

Case Study: Post-Placement Analysis on Placement Satisfaction

Vrunda Gadesha

Faculty of Computer Science, Assistant Professor & Placement Officer, K.S. School of Business Management, Gujarat University, Ahmedabad, Gujarat, India

Abstract---- Placement is the high-ranking initiative of any institute. During the placement process there are major three parts (a) Training (b) Placement administration (c) Suitable Placement. After this process the key point is does the students are satisfied with their placements? With respect to IT-Jobs, the job satisfaction changes frequently. But when an institute place some student at a company, it is very important for the institute to analyze that how many students are satisfied with their placements? Which Companies should get priority at on-campus, which companies should be invited for off-campus, which companies are there who deals with mass-students, which companies are there who deals with skills of students, what kind of training module do these companies are offering, what would be the lower and upper limit of the package offered by companies and how many companies should be invited for the placements in single session. These all parameters take place after the session of placement ends in the final year batch. These parameters can be common for wide streams but the result would be very-very different and very according to stream and market requirement. Thus it is required to perform this analysis after each session which can be conducted as post-placement analysis. This Paper represents the post-placement Analysis for Institute based in Gujarat (*not to mention name) for MSc-CA & IT program, which is performed on the dataset of “Placemet-2019-20” using Correlation in python between ‘students rating for training at companies and Package offered by the companies.

Keywords--- Data Analysis, Placement Process, Placement Analysis, Correlation in python

I. INTRODUCTION

Post-Placement Analysis can be performed at least after the successful placement of one batch. This process is having successive progress after 3-4 iterations of batches as the dataset will increase gradually. This is the first iterations with the dataset of 124 students rating within 75 companies landed at the campus. Following specification shows the summary of the dataset taken under the analysis of the placement-19-20.

1. 75 Companies offering packages in the range of 1.8 Lacs/Annum to 4.6 Lacs/Annum
2. Student rating on the scale of 10 with respect to the training they offered by the companies.

This study targeted to find the most impactful companies which can be get the higher position in the priority list of campus placement. In addition, Institute – Placement Cell want to fetch the companies with 0 impacts at placement in terms of training, package and process. This study relates the training factor and package both to calculate the student satisfaction ratio.

Here the method Pearson Correlation is used to find the correlation between two different variables. A correlation coefficient of 1 means that for every positive increase in one variable, there is a positive increase of a fixed

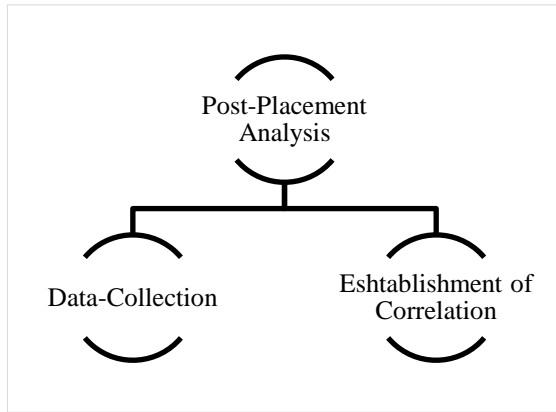
proportion in the other. For example, shoe sizes go up in (almost) perfect correlation with foot length. A correlation coefficient of -1 means that for every positive increase in one variable, there is a negative decrease of a fixed proportion in the other. For example, the amount of gas in a tank decreases in (almost) perfect correlation with speed. Zero means that for every increase, there isn't a positive or negative increase. The two just aren't related. The coefficient of correlation r_{xy} between two variables x and y , for the bivariate dataset (x_i, y_i) where $i = 1, 2, 3, \dots, N$; is given by –

$$r_{(x,y)} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

Where, $cov(x,y)$: the covariance between x and y and σ_x and σ_y are the standard deviations of the distributions x and y .

II. METHOD AND PROCESS:

Post-Placement analysis is divided into two major process shows below:



[Figure 1 - Processes of Post-Placement Analysis]

A. Data Collection

With respect to dataset, the data has been collected from two entities, students and Companies. While taking data from companies, we have taken-up following fields and we used Questioner to get the data from companies.

Table 1 Questioner for campus before the placement process

Field Name	Description
Name of the Company	Identify the Company
Company Address	Track the Location – Geographic Profile
Company's Director/CEO/Founder Name	Analyze Web Appearance of the person
Company's Director/CEO/Founder Contact Number	Administration purpose
Company's HR Name	Analyze Web Appearance of the person
Company's HR Contact Number	Administration purpose
Number of Vacancies	Placement Purpose and to analyze the Length of Employees of the Companies.
Position Details with Package	Analyze the Technology Trend and Market Value of the Trend

The Second parameter of this process is Student Rating, which is taken from the student with individual Question-answer method, the question for each student was same: “Rate your company out of 10 according to the training and environment you are getting science last 90 days.”

