



BigDataSchool 5.0

Курсовий проект

upgraded team 8

Семенченко Тарас

Прокопенко Андрій

Олена Мельник

Дмитро Лиховський





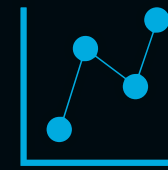
суть аналізу

аналіз мобільності
Україна та м.Ужгород



період аналізу

04.2019-09.2019
04.2020-09.2020



суттєва зміна умов

карантинні обмеження



Задача 1

Порівняти показники мобільності для кожного з виявлених сегментів

Задача 2

Провести аналіз яким чином кожен із сегментів в'їжджав та виїжджав з міста Ужгород та з якою метою

Задача 1 – аналіз даних



50 тис. абонентів



7 млн. поїздок



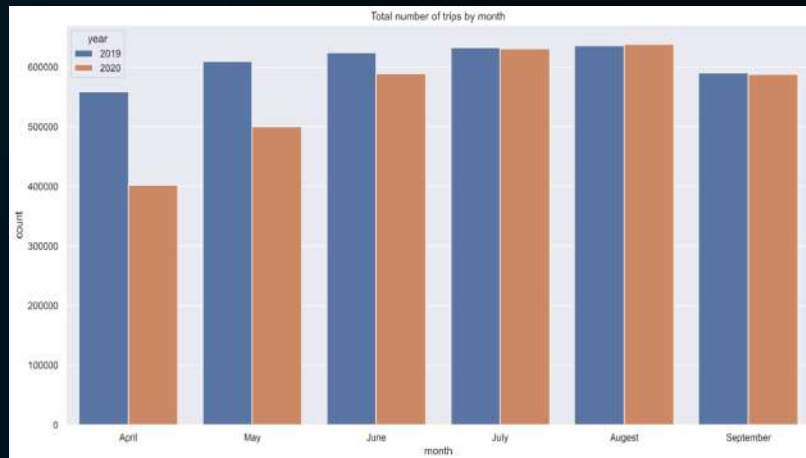
517 районів, що покривають всю країну



11 млн. год. в дорозі

-9% падіння кількості поїздок 2020 проти 2019

88% падіння припадає на період локдауну з поступовим відновленням протягом червня

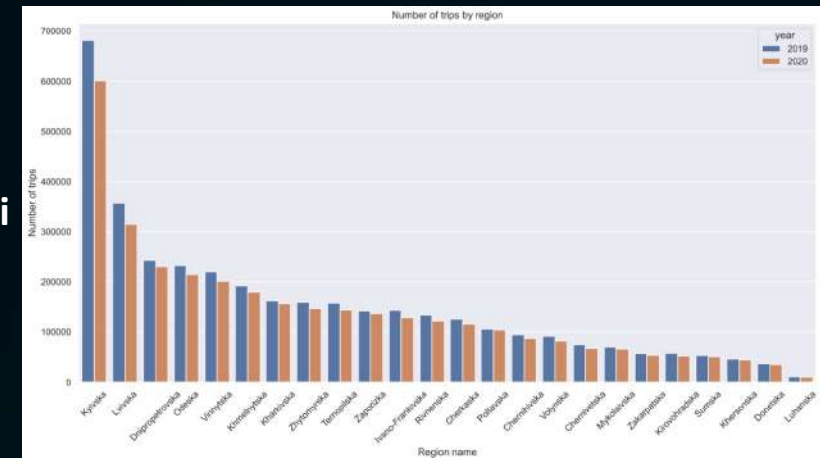
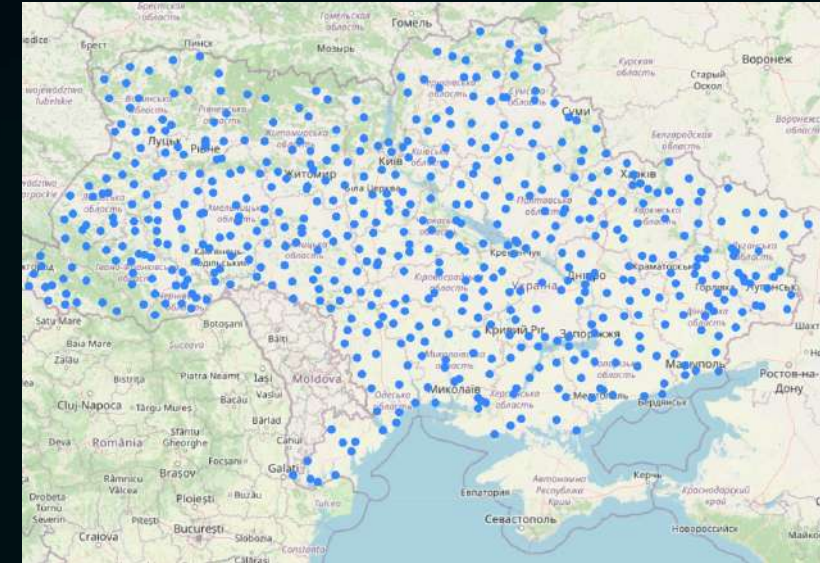


Найбільше постраждали в період локдауну Львівська, Київська, Кіровоградська, Чернівецька та Івано-Франківська області.

