REPORT ON IBM JOB SATISFACTION

SUBMITTED BY

DEEPANSHU MAHAJAN
HARIHARAN MURALI KRISHNAN
SAGAR BILLORE
SHAILY SAIGAL

EXECUTIVE SUMMARY

Research published in 2017 from the Smarter Workforce Institute at IBM revealed that the number one reason people left their last job was due to a lack of job satisfaction (40%), with only 14% of respondents saying they had left their previous job because of their manager. The inspiration for this project is taken from the fact that top companies value job satisfaction of the employees as it correlates with the productivity of the employees. So, we want to model the IBM data on employees to model their job satisfaction to understand and quantify the effects of the important parameters like years since last promotion, work-life balance, percent salary hike, Years at company, overtime and training numbers on satisfaction. This analysis aims to provide the HR department a realistic model for job satisfaction and valuable recommendations that could improve the job satisfaction of the employees

Based on the model we want to understand whether increase number of training will improve the job satisfaction of the employee. We also want to suggest the target group for our recommendation from our model. We built an ordinal logistic model to understand the employee satisfaction.

Our analysis suggests that increasing the number of training for employees with less work experience in the company increases their job satisfaction. Other significant findings of the analysis include,
Percentage salary hike is a stronger metric for improving job satisfaction than the monthly income. The percent salary hike if increased by 1% increases the odds of job satisfaction by 2 percent.
Environment Satisfaction was not found very significant parameter for modelling job satisfaction. So the company need not concentrate more on improving on extra facilities to improve environment satisfaction as this doesn't have a very significant effect on the job satisfaction.

TABLE OF CONTENTS

1. PROBLEM SIGNIFICANCE

In the discussions on organizational success, managers often say that employees' morale is one of the crucial factors for success. Employees who have an increased level of job satisfaction love their job and feel secure. This pushes the employee to go the extra mile making them creative, committed and productive. In recent times there have been several types of research to check the effects of job satisfaction level of employees on the organizational behavior, and most of the researches have proved that job satisfaction level has a positive and significant role in the success of an organization.

Recent industry trends have also shown that big organizations analyze job satisfaction level of their employees. The recently released list from Forbes and Business Insider suggests top companies like Facebook, Google, Mckinsey are rated highest in terms of job satisfaction of their employees. This shows the importance these companies give to employee satisfaction in the company. Higher satisfaction metric also drives job seekers to choose a company from multiple job offers. Organizations with lower job satisfaction rate often have negative reviews and job seekers would prefer other options over these organizations. This may lead to organizations loosing talented employees which would certainly affect the performance of the organization. So organizations put a lot of time and efforts in the increasing the satisfaction level of their employees as this increases their productivity and attract more talent.

Our analysis is aimed to provide a detailed report to the HR department of an organization. The analysis is on the job satisfaction level of the employees in the organization for finding out the main factors affecting the job satisfaction level and providing actionable recommendations to improve the job satisfaction level of employees.

2. DATA SOURCE AND PREPARATION

2.1 DATA SOURCE

The authenticity of the data is a very important aspect that has to be considered. The data we used for this project was sourced from IBM Watson. IBM Watson data platform is a collection of services which is used to store, prepare, ingest and analyze data.

2.2 DATA CLEANING AND PREPARATION

The dataset we got had approximately 23500 instances with 37 data attributes. It is based on employee feedback in an organization. The dataset was a .csv file which required data cleaning, for example, some instances had "1234" as educational qualification which required to removed. We worked on the data to remove data inconsistencies and missing values to clean the data as this would impact the performance of the model. Many of the observations had missing values. Also we notice that most of the observation with missing values are the ones which inconsistent and

erroneous values for the variables. Since we had a big dataset for building the model we ignored the observations with missing values which in turn removed most of the inconsistencies in the data.

The dataset contains details about the attributes of an employee which company believes to be influencing parameters to explain the job satisfaction of the employee. These include attributes of an employee which describe his/her work summary in the company and personal details like age, education, salary, work-life balance. We also have data on the environmental satisfaction and relationship satisfaction which the employees have provided as feedback.

