1. **Introduction**

---- What is 7cot and online emotional support services

---- Users can take on many different roles, e.g. user and listener

---- What makes a user decide to switch roles? This is an important question for a few reasons

---- Cannot survey or ask qualitative questions to users who switch, but a data driven approach is possible. Bring in cites.

---- Change point Analysis is a popular technique to detect change in mean and variance in time series data.

---- Maybe we should cite this ( [http://www.variation.com/cpa/tech/changepoint.html](http://www.variation.com/cpa/tech/changepoint.html" \t "_blank))

1. **Methodology**

…………..Interested in finding the actions that drive people to change roles

…………..change points – how different behavioural characteristics change around the time a user takes on a new role may be insightful

…………..We used CUSUM analysis for this purpose with setting as follows –

…………. Number of points to detect --at most one

………….. Penalty imposed is manual

………….. the penalty value used is variant of conventionally used SIC value (log(n) as in here. The penalty value settled on is 0.15\*(log(n))

(<https://www.jstatsoft.org/article/view/v058i03/v58i03.pdf>)

http://artax.karlin.mff.cuni.cz/r-help/library/changepoint/html/single.mean.cusum.html

1. **Implementation**

………….Change point analysis with the above parameters was applied to the total activity count recorded weekly for the total duration in which the user was active on the website

………….Subsequently, the change point analysis was applied to other activities.

…………..It was found that online conversation (ConvMessages) accounted for most of the activity for almost every user. Very few users had other activities such as Helpviews, Page views by app, Page views by web accounting for most of their activities on the website. Further, it was found that activity of conversation messages closely mirrored the change points occurring in the total activity count.

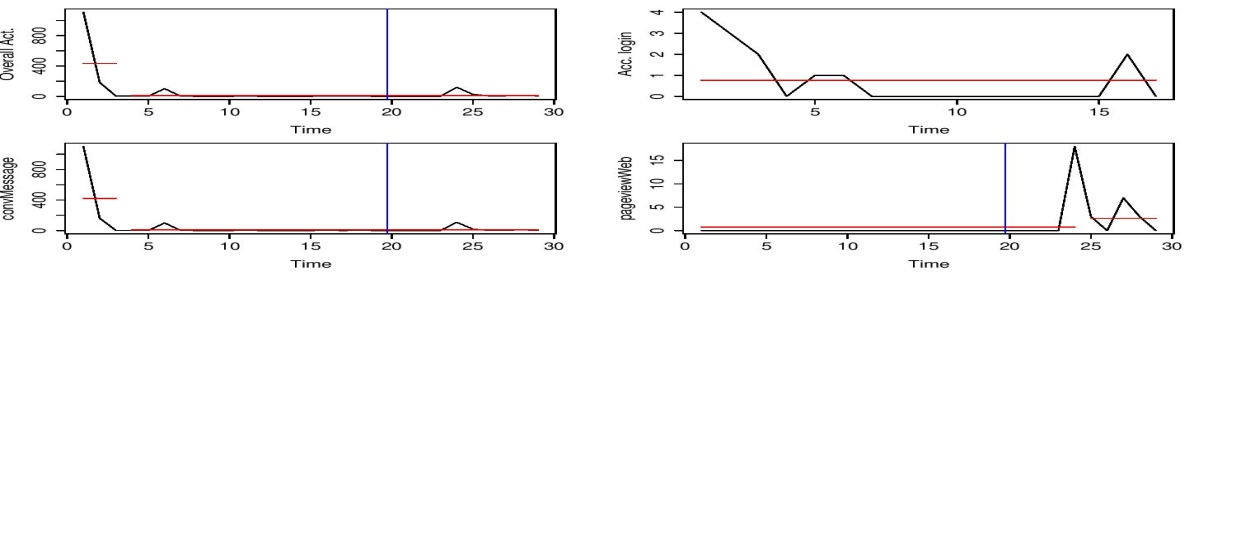


Figure 1:For a user with that we took as part of our study (This figure is also a representative figure for users taken into account for the study since they have the conversation messages serving for most of their activity on the website)