Найменше постраждали Полтавська, Херсонська, Харківська та Запорізька області

Найшвидше відновлення після локдауну відбулось в Полтавській області

Рятував ситуацію внутрішній туризм, що допомогло відновитись Херсонській, Закарпатській та Запорізькій областям



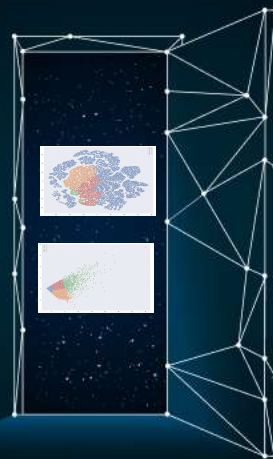
Задача 1 – сегментація абонентів



кількість районів,
які відвідав
абонент за період



умовна відстань, яку
проїхав абонент за
період



час витрачений в
дорозі за період

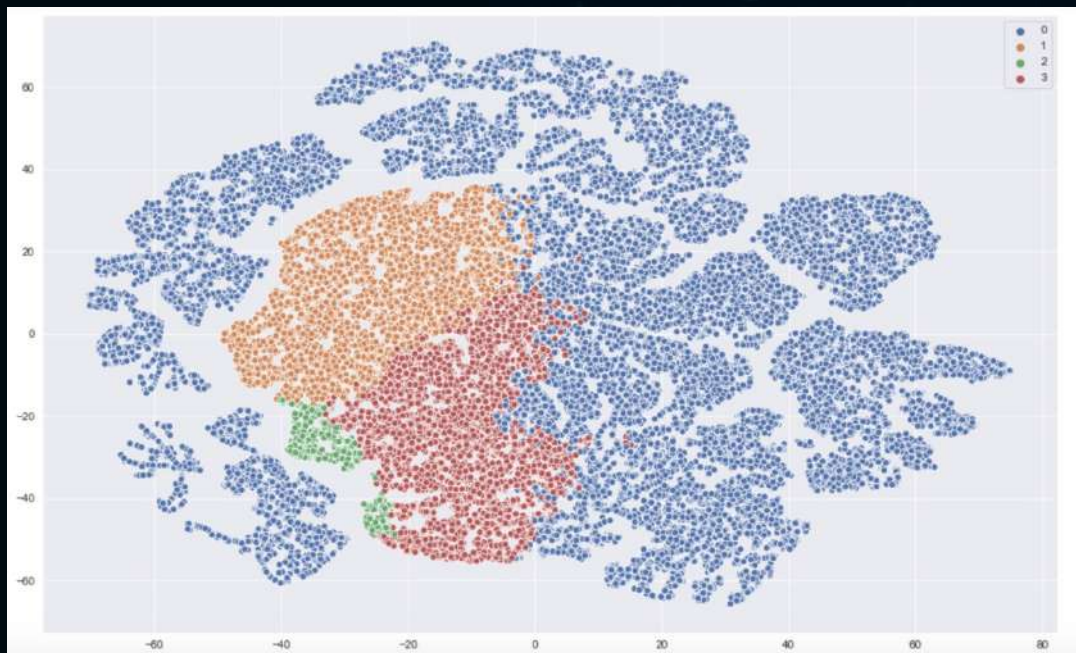


загальна кількість
поїздок в межах
області

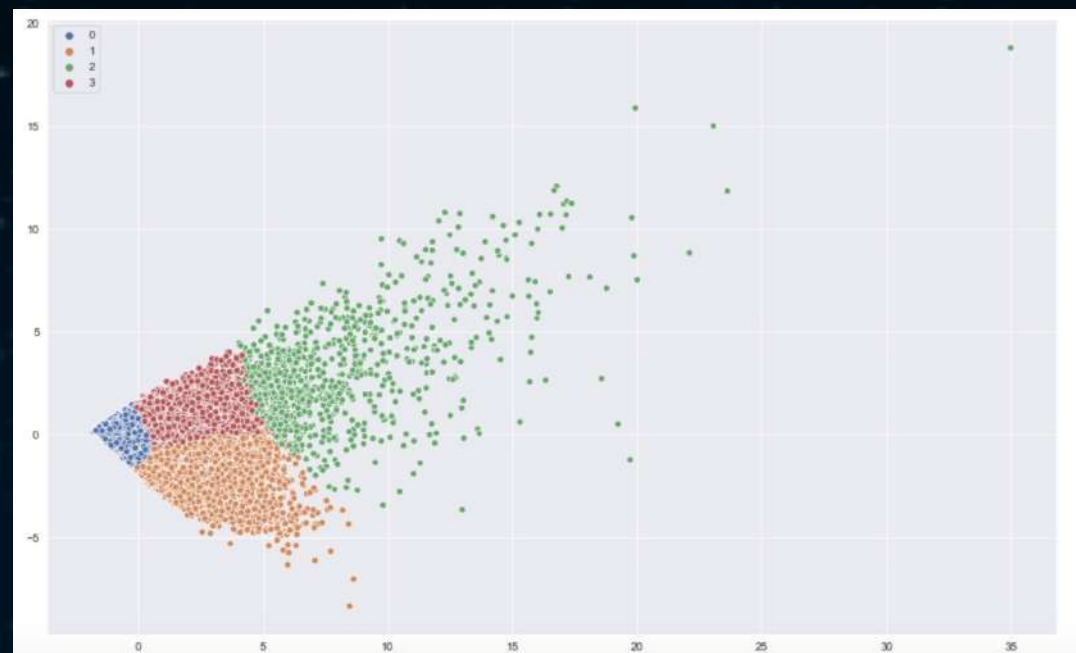


загальна кількість
поїздок за межі
області

Задача 1 – сегментація абонентів - кластери



t-SNE – візуалізація сегментації на 4 кластери



PCA – візуалізація сегментації на 4 кластери

00	70%	Живуть та працюють/навчаються в одному районі 3 РАЗИ на місяць виїжджають зі свого району	За пів року 1-2 РАЗИ були в іншій області
01	14%	Їздять в інший район в зв'язку з роботою майже щодня, їздять часто та недалеко (~32 км)	За пів року 2 РАЗИ були в іншій області
02	2%	Їхня робота в дорозі – 85 годин в місяць за кермом	За пів року 45 РАЗІВ були в іншій області