We are considering the significance of these attributes for including these attributes to model job satisfaction. We used the attrition attribute to get the list of current employees in the company since we wanted to model the job satisfaction of the current employees. So the data of employees who have resigned and terminated have been removed from the analysis.

The attributes considered for analysis can be classified into 3 types.

a. Continuous like age, salary etc.
b. Discrete ordinal variables: values like ‘1’, ‘2’, ‘3’, and ‘4’ which represented ‘Low’, ‘Medium', 'High' and 'Very High' respectively.
c. Nominal categorical variables like gender.

Apart from these over18 and Standard Hours were not considered in the analysis since they had no variance.

After going through the attributes in the dataset and taking into account the importance of the parsimonious approach for starting an analysis the following attributes were selected for analysis.

Categorical Variables

VARIABLE NAME	DEFINITION/RATIONALE
Job Satisfaction (Dependent Variable)	A measure of workers' contentedness with their job.
Work-Life Balance	Describes the balance between time allocated for work and other aspects of life.
Over Time	If an employee works beyond normal working hours or not.
Job Involvement	Describes the extent to which someone participates in his/her work.
Environment Satisfaction	Represents satisfaction levels of a workplace environment.
Relationship Satisfaction	Level of satisfaction that the employee holds with his peers.

Business Travel	Refers whether an employee travels frequently, rarely or not for business purposes.
Department	A functional area within an organization.
Job Role	Defines what a person does in the company.
Stock Option Level	Grants specified employees of a company the right to buy a certain amount of company shares.

Continuous Variables

VARIABLE NAME	DEFINITION/RATIONALE
Percent Salary Hike	The percentage of increase in salary.
Years Since Last Promotion	The number of years since the employee got the promotion.
Years in Current Role	Number of years the employee spent in the current role.
No. of Trainings Last Year	Training pursued by the employee during the last year.
No. of Companies Worked	Companies the employee worked in previously.
Monthly Income	Income an employee earns in a month.
Distance from Home	The distance an employee has to travel.

3. CORE HYPOTHESES

The data has been cleaned and the appropriate variables have been obtained for building the model. The core hypotheses have been formulated based on the domain knowledge and research done on these attributes and the general effects these are perceived to have on the job satisfaction. These hypotheses will be tested using the model to support our claims

The following hypotheses have been laid out for this project:

1. *H1a:* $\beta_{WorkLifeBalance} > 0$
   Work-life balance has a positive impact on employee's job satisfaction since people with a balanced family and works lives are satisfied with their work and their personal life.

2. *H2a:* $\beta_{PercentSalaryHike} > 0$
   Higher Percent salary hike will make employee more satisfied with the job. Since the percent hike is a gesture from the organization to acknowledge the efforts put into the job.

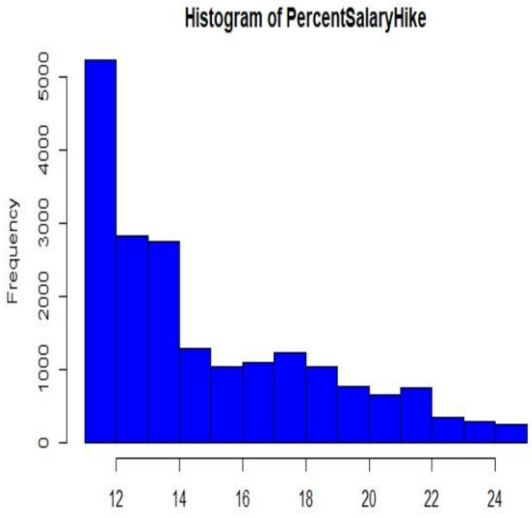
3. *H3a:* $\beta_{YearsAtCompany} > 0$
   Increase in Years at company parameter for an employee suggests that employee is satisfied with job.

