



A MOL Bubi „közbringa” rendszer forgalmi adatainak adatbányászati és térinformatikai elemzése üzemeltetési és fejlesztési szempontból



Béres Orsolya Katalin
2018. január

Konzulens:
Dr. Tettamanti Tamás



Tartalom



- A „közbringa” rendszerekről általánosságban
- A MOL Bubi rendszer
- Az adatbányászat módszere
- Felhasznált adatok és szoftverek
- Eredmények
- Az eredmények felhasználása



A közbringa rendszerekről általánosságban



**A városi
kerékpározás, mint
közlekedési forma
újra népszerű, káros
környezeti
hatásoktól mentes.**



**A „közbringázás” az egyik
legdinamikusabban fejlődő
közlekedési alternatíva.**

A MOL Bubi rendszer



A MOL Bubi közbringa rendszer...

- Budapest első, a nagyközönség számára használható közbringa rendszere.
- 2014. szeptember 8. óta üzemel.
- jelenleg 124 gyűjtőállomásból és 1486 darab kerékpárból áll.



Az adatbányászat módszere



Az adatbányászat...

- nagy mennyiségű adatban rejlő **összefüggés** gépi tanuló algoritmusok által feltárható.
- **trendeket** és **mintázatokat** keres.



A csoportosítás – klaszterezés...

- során a csoportképző algoritmusok olyan **csoportokat** hoznak létre, amelyekben az objektumok egymáshoz viszonyított **hasonlósága nagymértékű** és a **különböző csoportok hasonlóságának mértéke alacsony**.

Felhasznált adatok és szoftverek I.



Adatok

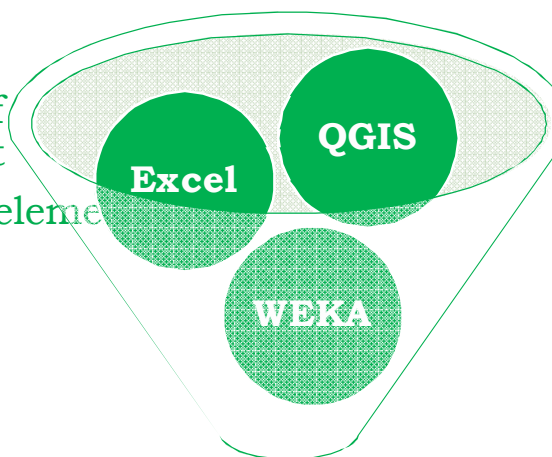


**Open Street Map
adatok**

Point of Interest

Vonali elemek

Felületi
elemek



**Archív budapesti
időjárás adatok**

Napi minimum,
maximum hőmérséklet

Időjárás jellege naponta

Szélsebesség

MOL Bubi adatok

Bérlési adatok (2015. év)

Regisztrált felhasználók adatai

Eredmények

Szám	Szám	Szám
100	1000000000	1000000000
101	1000000000	1000000000
102	1000000000	1000000000
103	1000000000	1000000000
104	1000000000	1000000000
105	1000000000	1000000000
106	1000000000	1000000000
107	1000000000	1000000000
108	1000000000	1000000000
109	1000000000	1000000000
110	1000000000	1000000000
111	1000000000	1000000000
112	1000000000	1000000000
113	1000000000	1000000000
114	1000000000	1000000000
115	1000000000	1000000000
116	1000000000	1000000000
117	1000000000	1000000000
118	1000000000	1000000000
119	1000000000	1000000000
120	1000000000	1000000000
121	1000000000	1000000000
122	1000000000	1000000000
123	1000000000	1000000000
124	1000000000	1000000000
125	1000000000	1000000000
126	1000000000	1000000000
127	1000000000	1000000000
128	1000000000	1000000000
129	1000000000	1000000000
130	1000000000	1000000000
131	1000000000	1000000000
132	1000000000	1000000000
133	1000000000	1000000000
134	1000000000	1000000000
135	1000000000	1000000000
136	1000000000	1000000000
137	1000000000	1000000000
138	1000000000	1000000000
139	1000000000	1000000000
140	1000000000	1000000000
141	1000000000	1000000000
142	1000000000	1000000000
143	1000000000	1000000000
144	1000000000	1000000000
145	1000000000	1000000000
146	1000000000	1000000000
147	1000000000	1000000000
148	1000000000	1000000000
149	1000000000	1000000000
150	1000000000	1000000000

Felhasznált adatok és szoftverek II.



