

## ZΗΤΗΜΑ Α

1) Εφαρμογή του *K-means clustering* αλγορίθμου για την συσταδοποίηση του δείγματος της *telco*, μέσω της πλατφόρμας *WEKA*.

- Δοκιμές (cluster no. 2-14) για τον εντοπισμό του καταλληλότερου πλήθους συστάδων μέσω του SSE (Sum of Squares Error).

#1                                      **2clusters**  
Within cluster sum of squared errors: 2421.92045806835

#2                                      **3clusters**  
Within cluster sum of squared errors: 1728.007107505676

#3                                      **4clusters**  
Within cluster sum of squared errors: 1386.4326675022298

#4                                      **5clusters**  
Within cluster sum of squared errors: 1284.4540180899392

#5                                      **6clusters**  
Within cluster sum of squared errors: 1228.9707097368434

#6                                      **7clusters**  
Within cluster sum of squared errors: 1166.236168834846

....

#11                                     **12clusters**  
Within cluster sum of squared errors: 970.270873217587

#12                                    **13clusters**  
Within cluster sum of squared errors: 982.1064248643329

### ❖ Επιλογή αριθμού συστάδων 5.

*Διότι, η τιμή του σφάλματος μέχρι τότε μειώνεται με μεγάλο ρυθμό, ενώ από αριθμό συστάδων 6 και πάνω ο ρυθμός μείωσης είναι σταθερός και σε κάποιες περιπτώσεις αρνητικός.*

*Επίσης αυξάνοντας παραπάνω το πλήθος δημιουργούνται μικρά σύνολα που πιθανόν αντιπροσωπεύουν ακραίες τιμές.*

## Αποτελέσματα συσταδοποίησης με πλήθος συστάδων 5.

=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0  
-t1 -1.25 -t2 -1.0 -N 5 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: telco

Instances: 1057

Attributes: 19

### Ignored:

Region, Marital, Gender, Custcat, Churn

Test mode: evaluate on training data

=== Clustering model (full training set) ===

K-Means

=====

Number of iterations: 6

Within cluster sum of squared errors: 1284.4540180899392

Missing values globally replaced with mean/mode

Final cluster centroids:

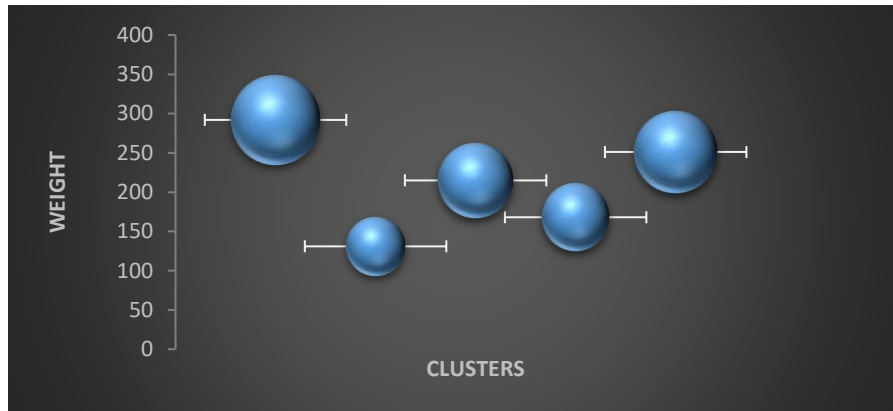
Attribute	Cluster#					
	Full Data	0	1	2	3	4
	-1057	-292	-131	-215	-168	-251
=====	=====	=====	=====	=====	=====	=====
longmon	12.8391	13.9885	20.1984	13.2752	10.3148	8.9773
tollmon	14.753	26.1062	5.521	27.3951	2.3997	3.8029
equipmon	15.6165	5.1109	4.29	38.4985	29.2505	5.0241
cardmon	15.2767	20.2417	17.8435	21.3528	8.1305	7.7394
wiremon	12.8574	9.1786	0.6869	44.8473	6.0968	0.6125
multline	No	No	Yes	Yes	Yes	No
voice	No	No	No	Yes	No	No
pager	No	No	No	Yes	No	No
internet	No	No	No	Yes	Yes	No
callid	No	Yes	No	Yes	No	No
callwait	No	Yes	No	Yes	No	No
forward	No	Yes	No	Yes	No	No
confer	Yes	Yes	No	Yes	No	No
ebill	No	No	No	Yes	Yes	No

Time taken to build model (full training data): 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

