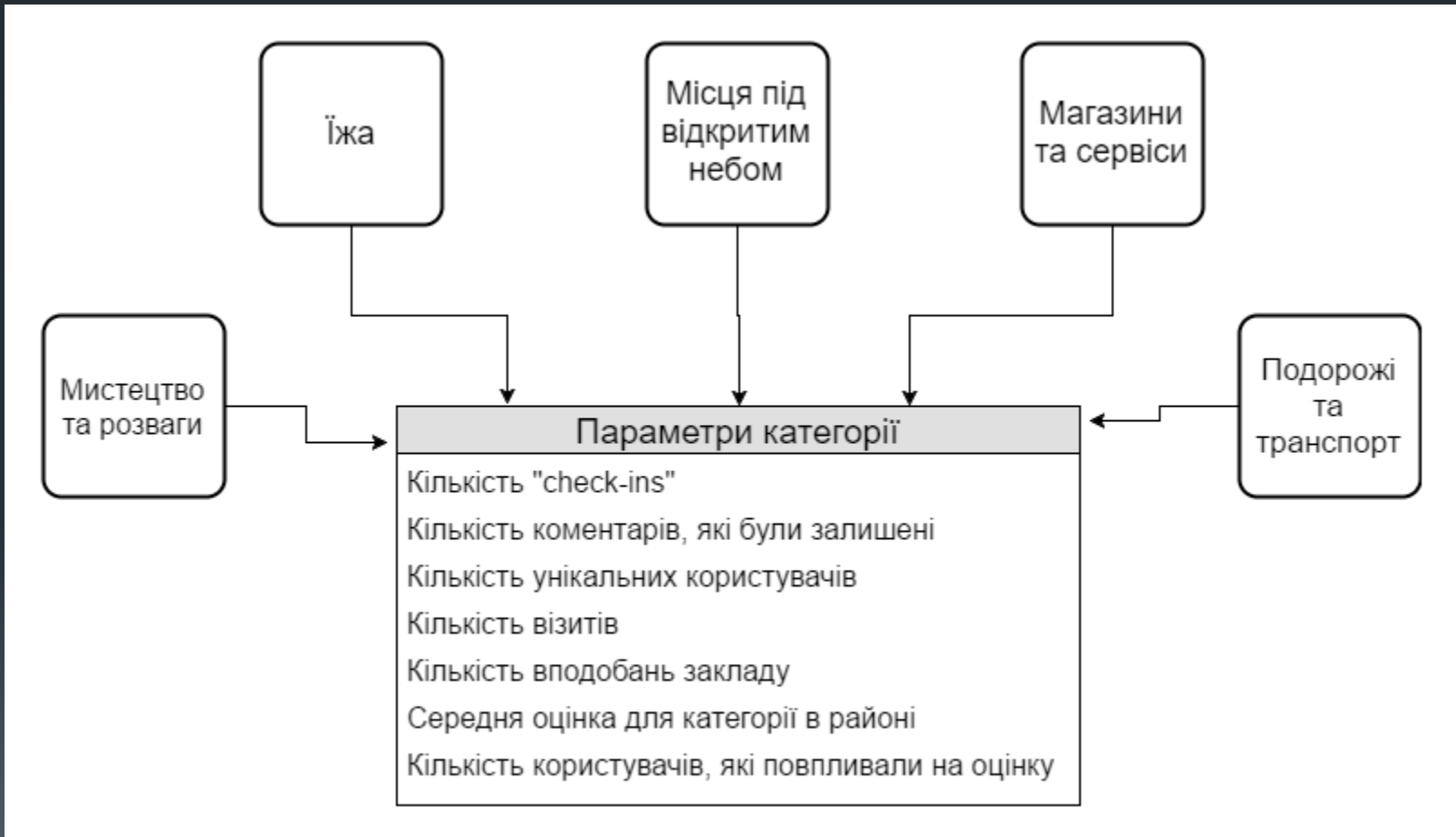
$$\begin{aligned}(I_n - \rho W)y &= X\beta + WX\gamma + \varepsilon, \\ y &= \rho Wy + X\beta + WX\gamma + \varepsilon, \\ y &= (I_n - \rho W)^{-1}X\beta + (I_n - \rho W)^{-1}WX\gamma + (I_n - \rho W)^{-1}\varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n)\end{aligned}$$

