EXPERIMENTS IN BUSINESS ANALYTICS - BUSM160

Topic: Does Children or Family help in overall job satisfaction?

SHALINI NAYAK
QUEEN MARY UNIVERSITY OF LONDON

1. Introduction

This study is based on a research paper "DOES WORKING FROM HOME WORK? EVIDENCE FROM A CHINESE EXPERIMENT*" where the results of a Work From Home experiment at Ctrip were reported. Ctrip is a NASDAQ-listed Chinese Travel Agency with a strength of 16000 employees.

The call centre employees who worked in the airfare and hotel booking departments at Ctrip's Shanghai office volunteered for WFH and were assigned to either work from home or work from the office for 9 months for the experiment. A total of 503 employees volunteered for this experiment, out of which 249 employees who had a good broadband connection, a private room and a tenure of 6 months were made eligible for the experiment. Employees with even birthdays who were selected to work from home acted as a treatment group and employees who have odd birthdays were selected to work from the office acted as a control group. In either case, other than location change all office and home employees had the same IT equipment and work order flow from Ctrip's central server.

The experiment concluded many amazing results, there was a drastic improvement and increase in satisfaction, and productivity and 13% increase in performance, and work hours and a 50% reduction in the attrition level of employees in the treatment group and then the controlled group. Also, employees took lesser also breaks and sick leaves, all because of the convenience of being at home. However, there was one demerit of what this experiment concluded that is promotions dropped by 50% associated with performance.

Elaborating more on satisfaction, the employees WFH has a more positive attitude and a greater level of satisfaction. The data collected from the Chinese experiment are several like performance, attrition, promotions, satisfaction etc. This report shall focus on data analysis and descriptive statistics of satisfaction.

2. Question

The research question for this report is **"Does children improve the satisfaction of employees working from home?".**

There are many factors such as age, gross wage or commute time in minutes which affect the satisfaction levels of employees working from home but these are not focused on in this report analysis. WFH has its own share of merits and demerits. According to the study by Novaes, 2020, employees have reported loneliness and isolation as demerits and others have cited that WFH has been very beneficial to them as they could spend time with their families, and felt less tired because of less travelling and reduced exhaustion saving money and time both.

Work-related concerns, such as a lack of work-life balance, stress, and burnout, can occasionally have a negative impact on family life. If work-family conflicts exacerbate emotional exhaustion, then it's important to consider whether or not having a family can help a worker deal with feelings of isolation, exhaustion, and stress on the job. For the purposes of this research, an employee who has children is considered to have a family.

Hence this report will focus on the analysis and descriptive statistics of how having children will help in the overall satisfaction of an employee.

3. Hypothesis and Methods

With an in-depth study of the research paper by "DOES WORKING FROM HOME WORK? EVIDENCE FROM A CHINESE EXPERIMENT* "Bloom,2014 and data analysis of our Satisfaction dataset provided for this report, we are going to run three pairs of hypotheses and test their significance for our assumptions.

The first pair is to check if children have a significant effect on the satisfaction level of WFH employees. The second pair is to check if the family has an effect on WFH employees and the third pair is to check of working from home has an overall impact on employees than the employees who do not work from home.

1st pair:

Ho: Employees with no children are, on average, more satisfied than employees with children

Ha: Employees with children are, on average, more satisfied than employees with no children

2nd pair:

Ho: There is no significant effect of family on job satisfaction for WFH employees

Ha: There is a significant effect of family on job satisfaction for WFH employees

3rd pair:

Ho: Employees people WFH, on average, satisfied than people not WFH

Ha: Employees people not WFH, on average, satisfied than people WFH

Along with linear regressions and density plots, we have run t-tests for both null and alternative hypotheses in consideration of the p-value. These tests are based on the comparison of within-group variance and between-group variance. If the between-group variance is very large so that there is little or no overlap between groups then it will be reflected by a low p-value. Alternatively, if there is a high within-group variance and low between-group variance then the p-value will be very high in the statistical test.

To ensure that the outcomes are consistent with statistical analysis using the t-test, linear regression on both analyses was conducted.

In this study, employees working from the office, the controlled group, are not considered as it is more likely that they will not be with their children, family or spouses during the time or experiment hence, the employees working from home are also called the treatment group shall be considered on as it is very likely that these employees will be in the nearby vicinity of their children, spouses or families

while working from home during the experiment duration. Employees from the control group are omitted from this study because of various reasons – assuming they are not happy and their results might create a bias with the satisfaction metrics as the academic article by Nicholas Bloom,2014, it is stated that most of them have stated dissatisfaction while working from the office.

