



Employee Performance: Heatmap Analysis

Jyothi M B

BE, Department of ECE, DBIT, Bangalore, India

Abstract: Mining of data from large data sets and the process of discovering patterns using statistics, machine learning, data correlation, data plotting or data visualization and data evaluation are called data mining. Data analytics and data mining are a subset of Business Intelligence (BI). In our previous paper titled “Data Science to Analyse Employee data” the process of data pre-processing was demonstrated by writing a program in Python. Libraries like pandas, numpy, seaborn and matplotlib of Python provide platform for computing, evaluation and visualization of acquired data. [1-2] In this paper we demonstrate three analytical tools- plotting and evaluating, correlation and data prediction/ Machine learning which are involved in data mining and analytics of company’s data. The company wants to understand the factors contributing to employee turnover and to think of various retention strategies.

Keywords: Python, analytical tools- plotting and evaluating, correlation and data prediction/Machine learning.

I. INTRODUCTION

A large company functional in the 21st century world consists of many departments ranging from manufacturing, R & D, to sales. Each department consists of several workers rewarded with salary and facilities. Employees form the backbone of an industry and their behaviour must be kept in track. In this paper a code is written in Python for data acquisition, data pre-processing, [3-6] plotting and evaluation. Data acquisition and data pre-processing are explained in our previous paper titled “Data Science to Analyse Employee data”.

```
#Checking the Correlation
corr = df.corr()
corr = (corr)
sns.heatmap(corr,
             xticklabels=corr.columns.values,
             yticklabels=corr.columns.values)

corr
```

Out [13]:

	Exit	Satisfaction	Evaluation	ProjectCount	AverageMonthlyHours	YearsAtCompany	WorkAccident
Exit	1.000000	-0.388375	0.006567	0.023787	0.071287	0.144822	-0.154622
Satisfaction	-0.388375	1.000000	0.105021	-0.142970	-0.020048	-0.100866	0.058697
Evaluation	0.006567	0.105021	1.000000	0.349333	0.339742	0.131591	-0.007104
ProjectCount	0.023787	-0.142970	0.349333	1.000000	0.417211	0.196786	-0.004741
AverageMonthlyHours	0.071287	-0.020048	0.339742	0.417211	1.000000	0.127755	-0.010143
YearsAtCompany	0.144822	-0.100866	0.131591	0.196786	0.127755	1.000000	0.002120
WorkAccident	-0.154622	0.058697	-0.007104	-0.004741	-0.010143	0.002120	1.000000
Promotion	-0.061788	0.025605	-0.008684	-0.006064	-0.003544	0.067433	0.039245

Figure 1 shows the Python code to check for correlation and the output.



II. PROBLEM STATEMENT

A company wants to understand the factors contributing to employee turnover and to create a model that can predict if a certain employee will leave the company or not. The goal is to create or improve different retention strategies on targeted employees. The implementation of this model will allow management to create better decision-making actions.

III. METHODOLOGY

A. Checking for Correlation

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate with respect to each other.

Positive correlation = extent to which variables increase or decrease in parallel.

Negative correlation = indicates the extent to which one variable increases as the other decreases.

Figure 1 shows the Python code to check for correlation and the output. Figure 2 shows the heatmap. From the heatmap, there is a positive (+) correlation between projectcount, averagemonthlyhours and evaluation. This could mean the employees who spent more hours and did more projects were evaluated highly. For negative (-) relationships, turnover and satisfaction are highly correlated, assuming people tend to leave the company more when they are less satisfied. [7-9]

B. Salary v/s Exit

Figure 3 shows a graph of salary v/s exit of employees and a python code to display it. The graph is divided into three regions

- High salary
- Medium salary
- Low salary
- Exit: blue=0 and orange=1[9-13]

C. Department v/s Exit

Figure 4 shows the trend of employee exit in each department. In this company the notable departments are sales, accounting, HR, technical support, management, IT, marketing and R & D. [10-12] [14][15]