Для отримання алгоритму обчислень, нам достатньо просто замість матриці  $X$  розглянути матрицю розширену матрицю  $\tilde{X} = (X|WX)$ .

$$\hat{\beta} = \hat{\beta}_1 - \rho\hat{\beta}_2 = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

А оцінки для  $\hat{\beta}_1$  та  $\hat{\beta}_2$  знайдемо, застосувавши МНК до моделей  $y = \tilde{X}\beta_1 + \varepsilon_1$  та  $Wy = \tilde{X}\beta_2 + \varepsilon_2$  відповідно. Аналогічно обчислюються нев'язки  $e_1 = y - \tilde{X}\hat{\beta}_1$  та  $e_2 = Wy - \tilde{X}\hat{\beta}_2$ , і з логарифмічної функції правдоподібності знаходиться оцінка коефіцієнта  $\rho$  та  $\sigma^2$ .

# Опис даних

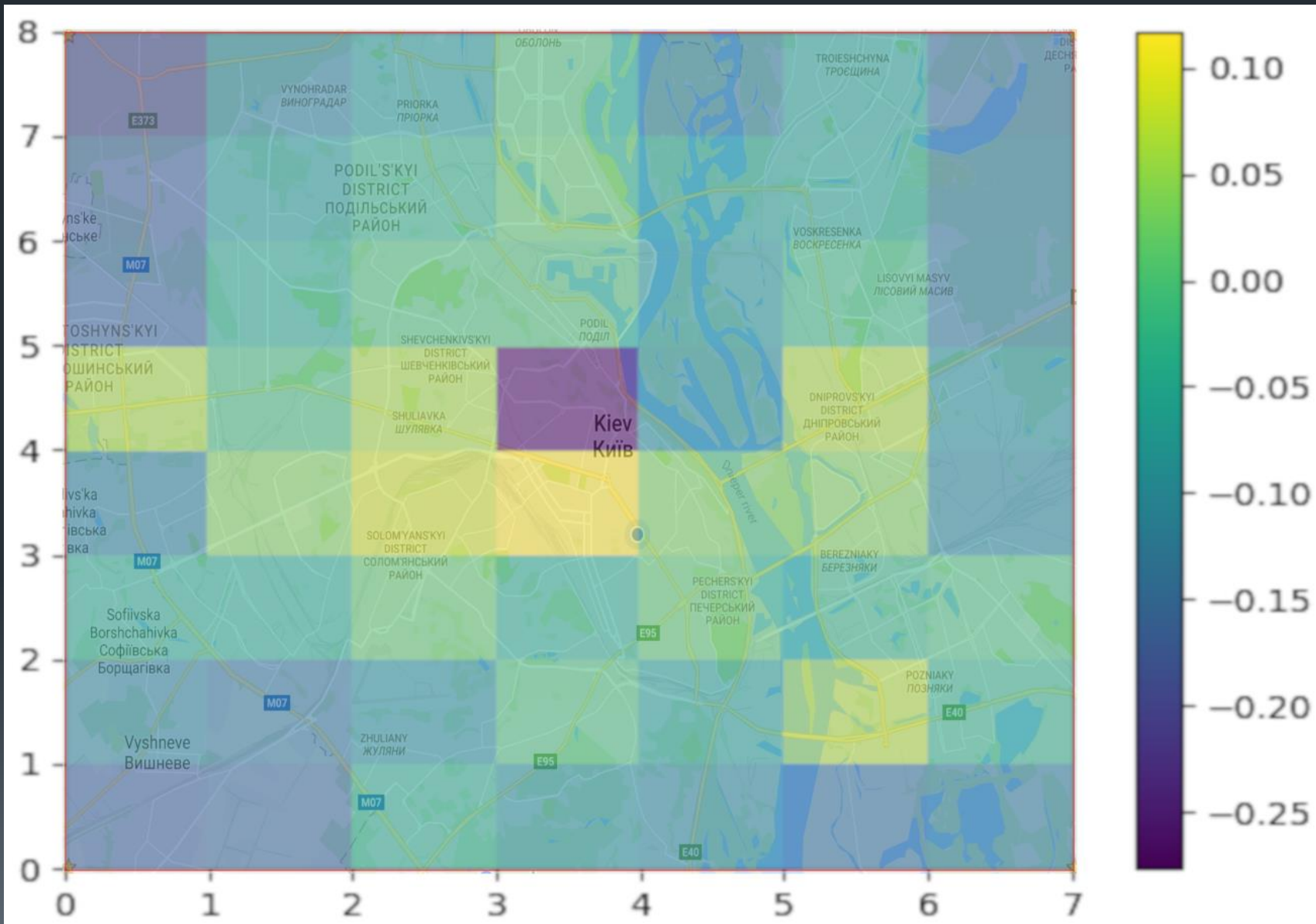


# Індекс Морана I та індекс Гірі C

	Всі	Мистецтво і розваги	Їжа	Вуличні місця	Магазини і сервіси	Подорожі і транспорт
<i>I</i>	-0.055	-0.058	-0.0522	-0.048	-0.055	-0.055
$P(0)$	0.025	0.017	0.042	0.069	0.025	0.026
<i>Z</i>	-2.233	-2.370	-2.025	-1.817	-2.238	-2.214

	Всі	Мистецтво і розваги	Їжа	Вуличні Місця	Магазини і сервіси	Подорожі і транспорт
<i>C</i>	0.177	0.209	0.195	0.110	0.171	0.188
$V(C)$	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
$P(0)$	0	0	0	0	0	0
<i>Z</i>	6.883	8.111	7.581	4.270	6.638	7.277

# Локальний Індекс Морана I



# Оцінка моделі СЕМ

	Всі	Мистецтво і розваги	Їжа	Вуличні міця	Магазини і сервіси	Подорожі і транспорт
$\lambda$	-0.2360	-0.2360	-0.2360	-0.2360	-0.0010	0.0078
$E_y$	687.875	54.892	127.392	92.196	355.964	62.107
$D_y$	750.547	75.415	159.235	78.147	390.402	67.218
$R^2$	0.984	0.976	0.995	0.947	0.990	0.970
$\sum_u$	498653	7682	6396	18190	80255	7878