The data analysis and descriptive statistics include simple linear regressions, density plots, trend analysis, correlation plots and t-tests for hypothesis testing in R language to check the effect of children on job satisfaction of employees' WFH and check if other factors affecting it, if so. The values of R square have also been taken into consideration to understand the dependent variable which is "satisfaction" in this report.

4. Data Description

The dataset Satisfaction.dta is run in R Studio using the read data function and then converted into a data frame. The necessary packages needed for this analysis are installed. With the view table function, we are able to view the overall summary of our dataset. With experiment data, employees are surveyed five times and there are 855 observations in this dataset. So it's 855/5 that is 171 people whose satisfaction information we have got to analyse. Observations which have five 0s mean the person never works from home.

Statistic	N	Mean	St. Dev.	Min	Max
surveyno	855	3.000	1.415	1	5
satisfaction	855	4.695	1.352	1	7
general	855	72.775	11.742	32	100
life	855	21.460	7.353	2	38
expgroup_treatment	855	0.501	0.500	0	1
age	855	24.661	3.623	18	35
tenure	855	28.289	22.618	2.000	96.000
grosswage	855	3.076	0.810	1.388	6.221
children	855	0.175	0.381	0	1
bedroom	855	0.971	0.169	0	1
commute	855	110.939	62.413	2.500	300.000
men	855	0.468	0.499	0	1
married	855	0.263	0.441	0	1
volunteer	855	0.877	0.328	0	1
high_educ	855	0.357	0.479	0	1
T_pid	855	0.200	0.400	0	1

Figure 1: Summary Statistics of the Satisfaction dataset

The scale of overall satisfaction is from 1-10, where we can see min is 1 and max is 7 with 4.6 as the mean value. . We see that satisfaction score is skewed with scores of 4-6 accounting for 75% of the data, while the maximum goes till 7. The general satisfaction of employees has a range of 32-100 with a mean value of 72.2. From the data, the minimum cost of transportation is 2.50 while the maximum is 300.00.

The variables children, bedroom, commute, gender be it man or woman, married, volunteer, and higher education attainment among others are factored into two binary outcomes as either 0 or 1 where 0 indicates False while 1 is True. The mean of the binary variables indicates the proportion of employees

within the group denoted by 1. For instance, 97.1 per cent of the sample population live in a house with a separate bedroom. 17.5% of the sample study have children, 26.3percent are married and 46.8 per cent are men. It is, therefore, an assumption that 53.2% are women taking into the assumption that this study only identified with the two gender groups. The study introduced an experiment group which accounted for 50.1% of the sample.

We have introduced a new variable in the data set known as the expgroup variable. This variable indicates whether an individual was part of the treatment group and was working at home during the course of the experiment. Using the stargazer out the function we were able to extract the summary dataset.

Statistic	N	Mean	St. Dev.	Min	Max
surveyno	855	3.000	1.415	1	5
satisfaction	855	4.695	1.352	1	7
general	855	72.775	11.742	32	100
life	855	21.460	7.353	2	38
expgroup_treatment	855	0.501	0.500	0	1
age	855	24.661	3.623	18	35
tenure	855	28.289	22.618	2.000	96.000
grosswage	855	3.076	0.810	1.388	6.221
children	855	0.175	0.381	0	1
bedroom	855	0.971	0.169	0	1
commute	855	110.939	62.413	2.500	300.000
men	855	0.468	0.499	0	1
married	855	0.263	0.441	0	1
volunteer	855	0.877	0.328	0	1
high_educ	855	0.357	0.479	0	1
T_pid	855	0.200	0.400	0	1
expgroup	855	0.626	0.484	0	1

Figure 2: Summary of the dataset with exp group. Overall analysis.

We have got 3 variables on the satisfaction that gives us the satisfaction of employees - satisfied, whether generally satisfied or whether they're generally satisfied with their life. The data is taken by giving employees question banks that have a list of questions curated by scientists in their previous work which together indicate whether people are satisfied, whether generally satisfied, and whether they're generally satisfied with their life which generates the score in the table above. Person ID is not considered as a variable and has been omitted it is not a descriptive variable and is a convenient variable. The dependent variable in our study is the “satisfaction” score.