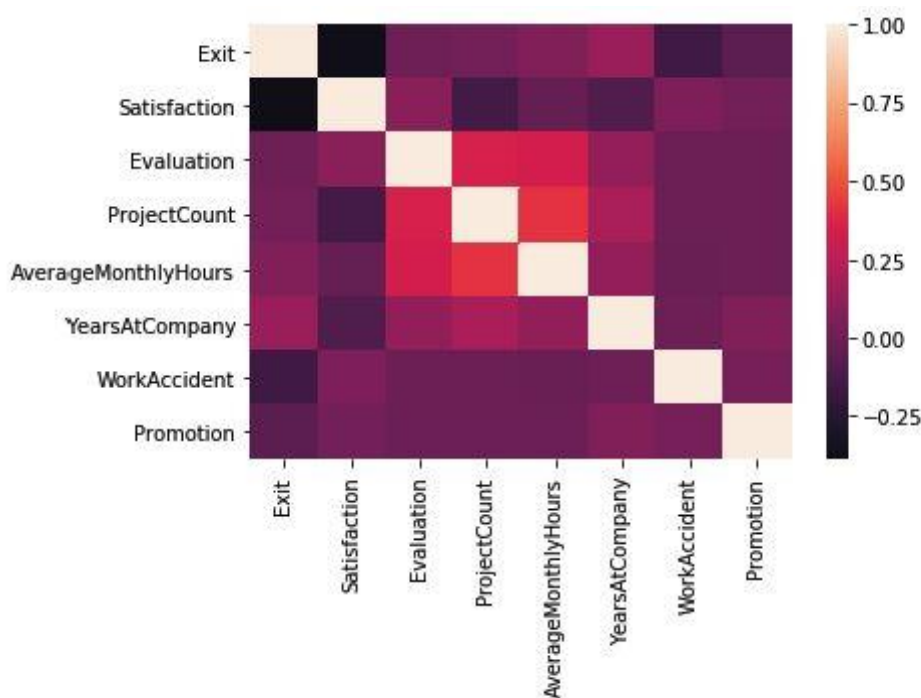


Figure 2 shows the heatmap.



IV. OBSERVATIONS

In the salary v/s exit plot, the employees with low salary tend to leave the company more. So salary can be one of the key reasons for dissatisfaction among the workforce. The sales department recorded the highest number of employee exit count as compared to other departments. Retention strategies planned by the company must focus on retaining the employees by various rewards, promotions and encouragement.

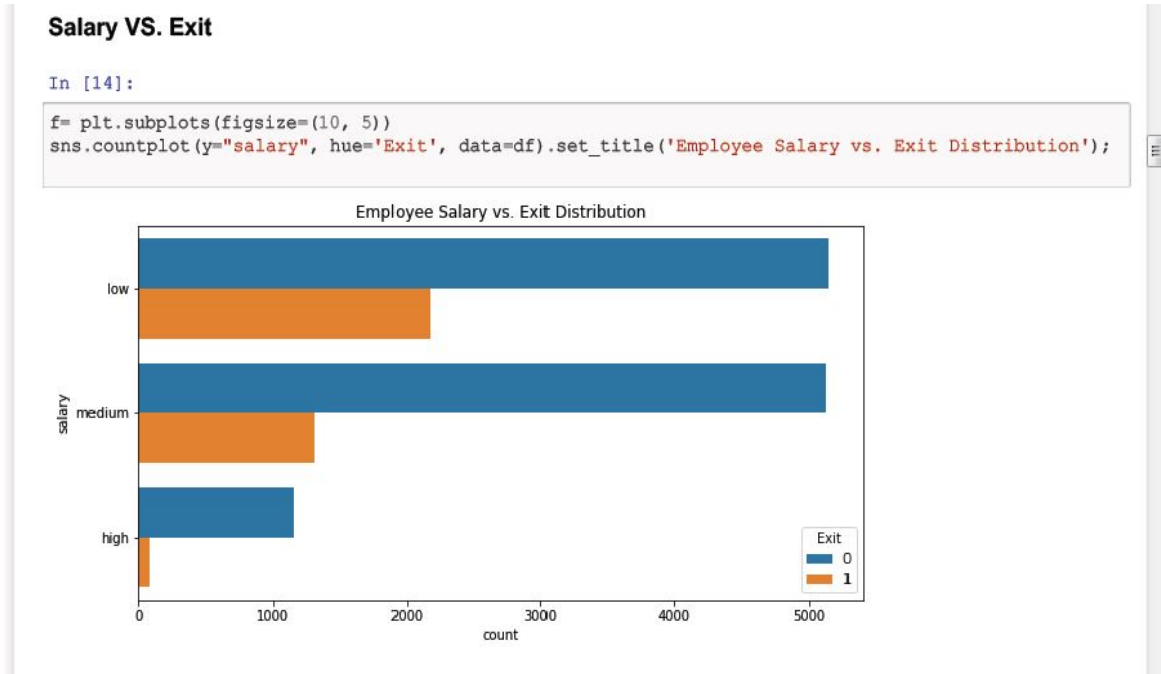


Figure 3 shows a graph of salary v/s exit of employees and a python code to display it.



