The attrition of Shanghai Ctrip call centre

I. Introduction

Attrition which always is an issue in most companies has become increasingly serious in recent year. When employees choose to work or consider leaving, the salary is not only one of the cause of it. There are numerous aspects lead employee to quit the job such as commute, satisfaction, promotion and so on. Although varieties of studies about attrition already are published, the main reasons behind the employee attrition in every industry or company are also widely different. Therefore, this report would explore the effect of age and marriage on the relationship between attrition in Shanghai Ctrip call centre. Through logistic regression and marginal effects, it finds that age has an important effect on attrition. In contrast, there is no or little effect on having children. In the first section, it will be introduced the definition and the other pieces of literature about attrition. In the second section, three hypotheses based on age, having children and attrition, and the regression models will be proposed. In the third section, it will use tabular and graphical description to explore data. In the last section, it will contain and explain results of analysis based on models.

II. Question

Before developing the hypotheses and methods, “attrition” which formed the main issue of this report should be defined and discussed. According to Business Dictionary (n.d.), attrition means that “the unpredictable and uncontrollable, but normal, reduction of workforce due to resignations, retirement, sickness, or death.” However, in this article, it is assumed that the employee initiatively quit the job without considering the retired, healthy, and death situations. In today’s fiercely competitive business environment, there is no denying that the impact of attrition on a company should be regarded as one of the urgent problems. In addition, the main reason why the attrition be used to analyse is that the turnover rate of Ctrip call centre is significantly higher than in other industries. In China, Hewitt (2017) indicates that the average voluntary turnover rate is 14.9 per cent in all industries. Compared with that, there is about 50 per cent of the turnover rate in Ctrip call centre, which was common in call canter industries (Bloom et al., 2015). Following that, the high percentage of resigning employee brings the cost increasing for Ctrip call centre such as hiring, onboarding and training. Bloom et al. (2015) point out that it costs around eight weeks’ wage to train a new worker, which even does not include the costs of recruiting, screening, interviewing and so on. At the same time, the company’s productivity and efficiency will be decreased until the new employees adapt to the job. Therefore, in order to lower the rate of attrition to alleviate these problems, there have been various relevance literature and analysis on it over the past few years

around the world. Of course, the result of attrition is dependent on a wide range of factors like salary, exhaustion or satisfaction. For example, according to Batt, Doellgast and Kwon (2019) investigates that due to dissatisfaction with the routine and boring standardized procedures, many employees decide to leave the workplace in America call centre. Also, Pawan et al. (2007) discuss that the labour turnover rate is because of the work-life balance and lack of promotion chances in India call centre. Although these pieces of literature show that it is worth exploring this issue from different aspects, it is hard to analyse on each factor through the limited space and dataset in this report. Thus, it is going to focus on whether the age and children have and influence on attrition or not in Ctrip call centre.

III. Hypotheses and Methods

As mentioned above, the hypothetical relationship between age, children and attrition will be created, and then the analytic method will be introduced in this section. The purpose of the research is to help Ctrip's human resource department to reduce the attrition rate and prevent employee turnover when they recruit new staffs. Firstly, for the recruiter, obtaining personal information about age and children are quite easier than performance, satisfaction and so on. If the below assumption is true, the manager can roughly predict the people's probability of quit and decide to let him take office or not. Moreover, if Ctrip tends to hire these people, it can make an intensive study to understand the reason which connects with the age and having children to improve the working environment. As a consequence, we formulated the following three hypotheses to guide the analyses:

Hypothesis 1: Younger employees have higher turnover intentions than older employees.

$$Attrition = \alpha Age + \gamma Married + \varepsilon Men$$

In general, because young workers have more flexibility to choose different careers and have less family or financial obligations, they are more unstable and have higher quit rates in the workplace. Gallup (2016) claims that only 29 per cent of younger people are engaged at work, but 50 per cent of them are open to a new job opportunity. Many firms notice this situation and they say "The young people on my team are great but I can't seem to keep them for long enough. As soon as they start to get really good at something, they leave." (Lambert, 2017).

Hypothesis 2: Employees who have children have higher turnover intentions than those who do not have.

$$Attrition = \beta Children + \gamma Married + \varepsilon Men$$

Having children people usually put more emphasis on work-life balance. They want to have more time to connect and accompany with their children. Hence, if the firm cannot provide enough employee benefits such as parental leave and parental subsidy, the one parent might quit the job for the children. South China Morning Post (2015) reports that around 35 per cent of 6100 families surveyed had a parent staying home to take care of their children. It also indicates that employees needed to more policies to support they could better strike a balance between children and work in China.