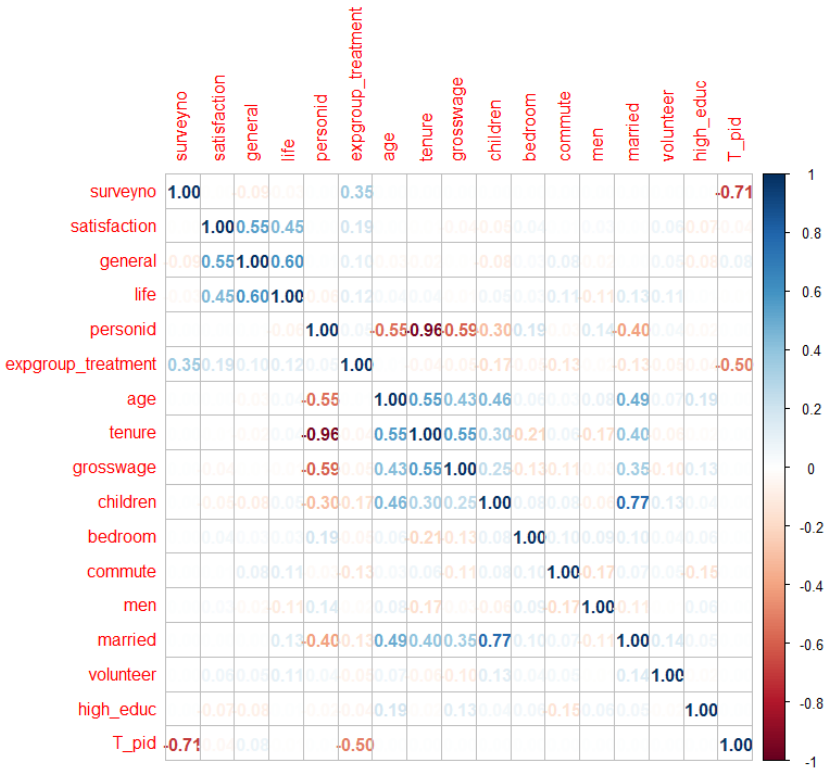
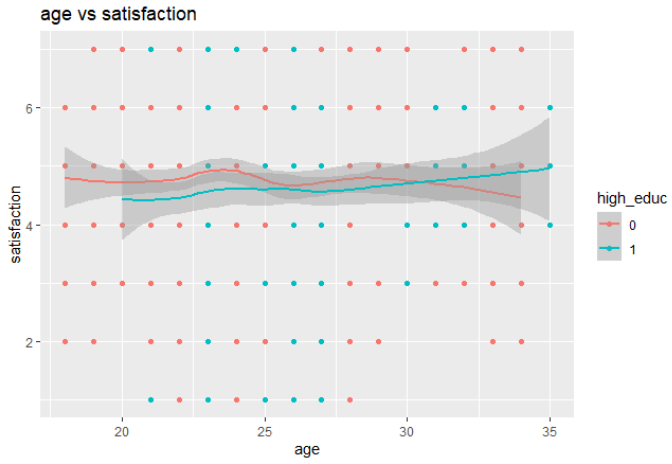


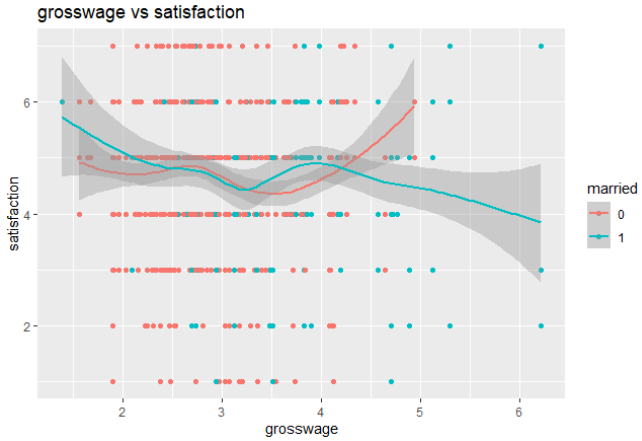
Figure 3: Correlation plot.

We can see that tenure and age (0.55), married and children (0.77) have a very strong correlation. The correlation matrix has a value between -1 and 1 where: -1 indicates a perfect negative linear correlation between two variables. 0 indicates no linear correlation between the two variables. 1 indicates a perfect positive linear correlation between two variables and we can also see general and life are highly positively correlated to satisfaction with values of 0.54 and 0.44. Tenure, gross wage, children, high_educ and T_pid indicate a negative correlation.