03	14%	Для виконання роботи регулярно виїжджають в інший район	За пів року 12 РАЗІВ були в іншій області

Задача 1 – сегментація абонентів – дані за 04-09.2019

70%



Назар живе та працює в м. Рівне. Декілька разів на місяць (18 разів за півроку) їздив в сусідні райони в межах області в основному на короткий час. 3 рази їздив в Львів та разок на 2 дні в Яремче.

2%



Вадим, живе в м.Бердичів, хоча його реальне життя в дорозі. За півроку Вадим 86 разів їздив за межі області і об'їхав велику частину України (Хмельницька, Київська, Запорізька, Одеська...). Ще й 68 разів їздив в інший район в межах області.

14%



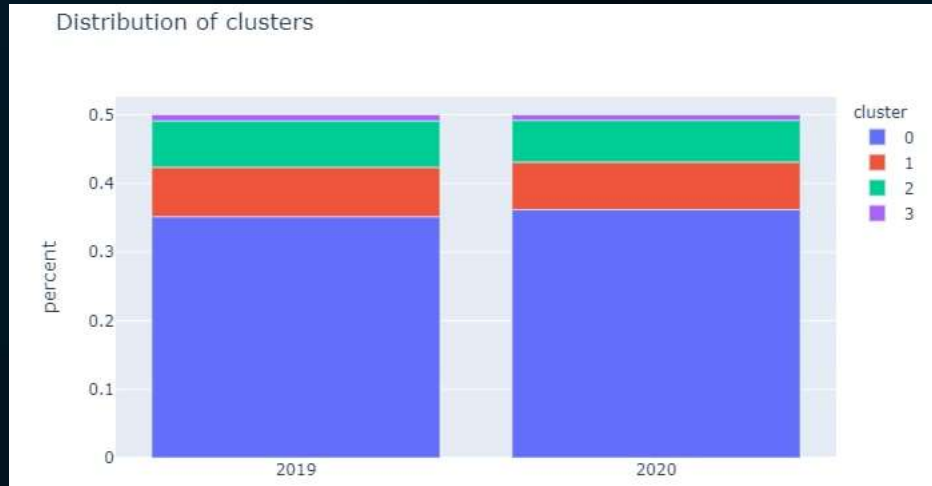
Юра, живе в Великобілозерівському р-ні Запорізької обл. Юра працює в сусідньому районі, тому дуже часто там буває. На короткий час Юра 2 рази їздив в Запоріжжя і 1 раз в Херсонську область.

14%



Юля живе і працює в Кременчуці. Хоча доволі часто буває в різних районах в межах області (43 рази за півроку). Ще й 4 рази їздила в інші області України, щоразу в нову область.

Задача 1 – аналіз динаміки змін в розрізі сегментів



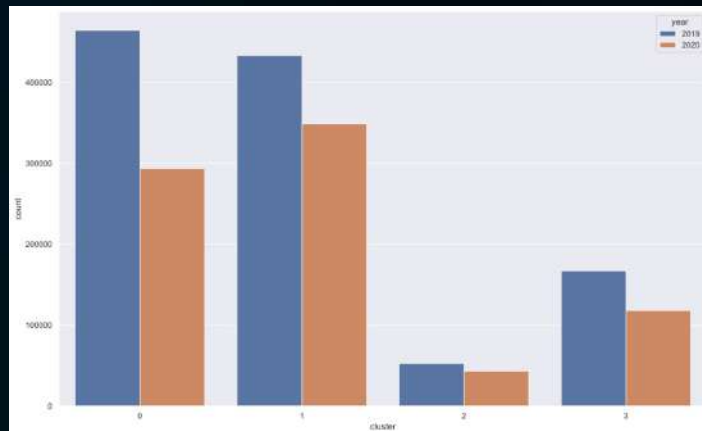
Підсумково, за 6 міс. структура кластерів змінилася мінімально.