1	292	28%
2	131	12%
3	215	20%
4	168	16%
5	251	24%



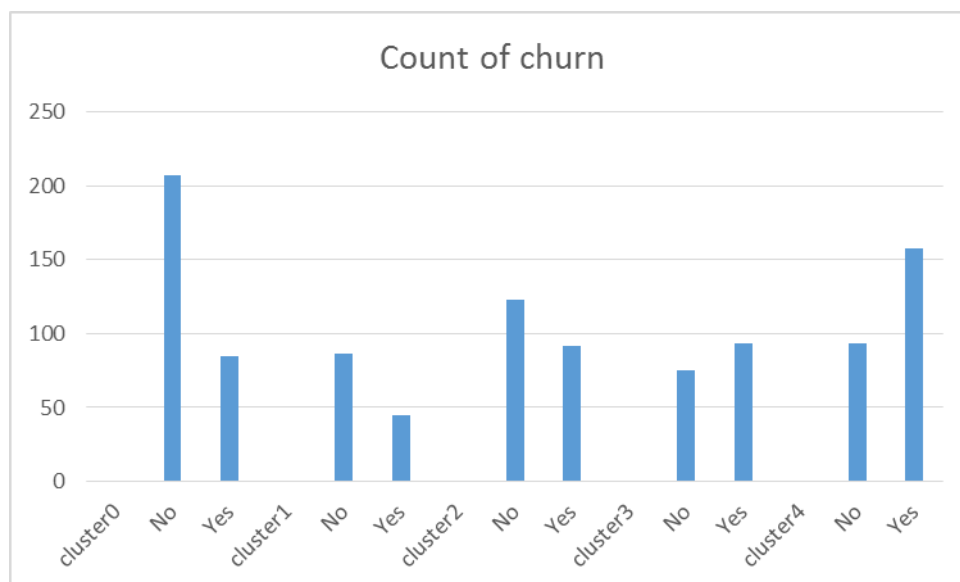
## 2) Ανάλυση

Telco clustering results (table - pivot + chart -demographic pivot + chart - pivot services + chart)



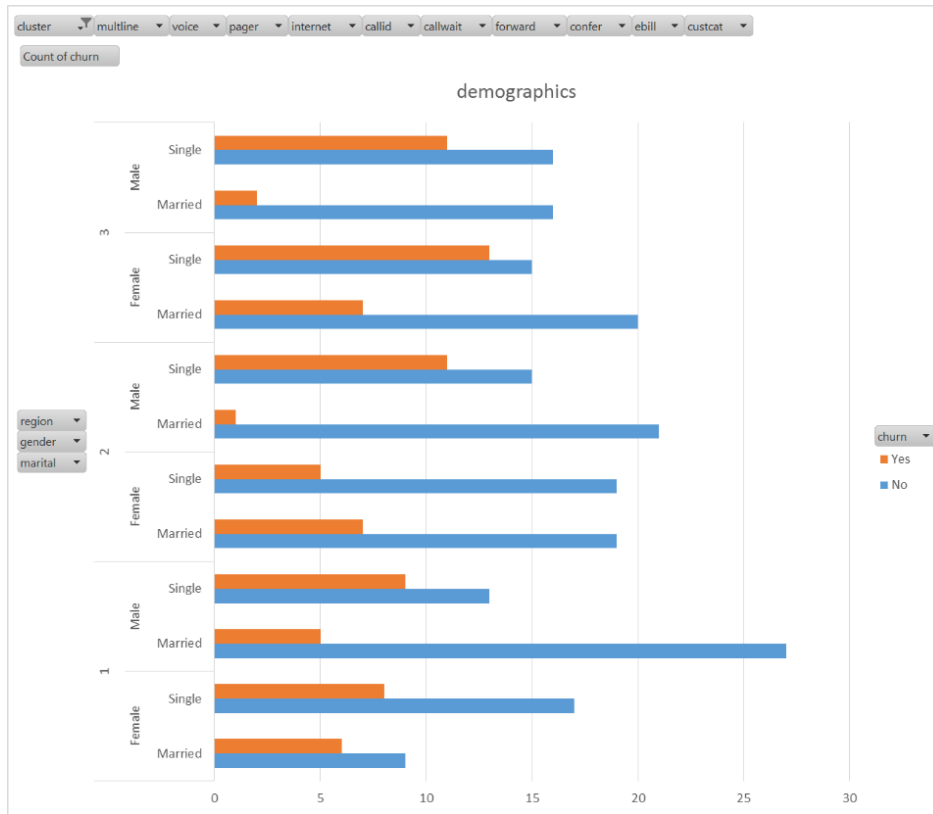
telco\_pivot.xlsx

### Ανάλυση Αποτελεσμάτων Clustering



**Από το παραπάνω γράφημα συμπεραίνουμε πως οι περισσότεροι που αποχωρούν από την εταιρεία Telco ανήκουν στην συστάδα cluster4, ενώ οι περισσότεροι που παραμένουν στην εταιρεία ανήκουν στην συστάδα cluster0**

