Bike sharing temporális adatok elemzése

Házi feladat specifikáció

Mérési útmutató

Készítette: Csilling Tamás U4E1EQ

Verzió: 1.0

2022

Budapesti Műszaki és Gazdaságtudományi Egyetem

Méréstechnika és Információs Rendszerek Tanszék

# Elemzés célja

A modern városok egyik legkörnyezetbarátabb közlekedési eszköze a bicikli. Ennek a közlekedési formának az elterjedését segítik a bicikli megosztó szolgáltatások, amik a kevésbé gyakorlott biciklisták számára is elérhetővé teszik ezt a közlekedési formát.

Ezek a szolgáltatások már egy ideje elérhetőek, így hosszabb időszakok is elérhetőek elemzési célokra. Az adatok alapján meg lehet határozni egy modellt a biciklizési szokásokról, ami segíthet a vásárlói igényeknek való megfelelésben, és az igények formálásában, a biciklik fejlesztési irányainak meghatározásában.

Az adatkészlet, amit választottam a 2015 óta történt összes bérlést leíró adat, illetve egy histórikus időjárás adat, amit fizikai mérőállomások Nearest Neighbour interpolációja alapján készítettem.

Az irodalomkutatás alkalmával számos regressziós modellt láttam, ami azt sejteti, hogy bizonyos paraméterek között lineáris kapcsolat van. Ezeknek a kapcsolatoknak a megmutatása, vizuális módszerekkel történő feltérképezése a házi feladat célja, illetve a bérlési adatok alapján idősorok elemzése, és vizsgálata.

# Kiválasztott adatkészlet

Az adatkészlet több adatforrásból származik, a fő adatforrás a Londoni közlekedési intézet adatai, és a Meteostat nyilvános időjárás adatai. Az adatok:

* Transport for London ( továbbiakban TfL) [1]:

A biciklihasználati adatok a TfL honlapján nyilvánosan elérhetőek, egy publikus S3 bucketből egyszerűen letölthetőek. Az adatok a TfL saját licencee alá esnek. Az adat a 2015 és 2022 naptári években között történt összer bérlést tartalmazzák több kevesebb hibával.

* meteostat.com [2]:

Publikus időjárási adatok, amikhez egy remek API is tartozik., python wrapperrel[3] Előnye a jellemzően használt meteorológiai modellekkel[4] szemben, hogy valós mért adatokból interpolál, így például a csapadékmennyiségre egy sokkal pontosabb, és histórikusan megbízhatóbb adatot ad. Az adatok a  Creative Commons Attribution-NonCommercial 4.0 International Public License (CC BY-NC 4.0)[5] licence alá esnek.

A bikesharing adatkészlet nagyságrendileg . 730.000.000 sorral rendelkezik, és összesen 9 oszloppal, míg az időjárási adat nagyjából 3000 sorral és 9 oszloppal büszkélkedhet.

# Adatkészlet alapvető statisztikai kiértékelése

BikeSharing:

Numerikus változók:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Leírás | mean | std | min | 25% | 50% | 75% | max |
| Duration | Bérlés hossza [s] | 1293.36 | 5567.72 | -3240 | 480 | 840 | 1320 | 1783920 |
| End Date | Bérlés vége | - | - | 2015.  01.02 | - | - | - | 2022.  12.08 |
| Start Date | Bérlés kezdete | - | - | 2015.  01.02 | - | - | - | 2022.  12.08 |

Kategórikus változók:

|  |  |  |
| --- | --- | --- |
| Leírás | típus | értékkészlet |
| Rental ID | uuid | R |
| Bike ID | uuid | R |
| EndStation ID | uuid | R, ~850 különböző |
| EndStation ID | szöveg | ~950 különböző |
| StartStation ID | uuid | R, ~850 különböző |
| StartStation ID | szöveg | ~950 különböző |

Időjárás:

Numerikus változók:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Név | Leírás | mean | std | min | 25% | 50% | 75% | max |
| tavg | Átlag hőmérséklet | 11.71 | 5.4 | -3.7 | 7.6 | 11.6 | 15.9 | 27.8 |
| tmin | Min hőmérséklet | 8.5 | 4.97 | -5.7 | 4.9 | 8.5 | 12.2 | 26.7 |
| tmax | Max hőmérséklet | 15.0 | 6.29 | -1.9 | 10.2 | 14.8 | 19.9 | 35.5 |
| prcp | Csapadék [mm] | 1.7 | 3.7 | 0 | 0 | 0 | 1.5 | 48 |
| snow(\*) | Hó [mm] | - | - | 0.0 | 0 | 0 | 0 | 20 |
| wdir | Szélirány [°] | 197.7 | 92.9 | 0 | 132 | 224 | 260 | 359 |
| wspd | Szélsebesség[km/h] | 14.9 | 5.8 | 1.7 | 10.7 | 14 | 18.2 | 45.9 |
| wpgt | Széllökés [km/h] | 32.8 | 10.7 | 13 | 25.9 | 31.5 | 38.9 | 98.2 |
| pressure | Légnyomás [Pa] | 1015 | 10.4 | 974.4 | 1009.7 | 1016.9 | 1022.8 | 1047.3 |

Kategórikus változók:

Maga a hó egy kategórikus változónak is tekinthető a (0,10,20) értékkészlettel.

# Kérdések/célok

Megfigyelhető-e összefüggés különböző időjárási körülmények, és a biciklik bérlésének gyakorisága között?

Lehet-e összefüggést kimutatni az évszakok változása között?

Milyen trendek figyelhetőek meg globálisan a bérlés változásában?

Milyen trendek figyelhetőek meg egy/két kiválasztott állomás bérléseiben?

A kiemelkedően jó időjárási körülmények között szignfikkánsan többen bicikliznek-e?

Milyen felhasználói csoportok azonosíthatóak a bérlések alapján?

(Extra: kimutatható-e amortizáció, biciklik állapota az adatok alapján?)

# Vizualizáció jellege, technológiája

A cél egy riport készítése, ami meghatározza, hogy a biciklik kölcsönzésében melyek a meghatározó faktorok, ami alapján például a biciklik ahsználtságára következtetni érdemes.

Ehhez egy jupyter notebookot szeretnék használni, Python környezettel, ahol a report ismeretterjesztő jelleggel a notebookban elhelyezett Markdown leírások között, a forráskódot is bemutatva szemlélteti a fakrotok jelentőségét. A vizualizációk kialakításához Seaborn[6], Plotly[7] plotokat illetve Prophet[8] könyvtárakat szeretném használni.

# Irodalomjgyzék

[1] „cycling.data.tfl.gov.uk”. https://cycling.data.tfl.gov.uk/ (elérés 2022. október 23.).

[2] „Formats & Units | Meteostat Developers”. https://dev.meteostat.net/formats.html#meteorological-parameters (elérés 2022. december 7.).

[3] „Python Library | Meteostat Developers”. https://dev.meteostat.net/python/#example (elérés 2022. december 7.).

[4] M. Giusti, „ERA5-Land”, *ECMWF*, 2020. február 4. https://www.ecmwf.int/en/era5-land (elérés 2022. december 7.).

[5] „Creative Commons — Attribution-NonCommercial 4.0 International — CC BY-NC 4.0”. https://creativecommons.org/licenses/by-nc/4.0/legalcode (elérés 2022. december 7.).

[6] „Overview of seaborn plotting functions — seaborn 0.12.1 documentation”. https://seaborn.pydata.org/tutorial/function\_overview.html (elérés 2022. december 7.).

[7] „Plotly Python Graphing Library”. https://plotly.com/python/ (elérés 2022. december 7.).

[8] „Prophet | Forecasting at scale.” https://facebook.github.io/prophet/ (elérés 2022. december 7.).