Основні відхилення спостерігалися в місяці локдауну

В 2020 збільшився кластер 0 на 3% в основному за рахунок кластера 3

В серпні відновилась мобільність та навіть дала приріст проти 2019

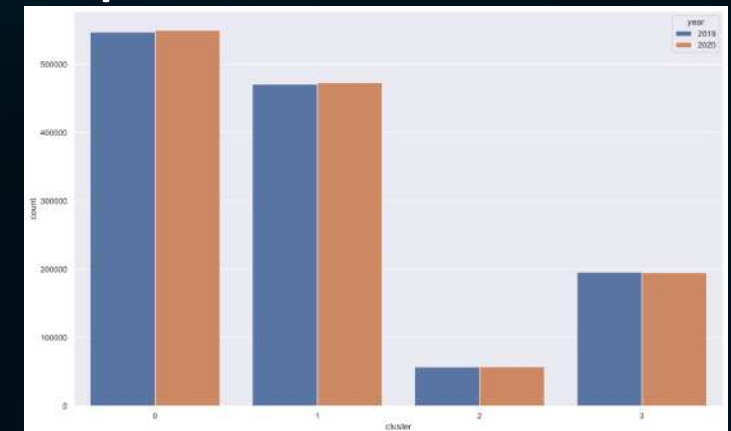
Квітень



Травень



Серпень



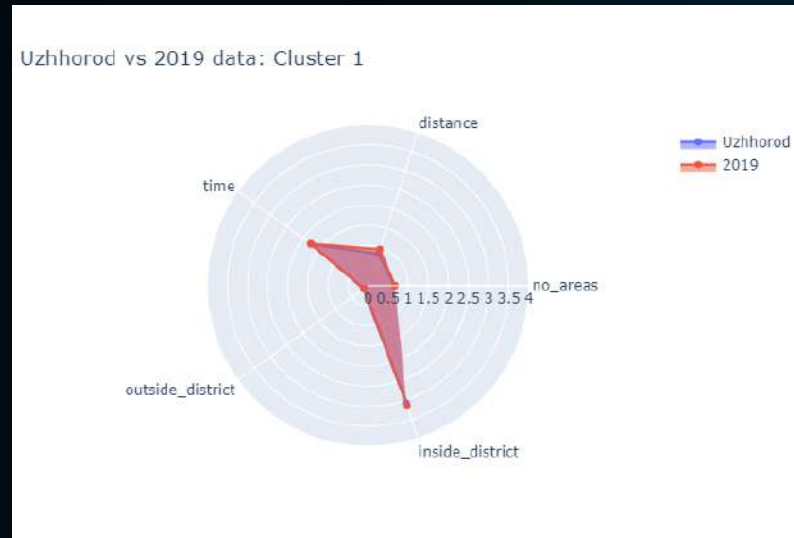
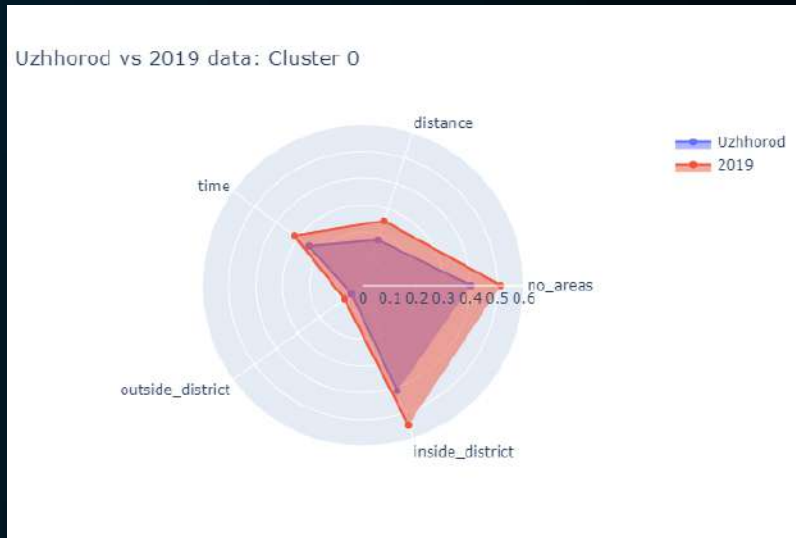
Кількість поїздок в розрізі кластерів за квітень, травень, серпень 2019 та 2020 років