Weka Explorer interface showing the results of a SimpleKMeans clustering algorithm. The interface includes a menu bar (Preprocess, Classify, Cluster, Associate, Select attributes, Visualize), a toolbar, and several panels.

Clusterer
Choose: SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 4 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S

Cluster mode
 Use training set
 Supplied test set
 Percentage split % 66
 Classes to clusters evaluation
(Nom) Cluster
 Store clusters for visualization
Ignore attributes
Start Stop

Clusterer output
Number of iterations: 8
Within cluster sum of squared errors: 4.378441485914202
Initial starting points (random):
Cluster 0: 7.934579, 3.715517, 11.650097, 2
Cluster 1: 2.306818, 2.103448, 4.410266, 16
Cluster 2: 3.177778, 2.15873, 5.336508, 4
Cluster 3: 3.706897, 3.708995, 7.415891, 3
Missing values globally replaced with mean/mode
Final cluster centroids:
Attribute Full Data Cluster#
 (98.0) (11.0) (13.0) (59.0) (15.0)

ejjel_avg_fel 3.3237 6.4667 3.7523 2.4359 4.1391
ejjel_avg_le 3.1862 5.2343 2.9868 2.6208 4.0805
ossz_ejjel 6.5098 11.701 6.7391 5.0568 8.2196
entertainment_sum 17.5714 25.8182 45.7692 10.2034 16.0667

Weka Clusterer Visualize: 19:10:10 - SimpleKMeans (entertainment_ejel)
X: ejjel_avg_fel (Num) Y: entertainment_sum (Num)
Colour: Cluster (Nom) Select Instance
Reset Clear Open Save Jitter

Plot: entertainment_ejel_clustered
A scatter plot showing the results of the SimpleKMeans clustering. The X-axis is labeled 'ejjel_avg_fel (Num)' and ranges from 1.36 to 9.33. The Y-axis is labeled 'entertainment_sum (Num)' and ranges from 0 to 72. The plot shows four distinct clusters of data points, each represented by a different color (green, red, blue, and purple). The clusters are well-separated, indicating good clustering performance.

Class colour
cluster0 cluster1 cluster2 cluster3

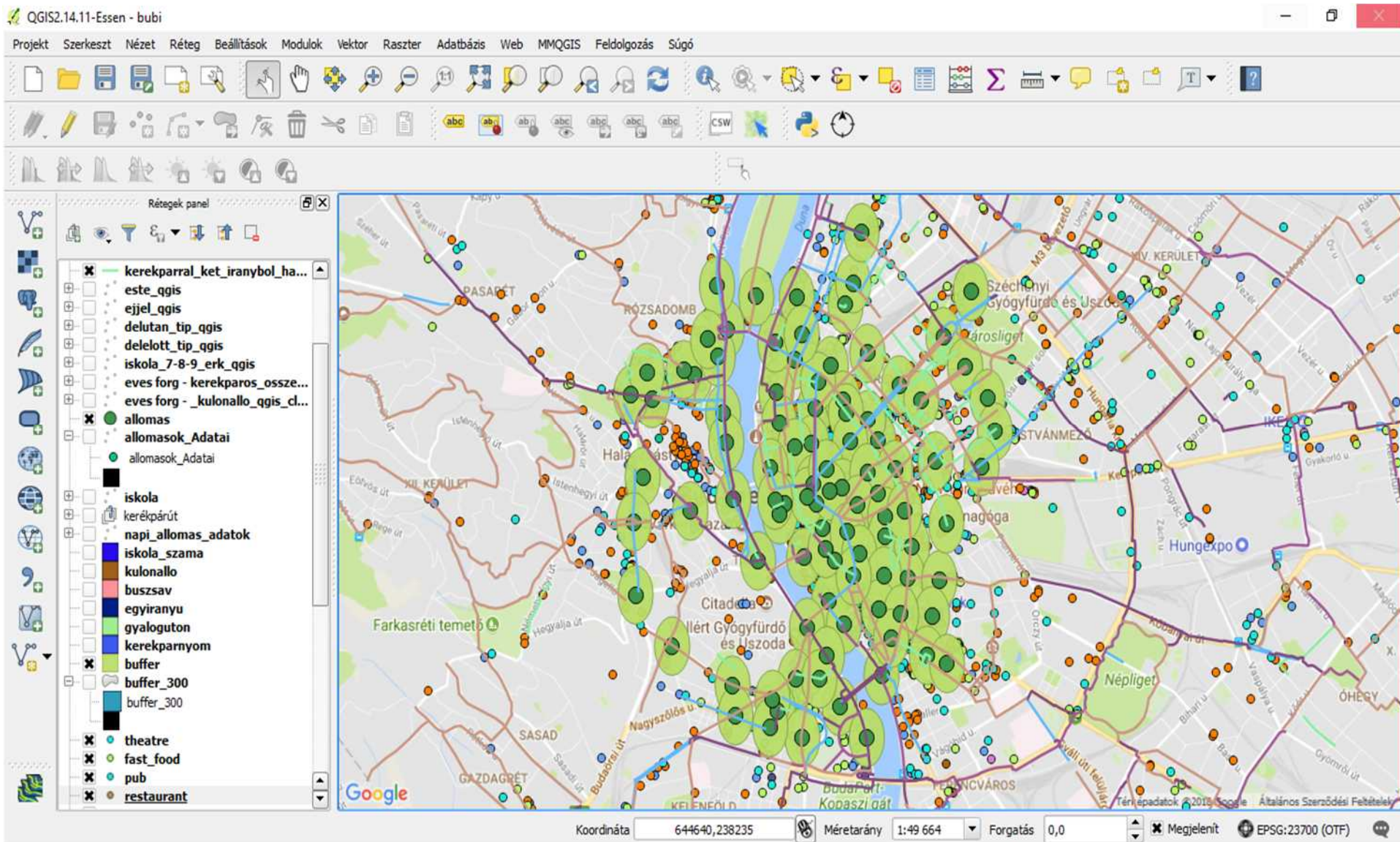
Result list (right-click for options)
19:05:27 - SimpleKMeans
19:05:42 - SimpleKMeans
19:05:56 - SimpleKMeans

