Statistics for Management (IDS 570)

ANALYSIS OF WORK/LIFE PROGRAMS' SATISFACTION BASED ON EMPLOYEE SATISFACTION SURVEY

University of Illinois at Chicago

Under the guidance of
Prof. Zack Kertcher

Team 1
MEMBERS:
Apurva Krishnan (UIN: 667895786)
Dhanashree Lele (UIN: 678876925)
Kousalya Dwarapudi (UIN:669348510)
Nikita Bhatia (UIN: 659623638)
Sarada Kannan (UIN: 657213957)

EXECUTIVE SUMMARY

The aim of the project is to conduct detailed analysis on the Data collected from the survey of employees of US Federal Government. Based on the extensive data provided in the survey, the team has identified Variables on which the analysis would be conducted specifically and the conclusion would be provided on the basis of that.

The data was originally collected through survey held in 2014. The survey had 84 questions inclusive of various aspects pertaining to employee attrition, work experience and employee satisfaction. The responses of questions are to be selected from given options only. Also, most of the questions are not mandatory.

The Dataset includes 16 other columns apart from 84 questions. These 16 columns include the information about the employees who took the survey. So, the dataset has total 100 columns. There are 392752 responses to the survey.

STRATEGY / APPROACH

Following steps/methods were followed in analyzing the dataset:

- The team identified Business scenario and selected set of variable on which the analysis is to be conducted. These variables included Independent and control variables.
- The survey did not include any numeric dependent variable. Hence, based on certain condition the dependent variable calculation methodology was identified based on Independent variables' values.
- The team then formulated Research Question and Hypothesis.
- Extensive Data cleaning was conducted on the dataset.
- Then Univariate, Bivariate analysis, hypothesis testing and regression testing was conducted.
- Based on the results fetched from given analysis conclusions were derived and recommendations were given.

INTRODUCTION

Research Question

What influence does the satisfaction level of the work life programs conducted by the Federal Government have on their employees?

Hypothesis

- Female employees are more satisfied with the work-life activity programs conducted in their organization.
- The average satisfaction index is same for all agencies.
- The satisfactory index is highly correlated with the federal government work experience.
- The satisfaction index does not influence employee retention.

Data variables

Dependent variable: Index (Work/Life Programs Satisfaction Index)

Independent variables:Q74, Q75, Q76, Q77, Q78, Q80, Q81, Q82, Q83, Q84

Control variables:DSEX, DSUPER, DFEDTEN, DLEAVING

Calculation steps of Dependent variable

Work/Life Programs Satisfaction Index

- To quantify how satisfied the employees were with the work - life balance programs organized by the federal agencies, the team calculated the work - life programs satisfaction index from the questions 80 - 84.
- To arrive at this index, we added values (responses of employees) from questions 80 - 84, and calculated its percentage.
- The final percentage represented the employee-satisfaction level on the whole.

Challenges:

- Since most of the questions in the survey were not mandatory, there were many NA values in the responses. Filtering out such values and Data Cleaning was a major challenge.
- Almost all the responses/Independent variables were in the form of factors and there was no numeric dependent variable.

- The team had to formulate the method to calculate numeric dependent variable(Work/Life Programs Satisfaction Index) based on independent variables.

Limitations:

There are certain limitations of Dataset and analysis conducted upon it, they are enlisted below:

- The survey responses gathered from employees can be momentary, which means their responses may have been influenced by the recent happenings with them and might not correctly represent the actual scenario. For example, a recent dispute with a supervisor can be reflected in the responses and will not represent the general scenario throughout employee's experience.
- There can exist other factors which may affect Work/Life program satisfaction index which might not be considered in the survey.

Data Cleaning:

Step 1: The dataset contained a total of 84 questions. We have included sets of questions from 74 - 78 and 80-84 for our analysis, so we excluded columns which included responses from questions 1 - 73 and 79 from our analysis.

Step 2: Questions 74 - 78 and 80 - 84. These sets of questions indicate the participation level and satisfaction levels of employees respectively.