Hypothesis 3: Employees who are younger and have children are most likely to quit the job.

$$Attrition = \alpha Age + \beta Childrem + \gamma Marriedgender + \varepsilon Men \quad (1)$$

After following the above hypotheses, hypothesis 3 combines two factors: age and children. If the hypothesis 1 and 2 are true, we can further infer that workers who are younger and have children are most likely to leave.

Next, in order to test these hypotheses, the variables are divided into three categories, and then the three steps to build the regression model.

(i) Dependent Variable

To answer these three hypotheses, the primary dependent variable is attrition that is affected by the independent variables.

(ii) Independent Variables

The independent variables are age and children which are manipulated in these hypotheses.

(iii) Controlled Variables

In previous research shows that individual characteristics are considered as the factor of turnover should be controlled. (Moynihan & Landuyt, 2008 ; Selden & Moynihan, 2000). As a result, the marriage and gender are the control variables.

Initially, using the Ordinary Least Squares (OLS) which is a linear regression model is one of the common methods to build the preliminary regression. However, since the dependent variable which is not the continuous variable takes only two values: leaving or staying in Ctrip call centre, results of the model are not as expected. Subsequently, facing such a binary dependent variable, we decided to try running a logistic regression which can calculate the probability value by a link-function. Through this function, the output will become any real values between 0 and 1. Mathematically, it is called the sigmoid function (Géron, 2017).

$$\text{Sigmoid } \sigma(z) = {e^z}/{1 + e^z}$$

Thirdly, nevertheless, the coefficient estimates in the generalized linear models (GLMs) usually cannot be explained directly. Consequently, marginal effects is often used in research in order to derive substantively meaningful quantities from regression estimates. In this report, we will interpret the logistic regression and refer as the result of this report through average marginal effects.

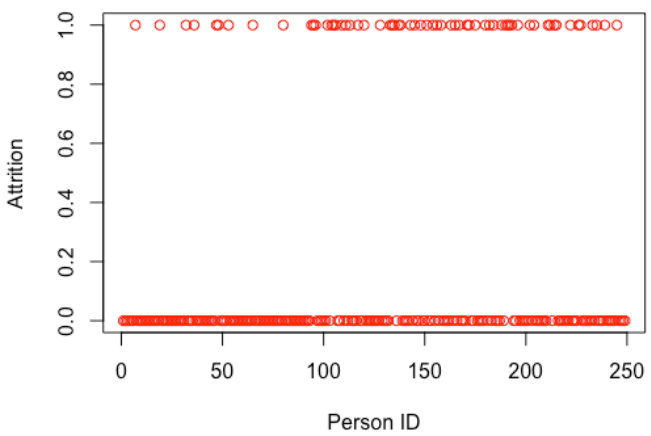
IV. Data Description

However, before building the model, the dataset should be explored and inspected. This dataset is collected from WFH(working from home) experiment in Ctrip's Shanghai call center from 2010 to 2011. At that time, among the 994 employees, the 249 employees participate in the experiment. In Table 1, the brief information of the dataset is given. It can be seen that there are totally 249 observations and 12 variables without any missing values. In addition, the dependent variable, independents and controlled variables are recorded in it.

Table 1 : Dataset Information

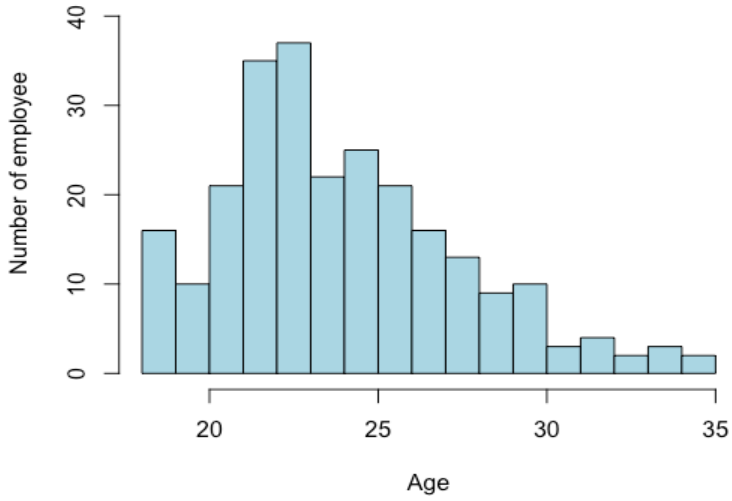
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
personid	249	29,619.550	11,844.940	3,906	19,470	39,748	45,442
quitjob	249	0.249	0.433	0	0	0	1
age	249	24.394	3.545	18	22	26	35
costofcommute	249	8.103	6.958	0	4	10	55
children	249	0.173	0.379	0	0	0	1
expgroup	249	0.526	0.500	0	0	1	1
men	249	0.466	0.500	0	0	1	1
married	249	0.269	0.444	0	0	1	1
perform10	249	-0.034	0.581	-1.260	-0.392	0.366	1.539
perform11	249	-0.202	0.739	-3.031	-0.758	0.275	2.115
perform10_expgroup	249	-0.015	0.448	-1.260	-0.063	0.000	1.539
perform11_expgroup	249	-0.032	0.531	-1.473	-0.123	0.000	2.115