Status
OK

K-means – k-átlag módszer

- a módszer alapját az adathalmaz elemeinek a távolságmértékei képzik.

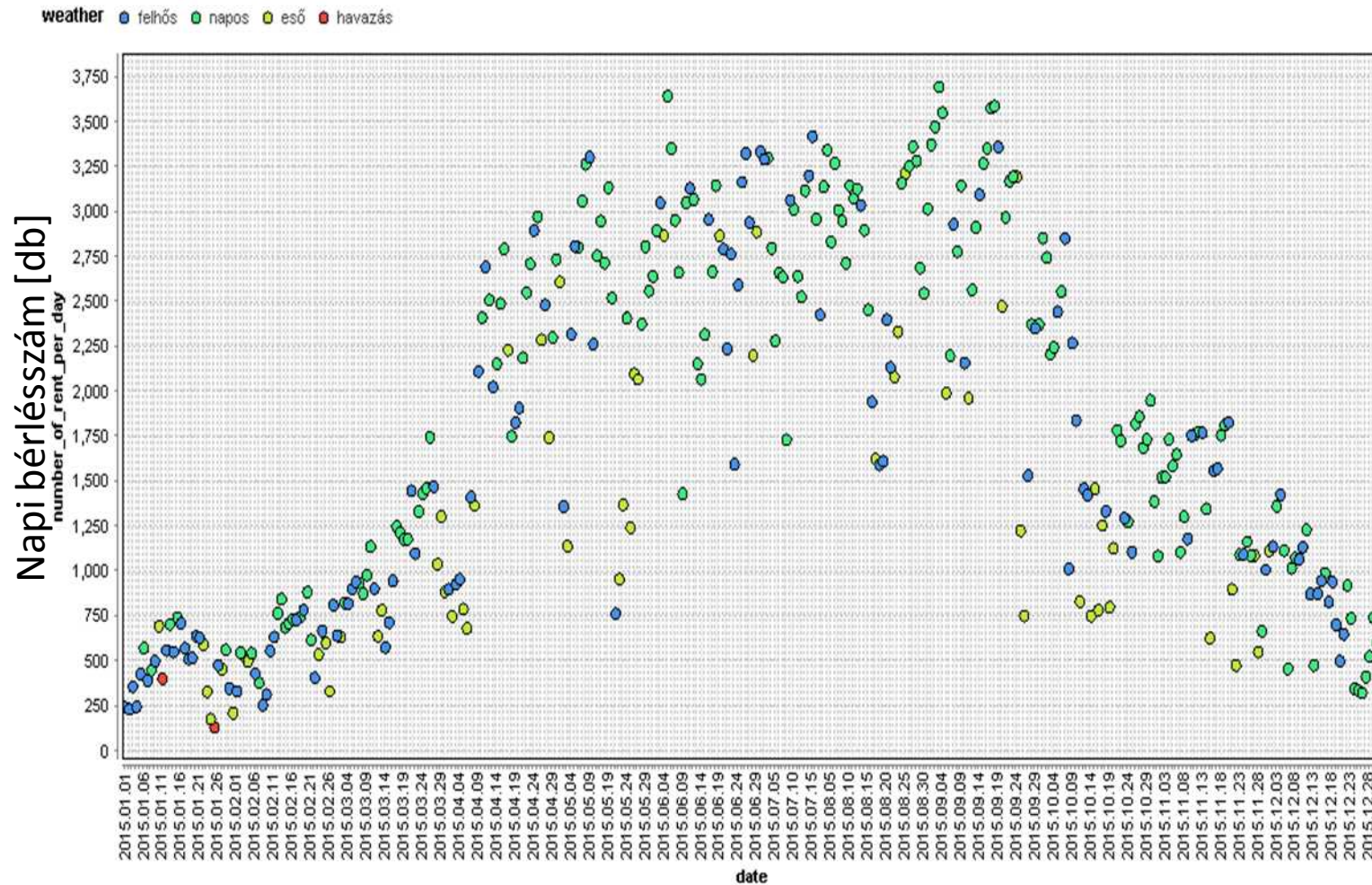
Felhasznált adatok és szoftverek III.