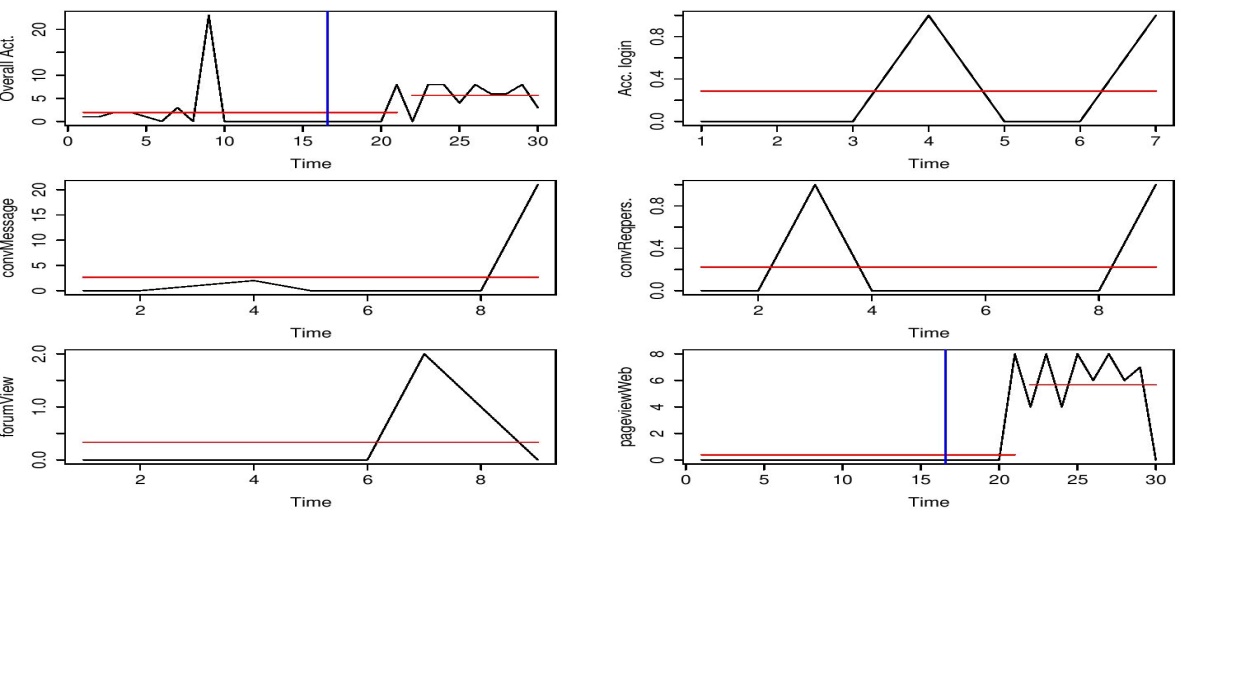


Figure 2: For a user that we did not take into account took as part of our study (This figure is also a representative figure for users not taken into account for the study since they do not have any change in their ConvMessage activity during their run on the website)

Can you pull in figures and statistics or quantitative measurements from your study to support each of the these claims in this paragraph? These users who ended up converting to listeners with other activities accounting for most of their activity on the website apart from Conversation Messages were not considered for further study since they presented outliers in our study (very small in amount) and reason for the conversion could have been random with hardly any bearing to most other users that were converting. Our study is about deciphering the major reasons behind the users taking up the additional role of a listener. Therefore, users that took up additional role without any substantial activity on the website were also discarded since them taking up additional role as a listener could hardly be linked to the website. Since weekly activities were recorded, users who ended up taking the additional role of a listener in their first week were discarded as well.

………..Since the interactions (ConvMessages attribute in figure 1 and figure 2) took place with the listeners on the website, we chose to further explore the activity of these users to see if interactions with a more active listener was one of the major reasons prompting these users in taking up additional roles. We had the total number of conversations for every listener on the website. We choose to define the activity of the listener on the website in terms of the total number of conversations that she had on the website. Since these conversations were collected over the same time frame (5/7/2013 -10/16/2015 ), more conversations directly translate into being more active on the platform. Figure? . Can we quantify this chance?

……..To see if initial distress levels had any co-relation with these conversions (users taking up additional role), we also took signup distress levels in account

……Time elapsed between a substantial change in the way user behaves on the website and them taking up the additional role as a listener is an indicatior of the user’s preference, inclination and readiness in taking up this additional responsibility (**say time a**). It was understood that in some cases this choice might just be a random fluke emerging from exploring the website; therefore users who only had substantial activity on the website were taken into account. It was necessary to have a uniform measure that also took into account the time period of users’ activity on the website. A user taking up additional role of listener within a week of altering its behaviour may have entirely different perspective then the one who took up more time to do the same thing for the same total period of activity. The amount of time user has been active on the website is also important from the point of view of gauging the users’ engagement with the website. In order to take this into account with **a**, **the time a** was divided with the total time user was active on the website as a member. This provided with a fraction between 0 and 1 that indicates the time between the user’s behaviour change and them taking up additional role of a listener relative to the total time they spent on the website.

