ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

Data Mining: Employee Turnover in a Company

Rohan AR[1], Shirsa Mitra[2], Anil Umesh[3]

Student, BE, Department of CSE, BNMIT, Bangalore, India [1]

Student, BE, Department of CSE, BNMIT, Bangalore, India [2]

BE, Department of CSE, SJCE, Mysore, India [3]

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). [1] In our previous paper titled "Data Analytics: Employee Turnover in a Company-1" the process of data pre-processing was demonstrated by writing a program in Python. Libraries like pandas, numpy, seaborn and matplotlib [2] of Python provide platform for computing, evaluation and visualization of acquired data. 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] plotting and evaluation. Data acquisition and data pre-processing are explained in our previous paper titled "Data Analytics: Employee Turnover in a Company-1". [4]

```
#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	WorkAc
Exit	1.000000	-0.388375	0.006567	0.023787	0.071287	0.144822	-0.15462
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	-0.00710
ProjectCount	0.023787	-0.142970	0.349333	1.000000	0.417211	0.196786	-0.00474
AverageMonthlyHours	0.071287	-0.020048	0.339742	0.417211	1.000000	0.127755	-0.01014
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

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

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.

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

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.

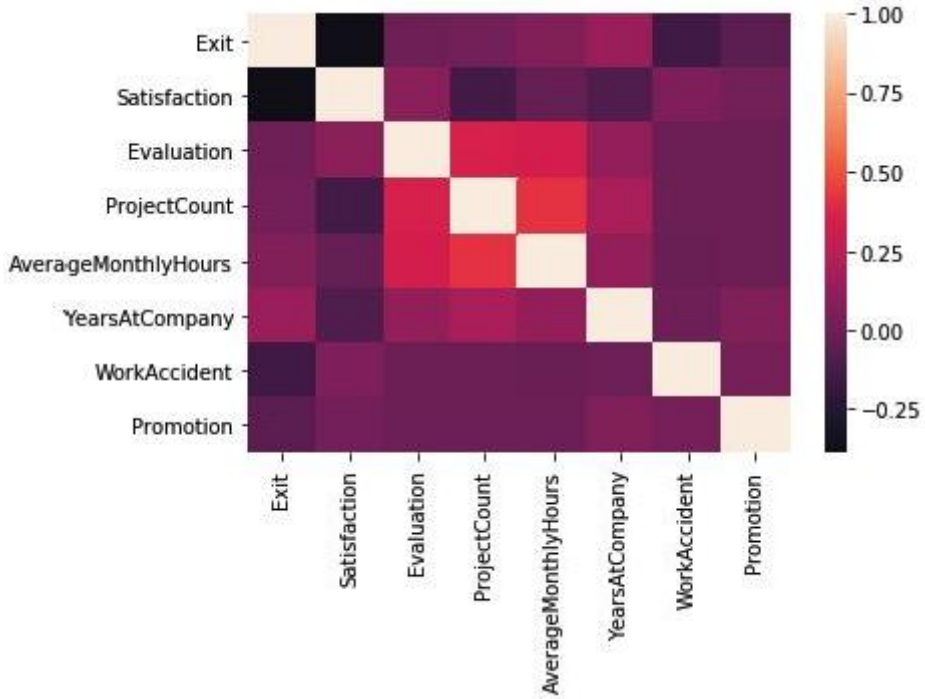

Figure 2 Shows the Heat Map

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.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

Salary VS. Exit

In [14]:

```
f= plt.subplots(figsize=(10, 5))
sns.countplot(y="salary", hue='Exit', data=df).set_title('Employee Salary vs. Exit Distribution');
```

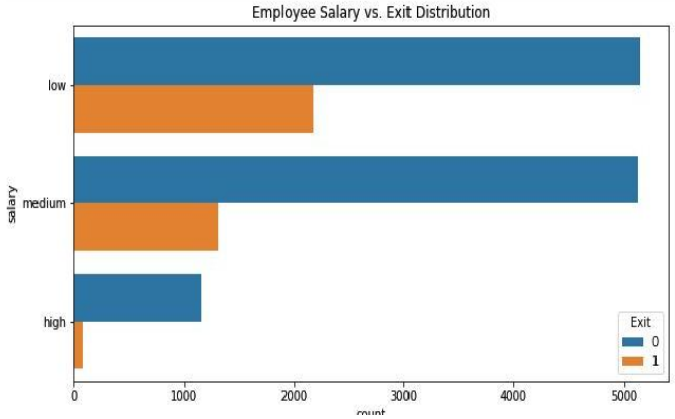

Figure 3 shows a graph of salary v/s exit of employees and a python code to display it.

In [15]:

```
color_types = ['#78C850','#F08030','#6890F0','#A8B820','#A8A878','#A040A0','#F8D030',
               '#E0C068','#EE99AC','#C03028','#F85888','#B8A038','#705898','#98D8D8','#7038F8']

# Count Plot (a.k.a. Bar Plot)
sns.countplot(x='Department', data=df, palette=color_types).set_title('Employee Department
Distribution');

# Rotate x-labels
plt.xticks(rotation=-45)
```

Out[15]:

```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]), <a list of 10 Text xticklabel objects>)
```

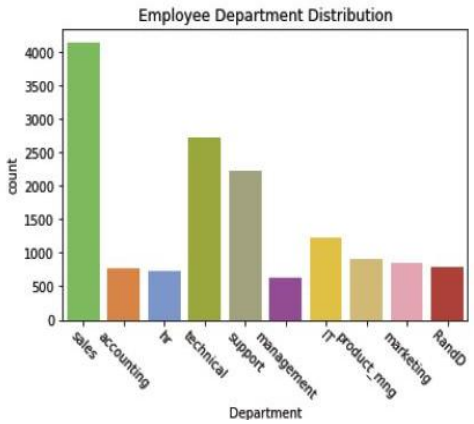

Figure 4 shows the trend of employee exit in each department.

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. 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.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 11, November 2018

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]. Principles of data mining DJ Hand - Drug safety, 2007 - Springer
[2]. The Python Standard Library — Python 3.7.1rc2 documentation-https://docs.python.org/3/library/
[3]. Research on Data Preprocess in Data Mining and Its Application- J Zhi-gang, JIN Xu - Application Research of computers, 2004 - en.cnki.com.cn
[4]. "Data Analytics: Employee Turnover in a Company-1"- https://ijarcce.com/wp-content/uploads/2018/10/IJARCCE.2018.7913.pdf

OUR GUIDE

VISHESH S born on 13th June 1992, hails from Bangalore (Karnataka) and has completed B.E in Telecommunication Engineering from VTU, Belgaum, Karnataka in 2015. He also worked as an intern under Dr. Shivananju BN, former Research Scholar, Department of Instrumentation, IISc, Bangalore. His research interests include Embedded Systems, Wireless Communication, BAN and Medical Electronics. He is also the Founder and Managing Director of the corporate company Konigtronics Private Limited. He has guided over a hundred students/interns/professionals in their research work and projects. He is also the co-author of many International Research Papers. He is currently pursuing his MBA in e-Business and PG Diploma in International Business. Presently Konigtronics Private Limited has extended its services in the field of Software Engineering and Webpage Designing. Konigtronics also conducts technical and non-technical workshops on various topics.