Table 2 : Attrition by Person ID



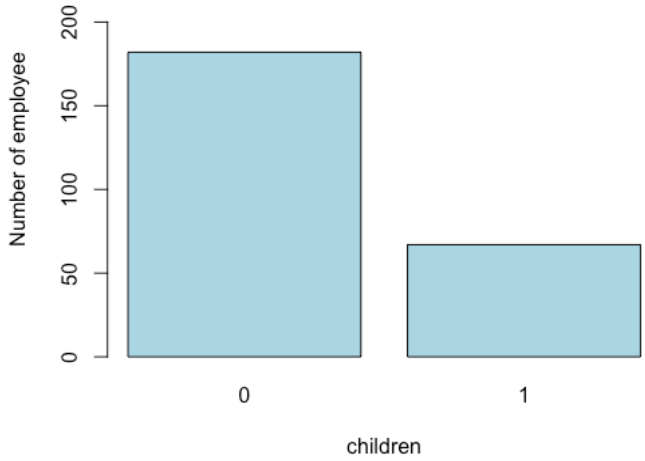
In order to further understand the required variables, the graph of quit job, age and children variables are plotted. In Table 2, it is clear that quitjob variable which takes only two values 0 and 1 is a binary dependent variable. 0 stands for staying the organization, and vice versa. It proves again the regression model should be built by using logistic regression.

Table 3 : Age distribution in company



In Table 3, the age distribution is between 18 and 35 years old in this organization. The largest number of employees who are relatively young are 20 to 25 years old.

Table 4 : Having children distribution in company



In Table 4, because children variable only show two values (0 and 1), it is the binary variable. 0 means that employee does not have children, and vice versa. And the number of employees who has children less than employees do not have.

Furthermore, Table 5 and Table 6 display the attrition situation by age and children. From the per cent of quit and stay by age, because the higher percentage of quit happens between 20 and 25 years old, it seems like the trend is similar to hypothesis 1. However, it does not have an obvious trend in Table 6.

Table 5 : The pecent of quit and stay by age

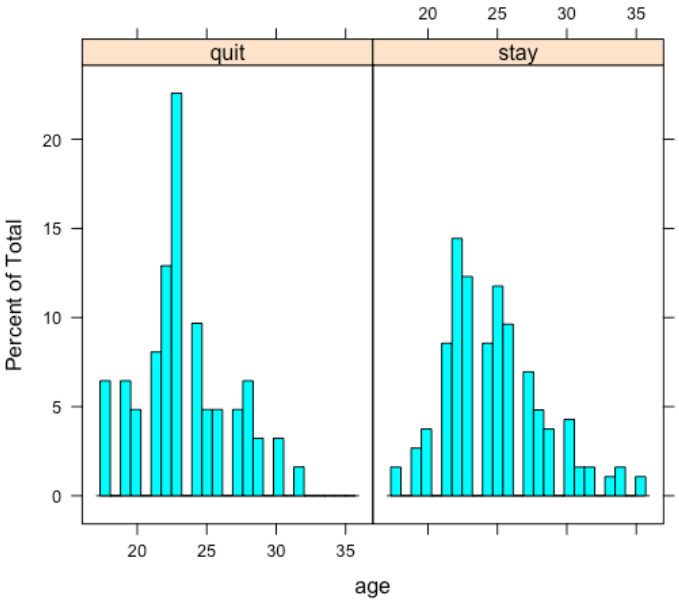
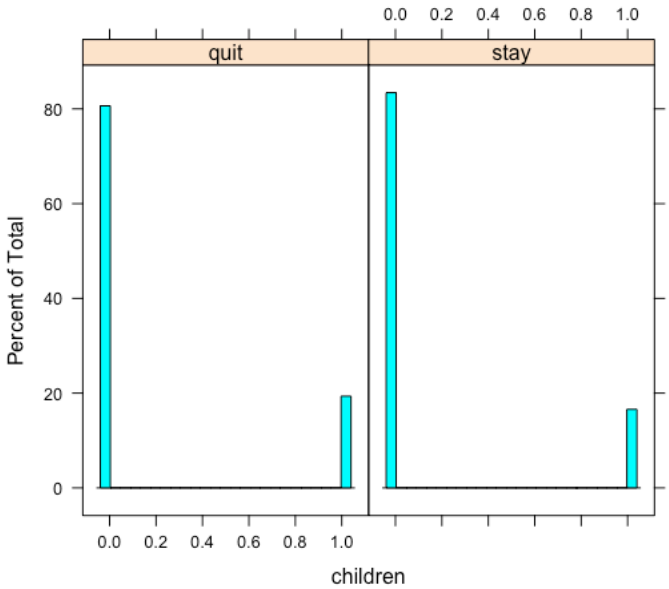


