International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 4 Issue 1, January 2017

ISSN: 2394-4404

Analysis Of Common Problems Among It Professionals

R. Aarthi

PG Student, Dept of Computer Science, JayarajAnnapckiam College for Women (Autonomous), Periyakulam, TheniDt, Tamil Nadu

S. IrudhayaAnanthi

Assistant Professor, Department of Computer Science, JayarajAnnapckiam College for women (Autonomous), Periyakulam, TheniDt, Tamil Nadu

Abstract: The potential and problems of IT employment and their implications in the socio–cultural and familial environment of the professionals in the Chennai context is examined in this study by taking three groups of couples and making comparisons among them with respect to their work-life balance. A narrative enquiry supported by quantitative measurement approach is adopted. The study results show that IT employment makes the employees financially content although it often leads to a cultural shift. The practices associated with child care and the career prospects of women employees are also dealt with in this study. The claim of equal opportunity employment does not seem to agree with the attitude of employers towards pregnant and lactating women. Overtime and overload of work are accepted as integral parts of this job by the employees. Family environment is found to be degenerating and the rate of marriage breakdown is reportedly high among IT professionals. Various stakeholders of the industry including government are found to be ignorant of these issues. The study ends with a number of policy recommendations to improve work-life balance among IT professionals.

Keywords: socio–cultural, common problems among IT, quantitative measurement, work-life balance.

I. INTRODUCTION

Theorizing the nuances of the given quote would be tantamount to defining the plural space of the individual within the standardizing corporate space. Information Technology, a double edged mechanics which both pioneered the present brand of capitalism and is its brand ambassador, has been in the headlines of the economic columns in this era, and India is believed to be the best IT destination. Perceiving the IT potential for employment, Chennai has invested a lot of money in government sector and also in public-private partnership mode in developing IT industry enabling Tamil people to work from Chennai.

It is estimated that there are about 35000 IT professionals who working in Chennai. (The Financial Express, 11th Nov., 2010.) The study leading to the present work began in January 2004 when the researcher started working as a technical recruiter. Since lateral recruitment is a process of picking a professional from one organization and placing him/her in one of the client organizations, a recruiter's job involves continuous interactions with the HR managers of client organizations and with candidates looking for a change in job.

Many IT employees would like to retire at an early age or change career due to stress, burnout, and dissatisfaction with work and family or personal problems. Women employees were expressing their interests in switching job were observed to be high in number.

II. RESEARCH METHODOLOGY

A. METHODOLOGY

The case study will consist of deferent stages, roughly following the cross industry standard procedure CRISP-DM. Firstly, the business understanding phase has to be carried out. In this phase, the project objectives and requirements are stated and reined and the resulting data mining problem is formulated. The results of this phase are summarized in the previous sections. Although the collection of additional data results in a richer data set and is therefore likely to give better results, model acting on a data set that is already automatically kept up-to-date is potentially a much more useful tool.

B. ALGORITHM USED

CLUSTER ANALYSIS

Cluster analysis is a multivariate analysis that attempts to form groups or "clusters" of objects(Sample plots in our case)that are "similar" to each other but which differ among clusters. The exact definition of "similar" is variable among algorithms. But has a generic basis. The methods of forming clusters also vary, but follow a few general blueprints.

K-MEANS CLUSTERING

The most common partitioning method is the K-means cluster analysis.

Conceptually, the K-means algorithm:

- ✓ Selects K cancroids (K rows chosen at random)
- ✓ Assigns each data point to its closest centroid
- ✓ Recalculates the centroids as the average of all data points in a cluster (i.e., the centroids are p-length mean vectors, where p is the number of variables)
- ✓ Assigns data points to their closest centroids
- ✓ Continues steps 3 and 4 until the observations are not reassigned or the maximum number of iterations(R uses 10 as a default) is reached.

IMPLEMENTATION DETAILS FOR THIS APPROACH CAN VARY

R uses an efficient algorithm by Hartigan and Wong (1979) that partitions the observations into k groups such that the sum of squares of the observations to their assigned cluster centers is a minimum. This means that in steps 2 and 4, each observation is assigned to the cluster with the smallest value of:

$$SS(K) = \sum_{i=1}^{n} \sum_{j=0}^{p} (X_{ij} - \bar{X}_{kj})^2$$

Where k is the cluster, Xij is the value of the jth variable for the ith observation, and Xkjnar is the mean of the jth variable for the kth cluster.