Eredmények I.



Időjárási jellegzetességek és a napi bérleti adatok összefüggése



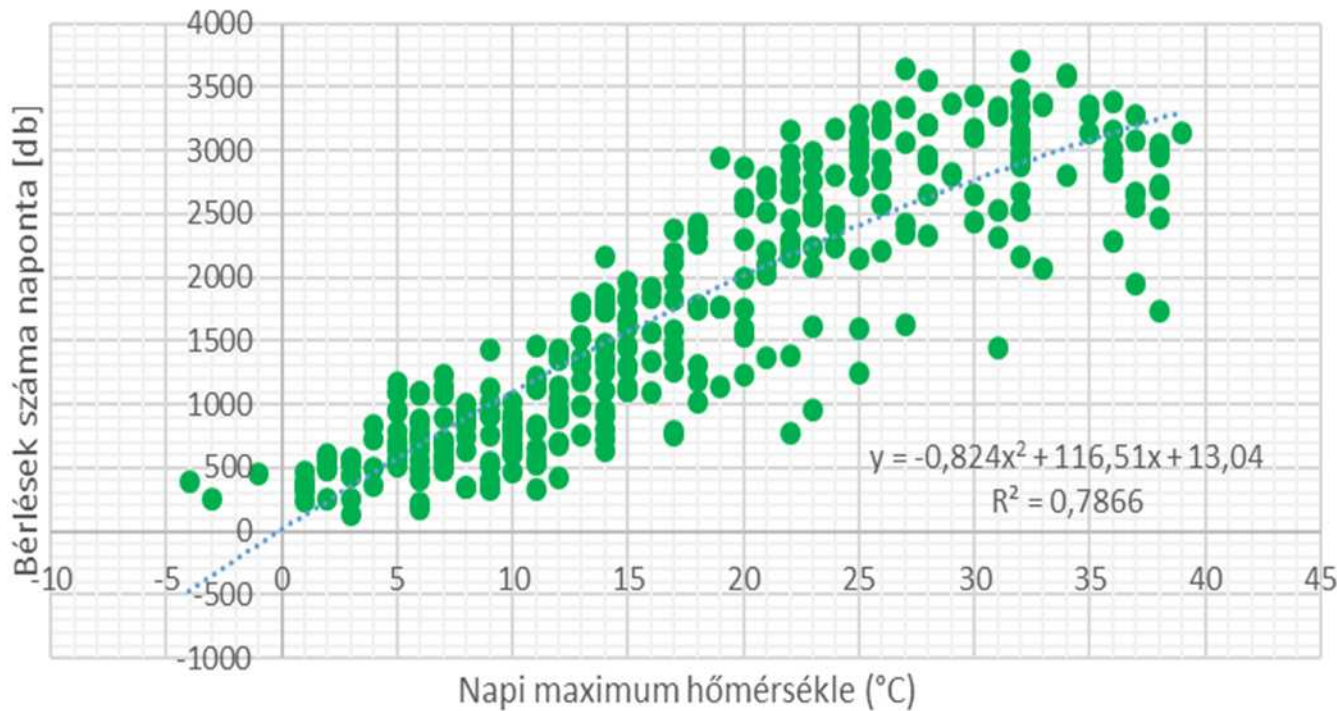
- Napos
- Felhős
- Esős
- Havazás

Eredmények II.



Napi maximum hőmérséklet és a napi bérlési adatok összefüggése

Napi maximum hőmérséklet és a napi bérlésszám összefüggése



Korrelációs együtthatók számítása

