Employee retention analysis using a logistic regression model based on Maslow's hierarchy of needs implemented using SPSS and Tableau.

An employee is an important asset and part of any business. Losing an employee equals to loss in terms of earlier investment made, training and talent, time and money spent on the employee and also another problem would be finding a suitable replacement. There can be a majority of factors or motivators which can influence an employee's decision to leave. In the project these factors are categorized as per Maslow's hierarchy of needs and a logistic regression model is applied to the various hierarchies to understand individual factors which stand out and their effects on attrition. The model can be used to analyze relationships between the employer and the employee or can also be used to understand motivational theories in practice. The results can be used to point out areas where the business is doing good as well as bad when it comes to employee work experience and further well-balanced decisions can be made by the management. More specifically depending on the HR and management, importance of certain needs can be understood, and decisions can be taken accordingly where the most important attributes can be paid more attention to improve employee satisfaction, performance and attrition ultimately benefitting the business by having motivated employees.

Keywords - Employee motivation theory, Logistic regression and hierarchy of needs model.

I. INTRODUCTION AND BACKGROUND OF DATASET

The relationship between an employer and an employee is important as it can affect the business in many possible ways. To understand this relationship and the underlying factors which affect employee motivation is best for better business and employee performance.

To better understand the motivators, Abraham Maslow one of the most influential phycologists of the twentieth century categorized the motivators on the basic of five categories of needs which are starting with the most important and ***basic physiological needs, safety needs, social needs, esteem needs and lastly lesser important self-actualization needs*** *[1]. Going forward we shall try to implement an employee motivating model/ attrition reducing model based on this hierarchy of needs.*

The dataset chosen is from Kaggle showing various attributes related to HR Analytics. The facts and numbers contained in the dataset is a fictional one simulated by IBM scientists for research purposes to understand various factors related to attrition and employee performance.[2] Even though the dataset is fictional the attributes contain factors which organizations conduct as a part of yearly feedback and also when employees decide to leave while submitting a final feedback for the organization.

The data dictionary picked from [2] are as follows: Dependent categorical variable – Attrition- 'yes' suggesting the employee quit the job and 'no' suggesting that the employee is still part of the business.

Independent predictor variables are as follows:

Group of attributes in relation to ***physiological needs*** *of an employee:*

1. *Age- age of the employee.*
2. *Marital Status- whether the employee is married, single or divorced.*
3. *Overtime - if employee works overtime or not.*
4. *Work life balance- score rating from 1 to 4, showing 1 as lowest and 4 as the highest score.*

Group of attributes in relation to ***safety needs*** *of an employee:*

5. *Monthly income- monthly salary.*
6. *Stock option level-varying from 0 to 3 with 0 being the lowest stock level and 3 being the highest.*
7. *Hourly rate-hourly income of employee*

Group of attributes in relation to ***social needs*** *of an employee:*

8. *Job involvement-varying from the lowest score of 1 to highest score of 4.*
9. *Relationship satisfaction- varying from the lowest score of 1 to highest score of 4.*
10. *Business travel-whether the employee travels frequently, rarely or not at all because of business.*

Group of attributes in relation to ***esteem needs*** *of an employee:*

11. *Years at current role-indicates the number of years spent by employee in the current role*
12. *Years since last promotion-number of years since the employee was promoted.*
13. *Performance rating- rating given to employee at the time of promotion varying between 3 and 4.*
14. *Years under current manager- number of years spent by employee under current manager*

15. *Job level-Varies from the lowest score of 1 to highest score of 5 for particular departments*

*Group of attributes in relation to **self-actualization needs** of an employee:*

16. *Training- indicates the number of times the employee underwent training varying from 0 to 6.*

17. *Environment Satisfaction- indicates the score representing the atmosphere, environment or work culture with 1 being the lowest and 4 being the highest score.*

18. *Job Satisfaction- score rating from 1 to 4, showing 1 as lowest and 4 as the highest score.*

II. *RELATED WORK*

The project is being done based on previous similar research conducted in the fields of HR analytics, employee retention/attrition and employee motivation/theories. However most of the research is done based on predicting attrition or which employee is the most likely to leave.

Research [3] has been conducted on the datasets for prediction of employee turnover or prediction of what kind of employees are likely to be quitting using classifier techniques such as Random forest, gradient boosting trees, K- nearest neighbors and logistic regression. Based on results random forest was found out to be the best technique for prediction whereas attributes heavily influencing attrition were found to be job satisfaction, time spend at work, number of projects worked on and performance evaluations.[3]

HR analytics uses employee data to make better management decisions while also aiming at improving the employee performance [4]. In this research predictive analytics have been used again to identify employees more inclined to leave. Logistic regression in SAS was used to carry out the research with attrition as the dependent variable. Conclusions were made that job satisfaction and performance rating scores were the most impactful attributes with attrition and also that employees having low salary while spending the most time at work are the most likely to be quitting the job.