K-means clustering can handle larger datasets than hierarchical cluster approaches. Additionally, observations are not permanently committed to a cluster. They are moved when doing so improves the overall solution. However, the use of means implies that all variables must be continuous and the approach can be severely affected by outliers. They also perform poorly in the presence of non- convex (e.g., U-Shaped) clusters.

The format of the K-means function in R is k-means(x, clusters) where x is a numeric dataset (matrix or data frame0 and centers is the number of clusters to extract. The function returns the cluster memberships, centroids, sums of squares (within, between, total), and cluster sizes.

Since K-means cluster analysis starts with k randomly chosen centroids, a different solution can be obtained each time the function is invoked. Use the set. seed () function to guarantee that the results are reproducible. Additionally, this clustering approach can be sensitive to the initial selection of centroids. The kmeans() function has an start option that attempts multiple initial configurations and reports on the best one. For example, adding nstart=25 will generate 25 initial configurations. This approach is often recommended.

Unlike hierarchical clustering, K means clustering requires that the number of clusters to extract be specified in advance. Again the NbClust package can be used as a guide. Additionally, a plot of the total within-groups sums of squares against the number of clusters in a K-means solution can be helpful. A bend in the graph can suggest the appropriate number of cluster.

C. CLUSTER ANALYSIS IN R

R has an amazing variety of function for cluster analysis. In this case study, I use three of the many approaches: hierarchical agglomerative, partitioning, and model base

DATA PREPARATION: Prior to clustering data, you may want to remove or estimate missing data and rescale variables for comparability.

#prepare Data

Mydata<-na.omit(mydata) # listwise deletion of missing

Mydata<- scale(mydata)

PARTITIONING: K-means clustering is the most popular partitioning methods. It requires the analyst to specify the number of cluster to extract. A plot of the within groups sum of squares by number of cluster extracted can help determine the appropriate number of cluster. The analyst looks for a bend in the plot similar to a screen test in factor analysis.

Determine number of cluster

>wss<- (nrow(mydata)-1)*sum(apply(mydata,2,var))

>for (i in 2:27) wss[i] <- sum(kmeans(mydata,centers=i)$withinss)

>plot(1:27, wss, type="b", xlab="Number of Clusters", ylab="Within groups sum of squares")

#K-means cluster analysis

> fit <- kmeans(mydata, 5)

get Cluster means

>aggregate(mydata,by=list(fit$cluster),FUN=mean)

append cluster assignment

Mydata<- data.frame(mydata, fit$cluster)

A robust version of K-means based on mediods can be invoked by using pam() instead of kmeans(). The function pamk() in the fpc package is a wrapper for pam that also prints the suggested number of cluster based on optimum average silhouette width.

HIERARCHICAL AGGLOMERATIVE: There are a wide range of hierarchical clustering approaches. I have good luck with ward's methods described below.

ward Hierarchical Clustering

d <-dist(mydata, method = "Euclidean")# distance matrix

fit<- hclust(d, method="ward")

plot(fit) # display dendogram

groups<- custree(fit, k=5) # cut tree into 5 clusters

drawdendogram with red borders around the 5 Clusters

Rect.hclust(fit, k=5, border="red")

Thepvclust() function in the pvclust package provides p-values for hierarchical clustering based on multiscale bootstrap resampling. Cluster that are highly supported by the data will have large p values. Be aware that pvclust clusters column, not rows. Transpose your data before using.

Ward Hierarchical Clustering with Bootstrapped p values

Library(pvclust)

Fit <- pvclust(mydata,method.hclust=”ward”, method.dist=”Euclidean”)

Plot(fit) # dendogram with p values

add rectangles around groups highly supported by the data

Pvrect(fit, alpha=.95)

MODEL BASED: Model based approaches assume a variety of data models and apply maximum likelihood estimation and bayes criteria to identify the most likely model and number of cluster. Specifically, the MClust() function in the mclust package selects the optimal model according to BIC for EM initialized by hierarchical clustering for parameterized Gaussian mixture model. One chooses the model and number of clusters with the largest BIC.