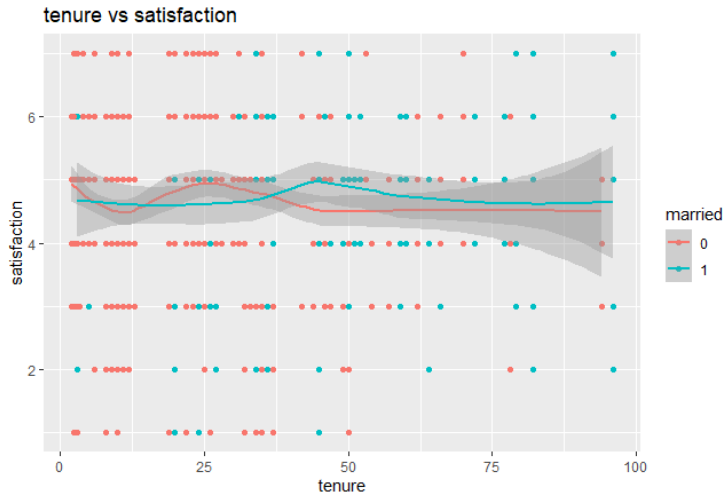


A scatterplot was created between satisfaction and age , we can see in the first graph that even if the age kept increases the satisfaction score on an average ten to be close to 4.5,

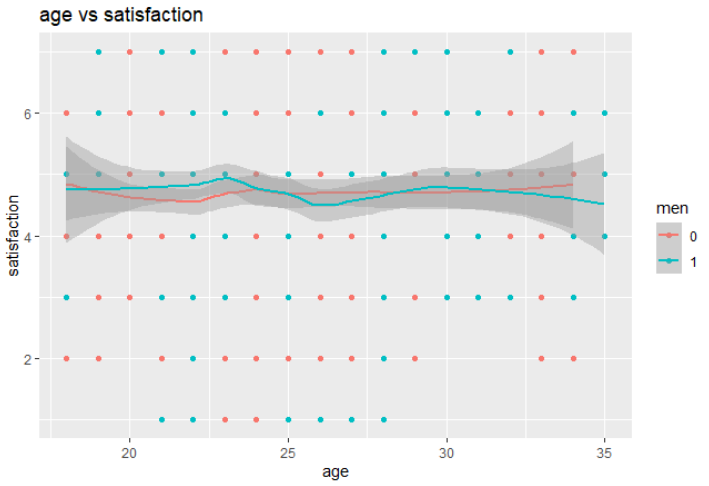
When compared with higher education, the satisfaction score increased when people pursued higher education and gradually decreased for high_educ = 0



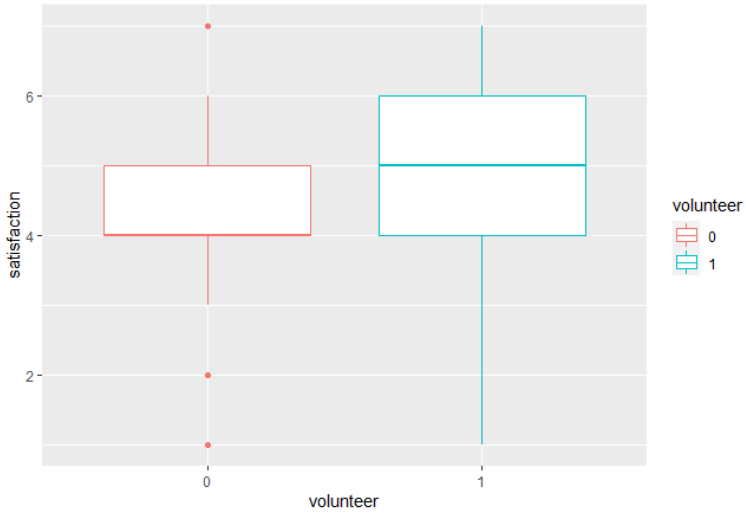
A scatterplot was created between satisfaction and gross wage with married as the group to check the difference in gross wage between married and single with regards to satisfaction. We can see that singles we more satisfied despite the increase in gross wage compared to married.