While taking ilisteners into account as per the total number of conversations that they had on the platform, they were ranked according to the number of conversations they had. So, the listeners having the most number of conversations were given the highest rank while the listeners having the least number of conversations were given the least rank. Further, these ranks were cumulatively added to give out a sum of ranks that would bear indication to not only the breadth of the conversations that users had on the website in terms of quantity but also the quality of conversations that depends upon the activity of listener that it interacted with

* 1. **Unsupervised Clustering**

First provide some context. Why are we going to use unsupervised clustering on the data, what is the intent? Why is clustering the right way to unearth what we are interested in?

Our hypothesis revolves around three attributes –

1. Users sign up distress levels.
2. Cumulative ranks of the listeners on the website that it interacted with. This is also an indication of the breadth of the conversations that a user had on the website and how active were the listeners on the website that it interacted with.
3. The relative time between a significant behavioural change of a user on the website relative to the total time that it spent on the website. This is a fraction between 0 and 1.

Clustering offers a very straightforward way to group together objects that are more similar to each other in one group than they are to another group. Since these attributes relate closely to a user’s behaviour, any similarity in user’s behaviour/preferences and inclination would also reflect in the above attributes. Moreover, it would also help decipher the different ways that user’s ended up taking additional roles. For a good clustering result, different user’s in different clusters would obviously differ in these attributes and two users in two different clusters with different values of these attributes may be entirely different in terms of their initial situation/expectations (this could be seen from sign up distress levels), breadth and quality of interaction they had on the website (using cumulative ranks of the listener, in other words the service provided by the website to these users) and finally their reluctance/inclination in taking up additional role of a listener (this reflects in the relative time that these users took in taking up the additional role in comparison to the total time they spent on the website).

Unsupervised clustering also lets one free of the clutches of making any assumption on the data and lets the data decide for itself, thus providing another reason to be useful in this situation.

The resulting data set was separated into two sets –

1. Users who had a change in their activity before they took up the additional role of a listener.
2. Users who had a change in their activity after the took the additional role of listener

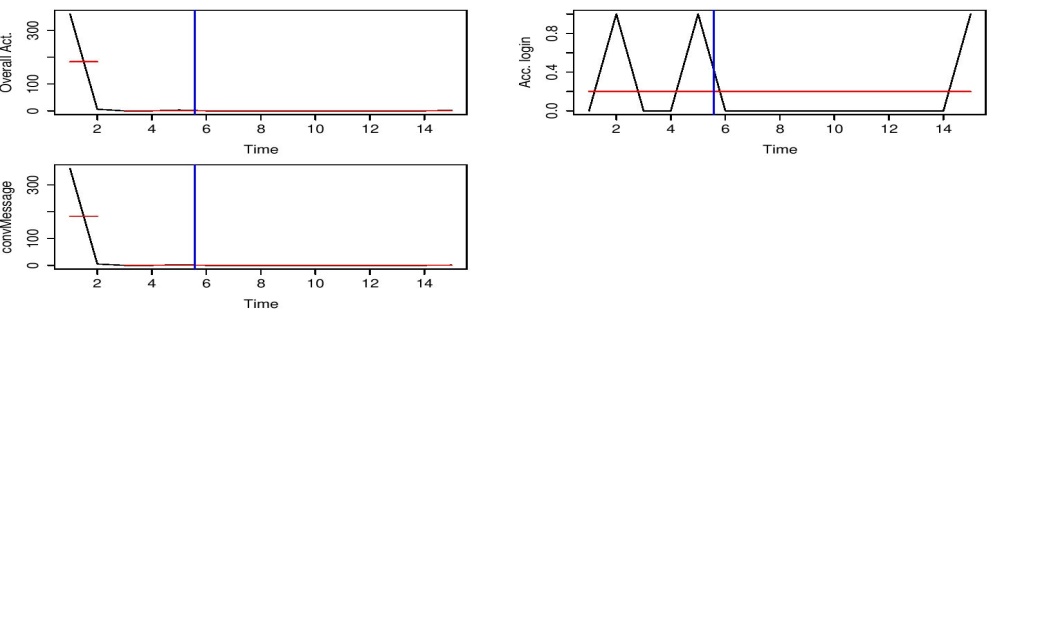
users in the 1st group –

Distress levels ranks fraction

6.333333 6861852.8 0.2006790

6.105882 428287.4 0.4299553

As evident for almost similar distress levels, the cumulative sum of the ranks of the users differed to a large extent and this clearly reflects in the fraction of the time they took to take additional role of listener after registering the change. This fraction is relative to the total time these users spent on the website. Therefore, for almost the similar distress levels, interactions with more number of listeners or interactions with just more interactive listeners considerably reduced the relative time these users spent on the website after which they took up the additional role. Figure below is an example of such users.



