

Vasútmenetrend

keresések elemzése

Rendszer

- iMenetrend, TElvira
- kb 300.000 felhasználó
- naponta 25.000 keresés

Feladat

- az éppen kialakított reklámhálózat (adnetwork) finomítása
- felhasználói korrelációk felderítése

Adatbázis

- gyűjtés időpontja: 2015. 01. 24.
- ~2300 rekord
- honnan-hová adatok
- dátum
- távolság
- életkor (Google Adwords adat)
- eszköz típusa (android / ios)

Legkedveltebb induló állomás

Budapest	878
----------	-----

Győr	68
------	----

Vecsés	61
--------	----

```
as.data.frame(rev(sort(table(db$from)))[1:3])
```

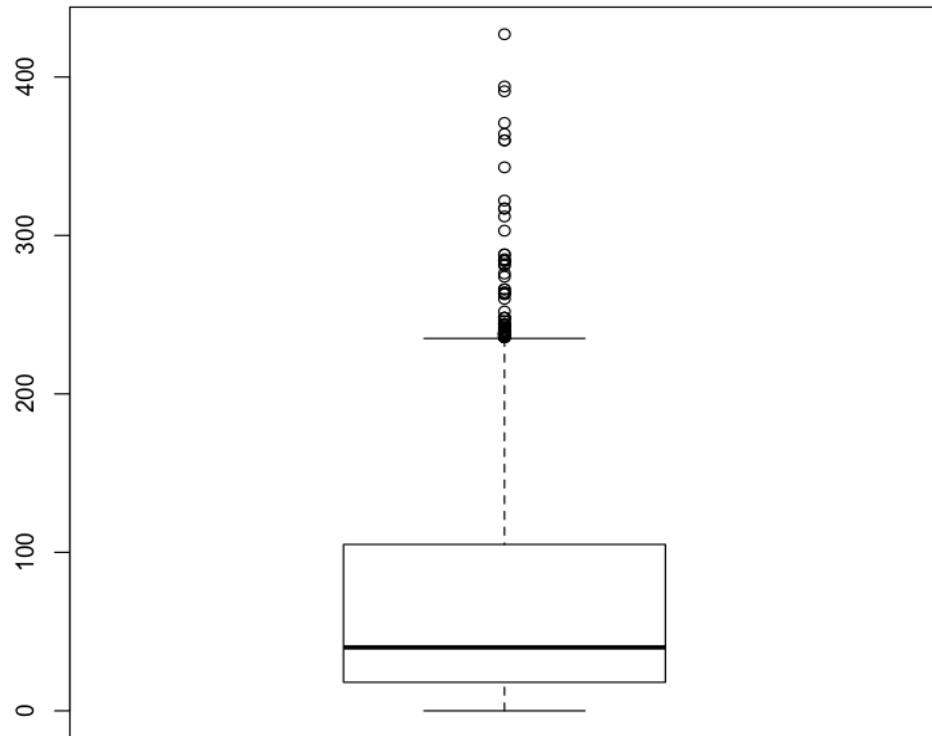
Legkedveltebb célállomás

Budapest	919
Győr	64
Debrecen	60

```
as.data.frame(rev(sort(table(data$to)))[1:3])
```