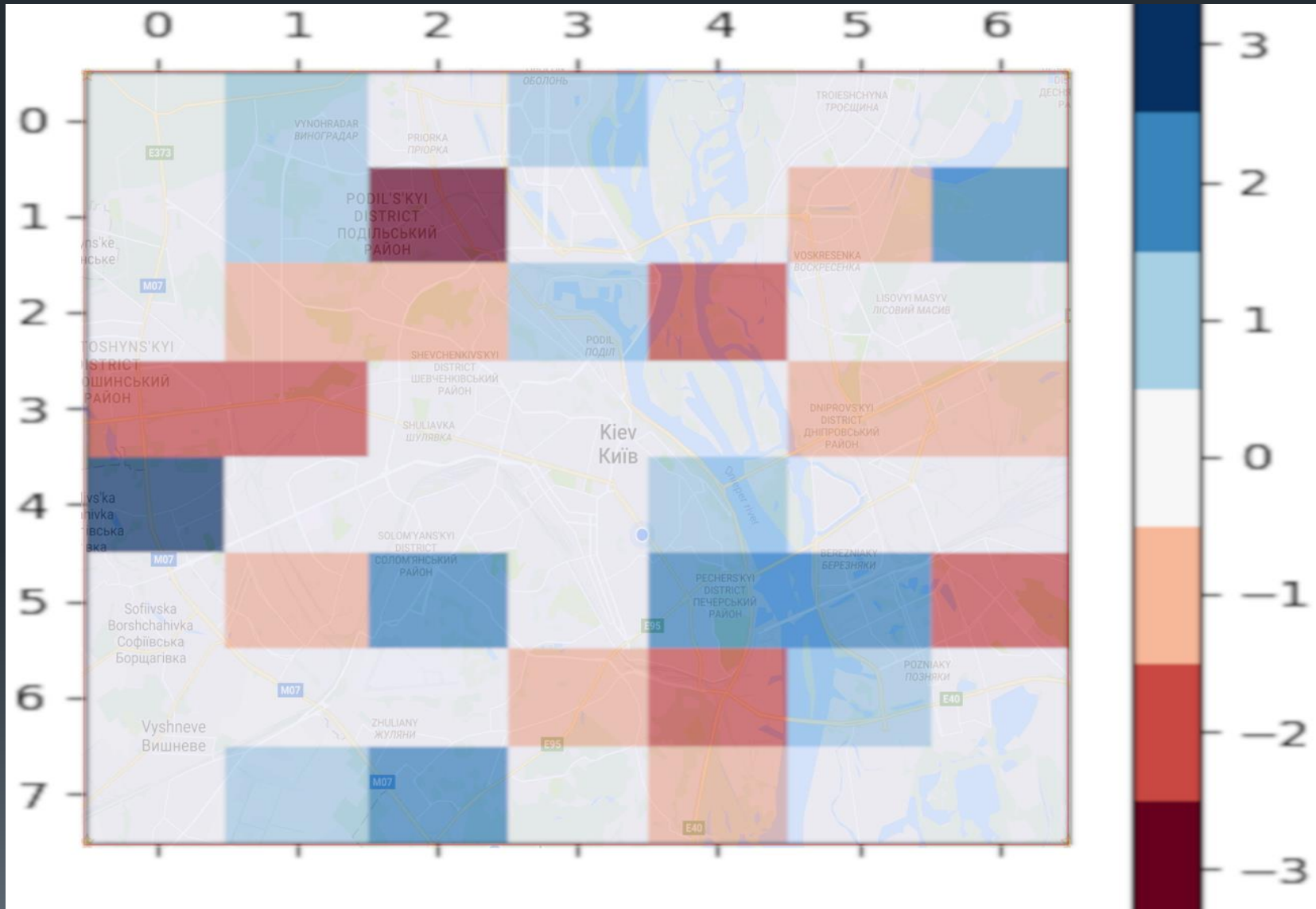


# Оцінка моделей CAR і SDM

	Всі	Мистецтво і розваги	Їжа	Вуличні місця	Магазини і сервіси	Подорожі і транспорт
$\rho$	0.0045	0.0048	-0.0023	0.0083	-0.0010	0.0078
$Ey$	687.875	54.892	127.392	92.196	355.964	62.107
$Dy$	750.547	75.415	159.235	78.147	390.402	67.218
$R^2$	0.988	0.980	0.996	0.958	0.991	0.975
$\sum u$	371670	5966	5274	13792	68155	5976

	Всі	Мистецтво і розваги	Їжа	Вуличні місця	Магазини і сервіси	Подорожі і транспорт
$\rho$	-0.5364	-0.8629	-0.9183	-0.9062	-0.8520	-0.8303
$Ey$	687.875	54.892	127.392	92.196	355.964	62.107
$Dy$	750.547	75.415	159.235	78.147	390.402	67.218
$R^2$	0.995	0.999	0.999	0.996	0.999	0.997
$\sum u$	142310	121	100	1215	2433	590

# Оцінка моделі SDM для категорії мистецтво та розваги



# Висновки

У роботі був виконаний порівняльний аналіз методів просторової економетрики для розв'язання задачі рекомендації у виборі місця відкриття закладу відпочинку, або місця проведення дозвіля.

- Після збору даних для проведення аналізу, їх було досліджено на наявність просторових ефектів, щоб визначити, чи взагалі є сенс у використанні засобів просторової економетрики.
- Протягом всієї роботи використовувався запропонований метод обчислення вагової матриці, який враховував не тільки адміністративно-територіальний устрій міста Києва, географічні координати локацій, але і значення параметрів, які ми оцінюємо для всіх п'яťох категорій в різних місцевостях.
- Проаналізувавши методи, було визначено, що найкращу оцінку дає просторова модель Дарбіна. При чому, також було показано, що всі методи дають незміщену, ефективну та конзистентну оцінку, якщо використовувати метод максимальної правдоподібності. Паралельно з цим було складено обчислювальний алгоритм, який базувався на методі найменших квадратів та методах одновимірної оптимізації.

Дякую за увагу!

