

Twitter felhasználók kirándulási szokásai

Szeiler Pál

Témavezető: Béres Ferenc, Molnár András József, Benczúr András

1 Bevezetés

Témavezetőm, Béres Ferenc, 2020 augusztus óta gyűjt geolokációval ellátott, kiránduláshoz köthető tweeteket. A félev során ezen adathalmaz megismerése, tisztítása és elemzése volt a feladatomban. Az elmúlt években egyre több cikk készült arról, hogy miként lehetne az emberek social média tevékenységével egy-egy terület látogatottságát mérni. Ennek a módszernek vannak limitációi [2], de ahogy a [1] cikk is megállapítja, a social médián elérhető adatok elemzése könnyebben kivitelezhető, mint egy helyszíni felmérés.

2 Az adatok előkészítése

Az adathalmaz tisztítását a spam fiókok kiszűrésével kezdtem. Ehhez a 100 legtöbbet posztoló fiók közül kizártam a botokat és az irreleváns tartalmat sugárzó felhasználókat. A felhasználók mozgásának megismerésében nagy szerepe volt annak, hogy a Twitteren minden usernek lehetősége van szöveges formában megadni a tartózkodási helyét (otthonát). Ezekhez a szöveges leírókhöz a Python wikipedia csomagjával rendeltem koordinátát. Előfordult, hogy egyes felhasználók érvénytelen lokációt adtak meg, vagy egyáltalán nem adtak meg tartózkodási helyet a fiókjukhoz. Így végül a userek 72,2%-hoz sikerült koordinátával ellátott tartózkodási helyet rendelni. Ezen felhasználók eloszlása a Twitter népszerűségének köszönhetően országokként jelentősen eltér (1. Ábra).

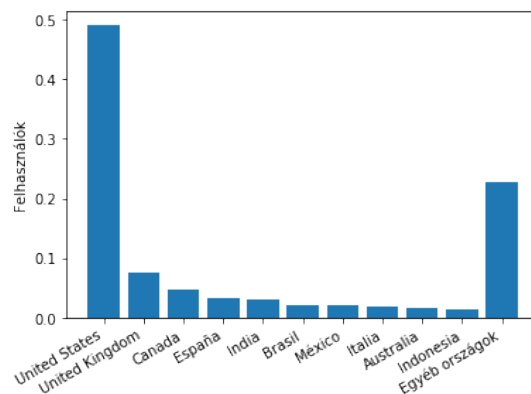
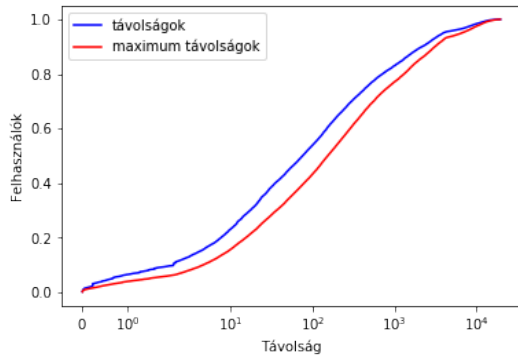


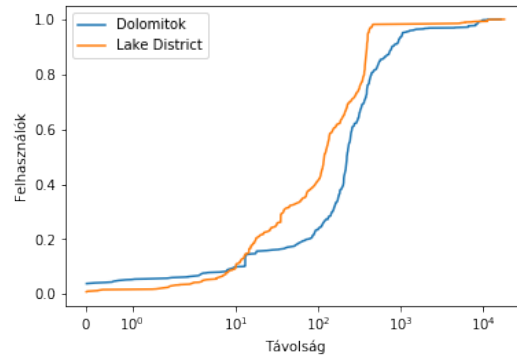
Figure 1

3 A távolságok eloszlása

Az érvényes lakhellyel rendelkező felhasználókhöz kiszámoltam, hogy mekkora távolságot tesznek meg egy-egy posztért. A 2.a ábrán jelenítem meg ezt a távolságeloszlást minden posztra (kék), illetve a felhasználók maximális poszt távolságára megszorítva (piros). A távolságeloszlások kiszámításához 83779 felhasználó 285302 tweetjét használtam fel.



(a) A felhasználók által megtett távolságok eloszlása.



(b) A Lake District és a Dolomitokhoz érkező túrázók távolságeloszlása a lakóhelyüktől.

Figure 2: Távolságeloszlások az adatban

Kiválasztottunk továbbá két olyan helyszínt, ahol megvizsgáltam, hogy milyen messziről érkeznek a látogatók (2.b ábra). A Lake District nemzeti park az Egyesült Királyságban található és több mint 1100 tweet készült itt az adatgyűjtés alatt. Az ide látogatók 96%-a lakik az Egyesült Királyságban. A Dolomitok Olaszországban található. Ide a felhasználók jelentős része Európából jött, az összes ide látogató 71%-a pedig belföldről. Innen nagyjából 1200 tweet érkezett. Nem mehetünk el amellett, hogy az adatgyűjtés a Covid alatt történt, így a lezárások jelentősen befolyásolhatták az utazási szokásokat.

4 További tervek

Az adatok több mint 1 évet ölelnek fel. Érdekes lenne megnézni, hogyan hatottak a Covid miatti lezárások a turizmusra, így szeretnénk az adatoknak az időbeli eloszlását is jobban megismerni. Ezen kívül a kirándulóhelyek jellege is különböző, egyes pontokat például nehezebb megközelíteni, mint másokat, ezért a tweetek eloszlását ilyen szempontból is meg fogjuk vizsgálni.

Irodalomjegyzék

- [1] Vuokko Heikinheimo et al. “User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey”. In: *ISPRS International Journal of Geo-Information* 6.3 (2017). ISSN: 2220-9964. DOI: 10.3390/ijgi6030085. URL: <https://www.mdpi.com/2220-9964/6/3/85>.
- [2] Emily J. Wilkins, Spencer A. Wood, and Jordan W. Smith. “Uses and limitations of social media to inform visitor use management in parks and protected areas: a systematic review”. In: *Environmental Management* 67.1 (2021), pp. 120–132. DOI: 10.1007/s00267-020-01373-7. URL: <https://doi.org/10.1007/s00267-020-01373-7>.