## DEMOGRAPHICS & churn

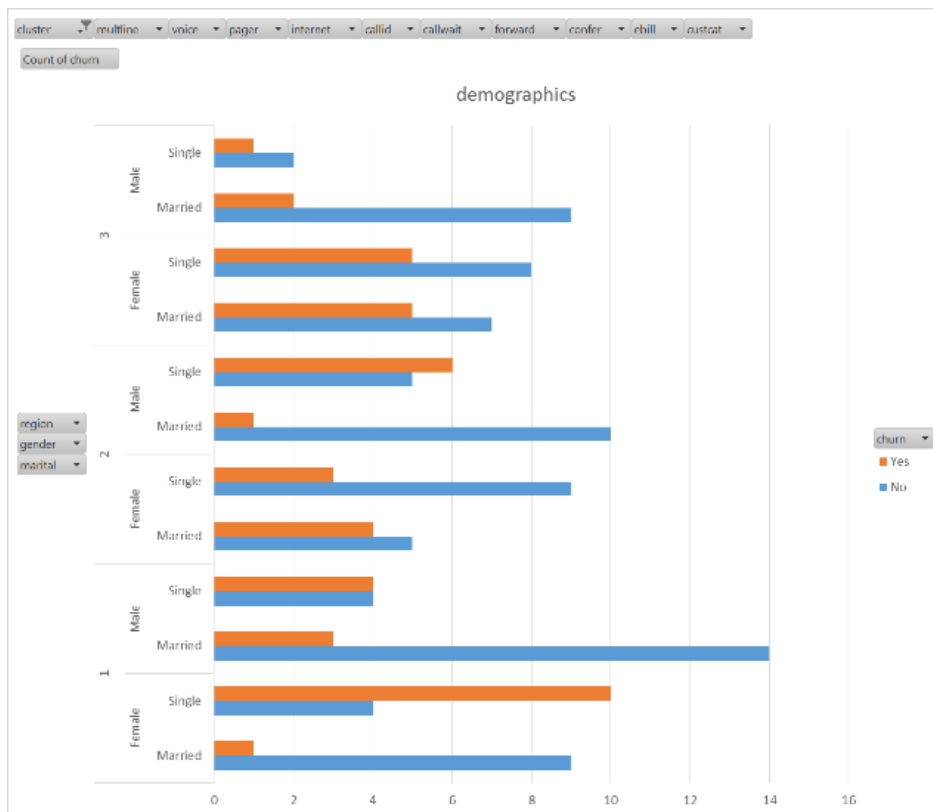


### Cluster 0

Στο Cluster0 παρατηρούμε ότι το μεγαλύτερο πλήθος αποχωρήσεων έρχεται από την περιοχή 3.

Όσον αφορά την αποχώρηση από την εταιρεία, άντρες και γυναίκες κινούνται στους ίδιους ρυθμούς με τους άντρες να είναι ελάχιστα πιο πιστοί στην εταιρεία.

Με μικρή διαφορά και βαθμό ανάλογα την περιοχή φαίνεται πως οι ανύπαντροι επιλέγουν πιο εύκολα την αποχώρηση.



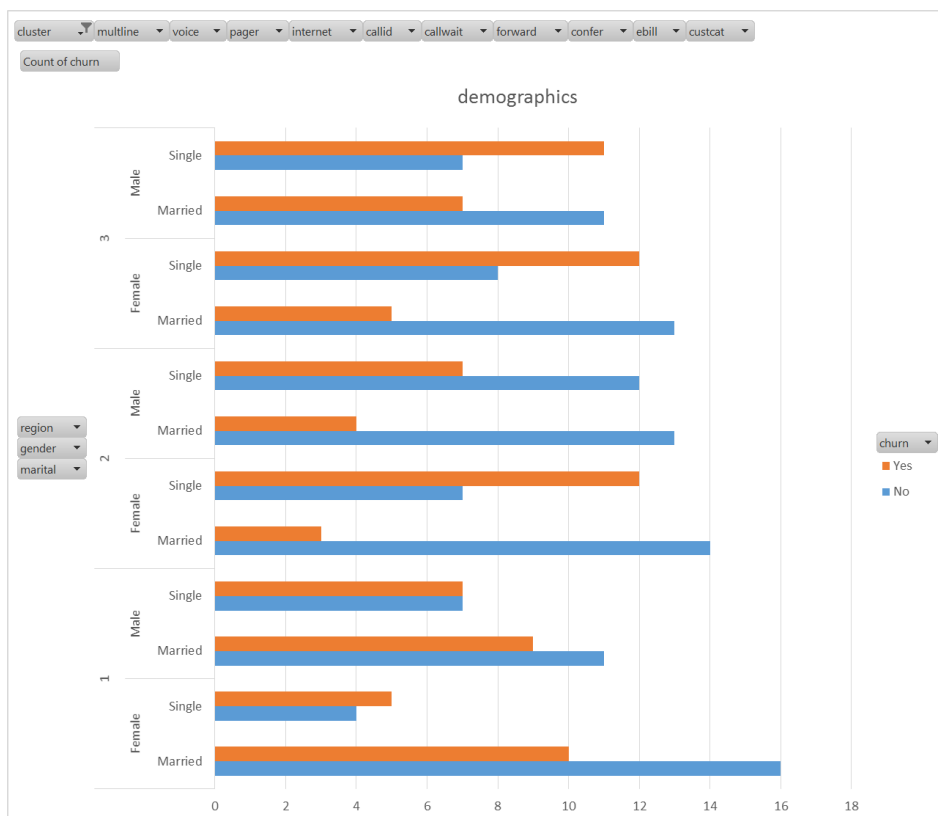
### Cluster 1

Στο Cluster1 παρατηρούμε κυρίως ότι το μεγαλύτερο πλήθος αποχωρήσεων έρχεται από την περιοχή 1.

Όσον αφορά την αποχώρηση από την εταιρεία άντρες και γυναίκες κινούνται στους ίδιους ρυθμούς με τις γυναίκες να αποχωρούν σε μεγαλύτερο βαθμό.

Με μεγάλη διαφορά φαίνεται πως οι ανύπαντροι επιλέγουν πιο εύκολα την αποχώρηση.

Στην συγκεκριμένη συστάδα οι παντρεμένοι και ειδικά οι άντρες φαίνεται να μην αποχωρούν εύκολα.