Questions 74 - 78 show whether employees have participated in the programs or not. The employees answered either "yes", "no" or "program not available to them".

Out of the five questions, we have included employees who have attended a minimum of three programs. This, we considered on the basis that, responses from employees who have attended very few programs would not be very useful for our analysis. Also, we cannot include just the employees who have attended all the five trainings. After this step, we got results that had some 21,000 rows approximately.

Step 3: Questions 80 - 84 show the satisfaction level of employees. The responses of these questions show how satisfied the employees are on a scale of 1 -5 with 5 being the most satisfied. The responses also contained values "X", which meant that the employees were not able to deduce the usefulness of the programs. The Xs and the nulls were not of much use for our analysis so, from the set of 21,000 rows we again filtered responses from employees who have given answers to atleast three questions. After this step, our dataset contained around 17,000 rows in total.

Data filtering:

- Satisfaction Index: Based on the responses for the questions from 80-84, we calculated a work-life programs satisfaction index. To calculate this index, we computed the sum of all responses

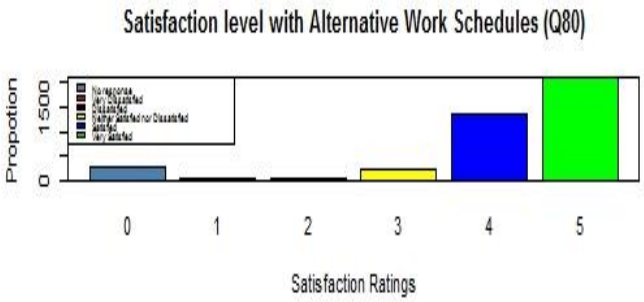
from questions 80 - 84 and calculated the percentage. This satisfaction index was considered as the dependent variable for the rest of our analysis.

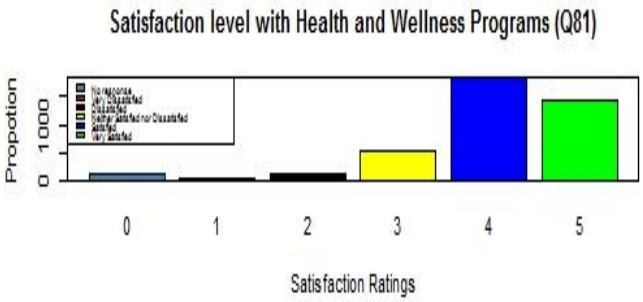
- To narrow down our analysis to specific agencies, we chose five agencies in total representing five categories; agency with the highest satisfaction index value, agency with the lowest satisfaction index, agency that has a satisfaction index equal to the mean, one with value between the lowest and the average and the final agency with value between the highest and the average satisfaction index.

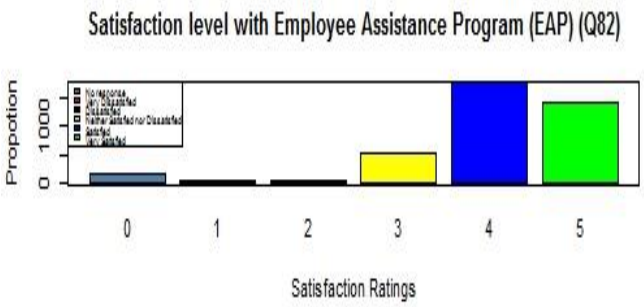
So, after applying the filters, we arrived at a dataset with **3973 rows and 27 columns**, which represented responses from employees working in the federal agencies: TR, IG, FM, CM and IN.

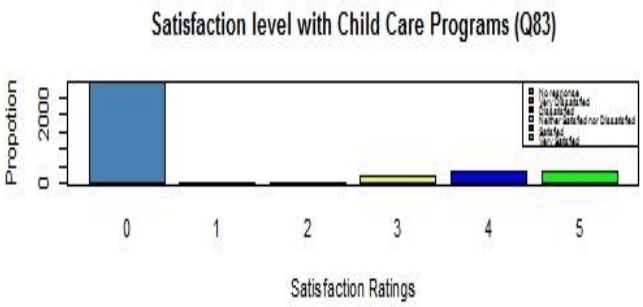
UNIVARIATE ANALYSIS

The univariate analysis is conducted upon Independent variables and control variables.