Pearson-féle együttható 0,8826

Spearman-féle együttható 0,8933

Regresszióvizsgálat

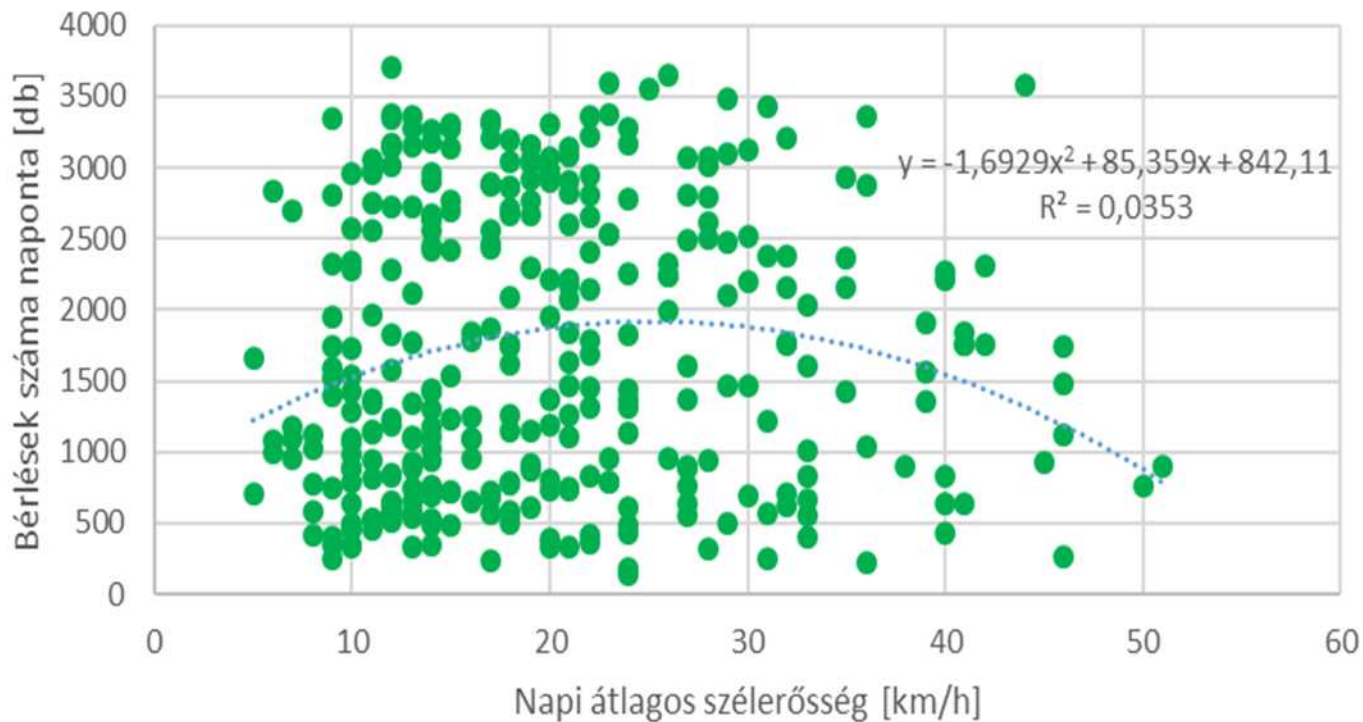
Regresszió	Egyenlet	R ²
Lineáris	$y = 85,274x + 219,67$	0,779
Nemlineáris - 2 Fokszámú	$y = -0,824x^2 + 116,51x + 13,04$	0,787
Nemlineáris - 3 Fokszámú	$y = -0,1687x^3 + 8,7052x^2 - 27,029x + 489,29$	0,825

Eredmények III.