УЖГОРОД

Аналіз в'їздів та виїздів
по сегментах

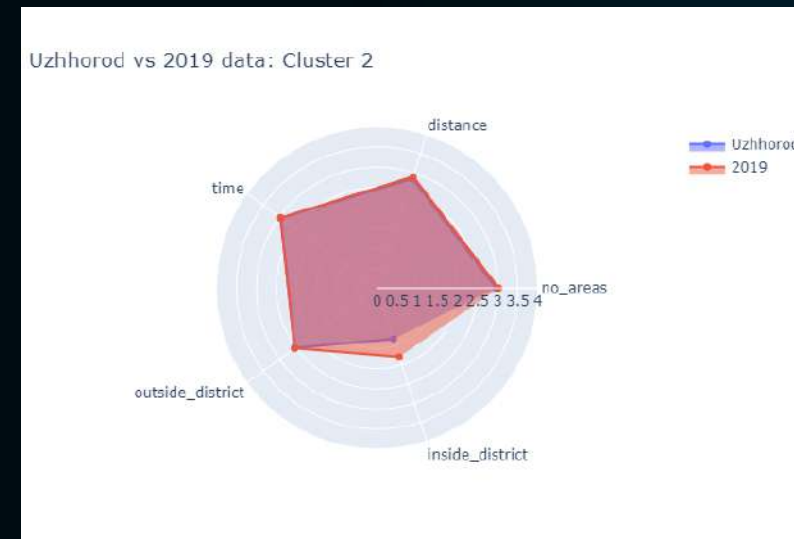


Задача 2 – порівняння сегментів

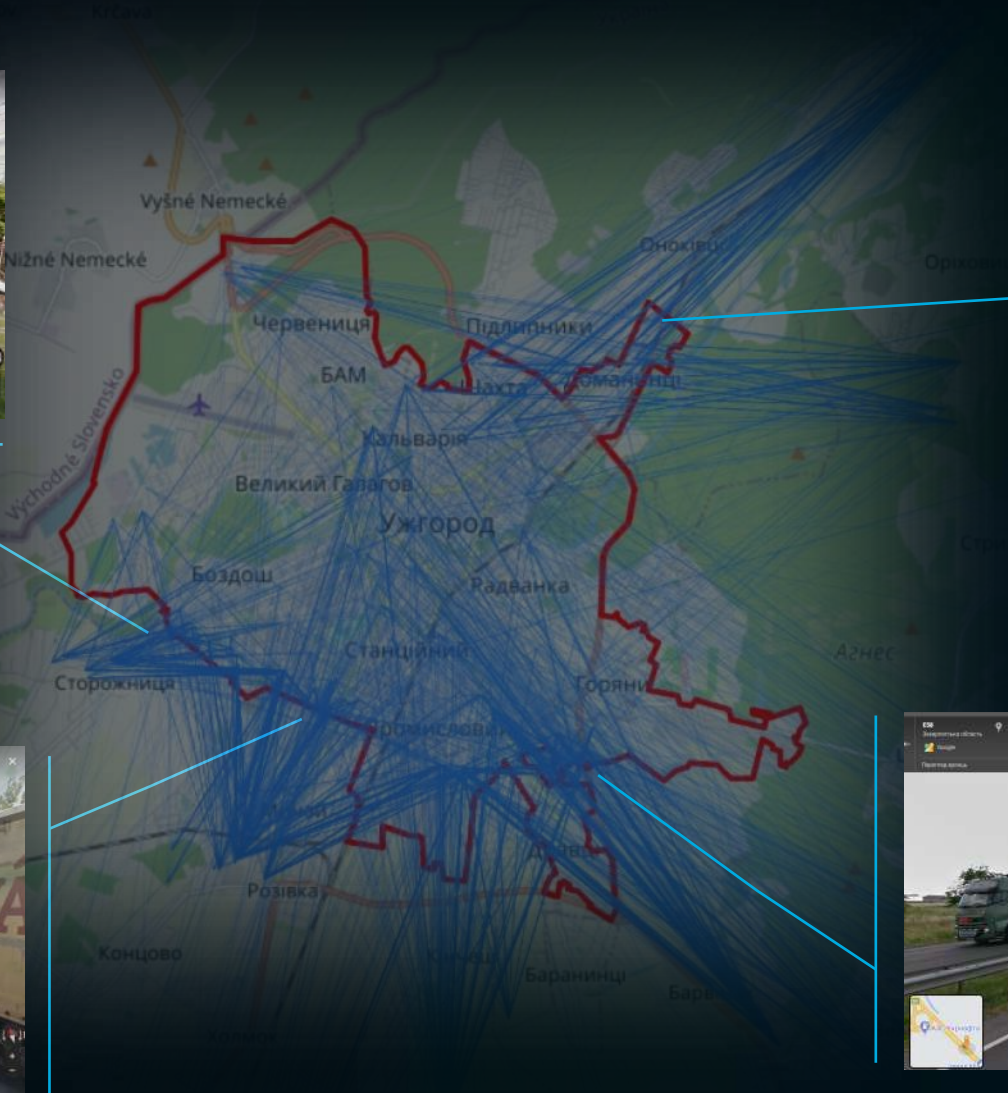


Аналіз не показує суттєвих розбіжностей між загальною сегментацією та для м.Ужгород за виключенням кластера 0.