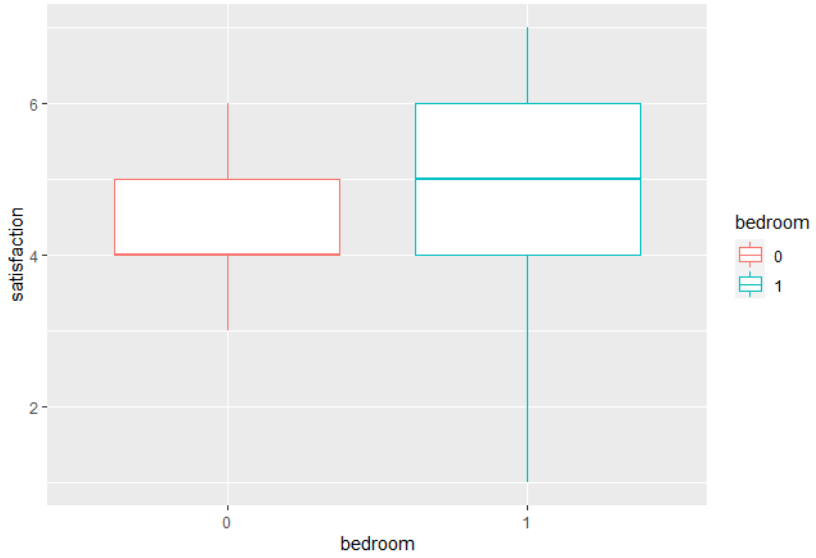
A scatterplot was created between satisfaction and tenure with married as the group to check the difference in tenure between married and single with regards to satisfaction. We can see that as tenure increases both married and single tend to have almost equal satisfaction.



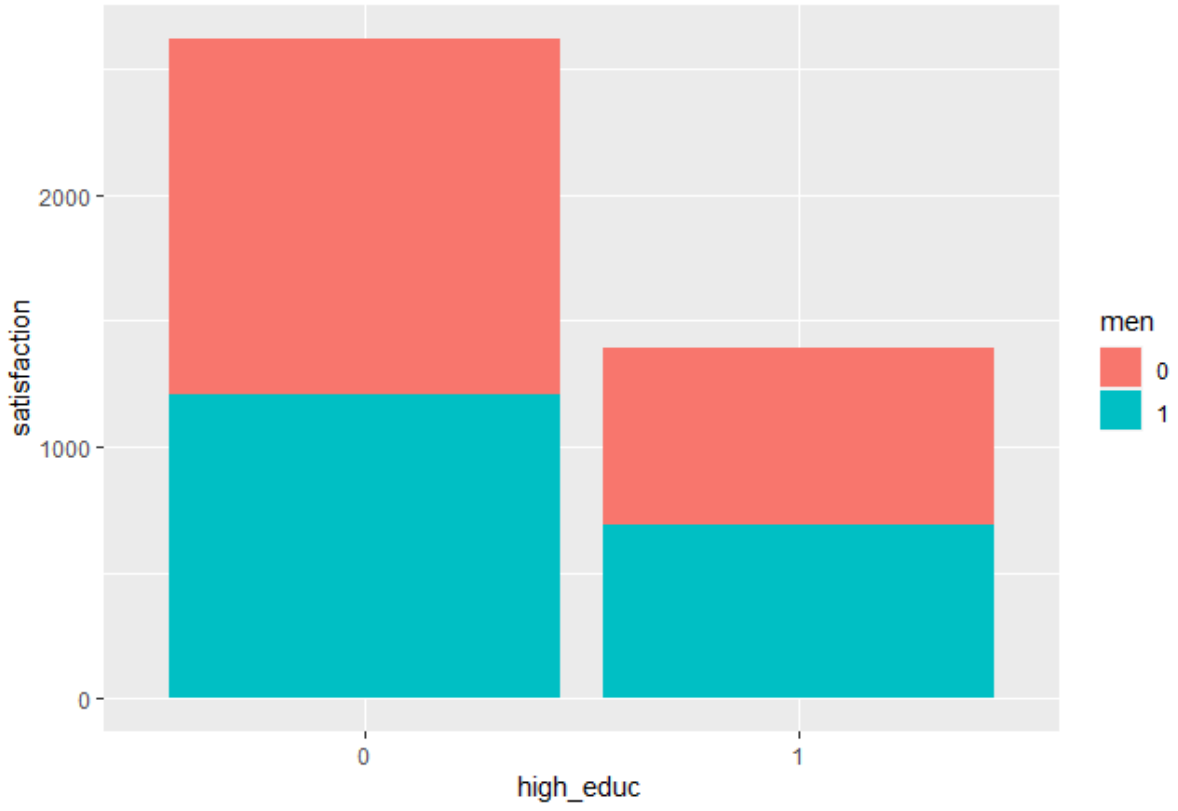
A scatterplot was created between satisfaction and age with men as the group to check the difference in age between men and women with regard to satisfaction. We can see that as age increases both men and women tend to have almost equal satisfaction but with men slightly dropping off their satisfaction