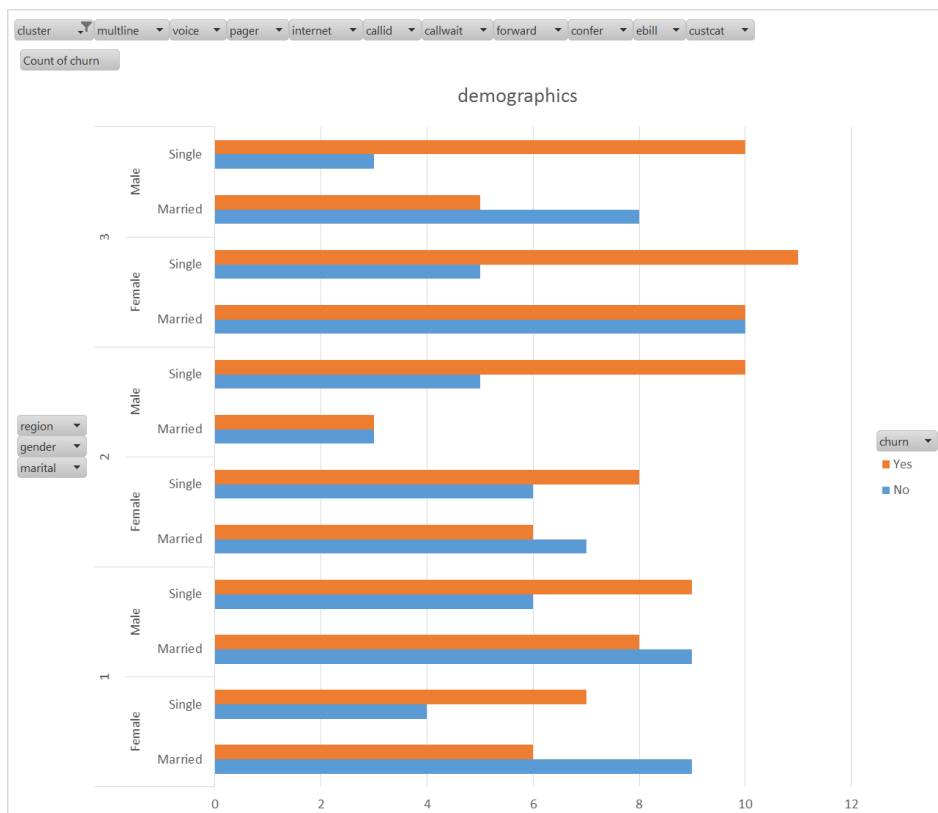
## Cluster 2

Στο Cluster2 παρατηρούμε κυρίως ότι το μεγαλύτερο πλήθος αποχωρήσεων έρχεται από την περιοχή 3 με μικρές διαφορές από τις υπόλοιπες περιοχές.

Πιο έμπιστη φαίνεται η περιοχή δύο.

Όσον αφορά την αποχώρηση από την εταιρεία άντρες και γυναίκες κινούνται στους ίδιους ακριβώς ρυθμούς.

Στην συγκεκριμένη συστάδα οι παντρεμένοι και ειδικά οι γυναίκες φαίνεται να μην αποχωρούν εύκολα.



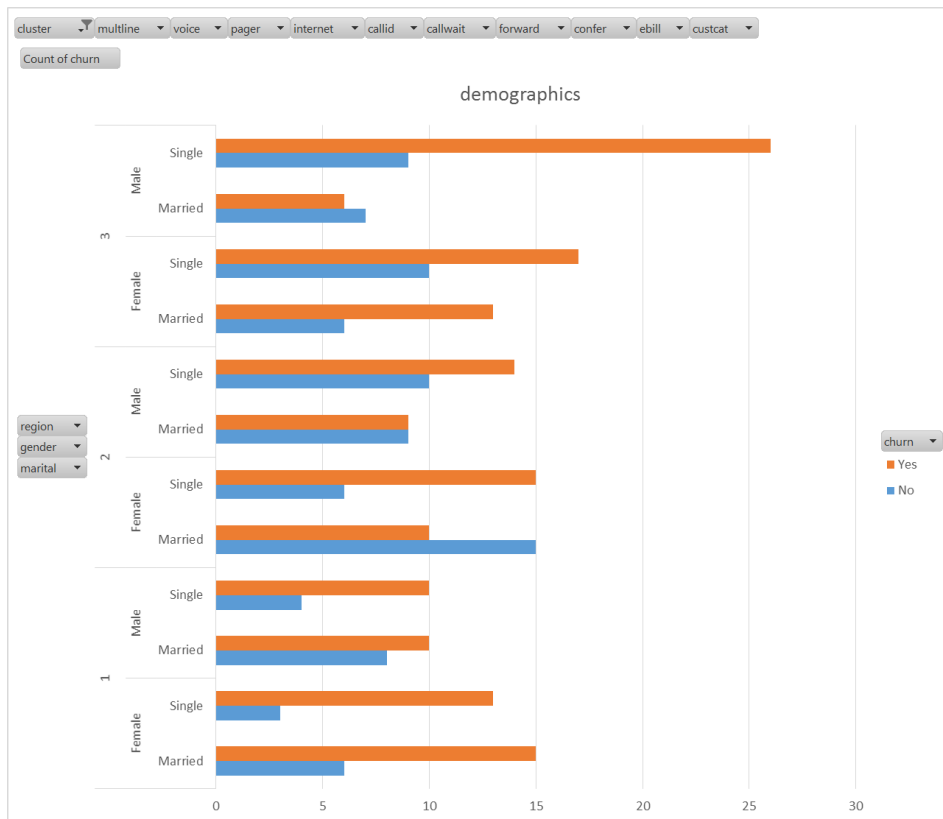
## Cluster 3

Στο Cluster3 παρατηρούμε το μεγαλύτερο πλήθος αποχωρήσεων να έρχεται από την περιοχή 3 με περιοχή εμπιστοσύνης να χαρακτηρίζεται η περιοχή 2.