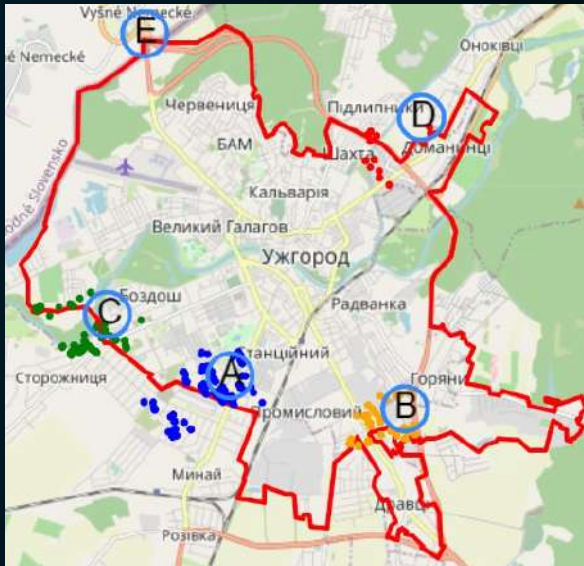
Для абонентів кластера 0 сегмент в Ужгороді має менші показники



Задача 2 – аналіз ключових в'їздів та виїздів



Задача 2 – аналіз в'їздів та виїздів



Виділені ключові точки в'їзду/виїзду

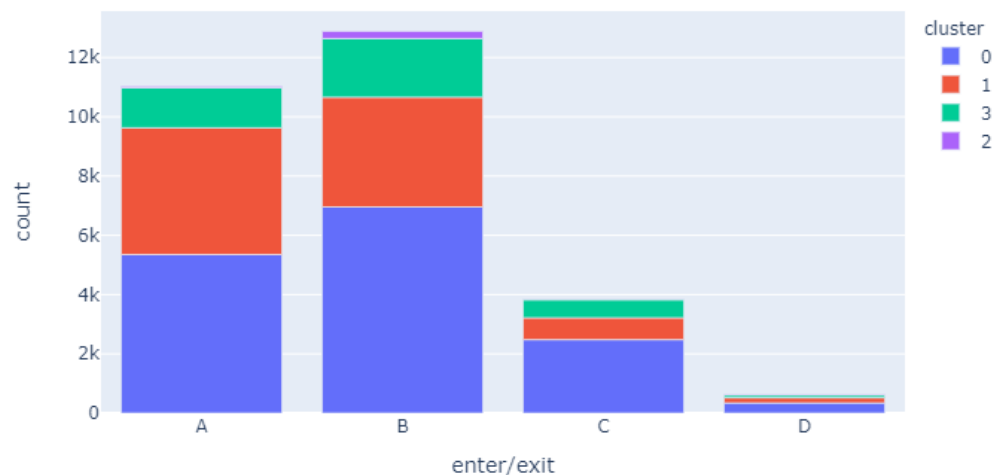
Точка А – Е573 (європейська траса) – припадає 38%

Точка В – Е50 (європейська траса) – припадає 46%

Точка С – Т-08-02 (регіонального значення) – припадає 14%

Точка D – Н-13 (національна дорога, веде до прикордонного пункту пропуску) – припадає 2%

Distribution of clusters in city entry/exit points

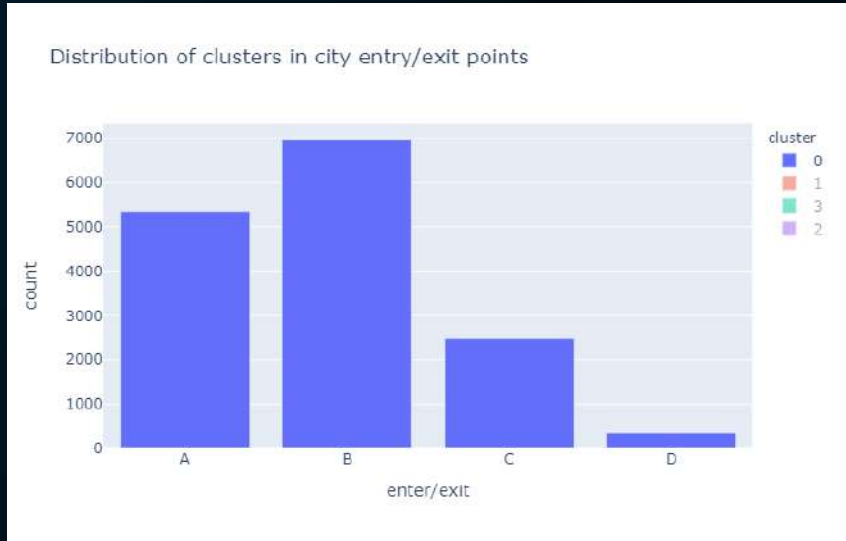


Пункт В – найбільш популярна точка в'їзду/ виїзду в/з міста.

