

Térinformatikai adattárolás

Műszaki leírás

A feladat bemutatása

A házi feladat célja egy térinformatikai adatbázis létrehozása volt, és azon 5 darab egyszerű, illetve 5 darab térbeli lekérdezés végrehajtása. A lekérdezések olyan eredményeit, amelyek geometriai típusúak QGIS-ben is ábrázolhatjuk.

Az adatok kiválasztása és bemutatása

A feladathoz felhasznált adatok kiválasztásánál fontos szempontot játszott, hogy a való életből merített problémát oldjon meg az adatbázis és nagyszámú mintaadat álljon rendelkezésre. Ezért úgy döntöttem, hogy a kaggle.com nevű oldalról használok fel egy közérdekű adatokat tartalmazó adathalmazt. Az adathalmaz Salt Lake City rendőrségi bejelentéseit tartalmazza a 2016-os évre vonatkoztatva. Az eredeti csv kiterjesztésű fájl az alábbi linken érhető el:

- <https://www.kaggle.com/foenix/slc-crime/data>

Az adatfájl a következő oszlopokat (kategóriákat) tartalmazza:

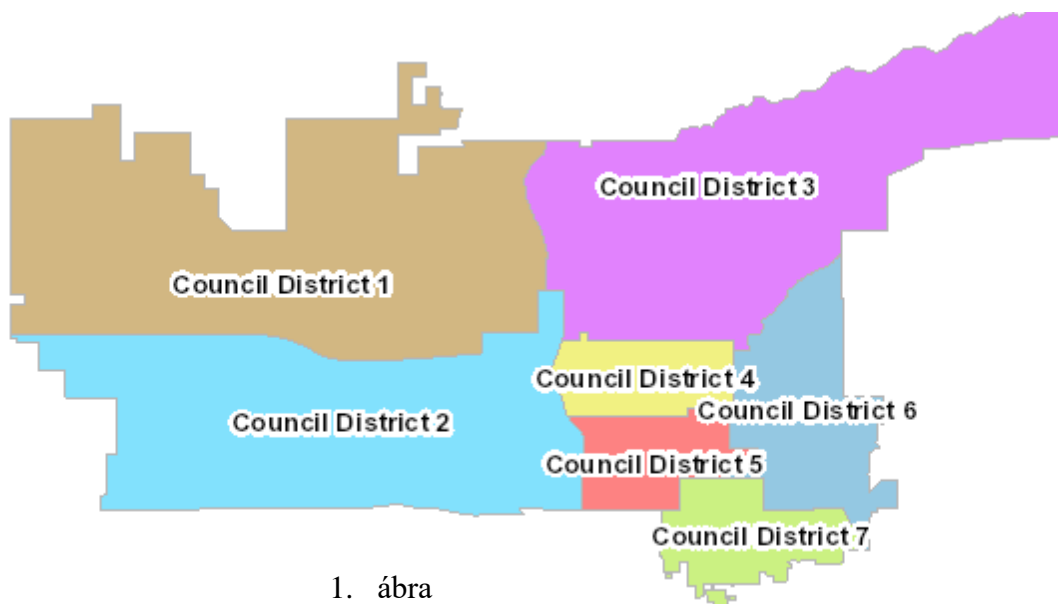
- Case number: A bűneset egyedi azonosítószámát tartalmazza.
- Offence Code for categorization: Azt határozza meg, hogy a rendőrség melyik osztályára kerüljön az eset.
- Description for context: Megnevezi a bűncselekmény típusát.
- IBR (National Incident-Based Reporting System Number): A könnyebb kategorizálás miatt létrehozott USA-beli azonosítószám.
- Occurrence Date: Mikor történt az eset.
- Report Date: Mikor történt a bejelentés.
- Day of the Week (1 = Monday, 7 = Sunday): A hét melyik napján történt az eset.
- Location: A legközelebb eső utcasarok azonosítja az eset helyét.
- City Council District: A város melyik körzetében történt az eset.
- Police Zones: Melyik rendőri zónához tartozik az eset.
- Grid: A rendőri zónához tartozó alazonosító.