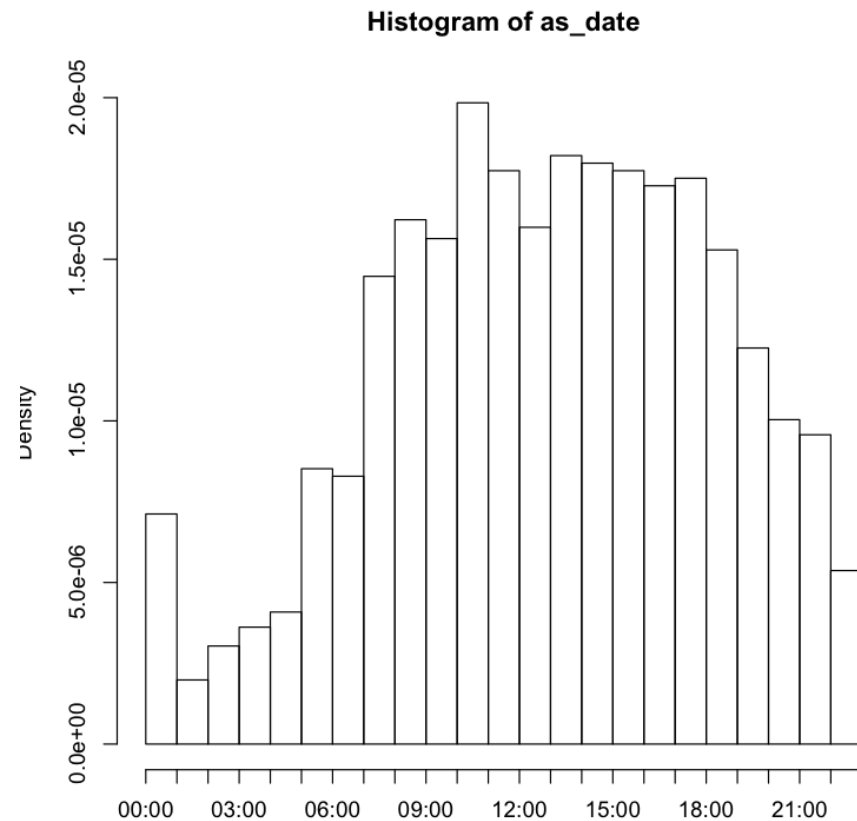
Outlierek száma

`boxplot(db$distance)`



Utazások eloszlása

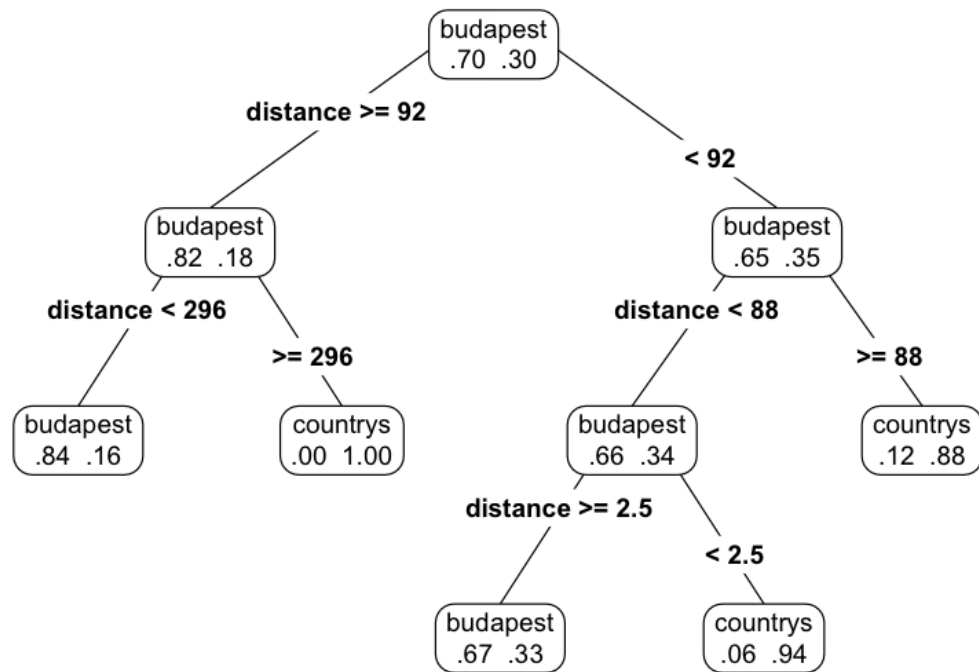
```
as_date <- as.POSIXlt(data$date, format="%Y-%m-%dT%H:%M:%S")  
hist(as_date, "hours")
```



Amit eddig láttunk:

- Budapest bázisú
- Sok szélsőérték, míg sokan főleg rövid távon utaznak
- Az utazások egész nap eloszlanak
- De hogy oszlanak el?

Budapest bázisú a közlekedés?



```
prp(rpart(countryside~distance_num,data=db
,control=rpart.control(cp=0,maxdepth=3)),type=4,extra=4)
```

További elemzések

- Eszköz típusa és az kiindulási állomásra illesztett regressziós modell
- $\text{Kor} \sim \text{Távolság} + \text{Eszköz-re illesztett regressziós modell}$ illetve az ezek közötti kapcsolat