Napi szélerősség érték és a napi bérlési adatok összefüggése

Napi átlagos szélerősség és a napi átlagos bérlésszám összefüggése



Korrelációs együtthatók számítása

Pearson-féle együttható 0,0275

Spearman-féle együttható 0,0807

Regresszióvizsgálat

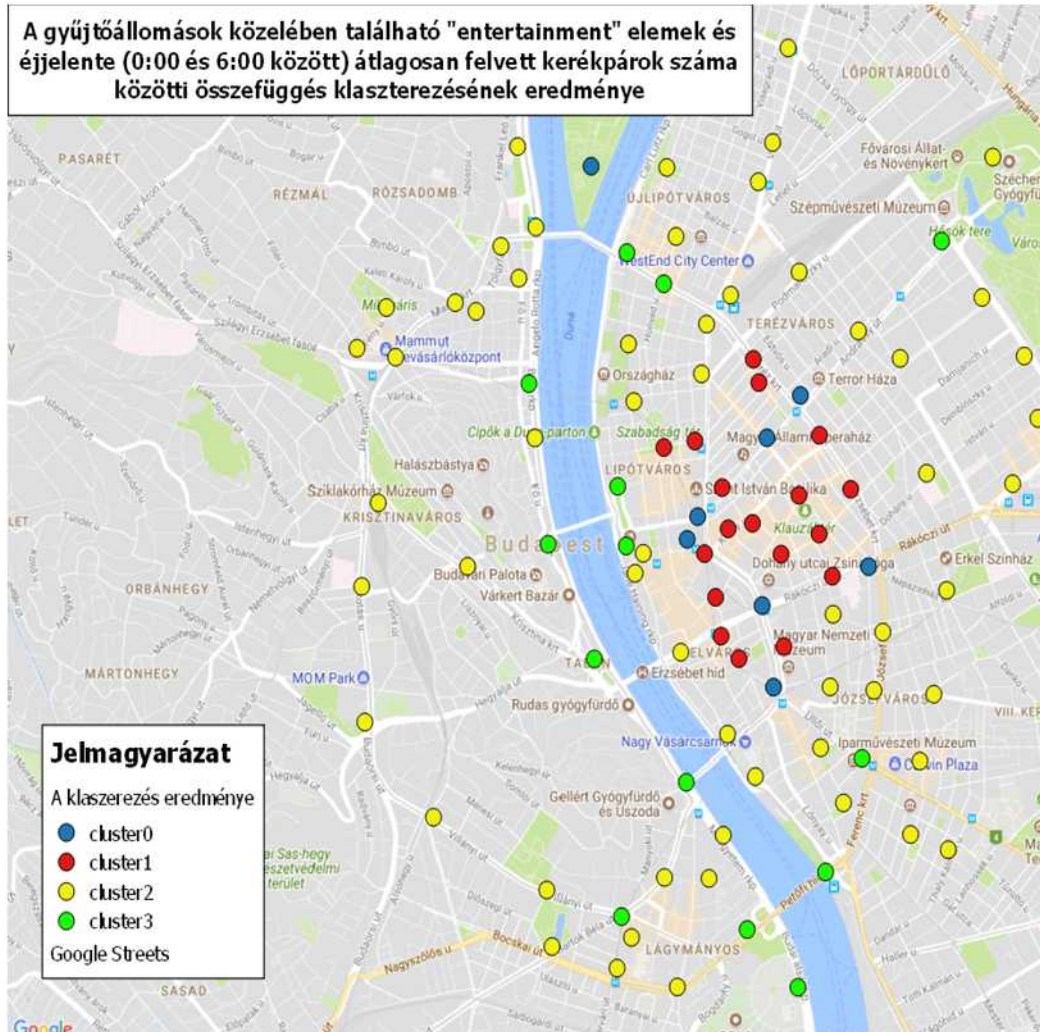
Regresszió	Egyenlet	R ²
Lineáris	$y = 2,898x + 1663,2$	0,001
Nemlineáris - 2 Fokszámú	$y = -1,6929x^2 + 85,359x + 842,11$	0,035
Nemlineáris - 3 Fokszámú	$y = 0,0379x^3 - 4,6393x^2 + 152,22x + 410,73$	0,038

Eredmények IV.