In another research [5] more relevant to analysis of employee and employer relationships, employee turnover was compared to various satisfaction needs and financial needs of a small-scale telephone company in Pakistan. The needs were classified into a hierarchy based on Maslow's theory and based on survey results it was found that happy employees were the best ambassadors for the company ,also if employees are required to go an extra mile for the customers it also is likely that the company will have to do the same for the employee in terms of monetary needs.

III. *HYPOTHESIS AND METHODOLOGY*

Hypothesis proposal:
Can a logistic regression model based on Maslow's hierarchy of needs be used to understand relationship between attrition/turnover and attributes under each hierarchy which in turn also affect performance, employee relationship and satisfaction?

Methodology:
The dependent variable would be attrition which is a dichotomous variable having outcome 'yes' suggesting the person has left or is leaving and 'no' suggesting the person is still a part of the employee. The independent variables are a group of variables based on each hierarchy of need.

Before labelling attributes into categories, the first step was to understand the concept behind the employee motivation hierarchy of Maslow.

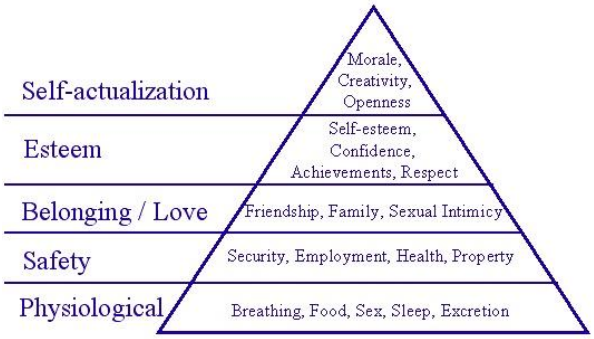

Figure: Hierarchy of need's [1].

As per this theory [6] the most basic needs or physiological needs for an employee would be aspects that would help the person survive such as air, food, shelter, sleep and sex. This would include factors such as comfortable working conditions and reasonable work hours. The attributes in this category of basic needs would be 1. Distance from home 2. Marital Status 3. Overtime 4. Work life balance.

The second most important level of needs is one related to safety [6] that provide a person with a sense of security which could be related to health and wealth. The attributes in this category are 1. Monthly income 2. Daily rate 3. Hourly rate 4. Stock level option.

The third level of need is one of social needs which related to feelings of love and belonging. [6] These needs are important so that people do not feel alone and depressed. Healthy work practices such as cooperative teamwork and friendly accessible supervisors can affect the employee in a positive way. The attributes in this category are 1. Job Involvement 2. Relationship Satisfaction 3. Business travel.

The fourth level is Esteem needs reflecting on self-esteem and respect [6]. Esteem needs can be met by providing recognition and awards when employees do well, and also providing promotions and additional responsibility to make the person feel he is important and of value to the business.

The attributes in this category are 1. Years in current role 2. Years since last promotion 3. Job Level.
The final needs would be self-actualization needs related to individual person needs and relates more to ambition and self-feelings of satisfaction and content. The attributes in this category are 1. Training 2. Job Satisfaction 3. Environment Satisfaction.

As per the theory once a need is satisfied it ceases being a motivator and aiming at fulfilling the next need begins unless the fulfillment is threatened.[7] The theory was based on American culture and the hierarchy may not follow the same pattern always. It can vary from culture to culture and also across individuals based on their personal choices however it remains the same in most cases.

After this a logistic regression was run on SPSS on the attributes under each 'need' against the dependent variable 'attrition' recoded as 'attrition rate', to understand the correlation between needs and attrition.

As per Maslow's theory in most ideal situations if the most basic needs were not fulfilled it would lead to more attrition or lead to employees quitting and if more of higher order needs such as self-actualisation needs were not fulfilled it would have comparatively lesser effects on attrition. This may always not be the case as particular attributes under each need may be more correlated to attrition and have a higher significance compared to other attributes. Going forward we shall see what can be interpreted from the results.

IV. *FINDINGS AND VISUALISATIONS*

Logistic regression is used for finding the strength of a relationship between dependent variable and an independent variable while controlling the effects of other variables.

While running logistic regression in SPSS the following output was processed:

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	1470	100.0
	Missing Cases	0	.0
	Total	1470	100.0
Unselected Cases		0	.0
Total		1470	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Based on figure above we can see that there is a large sample size ideal for logistic regression and no missing values as well.