Όσον αφορά την αποχώρηση από την εταιρεία άντρες και γυναίκες κινούνται παρόμοιους ρυθμούς.

Και σε αυτή τη συστάδα οι παντρεμένοι και ειδικά οι γυναίκες φαίνεται να μην αποχωρούν εύκολα.

Υψηλά επίπεδα αποχώρησης



#### Cluster 4

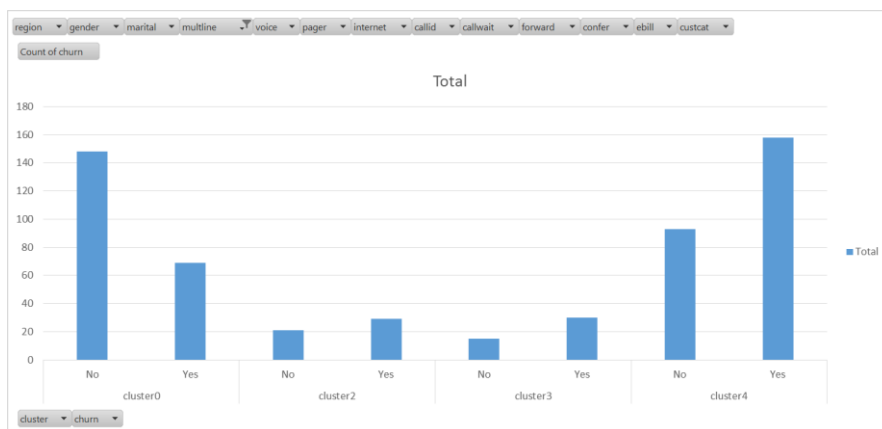
**Στο Cluster4 παρατηρούμε μεγάλο αριθμό αποχωρήσεων να έρχεται από την περιοχή 3.**

**Όσον αφορά την αποχώρηση από την εταιρεία άντρες και γυναίκες κινούνται στους ίδιους ρυθμούς με την παραμονή να υπερτερεί.**

**Με μεγάλη διαφορά φαίνεται πως οι ανύπαντροι επιλέγουν πιο εύκολα την αποχώρηση.**

**Στην συγκεκριμένη συστάδα υπάρχει αισθητά μικρότερο ποσοστό παραμονής.**

#### Services & churn (έξτρα)

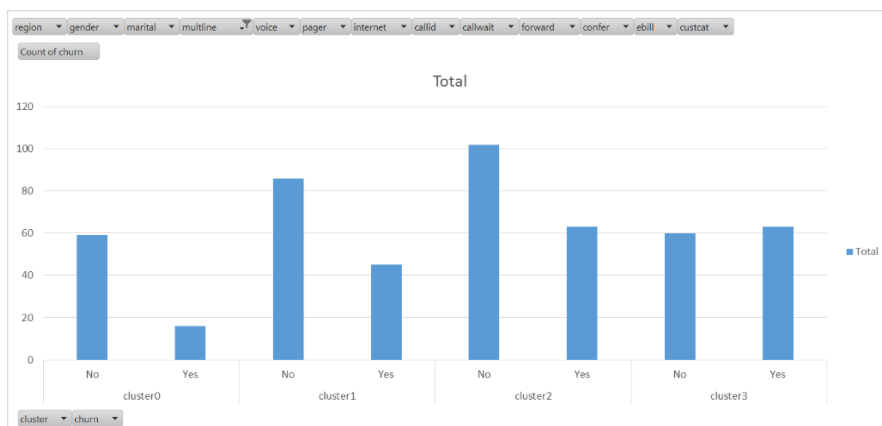


Multline = No

Σε αυτή την περίπτωση η συστάδα 4 εκφράζει τις περισσότερες αποχωρήσεις

Ενώ η συστάδα 0 εκφράζει τους πιο έμπιστους πελάτες από αυτούς που δεν χρησιμοποιούν πολλαπλές γραμμές.

Η συστάδα 1 δεν εμφανίζεται.



Multline = Yes

Σε αυτή την περίπτωση οι συστάδες 2 και 3 εκφράζουν τις περισσότερες αποχωρήσεις

Η συστάδα 1 εκφράζει τους πιο έμπιστους πελάτες από αυτούς που χρησιμοποιούν πολλαπλές γραμμές.

Η συστάδα 4 δεν συμμετέχει.