A boxplot was created between volunteer and satisfaction. The plot shows that there are more values for volunteers = 1. We can see that Volunteer = 1 has higher satisfaction score compared to Volunteer = 0, also volunteer = 0 has outliers



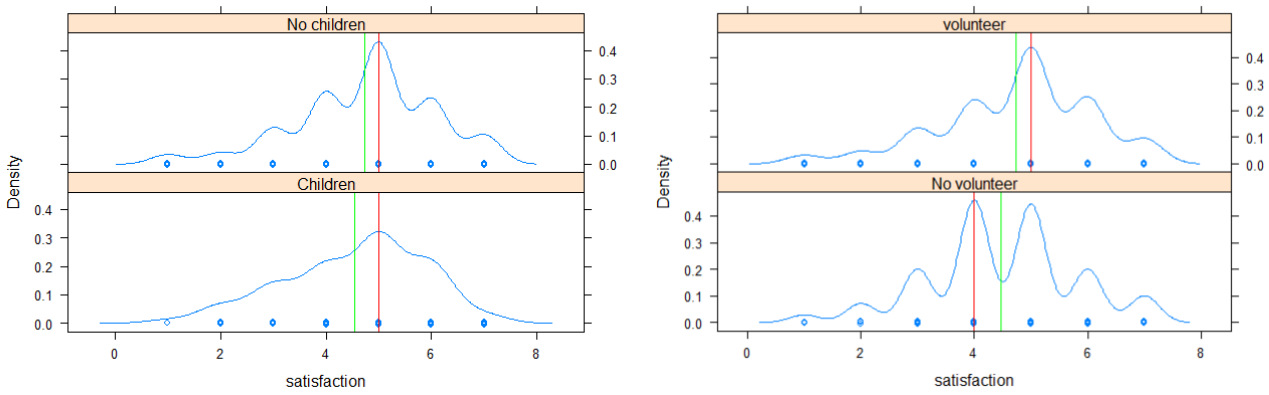
A boxplot was created between bedroom and satisfaction. The plot shows that there are more values for bedroom = 1 . We can see that bedroom = 1 has higher satisfaction score compared to bedroom = 0



A stacked bar chart was created between high_educ and satisfaction with men as a group. There are more people that don't prefer higher education irrespective of gender. The overall satisfaction score for higher education = 0 for both men and women is more when compared with high_educ = 1

5. Data Analysis

We now move to test our hypotheses mentioned above. We try to check our hypothesis with density plots first. A density plot is a smoothed version of a histogram. It uses a kernel density estimate to show the probability density function of the variable, in our case log of satisfaction. The density plot using the complete dataset shows that the mean (green line) and median (red line) of satisfaction are less for employees with children compared to that of employees without children indicating that children have a positive impact on satisfaction. A similar analysis is done on volunteers.



The t-test for both resulted in a p-value of more than 0.05 which means the null hypothesis cannot be rejected and hence employees working from home are more satisfied assumption has been correct. Similarly, people with no children are more satisfied than people with children. The Linear regression performed individual on volunteers and children gives R2 squared very less. The p-value here is very high which states that there is a significant difference in the family on satisfaction about children and volunteer.

	Dependent variable:	
	satisfaction	
	(1)	(2)
volunteer	0.249*	
	(0.141)	
children		-0.180
		(0.121)
Constant	4.476***	4.726***
	(0.132)	(0.051)
Observations	855	855
R2	0.004	0.003
Adjusted R2	0.002	0.001
Residual Std. Error (df = 853)	1.350	1.351
F Statistic (df = 1; 853)	3.135*	2.185
Note:	*p<0.1; **p<0.05; ***p<0.01	

Adding fixed effects that are more variables to the linear regression models does increase the value of R2 value but there is no significant difference in the figure mentioned below.

	Dependent variable:	
	satisfaction	
	(1)	(2)
volunteer	0.273*	
	(0.140)	
children		-0.360*
		(0.188)
married	0.069	0.335**
	(0.105)	(0.162)
high_educ	-0.175*	-0.181*
	(0.095)	(0.095)
expgroup_treatment	0.523***	0.499***
	(0.091)	(0.092)
Constant	4.237***	4.485***
	(0.145)	(0.080)
Observations	855	855
R2	0.045	0.045
Adjusted R2	0.041	0.041
Residual Std. Error (df = 850)	1.324	1.324
F Statistic (df = 4; 850)	10.061***	10.015***
Note:	*p<0.1; **p<0.05; ***p<0.01	

>

6. Data Analysis

The figure shows the density plot for the treatment and control groups which are checked during the hypothesis checking for this report. We can see that the data is skewed, negatively skewed. The plot depicts that having children or a family has a positive effect on the satisfaction of employees working from home. The mean indicated in the green line and median indicated by the red line states that the general satisfaction of employees is greater for those with children than the employees without children.