Also, Attrition is coded as yes = 1 for employees left or leaving the business and coded as no=0 for employees still part of the business.

Block 0: Beginning Block

Classification Table[a,b]

			Predicted		
			AttritionRate		Percentage Correct
	Observed		No	Yes	
Step 0	AttritionRate	No	1233	0	100.0
		Yes	237	0	.0
	Overall Percentage				83.9

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.649	.071	540.645	1	.000	.192

The above table Block 0 baseline model does not contain the independent variables and shows predictions are done based on the more frequently occurring category that is "No". Percentage shows that this approach to prediction without any variables is correct 83.9% of the time. This is also a significant predictor of the model since (p<.001)

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8.759	8	.363

Going forward the Hosmer and Lemeshow test shows that the model once in place is significant where p value is .363 and greater than .05 showing a significant model.

Classification Table[a]

			Predicted		
			AttritionRate		Percentage Correct
	Observed		No	Yes	
Step 1	AttritionRate	No	1221	12	99.0
		Yes	211	26	11.0
	Overall Percentage				84.8

a. The cut value is .500

Also, the accuracy of the model has increased from 83.9 to 84.8 after the dependent variables have been put in place.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	DistanceFromHome	.028	.009	10.056	1	.002	1.029
	MarriageStatus			50.845	2	.000	
	MarriageStatus(1)	.248	.226	1.204	1	.273	1.281
	MarriageStatus(2)	1.261	.222	32.226	1	.000	3.529
	Overtime1(1)	1.428	.154	86.465	1	.000	4.172
	WorkLifeBalance			18.529	3	.000	
	WorkLifeBalance(1)	-.917	.304	9.123	1	.003	.400
	WorkLifeBalance(2)	-1.173	.282	17.302	1	.000	.310
	WorkLifeBalance(3)	-.756	.346	4.785	1	.029	.470
	Constant	-2.083	.336	38.453	1	.000	.125

a. Variable(s) entered on step 1: DistanceFromHome, MarriageStatus, Overtime1, WorkLifeBalance.

The most important table for interpretation of results and for showing the joint association of attributes of basic needs with attrition is variables in the equation table as shown above or for showing the statistical significance of each predictor variable.

*In **Basic needs** hierarchy as per the table above:*

For one unit increase in distance from home, the odds of a person leaving the job increases by a factor of 1.029(exp[B]) or we can also say the odds of a person leaving is higher for a person who stays farther away from work.

Also, an additional hour of work or 1 hour of extra time at work increases the odds of a person leaving by a factor of 4.172 and is the biggest influencer in relation to attrition.

The major influencers in general, affecting attrition are overtime, marriage status and distance from home for basic needs.

*Running the regression analysis again for **safety needs** we get the below results of variables in the equation table.*

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	StockOptionLevel			50.966	3	.000	
	StockOptionLevel(1)	.417	.304	1.881	1	.170	1.517
	StockOptionLevel(2)	-.660	.322	4.212	1	.040	.517
	StockOptionLevel(3)	-.954	.418	5.217	1	.022	.385
	MonthlyIncome	.000	.000	30.989	1	.000	1.000
	HourlyRate	.000	.004	.000	1	.997	1.000
	DailyRate	.000	.000	3.705	1	.054	1.000
	Constant	-.610	.417	2.143	1	.143	.543

a. Variable(s) entered on step 1: StockOptionLevel, MonthlyIncome, HourlyRate, DailyRate.

Based on above results we can see stock option level are the main influencers to attrition.

For a lower stock level of 1 the odds of an employee leaving increases by a factor of (exp(b)) 1.517 and as the stock level increases it suggests a negative correlation suggesting that as stock level increases the odds of a person leaving decreases based on negative value of B with a person having the highest stock option having the lowest odds of leaving. This suggests stock level is a big motivator for a person to stay or not.

*Running regression on **social needs** we see the below results:*

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	JobInvolvement			28.766	3	.000	
	JobInvolvement(1)	1.729	.378	20.906	1	.000	5.637
	JobInvolvement(2)	.887	.322	7.578	1	.006	2.429
	JobInvolvement(3)	.541	.309	3.068	1	.080	1.718
	RelationshipSatisfaction			5.264	3	.153	
	RelationshipSatisfaction (1)	.396	.205	3.722	1	.054	1.487
	RelationshipSatisfaction (2)	-.026	.216	.014	1	.904	.974
	RelationshipSatisfaction (3)	.021	.191	.012	1	.913	1.021
	businesstravel1			24.931	2	.000	
	businesstravel1(1)	-.796	.318	6.260	1	.012	.451
	businesstravel1(2)	.651	.167	15.303	1	.000	1.918
	Constant	-2.503	.318	62.079	1	.000	.082

a. Variable(s) entered on step 1: JobInvolvement, RelationshipSatisfaction, businesstravel1.

