Csillag Renáta
CSRRABT.SZE

Statisztikai szoftverek esszé

Csillag Renáta
2011.

Csillag Renáta
CSRRABT.SZE

Helyzetfelmérés

Egy internetszolgáltató egy havi adatforgalmát vizsgáltam. A táblázatok az előfizetők letöltési forgalmát tartalmazzák, napi bontásban, ügyfélkód alapján rendezve – az első táblázatban a hónap első felében lebonyolított letöltések, a másikban pedig a hónap második felének adatai, valamint az, hogy ki mekkora sávszélességre fizetett elő. A letöltések értékei megabyte-ban vannak megadva. Adataim 2006 - os évet reprezentálják.

Elemzéseimmel egy átfogóbb képet kapok az internetszolgáltató adatforgalmáról és az ügyfelek internetezési szokásáról, valamint a felhasználók számáról.

Követelményspecifikáció

Az elemzés céljai a következők:

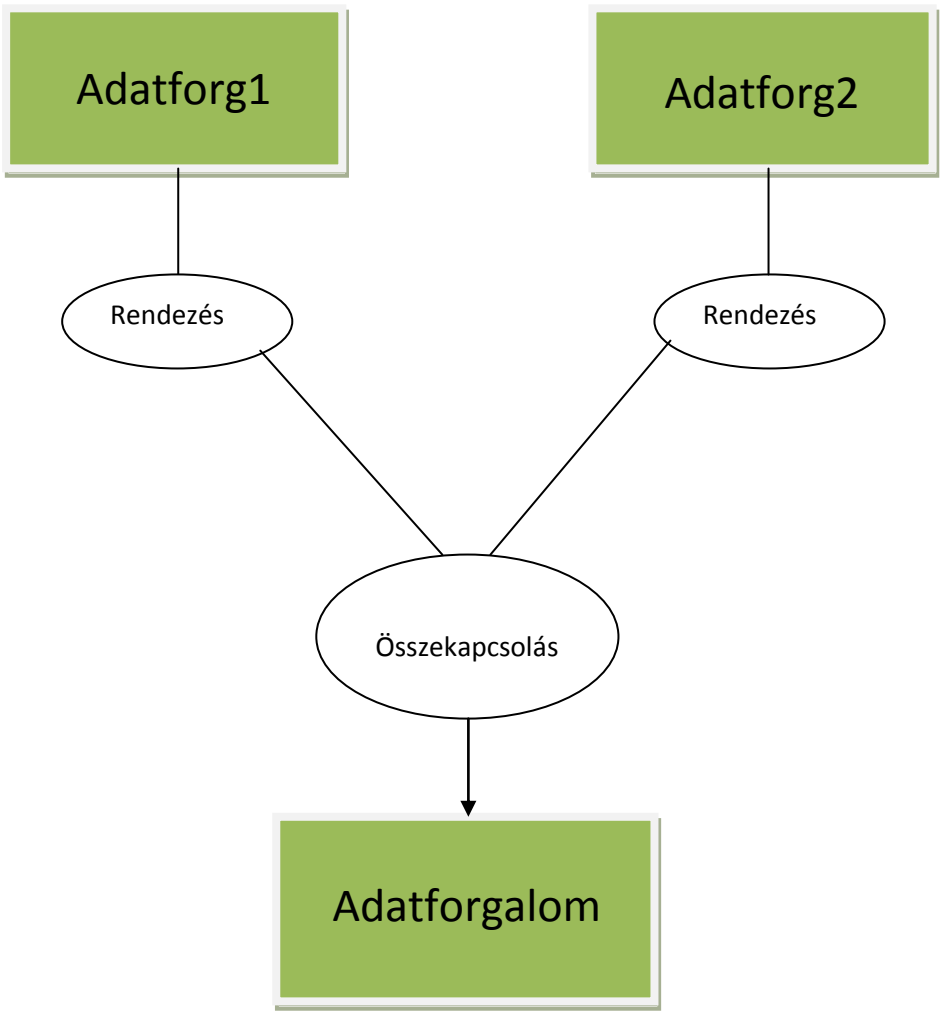
- Ügyfelek számának kiszámítása;
- Napi átlagok kiszámítása a hónap első felében;
- Napi összletöltések kiszámítása;
- Napi maximumok kiszámítása;
- Adott előfizető havi adatletöltésének kiszámítása;
- Adatkorlát-túllépések feljegyzése;
- Az egyes sávszélességeken belüli előfizetők száma.