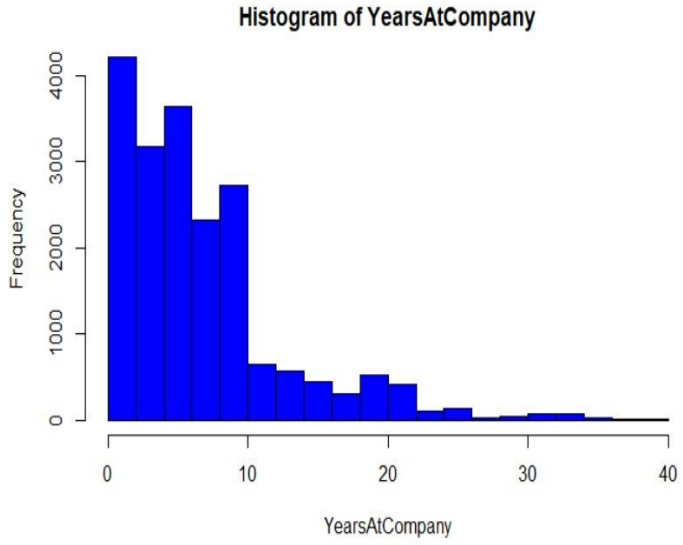
4. *H4a:* $\beta_{Overtime} < 0$
   Overtime has a negative relationship with job satisfaction as employees generally don't prefer to work more than the standard working hours.

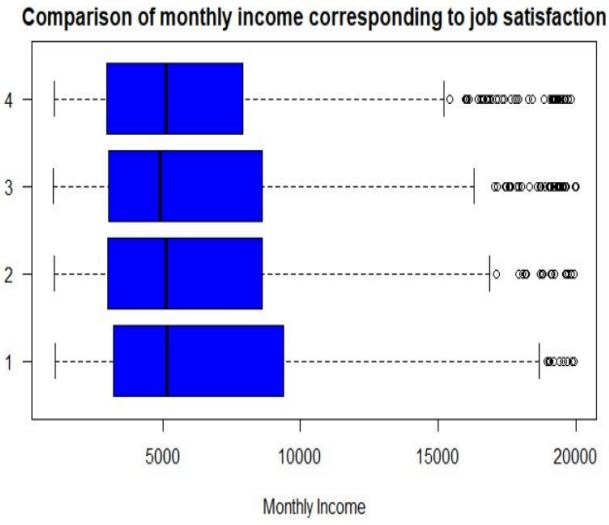
5. *H5a:* $\beta_{NoOfTrainings} > 0$
   No. of Trainings has a positive relationship with job satisfaction. This is based on the rational that a person's satisfaction increases as he/she attends more trainings because this help in better understanding of the significance of their work and motivates them stay appreciate the work they do.

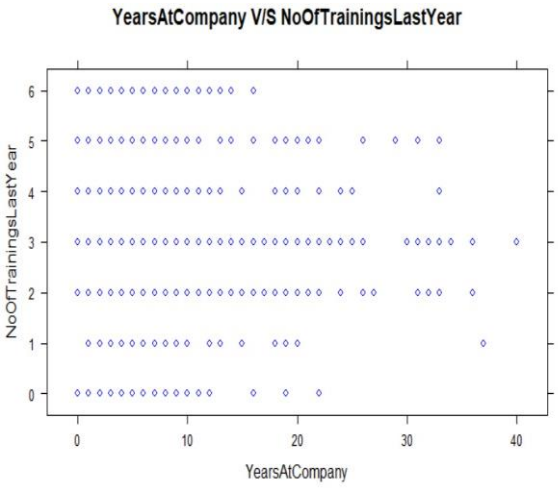
4. DESCRIPTIVE ANALYSIS

The following distributions have been plotted for all these variables or characteristics that are thought to have a potential impact on job satisfaction:









4.1 Percent salary hike

This graph represents how percent salary hike varies among the employees. A high percentage of people get a low hike in their salary whereas very few people get good increments.