The results show logic and is also in line with the hierarchy of needs. Based on output we can state that person having lower involvement at work is more likely to quit compared to a person who is required to be more involved.

*Running regression on **esteem needs** we see the below results:*

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	YearsAtCompany	-.057	.025	5.122	1	.024	.944
	YearsInCurrentRole	-.142	.037	14.939	1	.000	.867
	YearsSinceLastPromotion	.122	.034	12.529	1	.000	1.129
	PerformanceRating(1)	-.370	.319	1.346	1	.246	.690
	PercentSalaryHike	-.040	.032	1.550	1	.213	.961
	Constant	-.127	.726	.031	1	.861	.881

a. Variable(s) entered on step 1: YearsAtCompany, YearsInCurrentRole, YearsSinceLastPromotion, PerformanceRating, PercentSalaryHike.

Based on above results we can suggest that these factors have a strong relationship with attrition and also that as the values of these attributes increases the likelihood of person leaving also decreases based on negative value of B.

In general, based on results we can state that as the time spent by an employee in the company or the current role, or we can say as the performance rating and salary hike increases the likelihood a person leaving decreases for the mentioned factor under Exp(B).

*Running the final regression output for **self-actualization needs** we see the below results:*

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	TrainingTimesLastYear	-.137	.058	5.530	1	.019	.872
	JobSatisfaction			17.634	3	.001	
	JobSatisfaction(1)	.860	.205	17.526	1	.000	2.363
	JobSatisfaction(2)	.456	.220	4.284	1	.038	1.578
	JobSatisfaction(3)	.399	.197	4.091	1	.043	1.490
	EnvironmentSatisfaction			22.282	3	.000	
	EnvironmentSatisfaction (1)	.803	.197	16.659	1	.000	2.232
	EnvironmentSatisfaction (2)	.135	.218	.387	1	.534	1.145
	EnvironmentSatisfaction (3)	.025	.196	.017	1	.897	1.026
	Constant	-1.904	.242	61.913	1	.000	.149

a. Variable(s) entered on step 1: TrainingTimesLastYear, JobSatisfaction, EnvironmentSatisfaction.

The results again show logically meaningful outputs.

Based on above output we can say that the more a person undergoes training, the likelihood of a person leaving decreases by a factor of .872 and also a person who is not much satisfied with a job or the working environment and conditions, is more likely to leave compared to a person having more satisfaction at work.

Based on all above results individual attributes under each type of need, which are major contributors in influencing attrition can be found and better understood. Correlation among attributes across different levels of needs can also be studied to understand interdependency or influence.

Visualizations:

In basic needs we are comparing people who left with their marital status and the distance to the their work place.

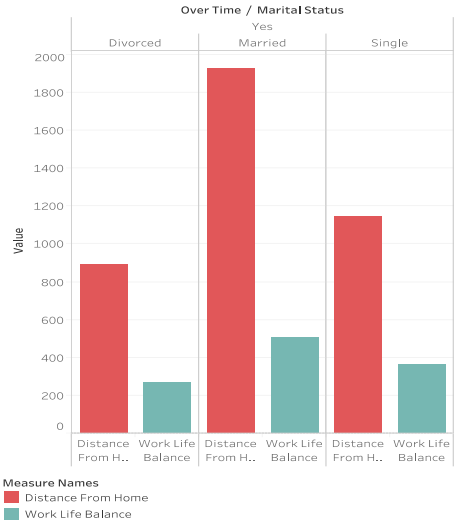

Figure : Basic needs.

The visualization clearly suggests that majority of the people who quit are married people and also those who staying the farthest from home. It shows the extent to which distance from work affects employee motivation.

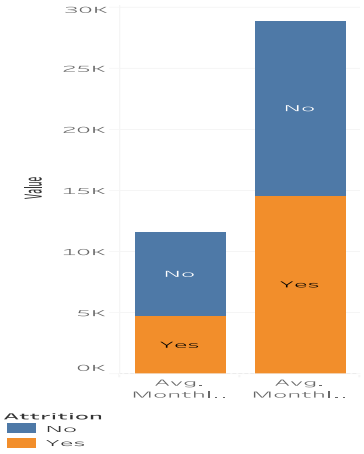

Figure : Safety needs.

The above figure in contrast to safety needs suggests that people having higher monthly rates and monthly income are less likely to quit.