Logikai rendszerterv

Első lépésként az Excel táblázatokat be kell importálnom a SAS programba, létrehozva ezzel az „Adatforg1” táblát, mely az egyes ügyfelek adatforgalmát tartalmazza napi bontásban az első 15 napon keresztül, majd az „Adatforg2” táblát, mely a következő 15 nap adatforgalmát és a sávszélességeket reprezentálja. (Az adatok MB-ban értendőek).

Az elemzések megkezdése előtt a két táblát ügyfélkód alapján rendeznem kell, hogy egyesíteni lehessen őket.

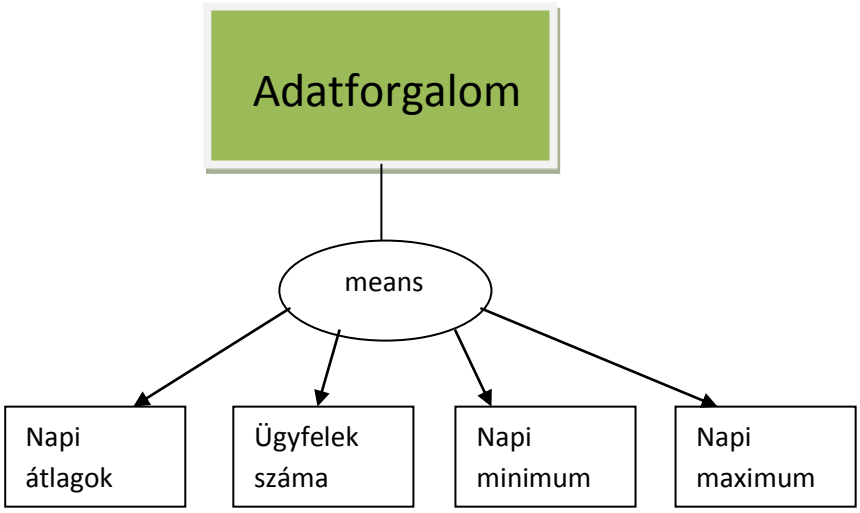
A rendezés után következhet az „Adatforg1” és az „Adatforg2” táblázatok összefésülése.



1.ábra
Az „Adatforgalom” tábla létrehozása

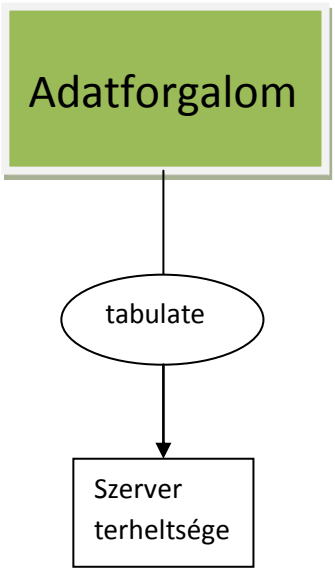
Ezen módosítások végrehajtása után már elvégezhetem a kitűzött elemzéseket.

Először kiszámítom, hány ügyfele van az internetszolgáltató cégnek, és naponta mekkora az átlag letöltési forgalom. Ezt úgy tehetem meg, ha egy napra kiszámítom az átlagot, és közben kiíratom a minták számát is. Ezzel a módszerrel megkapom a napi legkisebb, és legnagyobb letöltés méretét is.



