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 leginkább az időjárás, illetve naptári és évszaki időszakok, és a bicikli kölcsönzési kedv közötti kapcsolat leírását teszi lehetővé. Az adatkészlet óránkénti kummulált bérléseket mutatja, így geolokációs elemzésekre nem alkalmas. Nem tartalmazza az egy óránál hosszabb utakat, így természetes szűrést kaptunk a kiugró esetekre, mivel a bikesharing jellemzően rövid, városi távokra alkalmas.

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.

# Kiválasztott adatkészlet

Az adatkészlet több adatforrásból összeállított, előzetesen tisztított adatsor, amit a Londoni közlekedési intézet adatai, nyilvános időjárás adatok, és a brit ünnepnapok listájának segítségével alakítottak ki. Az adatkészletet a Kaggle honlapjáról [1] töltöttem le, az eredeti források a következőek:

* Transport for London ( továbbiakban TfL) [2]:

A biciklihasználati adatok a TfL honlapján nyilvánosan elérhetőek, publikus API-k is léteznek hozzá. Az adatok a TfL saját licencee alá esnek. Az adat a 2015 és 2016 naptári években történt összes bérlések számának óránkénti aggregációját mutatja.

* freemeteo.com [3]:

Publikus időjárási adarok amiknek a licence a Kaggle adatból nem derült ki. Várhatóan ingyenesen elérhető, és szabadon felhasználható adatok.

* Gov.uk bank-hoiday [4]:

A Brit kormány hivatalos oldalán szabadon elérhető adatok, nincsenek licenszhez kötve, letölthető naptár export formájában.

Az adatkészlet nagysgrendileg 17.000 sorral rendelkezik, és összesen 10 oszloppal, amiből 4 időjárást leíró adat, 3 pedig naptári eseményeket jelölő kategórikus adat.

# Adatkészlet alapvető statisztikai kiértékelése

Mivel az adatszett tisztított, ezért az adatok mindenhol teljesek, nincsenek hiányzó adatpontok.

Numerikus változók:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Leírás | mean | std | min | 25% | 50% | 75% | max |
| Bérelt biciklik száma (diszkrét értékek) | 1143.10 | 1085.10 | 0.0 | 257.0 | 844.0 | 1671.75 | 7860.0 |
| Hőmérséklet | 12.466 | 5.57 | -1.5 | 8.0 | 12.5 | 16.0 | 34.0 |
| Hőmérséklet érzés | 11.52 | 6.61 | -6.0 | 6.0 | 12.5 | 16.0 | 34.0 |
| Páratartalom | 72.32 | 14.31 | 20.5 | 63.0 | 74.5 | 83.0 | 100.0 |
| Szélerősség | 15.91 | 7.89 | 0.0 | 10.0 | 15.0 | 20.5 | 56.5 |

Kategórikus változók:

|  |  |  |
| --- | --- | --- |
| Leírás | típus | értékkészlet |
| Szünnap? | boolean | True/False |
| Időjárás kód | kategórikus | 8 kategória |
| Hétvége? | Boolean | True/False |
| Évszak | kategórikus | 4 kategória |

# Kérdések/célok

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

Van-e kimutatható részeg biciklizés péntek, szombat és vasárnap éjszakák után?

Lehet-e összefüggést kimutatni az évszakok változása között?

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

Igazolható hogy a temperature feel modell lineárisan függ a humidity, temperature, wind speed változóktól, mivel definíció szerint ezeknek lineáris kombinációjából áll elő.

# 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+HIP könyvtárakat szeretém használni.

# Irodalomjgyzék

[1] „London bike sharing dataset”. https://www.kaggle.com/datasets/hmavrodiev/london-bike-sharing-dataset (elérés 2022. október 23.).

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

[3] „freemeteo.hu”. https://freemeteo.hu/ (elérés 2022. október 23.).

[4] „UK bank holidays”, *GOV.UK*. https://www.gov.uk/bank-holidays (elérés 2022. október 23.).