Figure 4 shows the trend of employee exit in each department.



V. FUTURE SCOPE OF THE PAPER

The initial steps in data mining- data acquisition and data pre-processing was done using Python code. Libraries were imported for data manipulation. Noisy, incomplete and inconsistent data stream was filtered.[1-3] Data mining involves analysing the patterns and drawing conclusions out of these huge streams of data. In our next paper we will educate ourselves with Machine Learning- a branch of AI to predict data occurrence and fluctuations. Machine learning focuses on developing computer programs that can access data and use it learn for themselves. [16-18]

VI. CONCLUSIONS

A large company operational from a long period of time would mean several employees joining and leaving the company at random intervals. This would mean a large number of employees and a huge data stream of employee information. These data are stored and are acquired to analyse the patterns to predict the future of the company's work trend and employee satisfaction. Further efforts will be made in our coming assignments to highlight the application of Machine learning- logistic regression and random forest models.

REFERENCES

- [1] Ajit, P. (2016). Prediction of employee turnover in organizations using machine learning algorithms. *algorithms*, 4(5), C5.
- [2] W. C. Hong, S. Y. Wei, and Y. F. Chen, "A comparative test of two employee turnover prediction models", *International Journal of Management*, 24(4), 808, 2007.
- [3] Gao, Ying. "using decision tree to analyze the turnover of employees." (2017).
- [4] Mauricio A. Valle & Gonzalo A. Ruz (2015) Turnover Prediction in a Call Center: Behavioral Evidence of Loss Aversion using Random Forest and Naïve Bayes Algorithms, *Applied Artificial Intelligence*, 29:9, 923-942, DOI: 10.1080/08839513.2015.1082282.
- [5] Metz, Charles E. "Basic principles of ROC analysis." *Seminars in nuclear medicine*. Vol. 8. No. 3. WB Saunders, 1978.
- [6] Lessmann, Stefan, and Stefan Voß. "A reference model for customercentric data mining with support vector machines." *European Journal of Operational Research* 199.2 (2009): 520-530.
- [7] T. Fawcett, "An introduction to ROC analysis", *Pattern Recognition Letters* 27 (8), 861–874, 2006.
- [8] Raschka, S. (2015). *Python machine learning*. Packt Publishing Ltd.
- [9] Morgan, J.N., Sonquist, J.A.: Problems in the analysis of survey data, and a proposal. *J. Am. Stat. Assoc.* 58, 415–434 (1963)
- [10] Efron, B.S., Hastie, T.: *Computer Age Statistical Inference*. Cambridge University Press, Cambridge (2016)
- [11] Géron, A.: *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. O'Reilly Media (2017)
- [12] Zhao, Yue, et al. "Employee turnover prediction with machine learning: A reliable approach." *Proceedings of SAI intelligent systems conference*. Springer, Cham, 2018.
- [13] D. G. Allen and R. W. Griffith, "Test of a mediated performance – Turnover relationship highlighting the moderating roles of visibility and reward contingency", *Journal of Applied Psychology*, 86(5), 1014-1021, 2001.
- [14] D. Liu, T. R. Mitchell, T. W. Lee, B. C. Holtom, and T. R. Hinkin, "When employees are out of step with coworkers: How job satisfaction trajectory and dispersion influence individual-and unit-level voluntary turnover", *Academy of Management Journal*, 55(6), 1360-1380, 2012.
- [15] Sachdeva, Sujata. *Times of India Online*, 27 Nov. 2007, www.timesofindia.indiatimes.com. Accessed 6 May 2020.
- [16] Singh, Moninder, et al. "An analytics approach for proactively combating voluntary attrition of employees." *2012 IEEE 12th International Conference on Data Mining Workshops*. IEEE, 2012.
- [17] Yedida, R., Reddy, R., Vahi, R., Jana, R., GV, A., & Kulkarni, D. (2018). *Employee Attrition Prediction*. arXiv preprint arXiv:1806.10480.
- [18] Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. *Expert Systems with Applications*, 38(3), 1999-2006.