2.ábra
Az első elemzések végrehajtása

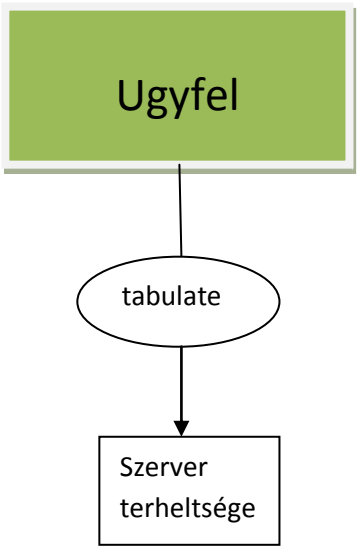
Ezután meghatározom a szerver napi terheltségét, melyet a napi forgalmak összeadásával kapok meg.



3.ábra
Második elemzés végrehajtása

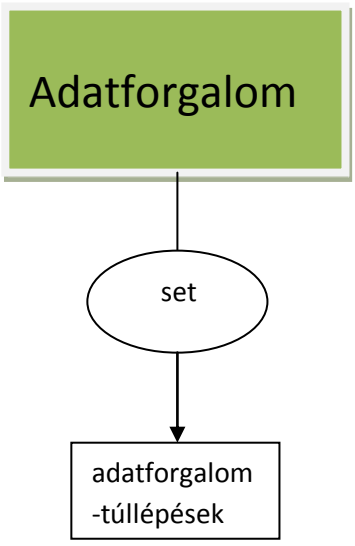
Ahhoz, hogy az egyes ügyfelek havi összletöltését is ki tudjam számolni, Excelben létre kell hoznom egy új táblázatot, ahol a változók már nem a napok, hanem az ügyfélkódok. De mivel körülményes lenne ennyi felhasználó adatforgalmát programmal kiszámolni, így csak néhány fő esetében hajtom végre az elemzéseket. A létrehozott „Ugyfel” táblázatot beimportálom a SAS programba, majd végrehajtom az elemzéseket.

Elsőként kiszámolom az öt kiválasztott előfizető havi adatforgalmát.



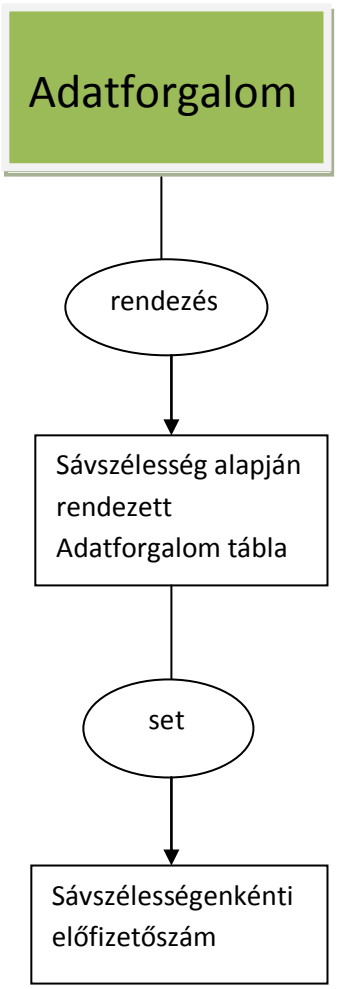
4.ábra
A szerver napi terheltségének kiszámítása

A szolgáltató egy 10 GB-os korlátot szabott meg a havi összletöltésekre. Az előző elemzésben megfigyelt öt előfizetőből is túllépte ezt a határt már két fő, és ha az előzőek szerint kiszámolom minden ügyfél adatforgalmát, akkor további kilenc ilyen esettel találkozunk. Ezen túllépéseket egy új oszlopban rögzítem az adatforgalom táblában.



5.ábra
A letöltési korlát túllépésének dokumentálása

Végezetül azt szeretném megtudni, hogy az egyes sávszélességekben hány előfizető van. Ehhez előbb sávszélesség alapján rendeznem kell az ügyfeleket, majd ezután kiszámolni ezen kulcs alapján, hogy hányan használják az adott sávszélességet.