4.2 Years at company

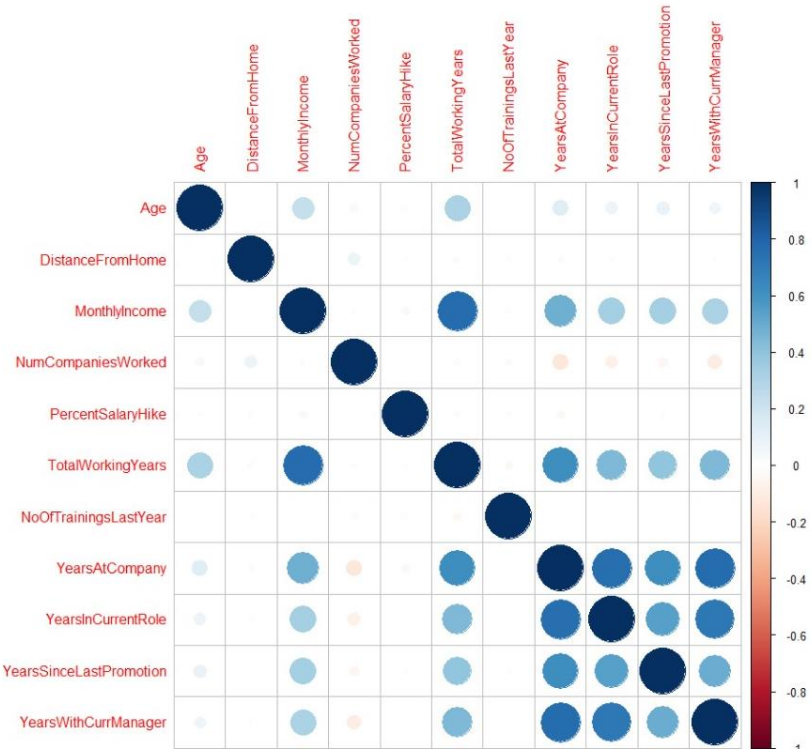
The graph represents that there are many people who work in the company for a short duration of time whereas not many people stay and work in the company for a longer span of time.

4.3 Monthly income and Job satisfaction

The box plot represents the average monthly income does not vary across the different class of job satisfaction.

4.4 Years at company and Number of training since last year

The above graphs depict the number of training an employee has taken last year, it can be observed here that as the experience increases the employees tend to participate in less number of training, whereas less experienced ones are participating in more training.



4.5 Correlation plot

The above plot shows the correlation matrix between different numeric variables. The plot represents that total working year and monthly income, years in current role and years at a company, years with current manager and years at a company are highly correlated.

5. MODELING

5.1 MODEL BUILDING

Since the target variable in our analysis is job satisfaction which is a discrete ordinal variable we have not considered building linear models for analysis. We have used ordinal logistic regression method which best suits out analysis. So Based on the above hypotheses and the descriptive analysis, we have build the following ordinal logistic models.

MODEL 1: Main

JOB SATISFACTION= f (Percent Salary Hike, Work-Life Balance, Number Of Trainings Last Year, Over Time, Years At Company)

MODEL 2: Main + Control

JOB SATISFACTION= f (Environment Satisfaction, Relationship Satisfaction, Percent Salary Hike, Business Travel, Job Involvement, Marital Status, Years In Current Role, Work-Life Balance, Gender, Number Of Trainings Last Year, Job Role, Age, Number of Companies Worked, Distance From Home, Over Time, Stock Option Level, Years Since Last Promotion, Department, Monthly Income)

MODEL 3: Main + Control + Interaction

JOB SATISFACTION= f (Percent Salary Hike , Job Role, Over Time, Years In Current Role, Number of Companies Worked, Number of Companies Worked, Stock Option Level, Marital Status, Age, Work Life Balance*Business Travel, Monthly Income*Years Since Last Promotion, Number Of Trainings Last Year*Years At Company, Environment Satisfaction*Relationship Satisfaction, Work Life Balance*Over Time, Job Involvement*Work Life Balance, Gender*Work Life Balance)

5.2 MODEL COMPARISON

Model 1 was built with the main variables to understand the effects of the main variables that we arrived by comparing co-relation plot and chi-square values. This model helps us to understand the impact these variables have on our dependent variable i.e. Job satisfaction. From the ordinal model, we observe the following points:

- The number of training last year is not significant.
- Overtime is significant and varies directly with job satisfaction.
- Work-life balance and years at the company are significant and are inversely proportional to job satisfaction.