Voice = No

Σε αυτή την περίπτωση η συστάδα 4 εκφράζει τις περισσότερες αποχωρήσεις,

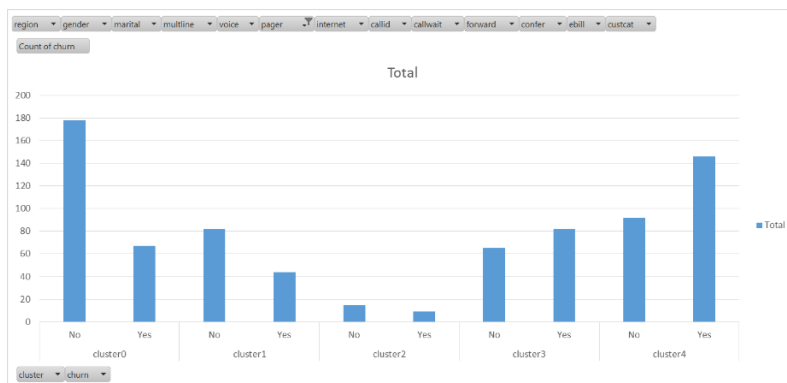
Η συστάδα 1 εμπεριέχει το μεγαλύτερο πλήθος αυτών που δεν χρησιμοποιούν φωνητικές υπηρεσίες και παραμένουν στην εταιρεία.

Στην συγκεκριμένη περίπτωση η συστάδα 2 φιλοξενεί μικρό πλήθος περιπτώσεων.



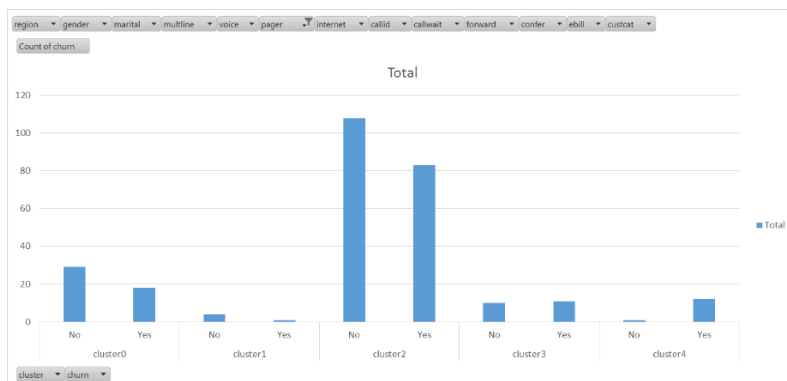
Voice = Yes

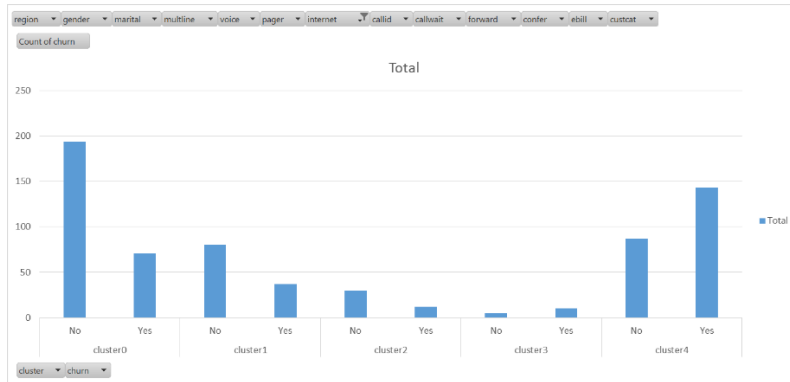
Η συστάδα 2 φιλοξενεί το μεγαλύτερο πλήθος περιπτώσεων, ισορροπημένο όσον αφορά την παραμονή και την αποχώρηση.



Pager

Ομοίως με τις φωνητικές υπηρεσίες





Internet = No

Η συστάδα 0 φιλοξενεί το μεγαλύτερο πλήθος παραμονών, όσων δεν χρησιμοποιούν υπηρεσίες ίντερνετ, σε αντίθεση με την συστάδα 4.



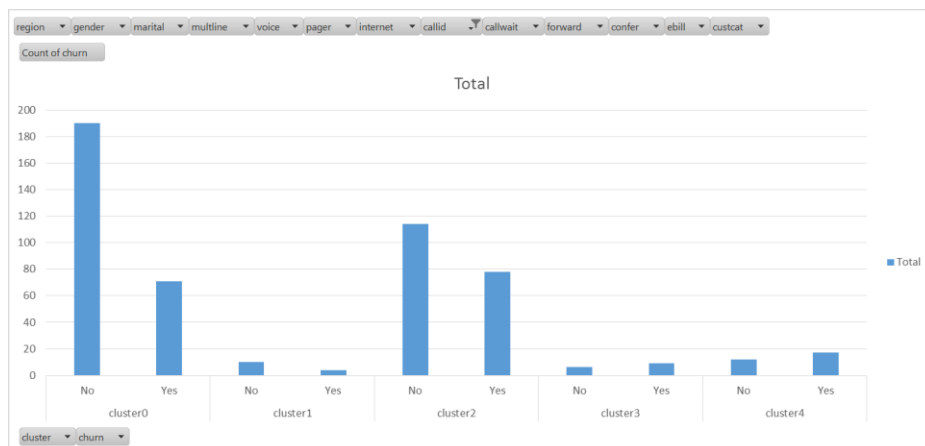
Internet = Yes

Σε αυτή την περίπτωση η συστάδες 3 και 4 αντιπροσωπεύουν το μεγαλύτερο πλήθος όσων χρησιμοποιούν υπηρεσίες ίντερνετ.