- x_coordinate: epsg:32043 –ben vett x koordináta.
- y_coordinate: epsg:32043 –ben vett y koordináta.
- x_gps_coors: x_coordinate áttaszformálva epsg:4326-ba.
- y_gps_coors: y_coordinate áttaszformálva epsg:4326-ba.

Ezen kívül készítettem még egy xlsx amely a város honlapjáról válogatta ki a legfontosabb információkat a kerületek képviselőivel kapcsolatban. A fájl neve:

- City_Council_Members_List.xlsx

Szükség volt továbbá egy képre is, arra vonatkozóan, hogy a város hogyan oszlik fel kerületeire. A későbbi területeket ezen kép alapján készítettem el:

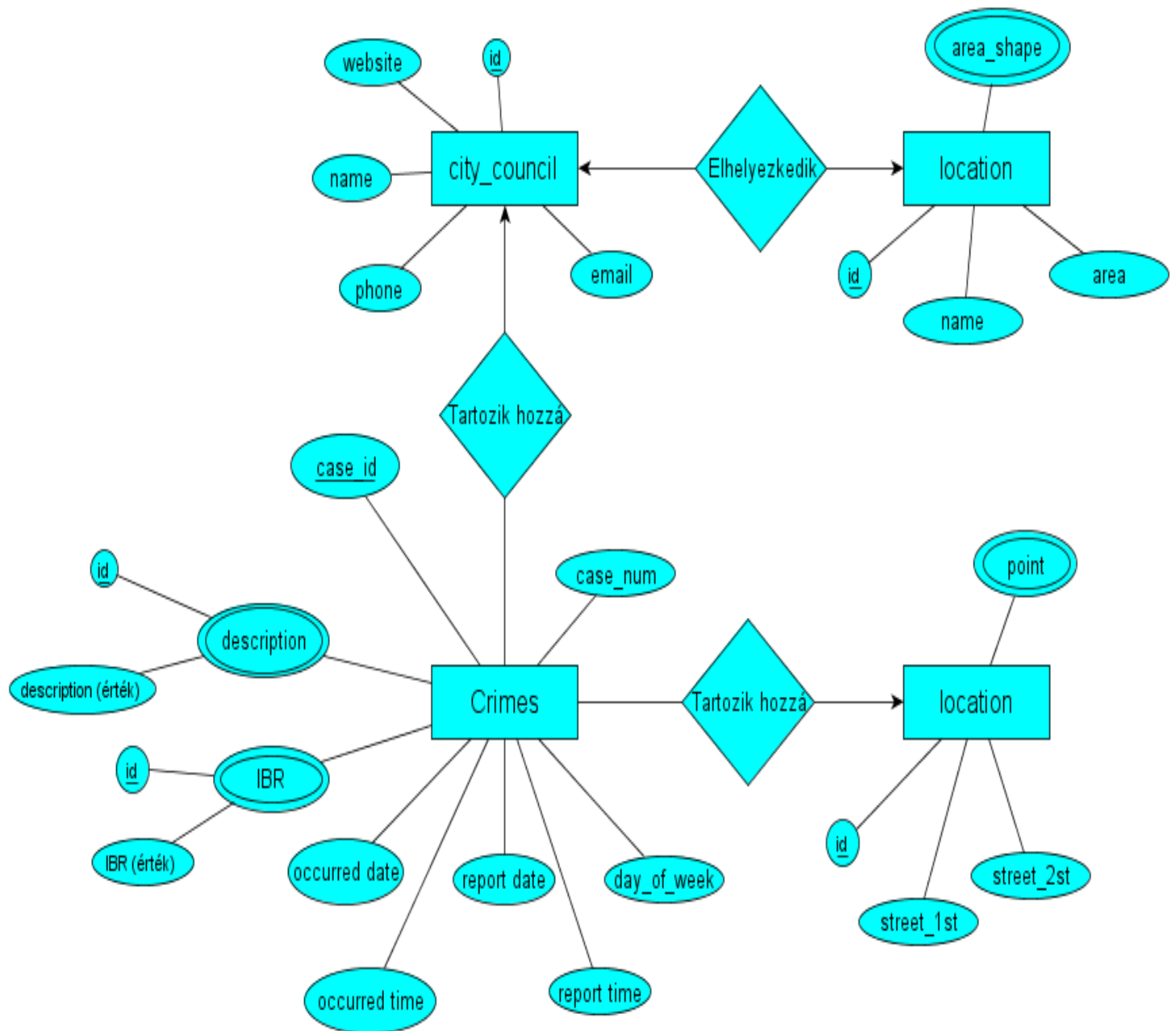


1. ábra

Az adatok tisztítása

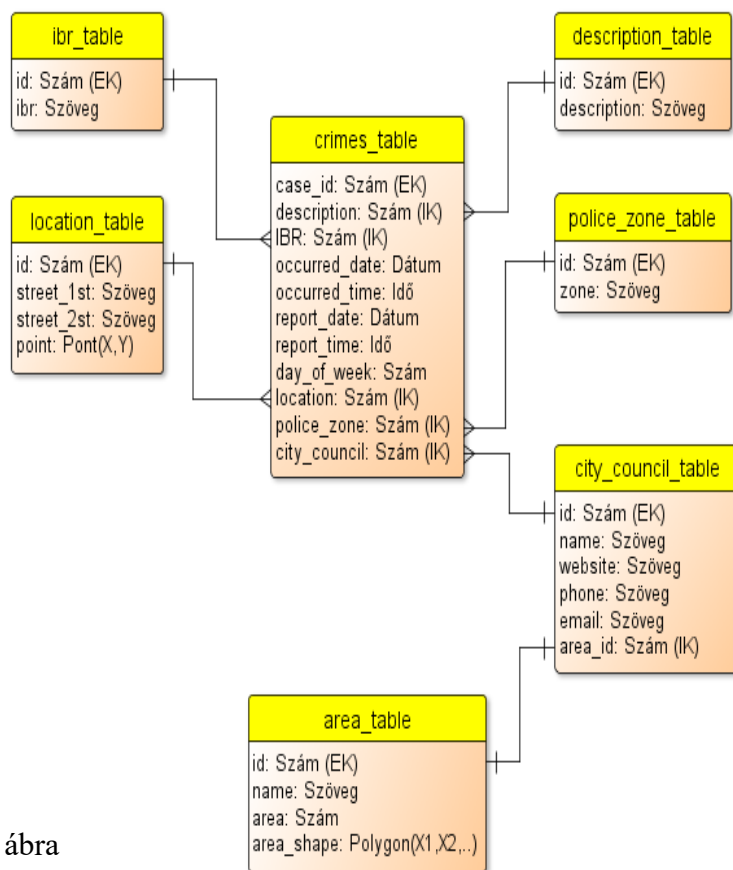
Mivel a bejelentéseket tartalmazó adatfájl részben nyers adatokat tartalmaz, ezért először szükség van az adatok megtisztítására. Ez az összes érték nélküli adatsor elvetését, valamint a számunkra jelentéktelen kategóriák törlését vagy ártérmezését jelenti. Erre a feladatra a legcélszerűbb eszköznek a python nyelvet találtam, amelynek beépített könyvtárai biztosították azt, hogy az érték nélküli adatsorokat kivegyem az adathalmazból, ezen kívül a felesleges koordináták is kikerültek, csak a később végig használt epsg:4326-os projekció koordinátáit tartottam meg. Továbbá itt kezdtem el olyan DataFrame típusú adatszerkezetekbe elhelyezni az adatokat, amelyek alapján a későbbi táblastruktúra is összeállt.

E/K diagram



2. ábra

Relációs adatbázis séma



3. ábra

A táblák elkészítése és feltöltése

A táblák elkészítéséhez ugyanazt a script fájlt használtam, mint amit az adatok elrendezéséhez. A python program, megpróbál kapcsolatot létesíteni a szerverrel, és ha létrejön a kapcsolat, akkor megvizsgálja, hogy léteznek-e a fenti táblák, ha nem, akkor létrehozza őket és fel is tölti a megtisztított adatokkal. Ahol nagyon sok sornyi adatot kell feltölteni, ilyen például a crimes_table tábla, ott egy pár milliszekundumos szünetet is beiktattam, hogy ne terhelje túl véletlenül sem a program a szerveret. Miután a program lefutott, lezárja a kapcsolatot.