Model Based Clustering

Library(mclust)

fit<- Mclust(mydata)

plot(fit) # plot results

summary(fit) # display the best model

PLOTTING CLUSTER SOLUTION: It is always a good idea to look at the cluster result

K-means Clustering with 5 Cluster

Fit <-kmeans (mydata, 5)

Cluster plot against 1st and 2nd principal components

vary parameters for most readable graph

Library(cluster)

Clusplot(mydata, fit$cluster, color=TRUE, shade=TRUE, labels=2, lines=0)

Centroid plot against 1st and 2nd discriminant functions

Library(fpc)

Plotcluster(mydata, fit$cluster)

VALIDATING CLUSTER SOLUTION: The function cluster.stats() in the fpc package provides a mechanism for comparing the similarity of two cluster solutions using a variety of validation criteria.

Comparing 2 cluster solution

Library(fpc)

Cluster.stats(d, fit1$cluster, fit2$cluster)

Where d is a distance matrix among objects, and fit1$cluster and fit2$cluster are integer vectors containing classification results from two different clusters of the same data.

D. STATISTICAL TECHNIQUES USED

a. *DATA SOURCES AND METHODOLOGY*

TARGET POPULATION: This survey covers all the IT employees of Chennai city

INSTRUMENT DESIGN: This questionnaire collects data on the attitude of the people regarding works of IT employees. The items and reasons on the questionnaire have remained unchanged for several years. However, should modifications become necessary, proposed changes would go through a review committee and a field test with respondents and users to ensure its relevancy?

SAMPLING: This survey is a census with cross-sectional design. Data are collected for all units of the target population, therefore no sampling is done.

DATA SOURCES: Responding to this, survey is mandatory. Data are collected directly from survey respondents. Data are compiled from the responses the researcher collected by the questionnaire. The researcher performs the data capture activities, and follow-up of non-respondents. Contact with respondents is maintained for subsequent follow-up.

ERROR DETECTION: There are edits built into the data capture application to check the entered data for unusual values, as well as to check for the logical inconsistencies. Whenever an edit fails, the interviewer is prompted to correct the information (with the help of the respondent when necessary). For most edit failures the interviewer has the ability to override the edit failure if necessary.

IMPUTATION: A 100% rate is attained; therefore imputation is not necessary.

QUALITY EVALUATION: Prior to the data release, combined survey results are analyzed for comparability; In general, this includes a detailed review of individual responses, general economic conditions, and historical trends. The data is examined at a macro level to ensure that the long-term trends make sense when compared to publicly available information in media reports, and etc., Large subject matter officers follows up with the academicians to conform the data and to document reasons for large fluctuations in sales or inventories.

DISCLOSURE CONTROL: Releasing any data would divulge information obtained under the statistics act that relates to any identifiable persons, business or organization without the prior knowledge or the content in writing of that person, business or organization. Various confidentiality rules are applied to all data that are released or published to prevent the publication or disclosed of any information deemed confidential. If necessary, data are suppressed to prevent direct or residual disclosed of identifiable data.

REVISIONS AND SEASONAL ADJUSTMENTS: Revisions in the raw data are required to correct known non-sampling errors. These normally include replacing imputed data with reported data, corrections to previously reported data, and estimates for new births that were not known at the time of the original estimates. Raw data are revised, on a monthly basis, for the month immediately prior to the current reference month being published. The purpose is to correct any significant problems that have been found that apply for an extended period. The actual period of revision depends on the nature of the problem identified, but rarely exceeds three years.

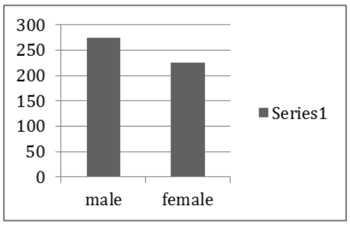
III. ANALYSIS OF DATA

The researcher collects more than 500 responses from samples all over chennai. She spent nearly 2 months to collect the fresh dada from end users. After collecting the information, all the details are fed into the software and checked for outlier. The cleaned data was analyzed using single attribute and multiple attributes. Name, Age, Gender, Company Name.