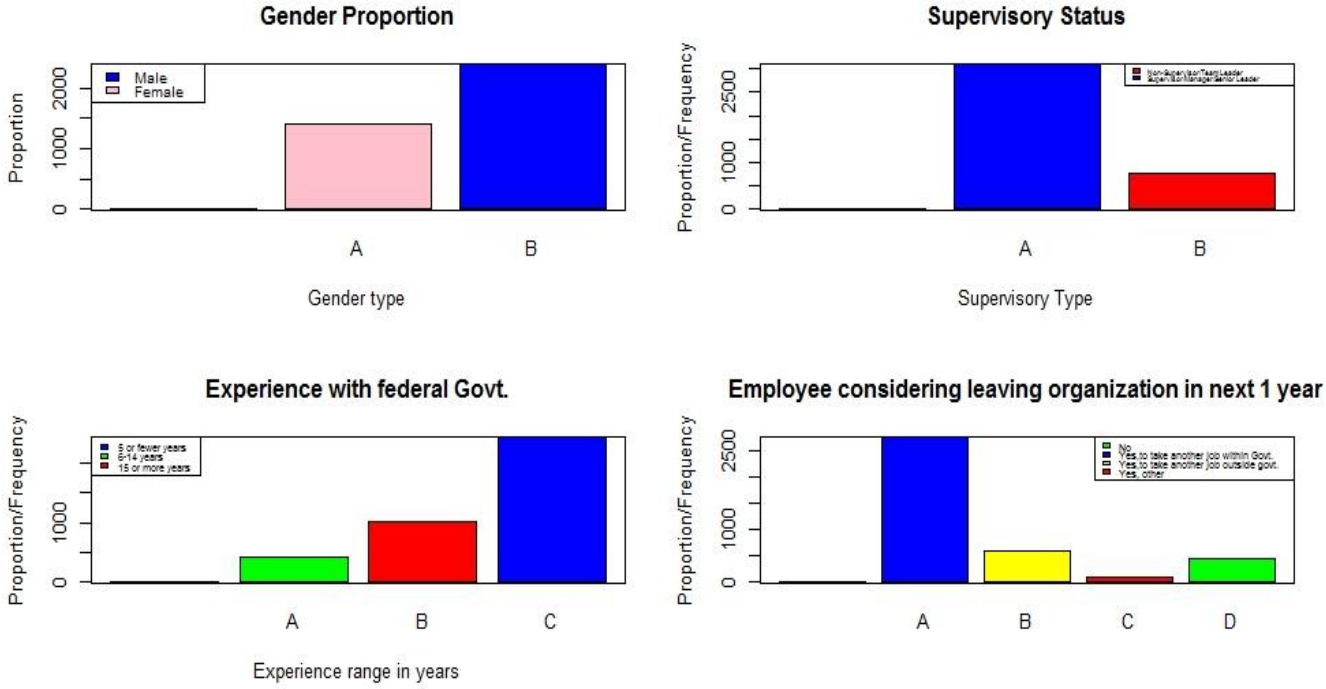




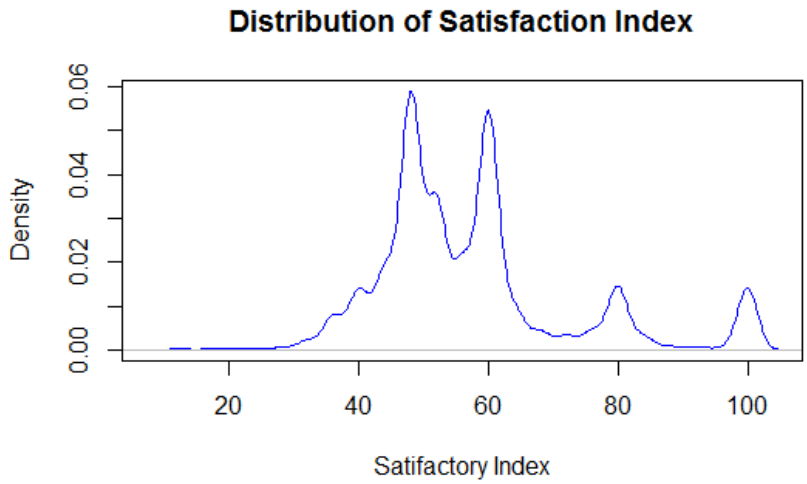


Independent Variables from Q80 – Q84 identify the satisfaction level of employees regarding the Work/Life programs they participated. Here, value 1-Very Dissatisfied,2-Dissatisfied, 3-Neither Satisfied nor Dissatisfied, 4-Satisfied, 5-Very Satisfied. Value 0 indicates that those many employees have not responded to that question. The X values- No Basis to Judge are filtered out for better analysis.

Analysis: Most of the employees are very satisfied or satisfied with Alternative work schedule program, Health and wellness program and employee assistance program. The other two programs have received highest no responses as both of them are not applicable to many of the employees.