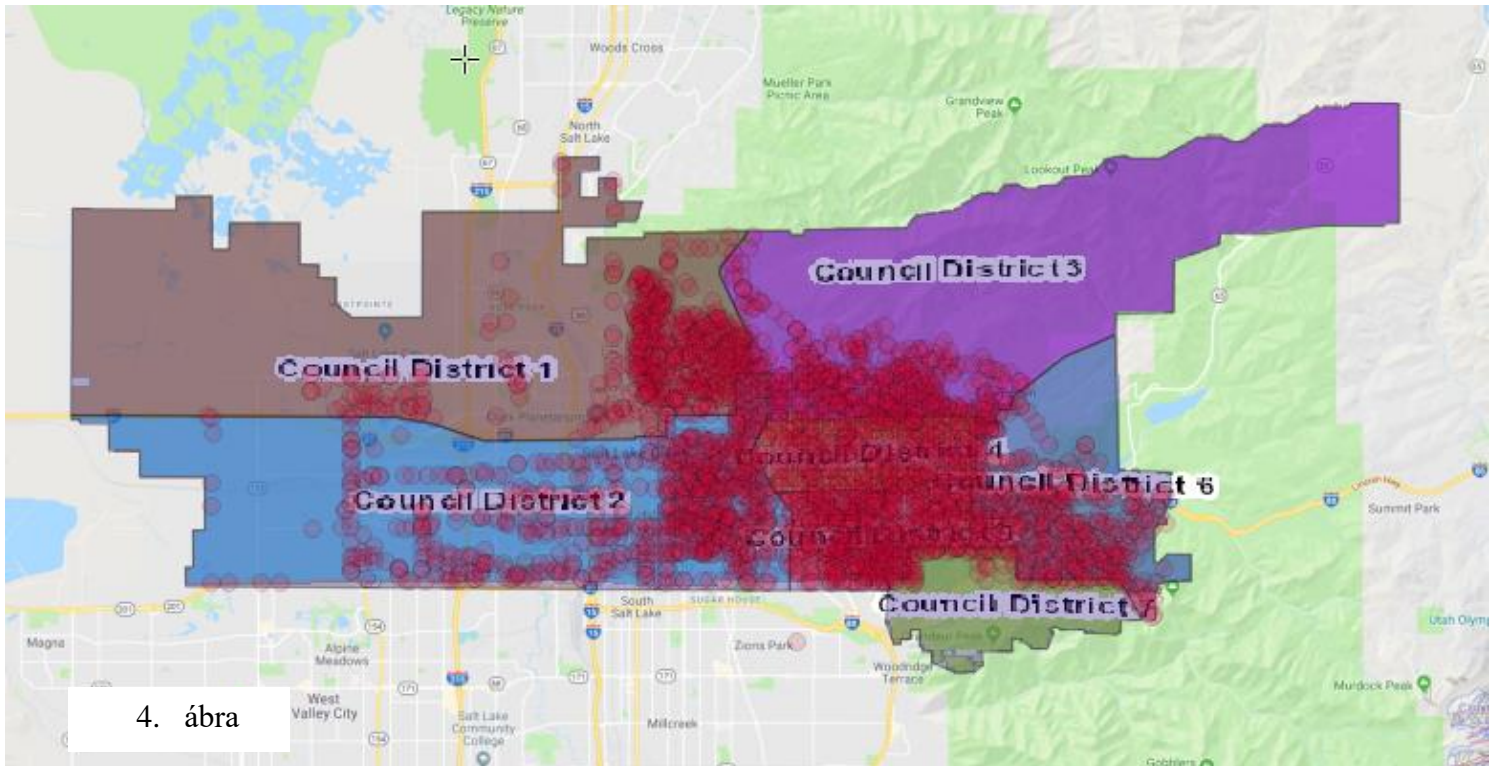
A tisztításért és feltöltésért felelős py kiterjesztésű fájl az alábbi néven található meg:

- Salt_Lake_tide.py

A feltöltés egy közbenső lépése volt, hogy le kellett határolni a kerületeket. Ehhez a fent már bemutatott 1. ábrát használtam fel, amit georeferáltam a Google Streets típusú térképére. A georeferáláshoz „Thin Plate Spline” típusú transzformációt használtam. A lefedés szinte tökéletesen sikerült.

A feltöltött adatok megjelenítése

A feltöltött adatokat úgy ábrázolom, hogy a pontok átlátszóság 85%-ra van állítva.



A kép alapján látható, hogy van olyan terület, ahol sokkal nagyobb volt a bejelentett bűncselekmények száma.

Egyszerű lekérdezések

A kérdéseket és a válaszokhoz szükséges postgresql lekérdezést az alábbi txt fájlban lehet megtalálni:

- queries.txt

Kérdések és válaszok

1. Mely 3 bűncselekmény típusból követték el a legtöbbet, és pontosan hányat 2016-ban?

A válasz:

Larceny (lopás) – 6719

Public order (közrend megbontása) – 5577

Drugs (kábitószerrel való visszaélés) – 4573

2. A hét egyes napjain hány bűnesetet követnek el?

1(Hétfő) – 5155

2 (Kedd) – 6183

3 (Szerda) – 6213

4 (Csütörtök) – 6429

5 (Péntek) – 6263

6 (Szombat) – 6676

7 (Vasárnap) – 5799

3. Van-e olyan bűneset, amelyhez több esemény is kapcsolódik az adatbázisban?

SL2016119037 kódhoz 2 eset köthető.

SL2016119114 kódhoz 2 eset köthető.

4. Melyik egy nap legveszélyesebb időszaka (órája) és hány bűneset történik ekkor?

A 17 és 18 óra közötti időszak, ekkor 2643 bűnesetet jelentettek be.

5. Kit és milyen telefonszámon kell hívnia annak, aki panaszt akar tenni a rendőrség gyatra munkája miatt, mivel nem találták meg a 2016 január 5.-én 00:00 és 07:00 között ellopott gépjárművét? (A kerületi képviselők az illetékesek a rendőrök helyrerakásában.)

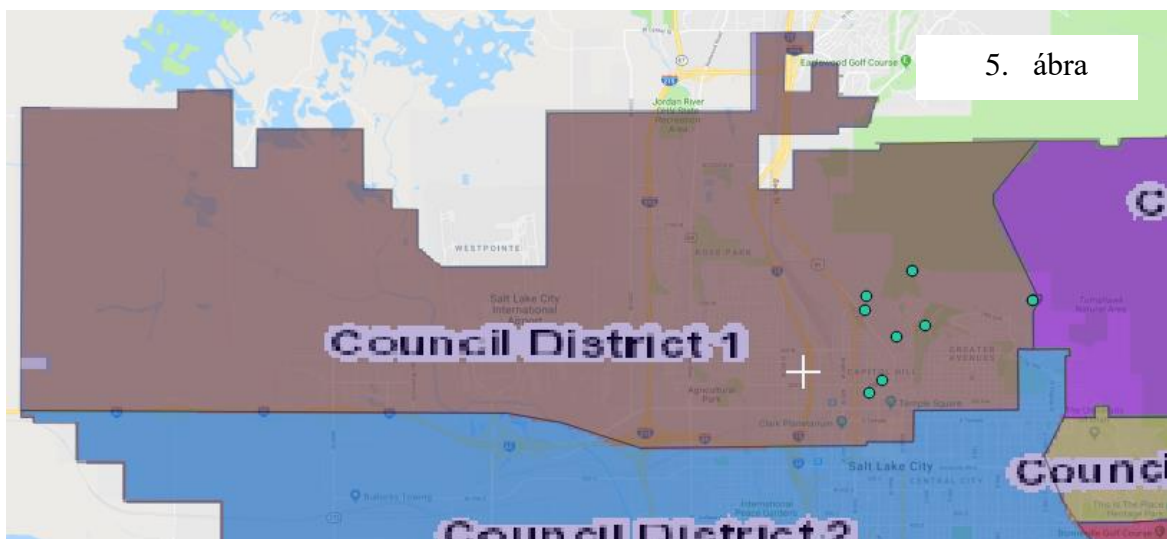
Andrew Johnston-t kell felhívnia a 801-535-7781 –es telefonszámon.