**Παρακάτω παρατίθενται οι υπόλοιπες περιπτώσεις μόνο διαγραμματικά..**

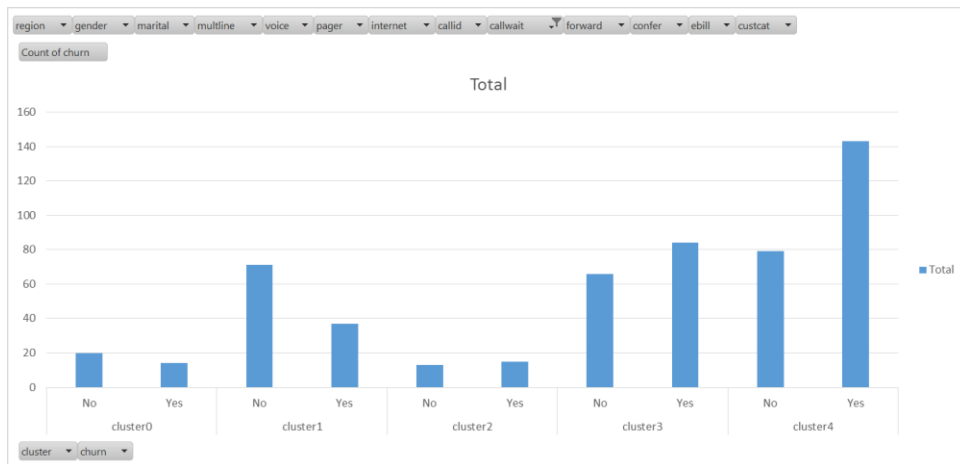


Callid = No

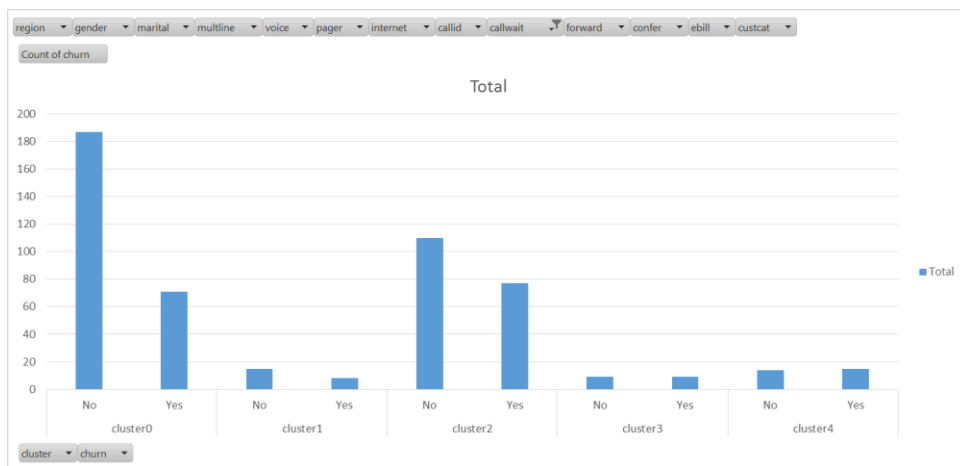


Callid = Yes

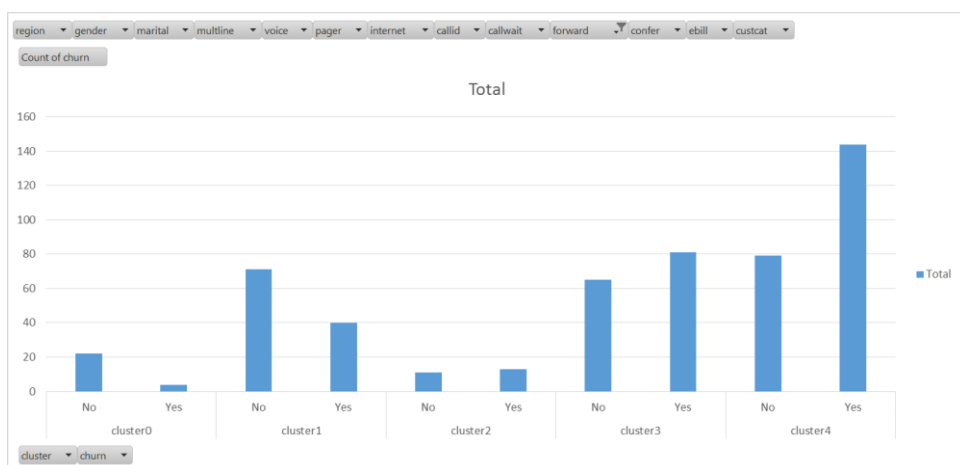




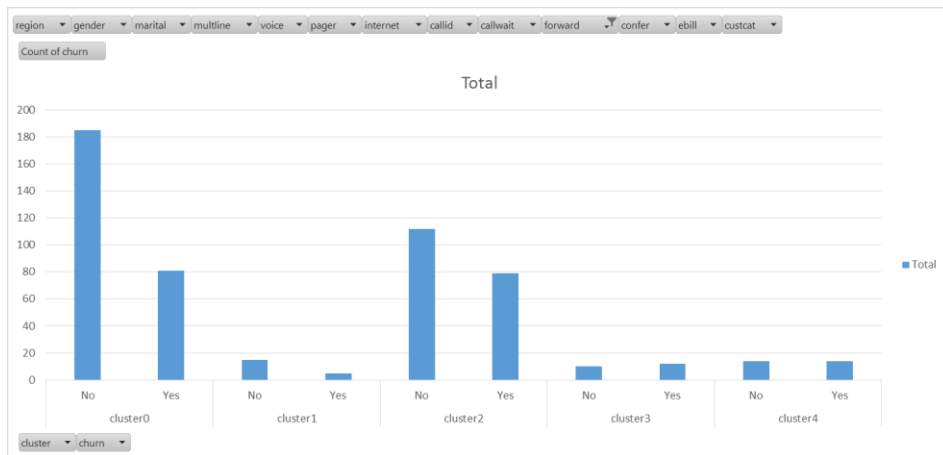
**Callwait = No**



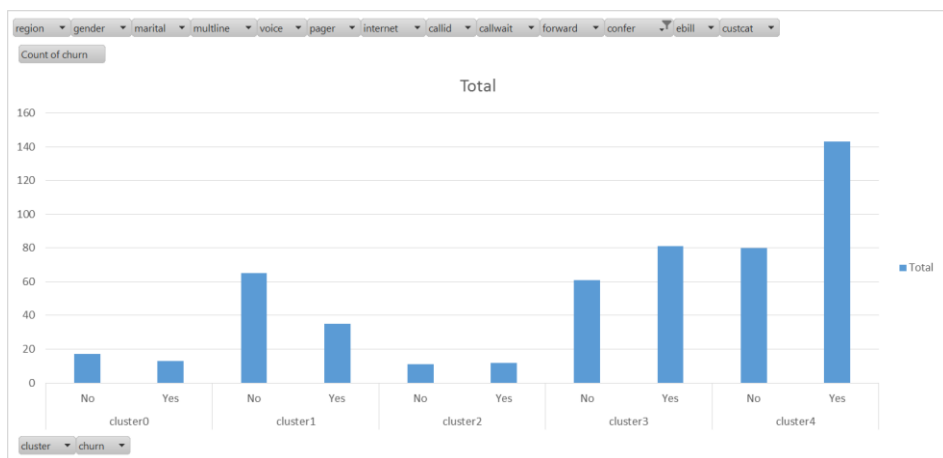