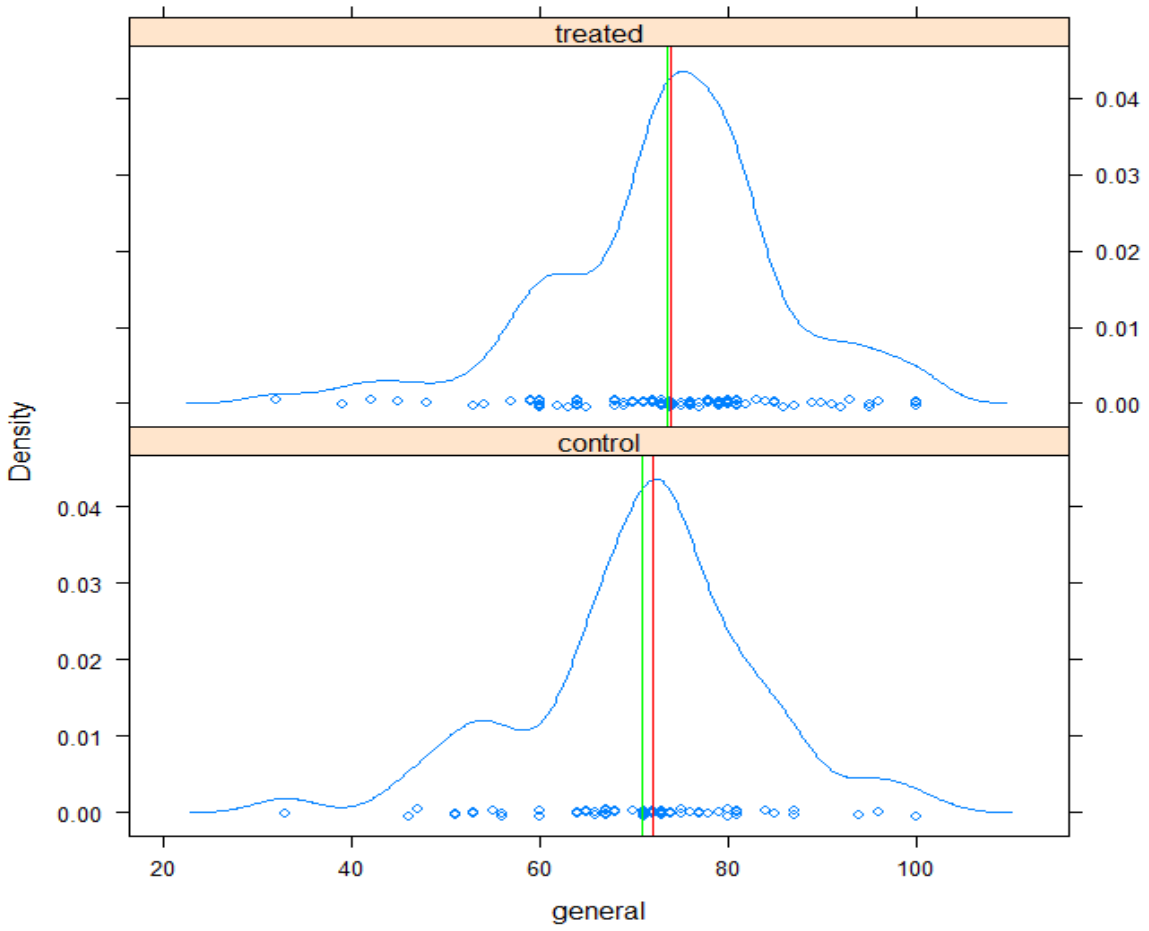


Figure 4: Density plot of treated and control group of Ctrip employees.

These plots do not always give a full indication or total significance of job satisfaction of the treatment and control groups. Due to this reason, we carry out our simple linear regression methods to carry out satisfaction levels with children to filter our results more precisely which will give a clear indication of our analysis.