Térbeli lekérdezések

1. Melyik kerületben történt a legkevesebb bűneset, és mennyi?

A 7-es körzetben és összesen 308 darab.

2. Mutassa meg a QGIS, hogy hol követtek el emberrablást az 1. kerületben?



A fenti képen a türkíz pontok jelölik a keresett helyeket.

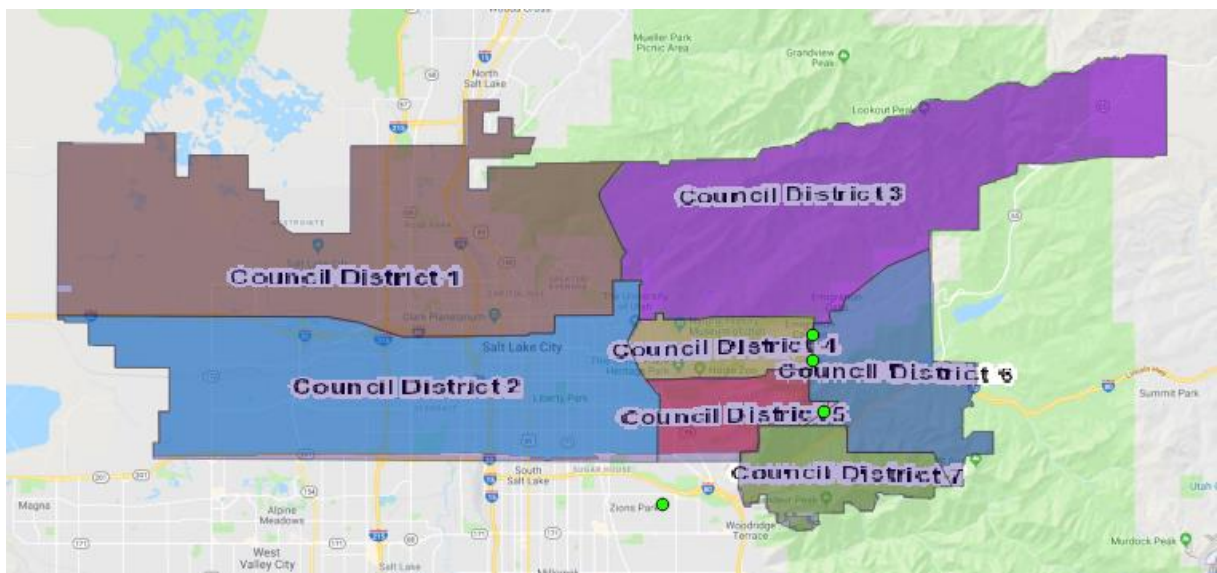
3. Ábrázolja a QGIS, hogy mely körzetekben fordult elő a sikkasztás bűnesete!



6. ábra

Mint az a képen pirossal lefedett területekből látható, a 2-es és 4-es körzetben történtek vélhetően ilyen bűnesetek.

4. 4. Mutassa meg a QGIS azokat az eseteket, amelyek egyik kerület területébe sem estek bele! (Vajon outlier vagy lefedési hiba okozza-e az ilyen pontok meglétét?)



7. ábra

Ahogy az az ábrán jól látható, 4 ilyen eset volt, abból 3 lehatárolási hiba, és egy outlier.

5. Mekkora a távolság méterben a legkeletibb és legnyugatibb eset pontjai között?

A távolság ezen két pont között: 22257.7441777 m \approx 22.26 km.