6. ábra
Előfizetők az egyes csoportokban

Csillag Renáta
CSRRABT.SZE

Megvalósítások

Excel fájlok importálása:

File/Import data paranccsal

A két táblázat rendezése:

```
proc sort data = m.Adatforg1;
by Ugyfelkod;
run;


proc sort data = m.Adatforg2;
by Ugyfelkod;
run;
```

A két táblázat egyesítése:

```
data m.Adatforgalom;
merge m.adatforg1 m.adatforg2;
by ugyfelkod;
run;
```

Napi átlagok, és az ügyfelek számának kiszámítása:

```
proc means data = m.Adatforgalom;
var a1 ;
run;
```

Az eredményül kapott táblázat:

Analysis Variable : a1				
N	Mean	Std Dev	Minimum	Maximum
100	279.6000000	163.0078389	0	589.0000000

A szerver terheltsége:

```
title 'A szerver leterheltsége a hónap elso napján' ;
proc tabulate data = m.Adatforgalom;
class Ugyfelkod;
var a1;
table all, a1, sum;
run;
```

Az eredményül kapott táblázat:

A szerver leterheltsége a hónap első napján

	Sum
a1	27960.00

A kiválasztott ügyfelek havi adatforgalma:

```
proc tabulate data = m.Ugyfel;
class nap;
var k5067666678;
table all, k5067666678, sum;
run;
```

Csillag Renáta
CSRRABT.SZE

Az eredményül kapott táblázatok:

	Sum
k5067666678	10466.00

	Sum
k2444950029	9443.00

	Sum
k464968584	11080.00

	Sum
k472041234	9392.00

	Sum
k170966016	9859.00

Csillag Renáta
CSRRABT.SZE

A túllépések dokumentálása:

```
data m.Adatforgalom2;
set m.Adatforgalom;
if ugyfelkod=3368856119 or
ugyfelkod=2816874836 or
ugyfelkod=664356754 or
ugyfelkod=6508228396 or
ugyfelkod=8315080518 or
ugyfelkod=4342413366 or
ugyfelkod=7111476821 or
ugyfelkod=5214755655 or
ugyfelkod=464968584 or
ugyfelkod=460121358 or
ugyfelkod=7381415400
then korlat='tullepte';
else korlat='ok';
run;
```

Az eredményül kapott táblázat részlete:

Ugyfelkod	korlat
6508228396	tullepte
7111476821	tullepte
7247419316	ok
7316788016	ok
7381415400	tullepte
7784955070	ok
7805534910	ok

Csillag Renáta
CSRRABT.SZE

Az egyes sávszélességek előfizetőinek száma:

```
PROC SORT DATA=m.Adatforgalom2 ;
          BY savszelesseg ;

DATA m.adatforgalom3(KEEP=Savszelesseg uszam)
         SET adatforgalom2 ;
BY Savszelesseg ;
IF FIRST.Savszelesseg  THEN  uszam=0 ;
uszam+1 ;
IF LAST.Savszelesseg ;
RUN;
PRINT ;
RUN ;
```