As per the univariate analysis of control variables it's evident that Male employees who took the survey are more than female employees. There are more employees who do not have managerial status and they are non-supervisors or just team leaders. And there are lesser employees who have very high experience with the federal govt. The employees who have less number of years of experience are among the highest who have taken the survey. There are highest employees in the category who want to leave the current govt. organization to join another govt. agency only.

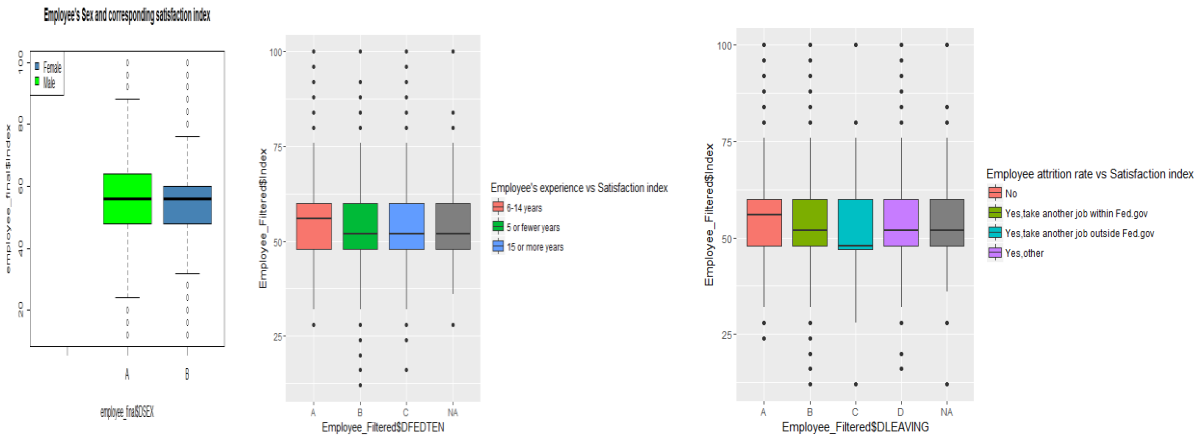


The satisfaction index is not very normally distributed. It tends to be right skewed.