Table 6 : The pecent of quit and stay by children



V. Data Analysis

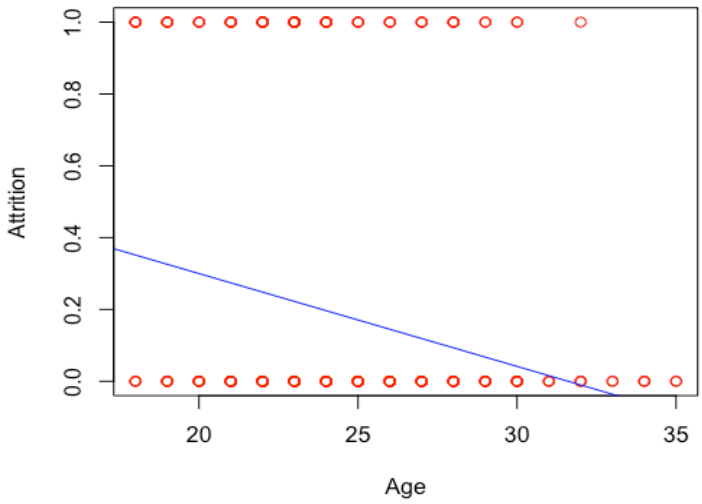
Finally, the model is built with the above-mentioned hypotheses and methods, and the following results which show in the table are obtained and analysed. In Tables 7 and 8, they illustrate that the OLS model is not suitable for this case. Although Column (1) is kept to the hypotheses and function, it can be found that the gap between each

observation and linear regression are quite large and even the values of the regression line are lower than 0 in Table 8.

Table 7 : Attrition in OLS model

	Dependent variable : quitjob		
	(1)	(2)	(3)
age	-0.026***		-0.029***
	(0.009)		(0.009)
children		0.147	0.211*
		(0.111)	(0.110)
married	0.083	-0.121	-0.040
	(0.072)	(0.095)	(0.096)
men	0.087	0.059	0.083
	(0.055)	(0.055)	(0.055)
Constant	0.817***	0.229***	0.893***
	(0.211)	(0.042)	(0.213)
Observations	249	249	249
R2	0.039	0.014	0.053
Adjusted R2	0.027	0.002	0.037
Residual Std. Error	0.427 (df = 245)	0.433 (df = 245)	0.425 (df = 244)
F Statistic	3.292** (df = 3; 245)	1.132 (df = 3; 245)	3.408*** (df = 4; 244)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 8 : Attrition by Age in OLS model



Thus, three models which are re-established are tested whether the influence of age and children on attrition existed by running the logistic regressions in Table 9. Theoretically, the results would be reasonable to use logistic regression on the binary dependent variable. To understand the intuitive results, in Table 10, 11, 12, the average marginal effects (AME) of the logistic model were calculated. The result of hypotheses would depend on it.