**Callwait = Yes**



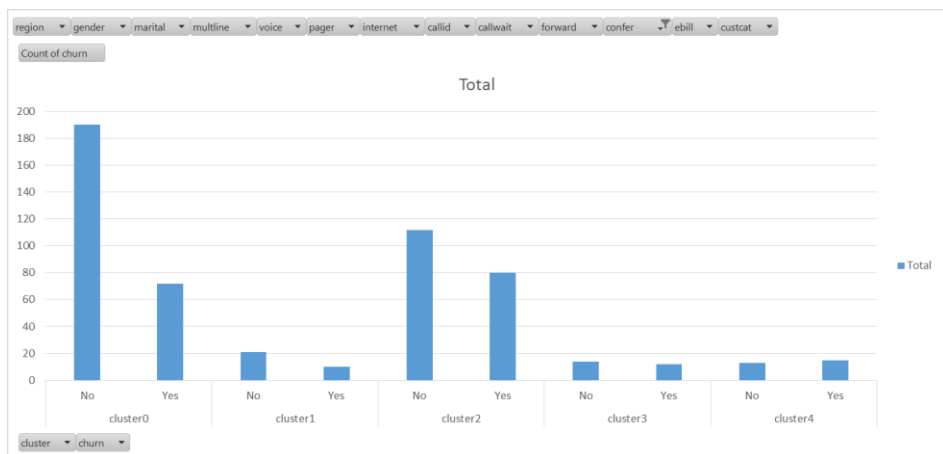
**Forward = No**



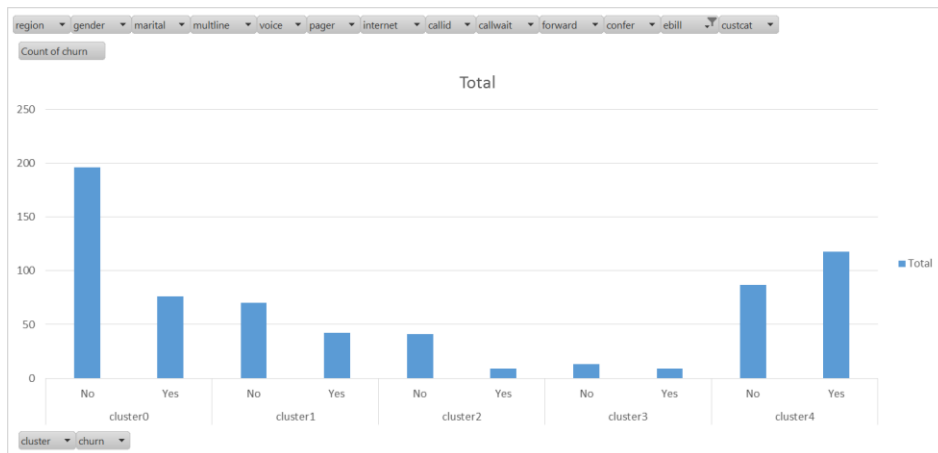
**Forward = Yes**



**Confer = No**



**Confer = Yes**



**Ebill = No**



**Ebill = Yes**

**Τέλος παρατίθεται διάγραμμα που δείχνει την σχέση μεταξύ των συστάδων και των δεικτών χρήσης Αστικών, Υπεραστικών, Εξοπλισμού, Κάρτα κλήσης και Ασύρματου δικτύου.**