BIVARIATE ANALYSIS

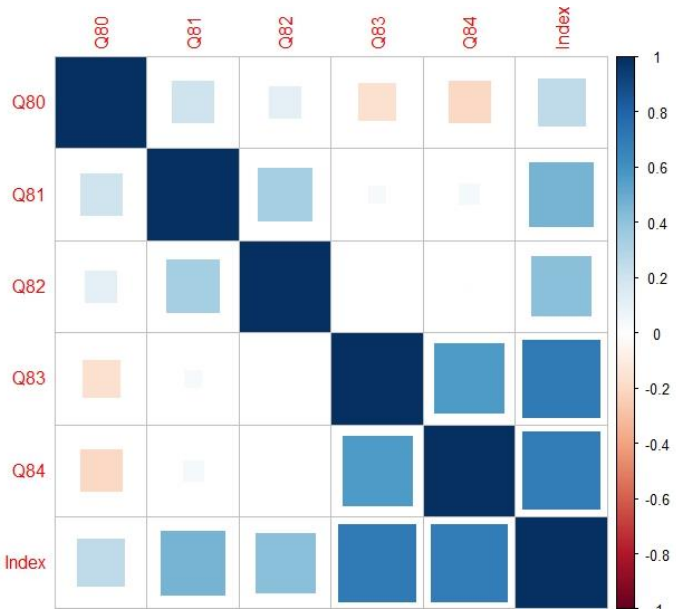
Bi-variate analysis was done between the control variables and the dependent variable (satisfaction index) as our first step of analysis. With the help of the graphs, we got a general idea as to how the respective control variables influence our dependent variable, and the following were our inferences:

Male staff, were better influenced and satisfied with the work-life programs offered by the agencies. Employees who did not have more experienced, were better influenced and satisfied with the programs offered. The work-life programs have considerable impact on the employees' retention rate; employees who appear to be satisfied with these programs have not shown interest in leaving the company in near future.



Correlation between the responses from q80 – q84 and the satisfaction index

The below correlation plot depicts the correlation between the questions (80 – 84) and the index, this shows how and why we chose these questions to calculate the satisfaction index, our dependent variable. The plot indicates that the responses to these diverse questions were highly correlated.



HYPOTHESIS TESTING

In this step, we will be evaluating how satisfaction index is influenced by the variables- DSEX, Agency, DLEAVING, DFEDTEN. We will be performing statistical tests to test the difference between mean values of two or more levels for each of these parameters.

As seen above, the distribution of our data seems normal and the independent variables are of type- FACTOR, hence we are using ANOVA Test to test the variance in the index. We have also used a post-hoc test- TukeyHSD to check the difference in the levels. We have reported our test results below:

ANOVA &TukeyHSD TEST RESULTS:

1)Variance in the index as per sex of employees:

Hypothesis: Female employees are more satisfied with the work-life activity programs conducted in their organization.

ANOVA Result:

Call:
aov(formula = Index ~ DSEX, data = step1)

Terms:
DSEX Residuals
Sum of Squares 19.99 55566.55
Deg. of Freedom 1 3809
Residual standard error: 3.819454
Estimated effects may be unbalanced
TukeyHSD Result:
Tukey multiple comparisons of means
95% family-wise confidence level
factor levels have been ordered
Fit: aov(formula = Index ~ DSEX, data = step1)
$DSEX
diff lwr upr p adj
Female-Male 0.1502249 -0.1013569 0.4018068 0.2417891

Conclusion:

Observing the difference column in the above result, we see that the difference in the means of satisfaction index of male and female employees is nearly 0.15, which is very low. The p-value is 0.24 which makes us reject our hypothesis. Hence, we conclude that satisfaction is not varied by the sex of the employees.

2)Variance in the index according to agencies:

Hypothesis: The average satisfaction index is same for all agencies.

ANOVA Result:

Call:
aov(formula = Index ~ agency, data = step1)
Terms:
agency Residuals
Sum of Squares 2857.69 55053.79
Deg. of Freedom 4 3968
Residual standard error: 3.724841
Estimated effects may be unbalanced
TukeyHSD Result:
Tukey multiple comparisons of means
95% family-wise confidence level
factor levels have been ordered
Fit: aov(formula = Index ~ agency, data = step1)
$agency
diff lwr upr p adj
IN-IG 4.191489 -5.97821882 14.361198 0.7934805
TR-IG 5.521454 -4.64597480 15.688882 0.5742826
CM-IG 6.726776 -3.44768109 16.901233 0.3711336
FM-IG 9.727273 -0.88993588 20.344481 0.0907233
TR-IN 1.329964 0.96003557 1.699893 0.0000000

CM-IN 2.535287 2.00630339 3.064270 0.0000000
FM-IN 5.535783 2.45595128 8.615615 0.0000096
CM-TR 1.205322 0.72214661 1.688498 0.0000000
FM-TR 4.205819 1.13352337 7.278115 0.0017748
FM-CM 3.000497 -0.09498014 6.095974 0.0626431

Conclusion:Observing the p-values obtained above, we can reject the null at 95% confidence interval. We also notice that there is a significant difference of nearly 9 points in the satisfaction index of different agencies, and hence we conclude that the average satisfaction index is not same for all agencies.