Table 9 : Attrition in Logit model

	Dependent variable : quitjob		
	(1)	(2)	(3)
age	-0.154***		-0.176***
	(0.055)		(0.057)
children		0.959	1.372*
		(0.706)	(0.733)
married	0.483	-0.827	-0.372
	(0.409)	(0.644)	(0.666)
men	0.492	0.317	0.469
	(0.307)	(0.297)	(0.308)
Constant	2.239*	-1.220***	2.751**
	(1.243)	(0.233)	(1.292)
Observations	249	249	249
Log Likelihood	-134.622	-137.892	-132.605
Akaike Inf. Crit.	277.244	283.783	275.209
Note:		*p<0.1; **p<0.05; ***p<0.01	

In our hypothesis, the first one emphasizes that the younger worker more easily to quit the job than the older. Table 10 shows the average marginal effects of age, married and men on attrition in AME column and their P-value in P column. As hypothesis 1, the age and attrition present a negative correlation and get 0.0034 in p-value. When P < 0.01 which means 1% significance, it proves that hypothesis 1 is valid. In addition, in AME, it shows that if age increase by one unit, the average attrition decrease 0.0277. In this model, age is a significant factors in this call centre.

For younger people, this job might too standard and no change which is mentioned previously, so they decide to leave.

Table 10 : Age and Attrition

```
summary(logit1_m)
 factor     AME     SE       z      p   lower   upper
    age -0.0277 0.0095 -2.9279 0.0034 -0.0463 -0.0092
married  0.0867 0.0729  1.1890 0.2345 -0.0562  0.2295
    men  0.0884 0.0543  1.6280 0.1035 -0.0180  0.1949
```

In Table 11, we test the second hypothesis which presents that having children workers are more likely to quit. However, it cannot be seen any significance this time. Our hypothesis 2 which is likely wrong is rejected. This might be because this dataset comes from the outcome of the WFH experiment. It could be argued that when parents can work at home, they might do not need to quit for children. Hence, this factor does not have an impact on attrition.

Table 11 : Children and Attrition

```
summary(logit2_m)
  factor     AME     SE       z      p   lower  upper
children  0.1769 0.1293  1.3679 0.1713 -0.0766 0.4304
 married -0.1525 0.1179 -1.2927 0.1961 -0.3836 0.0787
     men  0.0584 0.0544  1.0734 0.2831 -0.0482 0.1650
```

In the last model, we further join the age and children factors to test hypothesis 3. In Table 12, values are different from before. Interestingly, the age's p-value maintains the 1% significance, but the children's p-value gets the 10% significance. In the AME column, we also find that the factor of having children has a positive effect on this regression. If the children variable increase by one unit, the average attrition increase 0.2423. And the average marginal effect of age becomes stronger from -0.0277 to -0.0311. Consequently, the hypothesis 3 is also valid. Younger and having children employees are more likely to quit the job.

Table 12: Children and Attrition

```
summary(logit3_m)
  factor     AME     SE       z      p   lower   upper
     age -0.0311 0.0096 -3.2530 0.0011 -0.0499 -0.0124
children  0.2423 0.1272  1.9057 0.0567 -0.0069  0.4916
 married -0.0657 0.1176 -0.5587 0.5764 -0.2962  0.1648
     men  0.0829 0.0537  1.5421 0.1231 -0.0225  0.1882
```

VI. Conclusion

In conclusion, most companies are facing a higher turnover rate, especially in the call centre industry. Because of this feature, we have more interest in the reason for this situation. This report is paid attention to the relationship between age, having children and attrition. We hope that it can help the human resources department to easily identify suitable employees and reduce the cost of staff turnover through this report. Due to the binary dependent variable in this dataset, we should utilise the logistic regression instead of OLS regression and calculate the average marginal effect. For Shanghai Ctrip call centre, we find that attrition changes with staff's age. In our model, when the age increases one unit, the attrition decrease 0.0277 or 0.0311. In contrast, the effect of having children is uncertain in the model. However, this report does not take the WFH experiment into account. whether the bias exists should be tested deeply. Although attrition must be driven by several reasons, this is a direction that Ctrip can put more effort on the turnover rate of younger staff.