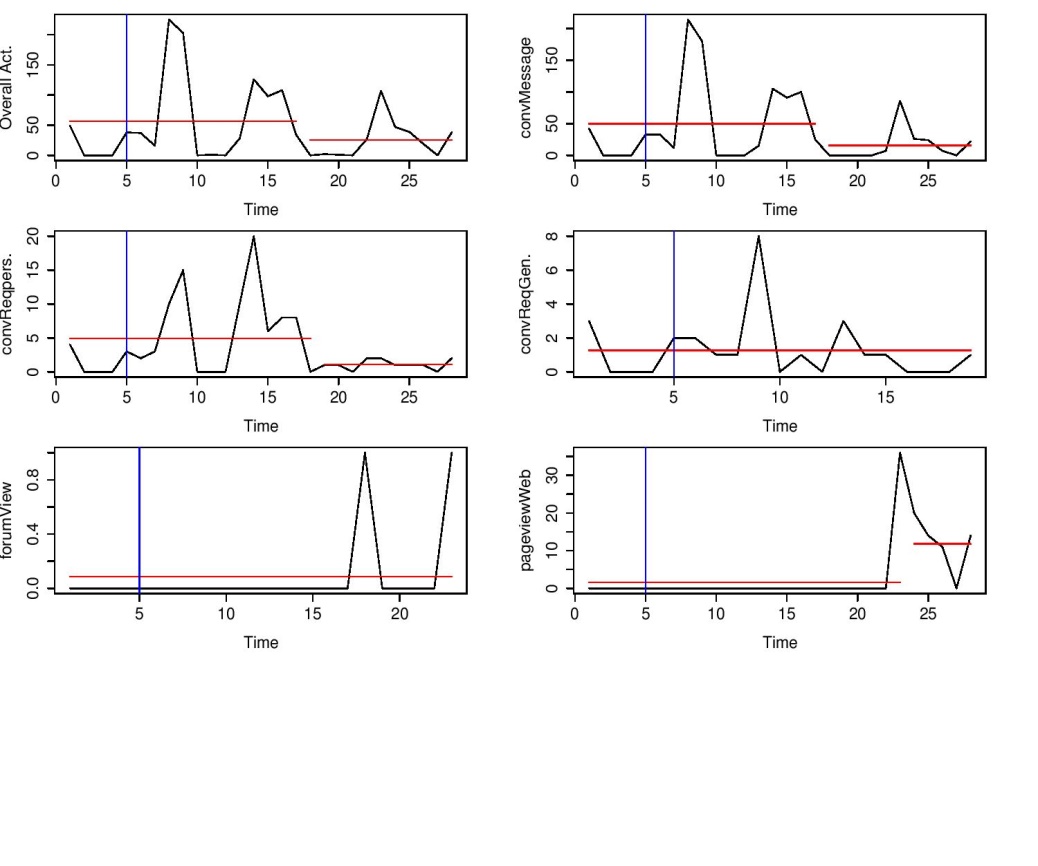
users in the 2nd group –

Distress level ranks fraction

6.583333 323050 -0.2843178

8.666667 2604893 -0.2247331

The distress level of the users that show change in their mean activity after they take up the additional role of the listener were found to be higher than the ones who had a change in their mean activity before the conversion. This is ample evidence to support the fact that the 7cot website caters well enough for the needs of the aggrieved users. Users with higher distress levels tend to engage more widely with the service (as in taking up the additional role of the listener). Their higher distress level makes them more susceptible to be influenced by listeners as evident that they turn to take up additional roles before they actually get their own needs met. In general, the overall activity of the users decreases on the website after the change so it is evident that them taking up additional role before gearing down in their run as a member proves thist. Figure below is an example of such users.



The website works to solve the problems of these users (first set of users) could be confirmed from the fact that even though the overall mean activity of the users fall, they return back to take the additional responsibility of listener (this may be confirmed from figure 1) while in second case they simply keep on the website even after taking up additional responsibility even after suffering from overall higher distress levels on their own part. Moreover, this is much more evident in their case. This confirms our hypothesis that the website is working in favour of reducing the users distress levels and is capable of bringing some sort of relief to them.

**Results**

The inference drawn from the above could be summarized as follows –

1. Users with high distress levels tend to fall in the category where they take up additional role before actually slowing down in their run as a member. This is good news for the website as it suggests that it is doing the job it is supposed to – engaging more distressed users rather than shunning them away as they take up additional responsibility relatively much before they even slow down in their own run as a member as evident from the figure above. According to above classification, these users lie in the second category of users.
2. Users with lesser distress levels choose to take more time relatively in taking up additional role of a listener.

3Exposure to more active listeners or to just more number of listeners considerably lessens the relative time that the users take to take on the additional responsibility of a listener as evident in the first group of listener.

4More distressed users either choose to interact with much more active listeners on the website or simply interact with a much wider breadth of listeners as evident in the second set of users.

At the end of the report summarize in a bulleted list the concrete findings of the investigation.