	Dependent variable:		
	satisfaction		
	(1)	(2)	(3)
expgroup_treatment	0.513***		
	(0.091)		
children1		−0.180	−0.467**
		(0.121)	(0.191)
married			0.321*
			(0.165)
Constant	4.438***	4.726***	4.692***
	(0.064)	(0.051)	(0.054)
Observations	855	855	855
R2	0.036	0.003	0.007
Adjusted R2	0.035	0.001	0.005
Residual Std. Error	1.328 (df = 853)	1.351 (df = 853)	1.349 (df = 852)
F Statistic	31.880*** (df = 1; 853)	2.185 (df = 1; 853)	2.997* (df = 2; 852)
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 5: Effect of the family on job satisfaction with omitted variables

Comparing the effects of the control and treatment groups, it is very important to add interaction terms to the experiment. It is evident that the availability of family and in particular children helps reduce job satisfaction by an average of 0.069. The p-value is greater than 0.01 hence the result is insignificant. Thereby indicating that children have no impact on the job satisfaction of employees WFH. Additionally, married employees have proven to increase job satisfaction for those WFH which does not conform with the initial assumption that marriage and children should all be categorized as a family. Nevertheless, the p-value is way bigger than the significance level and therefore we also do not reject the null hypothesis in this case. In summary, having a family does not have any impact on the job satisfaction level of employees who are working remotely. The inconclusiveness of the study can be explained by the value of the R-squared. The R-squared and adjusted R squared account for less than 5% which indeed shows that only 5% of the independent variable can explain the dependent variable. It is way below the threshold to make a conclusive judgement about the study.

The regression is likely to contain omission bias as some of the other variables have been excluded. Therefore, a run on the regression produces the output below:

	Dependent variable:									
	satisfaction									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
men	0.076									
	(0.093)									
commute		0.0002								
		(0.001)								
bedroom			0.345							
			(0.274)							
general				0.063***						
				(0.003)						
life					0.082***					
					(0.006)					
personid						0.00000				
						(0.00000)				
age							0.001			
							(0.013)			
tenure								−0.001		
								(0.002)		
grosswage									−0.066	
									(0.057)	
volunteer										0.249*
										(0.141)
Constant	4.659***	4.675***	4.360***	0.109	2.930***	4.691***	4.672***	4.711***	4.897***	4.476***
	(0.063)	(0.094)	(0.270)	(0.243)	(0.128)	(0.121)	(0.318)	(0.074)	(0.182)	(0.132)
Observations	855	855	855	855	855	855	855	855	855	855
R2	0.001	0.0001	0.002	0.299	0.200	0.00000	0.00001	0.0001	0.002	0.004
Adjusted R2	−0.0004	−0.001	0.001	0.299	0.199	−0.001	−0.001	−0.001	0.0004	0.002
Residual Std. Error (df = 853)	1.352	1.353	1.352	1.132	1.210	1.353	1.353	1.353	1.352	1.350
F Statistic (df = 1; 853)	0.666	0.060	1.579	364.691***	213.348***	0.001	0.005	0.075	1.320	3.135*
Note:										*p<0.1; **p<0.05; ***p<0.01

Figure 6: Improvement in R square value with fixed effects.

According to the table that is located above, which contains all of the study's data, the level of pleasure felt by WFH workers is increased by a number of other factors, including men, commute, bedroom, age, tenure, gross wage, and volunteerism. In addition, the increase in R squared provides more evidence that the factors are responsible for explaining work satisfaction.

7. Conclusion

This leads us to the conclusion that having children or a family contributes to employee sense of satisfaction and overall fulfilment in work. However, this study does not comment on whether or whether the level of pleasure differs between those who have children and those who do not, or between volunteers who have impairments and those who do not have disabilities.

Therefore the study shows that employees who have children and who are working from home are better satisfied than employees who do not have children. Also having a family or children tends to have company and it helps in reduced loneliness and helps improve job satisfaction of employees.

And if employees who have small children are more likely to be happy because they can be with their children and hopefully have a good holistic parenting and better relationship with their kids as well as work, therefore, having a better work-life balance. This study, therefore, supports the findings of other scholars that loneliness and lack of moral support cause emotional stress and reduce the productivity of employees in general.

The results of this are not so conclusive or obsolete because of the inconsistency of data and a robust conclusion cannot be drawn because there are many factors which influence the overall satisfaction of employees than the factors stated in the dataset. Criteria of health, large or small family, and relationship statuses like married, single or divorced are also very important to determine job satisfaction, not just hypothesis testing and descriptive analysis based on the r square, p, and t-test values.

8. References

Novaes, L. (2020). The pros and cons of working from home: Newspapers illustrators' views on remote work during the pandemic of Covid-19 in Brazil. International Journal for Innovation Education and Research

Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does working from home work? Evidence from a Chinese experiment.