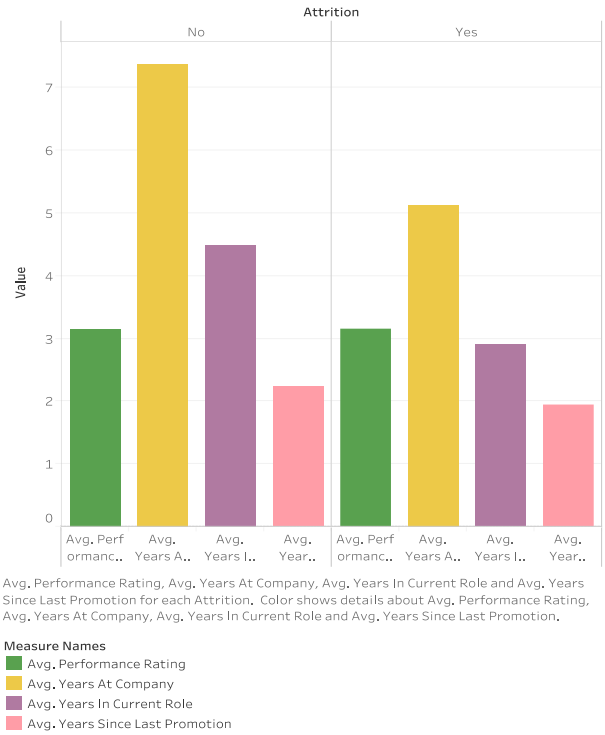

Figure : Esteem needs.

Coming to esteem needs we see that people having more experience, promotions and better performance rating tend to stay longer compared to people who quit due to lower ratings.

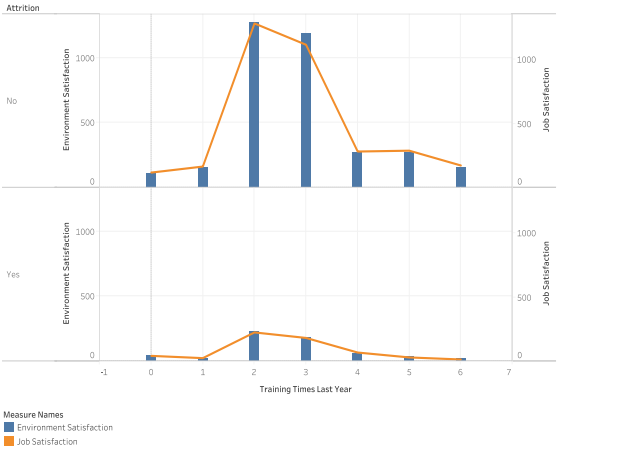

Figure : Self-actualization needs.

Coming to self-actualization needs we again see that more satisfied and content people are more likely to stay compared with people who quit and also receive less training times.

V. CONCLUSION

The model shows successful results and the data even though fictional, seems to follow a hierarchy of needs. Additionally, based on company capabilities and budgets the relationship and analysis can be used to provide better employee motivation where it actually can be provided. For example, if a company cannot provide higher compensations and salary hikes with regards to safety needs it can substitute it by providing shorter work timings or food compensation to target their basic needs or can provide training to target their self-actualization needs. Hence based on successful mapping of

relationships a CRM system can be implemented for HR analytics, understanding employee turnover, improving employee performance and most importantly in improving the relationship between an Employer and the employees.

VI. REFERENCES

[1] Joaquin Silva, "Abraham Maslow, His Theory and Contribution to Psychology", https://positivepsychologyprogram.com/abraham-maslow/ , February , 2019.

[2] Pavan Subhash, "IBM HR analytics employee performance and attrition", https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset, 2017.

[3] Imad Dabbura," Predicting Employee Turnover", https://towardsdatascience.com/predicting-employee-turnover-7ab2b9ecf47e , December 11,2019.

[4]Samuel Coyne, Yao Yao and Joseph Denney, "Prediction of Future Employee Resignations", https://www.slideshare.net/YaoYao44/prediction-of-future-employee-turnover-via-logistic-regression, January 12, 2018.

[5] Hanif Ayesha, Khalid Wajiha Khan and Tahir Nawaz, "Relating Maslow's Hierarchy of Needs with Employee Turnover and Retention: Case Study of Local Telco. ***International Journal of Human Resource Studies"*** *, [S.l.], v. 3, n. 2, p. Pages 51-68, June 2013. ISSN 2162-3058.*

[6]SHERRI HARTZELL," THE NEEDS THEORY: MOTIVATING EMPLOYEES WITH MASLOW'S HIERARCHY OF NEEDS",2019.

[7] Michael Hitt, Stewart Black and Lyman Porter, "Management : Pearson New International Edition", 2012.