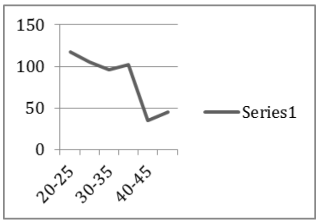
IV. FINDING, INTERPRETATION, RECOMMENDATIONS AND SUGGESTIONS

A. FINDING AND INTERPRETATIONS

Male and Female Division:

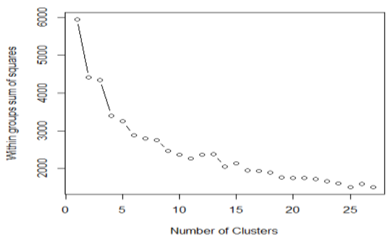


Density of data in Age wise:



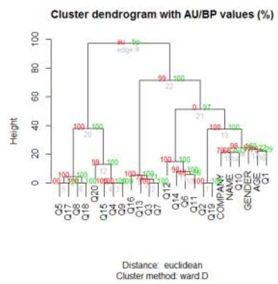
PARTITIONING FOR CLUSTERING

```
for (i in 2:27) wss[i] <- sum(kmeans(ram,centers=i)$withinss)
plot(1:27, wss, type="b", xlab="Number of Clusters", ylab="Within groups sum of squares")
```

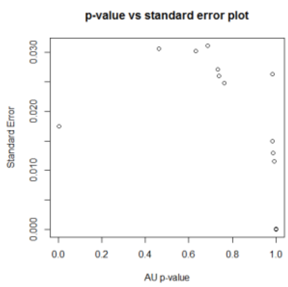


Hierarchical Agglomerative

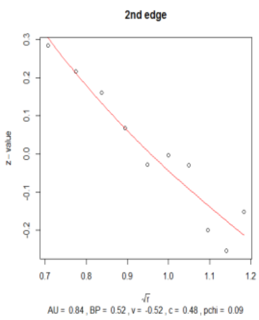
Plot(fit)



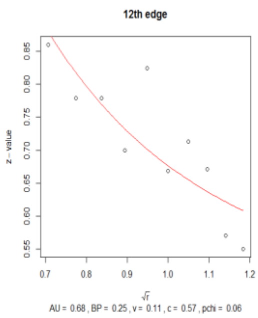
Finding standard Errors



FINDING PLOTS FOR EDGE



Msplot(fit, edges=2)

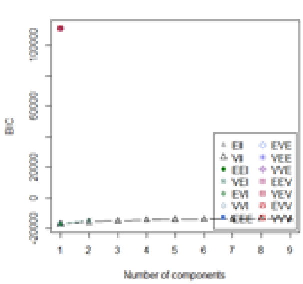


msplot(fit, edges=12)

MODEL BASED CLUSTERING

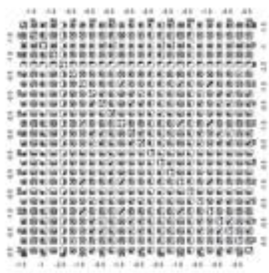
```
Library(mclust)
Fit <- Mclust(ram)
Plot(fit)
```

Selection: 1

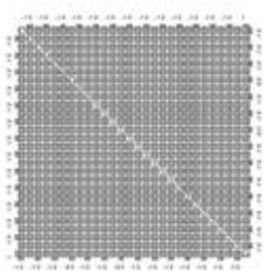


Selection: 2

Selection: 3



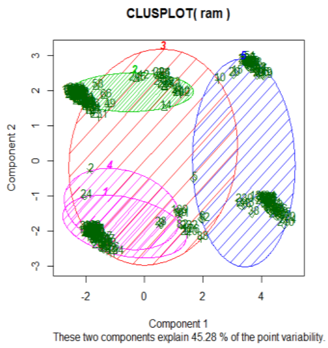
Selection: 4

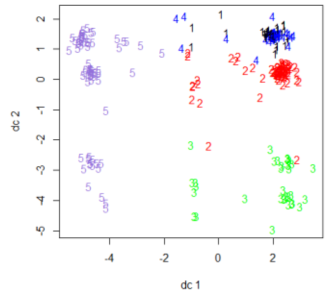


Selection: Enter an item from the menu, or 0 to exit

PLOTTING CLUSTER SOLUTIONS

```
Fit <- kmeans(ram, 5)
Library(cluster)
Clusplot(ram, fit$cluster, color=TRUE, shade= TRUE, labels= 2, lines=0)
```