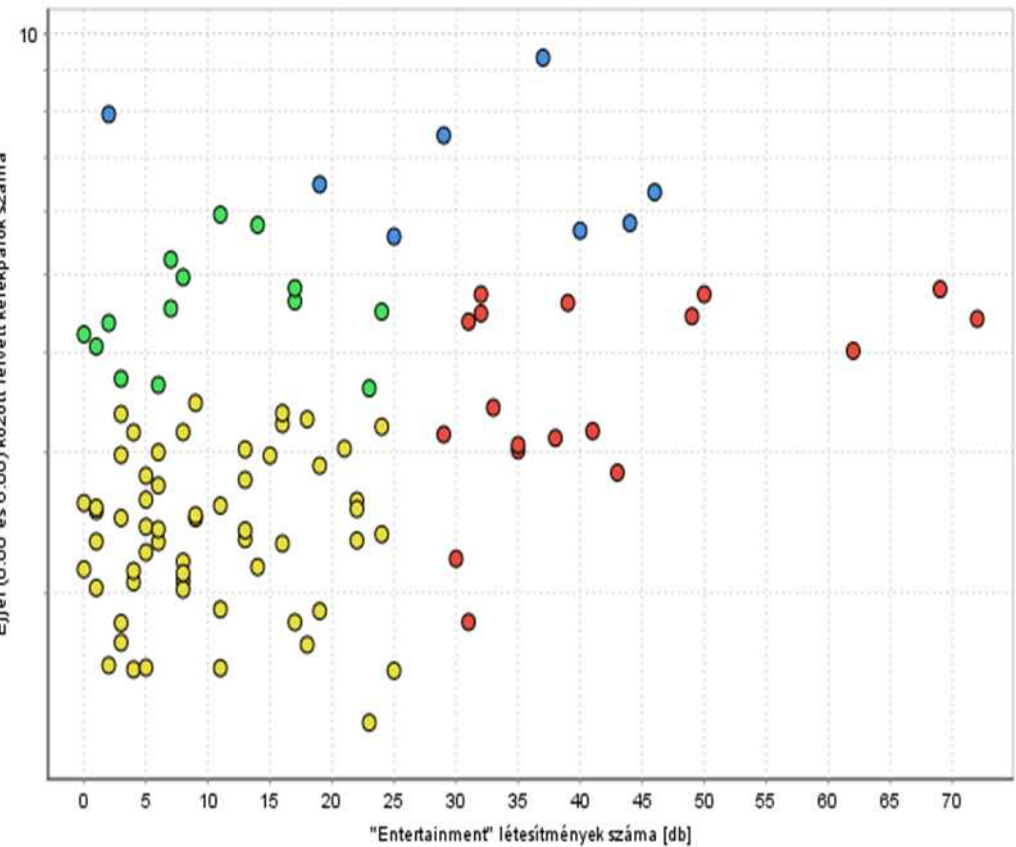
A gyűjtőállomások induló forgalma éjjel és a gyűjtőállomások közelében lévő szórakozóhelyek számának összefüggése

A gyűjtőállomások közelében található "entertainment" elemek és éjjelente (0:00 és 6:00 között) átlagosan felvett kerékpárok száma közötti összefüggés klaszterezésének eredménye



Éjjel indított bérletek száma [db]