Model 2 was built to understand the impact of main variables when controlled for control variables. This gives us an idea about how our main variables significance and impact job satisfaction. From the model, we observe the following points:

- Overtime and work life are inversely proportional to job satisfaction and not significant.
- A number of training and years a company are not significant and are inversely proportional.
- Percent salary hike is directly proportional and significant.

To make our model more realistic, we took into consideration the following interactions in **Model 3**:

- (Work-Life Balance: Job Involvement): This interaction is considered as high involvement in job my affect the work-life balance.
- (Work-Life Balance: Gender Male): This is included to understand the work-life balance and how it differs for men and women.
- (Work-Life Balance: Business Travel): This is to understand whether business travel affects his/her life outside of work.
- (Monthly Income: Years Since Last Promotion): This interaction is important because if an employee hasn't been promoted for a long and salary is less, then his job satisfaction might get affected.
- (Number of Trainings Last Year: Years at Company) : This interaction is considered as it give insight into how the experience of people in company how can interact with trainings.

This is to understand how the employee attitude towards training change as their term in the company increases.
From the model, the following points were observed:
After considering the interactions, we found that the number of training have become significant and is directly proportional to job satisfaction.

The following metrics have been used to compare all the above three models:

Models	**AIC**	**BIC**	**Pseudo R-squared**
Model 1	52774.27	52837.29	0.004
Model 2	52479.74	52739.69	0.023
Model 3	52368.13	52667.13	0.030

Based on the above analysis and results, model 3 having the lowest values for AIC, BIC and with highest values of pseudo-R-square has been selected and used for predictions.

Hypotheses	**Result**	**Impact**	**Statistical Significance**
H1a: $\beta_{WorkLifeBalance} > 0$	Rejected	-3.921e-01	Yes
H2a: $\beta_{PercentSalaryHike} > 0$	Accepted	1.945e-02	Yes

H3a: $\beta_{YearsAtCompany} > 0$	Accepted	1.993e-02	Yes
H4a: $\beta_{Overtime} < 0$	Rejected	1.025	Yes
H5a: $\beta_{NoOfTrainingsLastYear} > 0$	Accepted	3.067e-02	Yes

KEY FINDINGS

- We find that the percent salary hike is a stronger parameter than monthly income, which influences the job satisfaction level. This is because the higher percent salary hike makes him think that he is being valued in the company.
- In general, there is no gender bias in the company as job satisfaction is not significant for men and women.
- Employees with higher stock option level tend to be more satisfied with the job.
- Environment Satisfaction was not found very significant parameter for job satisfaction.

6. QUALITY CHECKS

In addition to low AIC and BIC values, the model has been built using the Ordinal Logistic Regression approach, the following assumptions need to be satisfied to ensure the quality of the Analysis:

PROPORTIONAL ODDS:

Each independent variable considered in our analysis has an identical effect at each cumulative split of the ordinal dependent variable which is job satisfaction.

INDEPENDENCE CHECK

Since all the observations in the data set are independent and the outcome of any particular employee is not dependent on the outcome of any other employee, this assumption has been satisfied with this analysis. Also to verify if Department in the company introduce hierarchy a random effects model was built. The random effects introduced were not significant. This suggests that there is no hierarchy in the data.

MULTI-COLLINEARITY

To check for multi-collinearity we checked the variation inflation factor for the models and was found to be less than 10 for all predictors Hence, this assumption has also been satisfied.

AUTO-CORRELATION

The analysis has been performed on the employee data sliced at a particular point in time(Cross-sectional data). So there is no Autocorrelation in the data.

7. RECOMMENDATIONS

From our analysis, we may make a few recommendations to the top authorities of the HR department which can be implemented to improve the job satisfaction among their current employees:

- In general, we find from our data that as the no of years spent in the company increases for an employee, the job satisfaction increases. For fresher/ new employees who are willing to learn and develop new skills, the company can increase the number of training in a year which will enhance their productivity and hence, increases their job satisfaction.
- Environment Satisfaction was not found very significant parameter for modelling job satisfaction. So the company need not concentrate more on improving on extra facilities to improve environment satisfaction as this doesn't have a very significant effect on the job satisfaction.