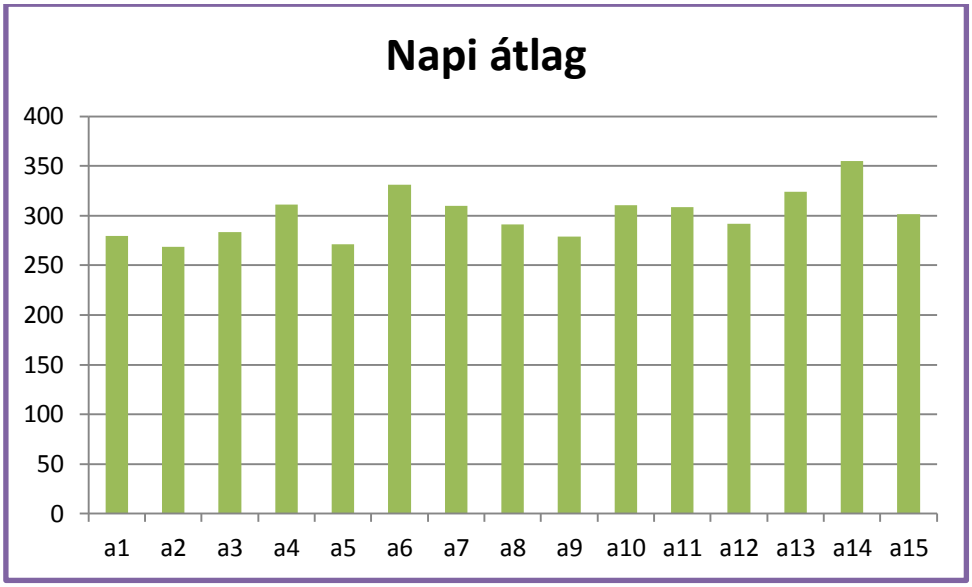
Az eredményül kapott táblázat:

Obs	Savszelesseg	uszam
1	256	43
2	512	32
3	1000	25

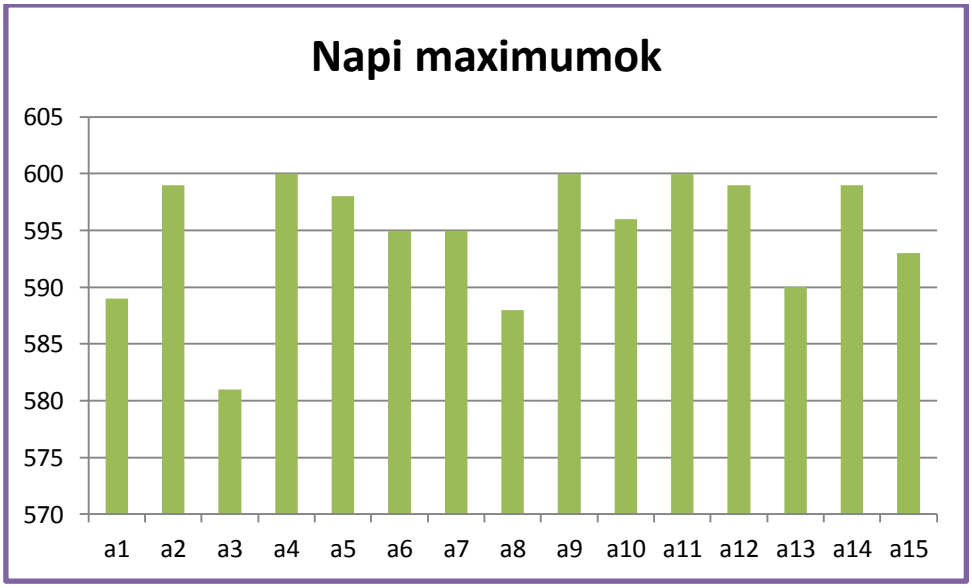
Csillag Renáta
CSRRABT.SZE

Összegzések:

Az első elemzésből megtudtam, hogy az adott internetszolgáltató 100 ügyféllel rendelkezik az elemzett hónapban. A letöltések napi átlagáról készítettem egy oszlopdiagramot.