Éjjel (0:00 és 6:00) között felvett kerékpárok száma

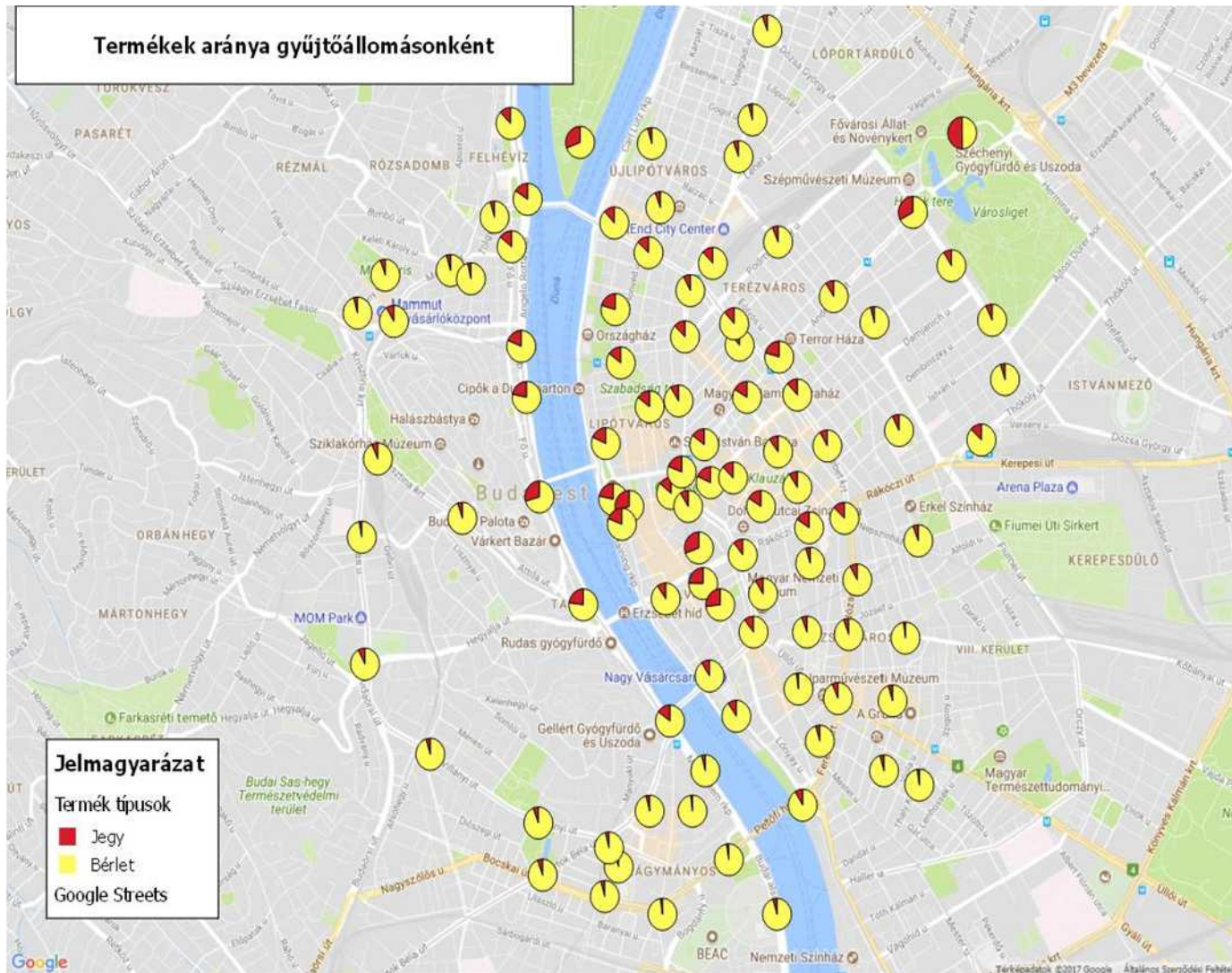


Szórakozóhelyek száma 300 méteres környezetben(db)

Eredmények V.



MOL Bubi termékek aránya gyűjtőállomásokonként



- Bérlet
- Jegy

Eredmények felhasználása



Forgalmi adatok és naptári, időjárási jellegzetességek összefüggése

Kerékpárok karbantartásának optimális ütemezése

Kerékpárok áthelyezési folyamatának optimalizálása

Bérlési adatokat előrejelző rendszer fejlesztése

Kerékpáros közlekedési szokások megismerése a fővárosban

Gyűjtőállomások viselkedéséből, elhelyezkedéséből származó összefüggések

Állomási kerékpárhiányok tervezhetősége

Logisztikai műveletek pontosabb tervezhetősége

Állomások működésének megfigyelésére monitorozó rendszer fejlesztése

Kerékpározási motivációk megismerése

Felhasználói adatokból származó összefüggések

Új gyűjtőállomások sikerességének megállapítása, megtérülés tervezése

Céltzott marketing tevékenységek

Turisztikai felhasználók tájékoztatása egyes állomásokon

Korosztályok kerékpározási szokásainak megismerése



**Köszönöm szépen a
figyelmet!**