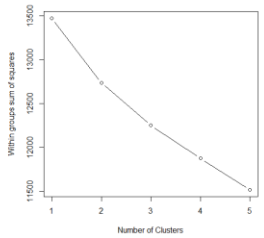

```
Library(fpc)
Plotcluster(ram, fit$cluster)
```

VALIDATING CLUSTER SOLUTIONS

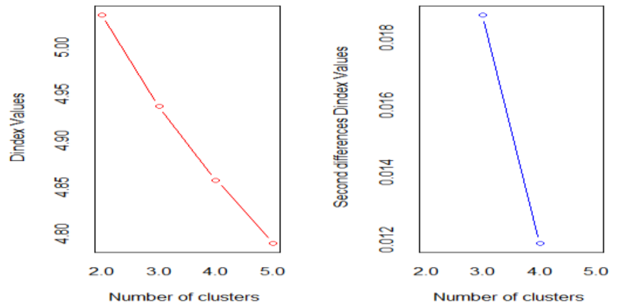
```
df<-scale(ram[-1])
```

International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 4 Issue 1, January 2017

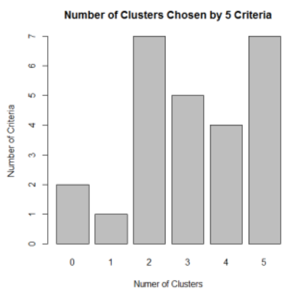
ISSN: 2394-4404



>wssplot(df)
>nc<-NbClust(df, min.nc=2, max.nc=5, methods="kmeans")



table(nc$Best.n[1,])
0 1 2 3 4 5
2 1 7 5 4 7
>barplot(table(nc$Best.n[1,]), xlab="Numer of Clusters", ylab="Number of Criteria", main="Number of Clusters Chosen by 5 Criteria")



B. CONCLUSION AND FUTURE ANALYSIS

Female representation is found to be higher than that of male representation in the entry cadres but the percentage goes down significantly in the middle and upper level positions. The quality of family environment goes down when both the partners are working in IT. The study could not establish a relation between the family environment and the designation of the employee. Knowledge updating is a mandatory requirement for sustaining the IT industry. To retain the competitive advantage, the organization should provide facilities for its employees to update their skills. It is observed that, a large percentage of employees are preoccupied about their lack of time in updating knowledge. In our study, some effective variables about store loyalty or patronage behavior such as, social expectations, the social responsibility of the IT professional as, the cultural structure of them were not included. For further studies, it is suggested that these variables can be taken into consideration.

REFERENCES

[1] Arun, Shoba., Heeks, Richard & Morgan, Sharon. (2004) "ICT Initiatives, Women and Work in Developing Countries: Reinforcing or Changing Gender Inequalities in South India?" *Development Informatics Working Paper Series*, Manchester: Institute for Development Policy and Management, University of Manchester http://www.sed.manchester.ac.uk/idpm/publications/wp/di/index.html

[2] Barrett, Rowena. (2001) "Labouring under an illusion? The labour process of software development in the Australian information industry." *New Technology, Work and Employment* 16(1), pp.18-34

[3] Heeks, R. B. (1996) *India's Software Industry: State Policy, Liberalization and Industrial Development*. New Delhi: Sage Publications.

[4] Hindu (2005) "IT professionals' work-life balance poor" *The Hindu*, October 15, 2005 http://www.hindu.com/2005/10/15/stories/2005101516950200.htm

[5] Kumar, Nirmalya & Puranam, Phanish (2008), "High wages, attrition compel IT sector to look beyond India," *The Economic Times*. Feb 14, 2008.

[6] Sharma, Dinesh. C. (2009). *The Long Revolution: The Birth and Growth of India's IT Industry*. India: Harper Collins Publishers.

[7] Kannan, Shanthi. (2011) "MAT a dampener on SEZs," *The Hindu*, March 1, 2011.

[8] D'Mello Marisa (2006) "Gendered selves and identities of information technology professionals in global software organizations in India" Information Technology for Development

[9] Dowse, S. & Underwood J. (2001) "Work Life Balance of IT Managers: Finding Time". *Proceedings of the Twelfth Australasian Conference on Information Systems*. Association for Information Systems, AIS Electronic Library (AISeL)