3)Variance in the index according to their federal government work experience:

Hypothesis: The satisfactory index is highly correlated with the federal government work experience.

ANOVA Result:

Call:
aov(formula = Index ~ DFEDTEN, data = step1)
Terms:
DFEDTEN Residuals
Sum of Squares 138.92 56557.77
Deg. of Freedom 2 3892
Residual standard error: 3.81206
Estimated effects may be unbalanced

TukeyHSD Result:
Tukey multiple comparisons of means
95% family-wise confidence level
factor levels have been ordered
Fit: aov(formula = Index ~ DFEDTEN, data = step1)
$DFEDTEN
diff lwr upr p adj
C-B 0.1285852 -0.20560739 0.4627778 0.6389915
A-B 0.6651367 0.15214957 1.1781237 0.0067296
A-C 0.5365514 0.07108885 1.0020140 0.0189321

Conclusion:We see that the difference in the satisfaction index is very low and hence we conclude that the satisfaction index does not change according to federal government work experience. So we have to reject the hypothesis that we claimed.

4)Variance in the index as per date of leaving their organization:

Hypothesis:The satisfaction index does not influence employee retention

ANOVA Result:

Call:
aov(formula = Index ~ DLEAVING, data = step1)
Terms:

```
              DLEAVING Residuals
Sum of Squares    462.18  55970.22
Deg. of Freedom        3      3891
Residual standard error: 3.792695
Estimated effects may be unbalanced
```

TukeyHSD Result:

```
Tukey multiple comparisons of means
95% family-wise confidence level
factor levels have been ordered
Fit: aov(formula = Index ~ DLEAVING, data = step1)
$DLEAVING
         diff       lwr       upr     p adj
B-C 0.60996633 -0.4436720 1.6636047 0.4449029
D-C 0.69873589 -0.3804670 1.7779388 0.3430039
A-C 1.31480783  0.3225169 2.3070988 0.0037254
D-B 0.08876956 -0.5231571 0.7006962 0.9823223
A-B 0.70484150  0.2639145 1.1457685 0.0002376
A-D 0.61607194  0.1171313 1.1150126 0.0082614
```

Conclusion:Upon analyzing above results, we can reject the null. We see there is a minor difference in the index levels for A and C, ie satisfaction index for employees who are not planning to leave the organization is slightly greater than the index of employees who are planning to take up ajob outside their organization.

REGRESSION TESTING

From our results in Hypothesis testing, we see that the satisfaction index varies with agency and date of leaving the organization. For this, we have built three regression models. The models and their results are as reported below:

1)Multiple linear regression of model considering agency and date of leaving

Results:

```
Call:
lm(formula = Index ~ agency + DLEAVING, data = step2)
Residuals:
    Min      1Q      Median      3Q       Max
-44.439  -9.812  -3.136   4.469  48.864
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  63.6395     0.6505  97.828  < 2e-16 ***
agencyIN    -10.0660     0.7755 -12.980  < 2e-16 ***
agencyTR     -4.7624     0.7087  -6.720 2.08e-11 ***
DLEAVINGB    -2.4379     0.6715  -3.631 0.000286 ***
DLEAVINGC    -4.8498     1.5101  -3.212 0.001331 **
DLEAVINGD    -2.0421     0.7595  -2.689 0.007200 **
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 14.82 on 3889 degrees of freedom
(78 observations deleted due to missingness)
Multiple R-squared: 0.05338, Adjusted R-squared: 0.05217
F-statistic: 43.86 on 5 and 3889 DF, p-value: < 2.2e-16

Conclusion:

Observing above results we see that the adjusted R squared is 5.2% which indicates weak correlation between satisfaction and agency and Date of leaving the organization, considered together.

We have attempted to remove the outliers and recalculate our model. Following are the results that we got:

Result:

Call:
lm(formula = Index ~ agency + DLEAVING, data = step2)
Residuals:
Min 1Q Median 3Q Max
-44.390 -10.014 -3.060 4.421 48.940

Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) 63.5791 0.6543 97.171 < 2e-16 ***
agencyIN -10.0254 0.7789 -12.871 < 2e-16 ***
agencyTR -4.6955 0.7123 -6.592 4.92e-11 ***
DLEAVINGB -2.4933 0.6737 -3.701 0.000218 ***
DLEAVINGC -4.8692 1.5172 -3.209 0.001341 **
DLEAVINGD -1.9746 0.7634 -2.587 0.009729 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 14.82 on 3857 degrees of freedom
(77 observations deleted due to missingness)
Multiple R-squared: 0.05321, Adjusted R-squared: 0.05198
F-statistic: 43.35 on 5 and 3857 DF, p-value: < 2.2e-16

Conclusion: Removal of outliers did not significantly increase our adjusted R squared value. We have further recomputed the model to include a sample of 200 values and we got following results:

Result:

Call:
lm(formula = Index ~ agency + DLEAVING, data = step2)
Residuals:
Min 1Q Median 3Q Max
-44.390 -10.014 -3.060 4.421 48.940
Coefficients:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  63.5791     0.6543  97.171  < 2e-16 ***
agencyIN    -10.0254     0.7789 -12.871  < 2e-16 ***
agencyTR     -4.6955     0.7123  -6.592 4.92e-11 ***
DLEAVINGB    -2.4933     0.6737  -3.701 0.000218 ***
DLEAVINGC    -4.8692     1.5172  -3.209 0.001341 **
DLEAVINGD    -1.9746     0.7634  -2.587 0.009729 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 14.82 on 3857 degrees of freedom
  (77 observations deleted due to missingness)
Multiple R-squared:  0.05321,           Adjusted R-squared:  0.05198
F-statistic: 43.35 on 5 and 3857 DF,  p-value: < 2.2e-16
```

Conclusion: We see that there is no improvement in the value of adjusted R square. Hence, from the above model, we conclude that there is no significant correlation between the satisfaction Index and factors- agency and date of leaving.

2)Multiple Linear Regression showing relationship between satisfactory index and Q78, Q81, Q82, Q83,Q84

In this model, we have tried to obtain the correlation between satisfaction index and employee feedback for work-life balance activities conducted by their organization. We have formulated a model which is a multiple linear regression model that identifies relationship between index and four of its closely correlated feedback response. Following are the results that we got:

Result:

```
Call:
lm(formula = Index ~ Q78 + Q81 + Q82 + Q83 + Q84, data = step2)
Residuals:
    Min      1Q  Median      3Q     Max
-18.119  -1.212   1.505   2.788  11.279
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.03781    0.55626  19.843  < 2e-16 ***
Q78          0.52214    0.18958   2.754  0.00591 **
Q81          5.02819    0.08494  59.195  < 2e-16 ***
Q82          4.25440    0.07761  54.815  < 2e-16 ***
Q83          3.78540    0.05401  70.082  < 2e-16 ***
Q84          3.64687    0.07454  48.925  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 5.022 on 3909 degrees of freedom
  (25 observations deleted due to missing-ness)
Multiple R-squared:  0.8918,  Adjusted R-squared:  0.8917
F-statistic:  6445 on 5 and 3909 DF,  p-value: < 2.2e-16
```

Conclusion: Observing above results we see that adjusted R square is 85% which shows significant correlation between index and feedback to work-life activity programs conducted by organizations.

CONCLUSION

Summing up all our above analysis, we can say that satisfactory index is not influenced by sex of employees and the number of years they have worked with Federal Government. It seems to vary slightly, nearly 15%, according to agency and their intention to stay or leave the organization. We saw from above results that employees who do not intend to leave the organization have given positive feedback for work-life activity programs conducted. Also, employees in departments like FM and CM seem to be more satisfied by such programs.