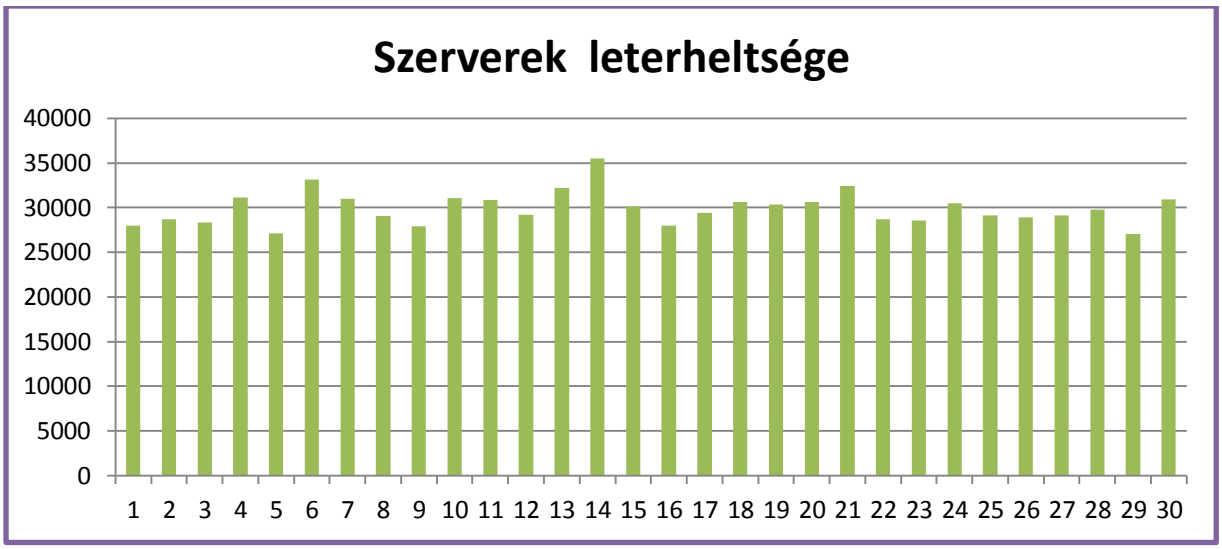
A diagram alapján kiderül, hogy vizsgált hónap első tizenöt napja közül a tizennegyedik napon volt a legnagyobb az adatforgalom átlaga, azonban nem kimagaslóan haladta meg a többi nap átlagát.
A napi maximum letöltéseket is diagramon szemléltetem.



A legnagyobb maximum a kilencedik és a tizenegyedik napon volt, 600 MB, amely csupán egy MB – tal haladja meg a diagramon többször is megjelenő 599 MB – os maximumot.

Csillag Renáta
CSRRABT.SZE

A következő elemzésem célja a szerver terheltségének vizsgálata volt. Az egyes napok adatforgalmát egy újabb oszlopdiagramon adom meg.



A terheltség a tizennegyedik napon érte el maximumát, a 35 532 MB – ot. Az átlag is ezen a napon volt a legmagasabb, és a maximumok közül a második adat is ezen a napon mutatkozott.

Az utolsó elemzésemben kiszámoltam, hány előfizetője van a szolgáltatónak az adott sávszélességeknél. A legtöbben (43 fő) – a felhasználók 43 %, azaz majdnem a fele - választotta a legkisebb szélességet. Ennek oka minden bizonnyal az, hogy a nagyobb sávszélesség nagyobb havidíjat von maga után, és hiába biztosít jobb internetezési lehetőségeket, a magas ár sok embert visszatart ennek a sávszélességnek választásától. Az is látható továbbá, hogy minél nagyobb szélességű az internetsáv, annál kevesebb az előfizető, ami szintén az árral hozható összefüggésbe.
Kördiagrammal szemléltetem az előfizetők sávszélesség szerinti megoszlását.

Csillag Renáta
CSRRABT.SZE

A megvizsgált öt előfizető közül a 464968584-as számú ügyfél bonyolította le a hónapban a legnagyobb adatforgalmat, túllépve ezzel a szolgáltató által szabott 10 GB – os korlátot. További elemzéseimből kiderült, hogy összesen tizenegy felhasználó követte el ezt a hibát – a legnagyobb mértékben a 8315080518-as kódú letöltő, aki 10 724 MB – os adatforgalmat bonyolított le, összesen 484 MB – tal átlépve a határt. Ezen ügyfelek az internetszolgáltató szabályai alapján a következő hónapban sávszélesség-csökkentésre számíthatnak.