Appendix:

```
rm(list=ls())

### Packages
install.packages("margins")    # to calculate marginal effects
library(foreign)
library(stargazer)
library(haven)
library(ggplot2)
library(Hmisc)
library(chron)
library(lattice)
library(dummies)
library(lfe)
library(sandwich)
library(lmtest)
library(miceadds)
library(multiwayvcov)
library(margins)

### load data
data1 <- read_dta("~/Desktop/Working from home-20190417/Attrition.dta")
View(data1)

### explore data
names(data1)
data2<-data.frame(data1)
stargazer(data2,type="text",title =,align=TRUE)

### explore the data graphically
plot(data2$quitjob,xlab ="Person ID", ylab = "Attrition", main= "Table 2 : Attrition
by Person ID",col="red")
#Dependent variable(quitjob) takes only two values 0 or 1.

hist(data2$age,xlab ="Age", ylab = "Number of employee", ylim=c(0,45),main=
"Table 3 : Age distribution in company ",col="light blue",breaks=24)
#hist(data2$children,xlab ="children", ylab = "Number of employee", main= "Table
4:Having children distribution in company",col="light blue",breaks =10 )
```

```
barplot(table(data2$married),col="light blue",ylim=c(0,249),xlab ="children", ylab =
"Number of employee", main= " Table 4 : Having children distribution in company")

data2<-within(data2,{D_quit<-ifelse(quitjob==1, "quit", "stay")}) #create dummy
histogram(~ age | D_quit, data=data2, nint =25,main="Table 5 : The pecent of quit
and stay by age") #young people more quit.
histogram(~ children | D_quit, data=data2, nint =25,main="Table 6 : The pecent of
quit and stay by children")

### run OLS
ols1<-lm(quitjob ~age+married+men    , data=data2) #negative effect
ols2<-lm(quitjob ~children+married+men    , data=data2) #positive effect
ols3<-lm(quitjob ~age+children+married+men    , data=data2)
stargazer(ols1,ols2,ols3,type="text", align=TRUE,title = "Table 7 : Attrition in OLS
model")
stargazer(ols1,ols2,ols3,type="html",align=TRUE,out="OLS.htm",title = "Table 7 :
Attrition in OLS model")
#control wheather sb is a man or marriage.

plot(data2$age,data2$quitjob,xlab ="Age", ylab = "Attrition",col="red",main =
"Table 8 : Attrition by Age in OLS model")
abline(ols1,col="blue")

### run logit
logit1<-glm(quitjob ~ age+married+men, data=data2,family=binomial(link="logit"))
logit2<-glm(quitjob ~ children+married+men,
data=data2,family=binomial(link="logit"))
logit3<-glm(quitjob ~ age+children+married+men,
data=data2,family=binomial(link="logit"))
stargazer(logit1,logit2,logit3,    type="text", align=TRUE)

### marginal effects
logit1_m<-margins(logit1)
logit2_m<-margins(logit2)
logit3_m<-margins(logit3)
summary(logit1_m)
summary(logit2_m)
summary(logit3_m)
```

Reference:

Batt, R., Doellgast, V. and Kwon, H. (2019). [online] Digitalcommons.ilr.cornell.edu. Available at: http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1005&context=cahrswp [Accessed 23 Apr. 2019].

Bloom, N., Liang, J., Roberts, J. and Ying, Z. (2015). Does Working from Home Work? Evidence from a Chinese Experiment *. *The Quarterly Journal of Economics*, 130(1), pp.165-218.

Budhwar, P. S., Varma, A., Singh, V., & Dhar, R. (2006). HRM systems of Indian call centres: an exploratory study. *The International Journal of Human Resource Management*, *17*(5), 881-897.

Business Dictionary. (n.d.). *What does attrition mean? definition and meaning*. [online] Available at: http://www.businessdictionary.com/definition/attrition.html [Accessed 25 Apr. 2019].

Gallup (2016). *Millennials: The Job-Hopping Generation*. [online] Gallup.com. Available at: https://www.gallup.com/workplace/236474/millennials-job-hopping-generation.aspx [Accessed 25 Apr. 2019].

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. [S.l.]: O'Reilly Media, Inc.

Hewitt, A. (2017). *China Posts an Average Salary Increase Rate of 6.7% and Turnover Rate of 20.8% in 2016 - HR in ASIA*. [online] Hrinasia.com. Available at: http://www.hrinasia.com/news/china-posts-an-average-salary-increase-rate-of-6-7-and-turnover-rate-of-20-8-in-2016/ [Accessed 25 Apr. 2019].

Lambert, M. (2017). *The Crisis of Purpose: Why Your Young People Are Leaving*. [online] Bts.com. Available at: https://www.bts.com/blog-article/business-insight/crisis-of-purpose-why-your-young-people-are-leaving [Accessed 24 Apr. 2019].

Moynihan, D. P., Landuyt, N. (2008). Explaining turnover intention in state government: Examining the roles of gender, life cycle, and loyalty. Review of Public Personnel Administration, 28, 120-143.

South China Morning Post. (2015). *Home freer: Chinese mothers quit jobs to care for the kids*. [online] Available at: https://www.scmp.com/news/china/society/article/1876853/home-freer-chinese-mothers-quit-jobs-care-kids [Accessed 24 Apr. 2019].

Selden, S. C., & Moynihan, D. P. (2000). A model of voluntary turnover in state government. *Review of public personnel administration*, *20*(2), 63-74.