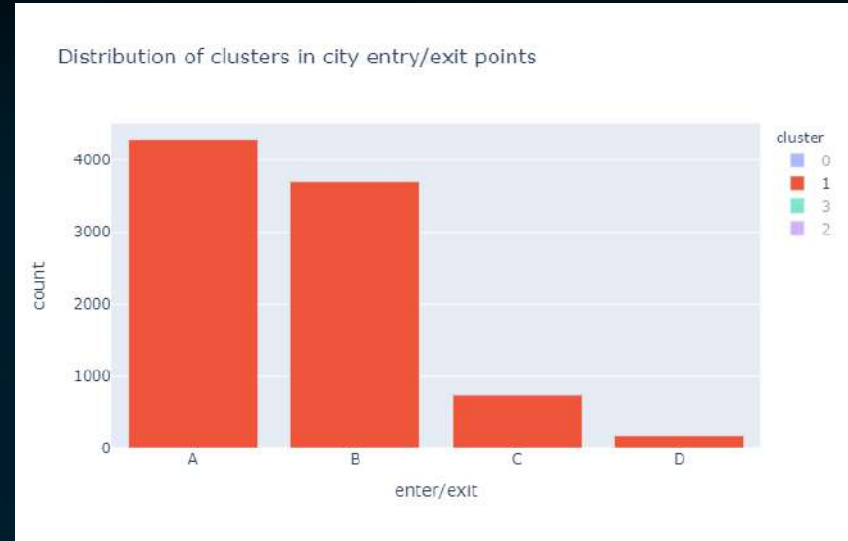
Основний пункт для всіх кластерів, окрім 1-го.

1-й кластер надає перевагу точці А

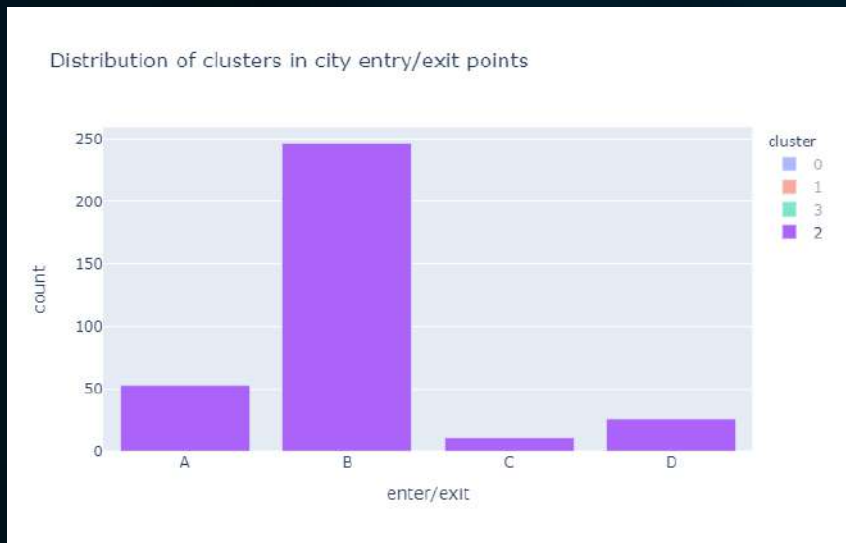
Задача 2 – аналіз в'їздів та виїздів



Кластер 0



Кластер 1



Кластер 2

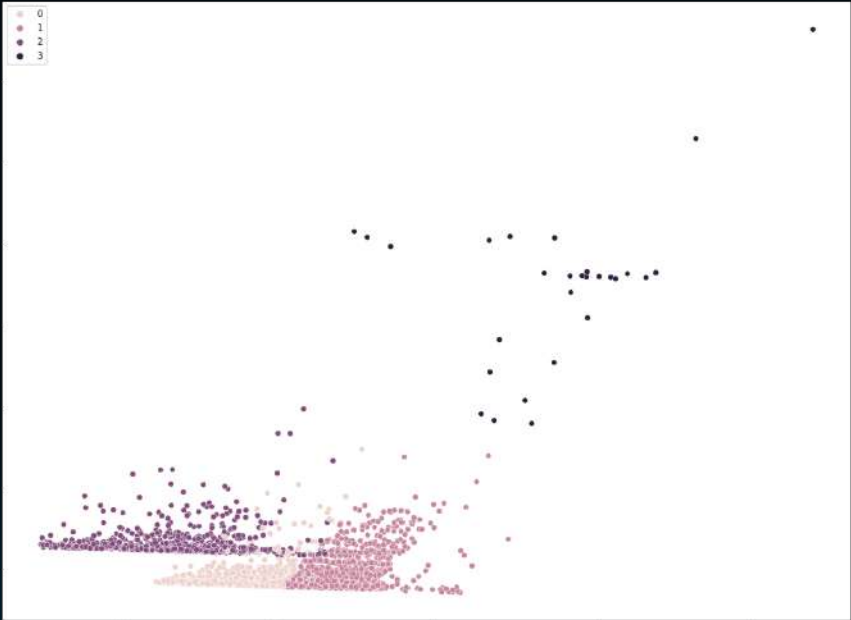


Кластер 3

Задача 2 – кластеризація за метою поїздки



t-SNE – візуалізація сегментації на 4 кластера



PCA – візуалізація сегментації на 4 кластера

00	27,02%	Приїжджають ввечері, в робочі дні
01	44,35%	Приїжджають зранку в будні дні з пригороду, ймовірно в цілях роботи
02	19,31%	Приїжджають у вихідні дні , ймовірно з ціллю покупок або відпочинку
03	0,67%	Приїжджають в різні дні після довгої дороги (~600 км), ймовірно в робочих цілях та туризму
*	8,65%	Абоненти, які проїжджають транзитом до кордону



Відстань



Будній вихідний



Час прибуття

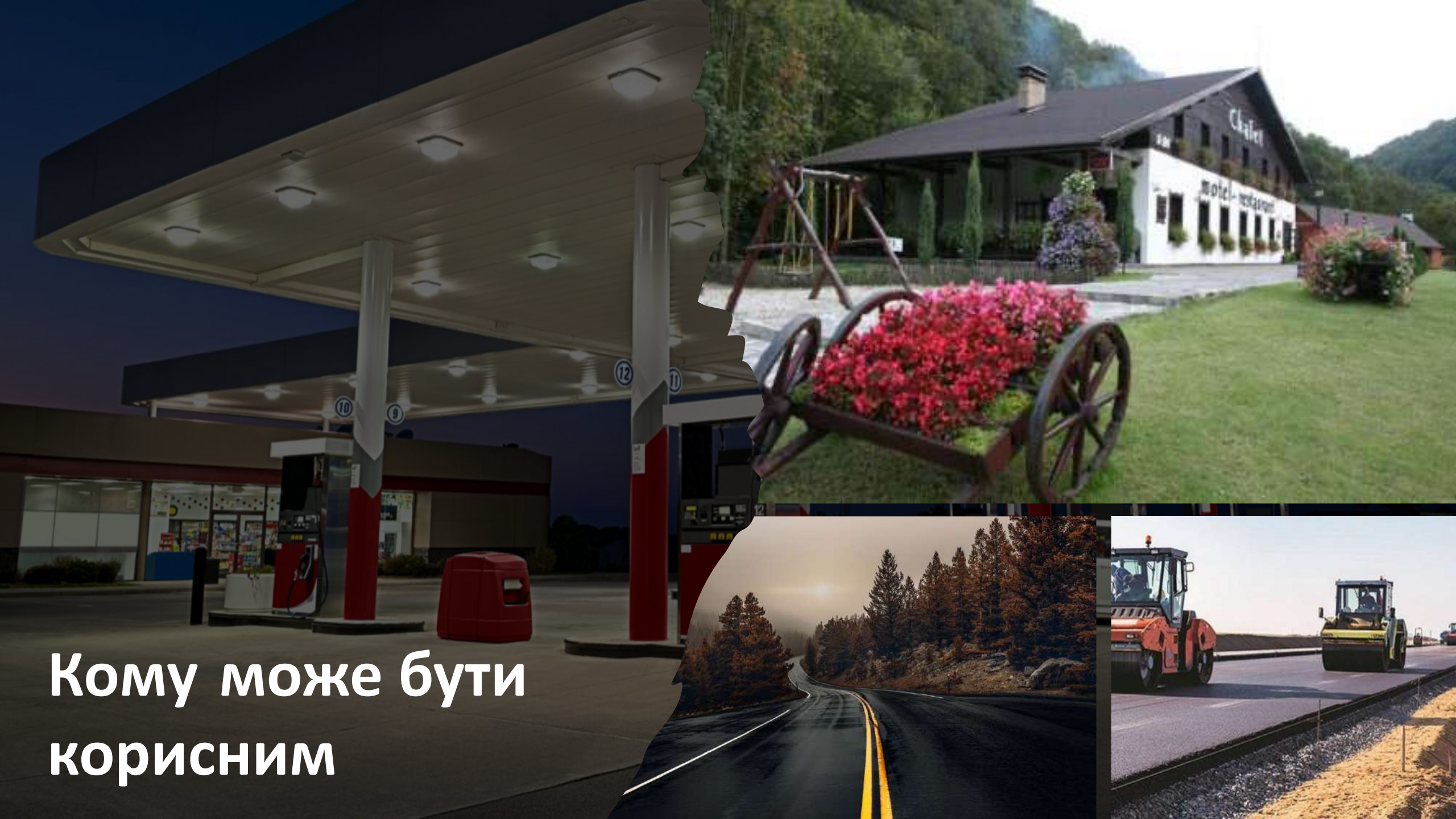
Задача 2 – аналіз поїздок, пов'язаних з кордоном



Структура поїздок в Ужгород, пов'язаних та не пов'язаних з кордоном



Регіональне походження абонентів, чиї поїздки пов'язані з кордоном



Кому може бути
корисним