This rating is taken blindly as student does not disclose their name or any other identity. Among 124 students 90 students has shared their opinion with us. So after this data Collection process we have 75 Companies with package details and 90 student ratings for their companies.

B. Establishment of Correlation

Before correlating the data, the variables have been defined as follows:

1. Company_Name – Showing the Company Title
2. Package – Showing the Company Package
3. R1, R2....R17 – Showing different ratings given by students.
4. Rate-mean – Average of Student Rating

C. Process Flow:

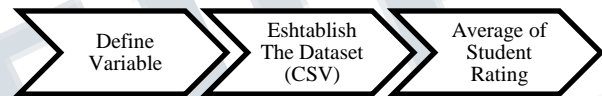


Figure 2 - Initial staeg of placement process



Figure 3 - Progressive stage of placement process

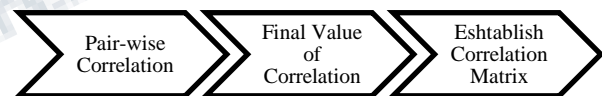


Figure 4 - Post Data Process Stage

In the Initial Stage, after defining the variables, the dataset is formed in required format – CSV and using python ‘DF’ feature – the whole dataset has been stored in single variable in python.



Figure 5 - Data Frame of the Placement dataset

**International Journal of Engineering Research in Computer Science and Engineering
(IJERCSE)
Vol 8, issue 5, May 2021**

Once the dataset has been stored, the “mean Function” has been applied over the Columns [R1:R17].



Figure 6 - Mean for each individual company

As shown in above screen-shot, the maximum number of rating for one company is 17, thus we declared 17 different variables, but many companies do have ratings less than 17, thus these records will be empty, while taking the mean, it will ignore the null values, and only calculate the mean for the non-empty record.

After this initial stage, the Correlation function has been applied between two columns: (a) Package (b) Rate_Mean The result is shown below:

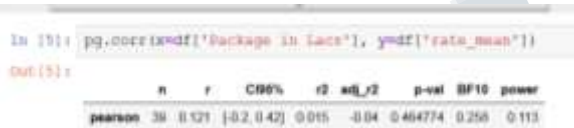


Figure 7 - Result of the Correlation

As shown in above screen-shot, we are getting only correlation coefficient (the r value) but also the p-value, which, in short, quantifies the statistical significance of the test. Following is the expiration of the result:

- ‘n’ – number of observations included in the calculation of the correlation coefficient – A Sample Size
- ‘R’ – correlation coefficient with the value 0.121 shows the positive relation.
- ‘CI95%’ - 95% confidence intervals around the correlation coefficient
- ‘r2’ – Square value of ‘r’
- ‘adj_r2’ - adjusted r-squared

For above two variables r2 and adj_r2, as its name implies, it is simply the squared r, which is a measure of the proportion of the variance in the first variable that is predictable from the second variable.

‘p-val’ – if we find this value below 0.05, we can say that our two variables ‘package’ and ‘rate_mean’ are not correlated with each other. Therefore we can say that there is a significant correlation between the two variables
‘BF10’ - Bayes Factor of the test, which measures the statistical significance of the test. It-directly measures the strength of evidence in favor of our initial hypothesis that

‘package’ and ‘rate_mean’ are correlated. Since this value is positive but small in number, it indicates that there the two variables are weakly correlated and they are conceptually different, thus ‘package’ with less value may have high rating in terms of training.

‘Power’ - The higher this value is, the more robust our test is. In this case, a value of 0.113 means that we cannot be much confident in our ability to detect the significant effect, but we can conclude that if we increase the data-size year by year, we can get the value of power around 1.

D. Data Visualization:

To make this data visualized, here we used “Seaborn Package”, to create scattered 2D plot shows plotting of “Package in Lacs v/s rate_mean”.

As Shown in above figure, very few companies do have rating between 7 and 8.5, majority of the companies do have rating above 8.5. The density of the plotting is darker between package range [2, 4].

The plotting shows correlation between package and rating. The distribution of the two variables is also shown on the margin. The continuous linear line reveals that package is not much the affective variable in fresher students in IT industry.

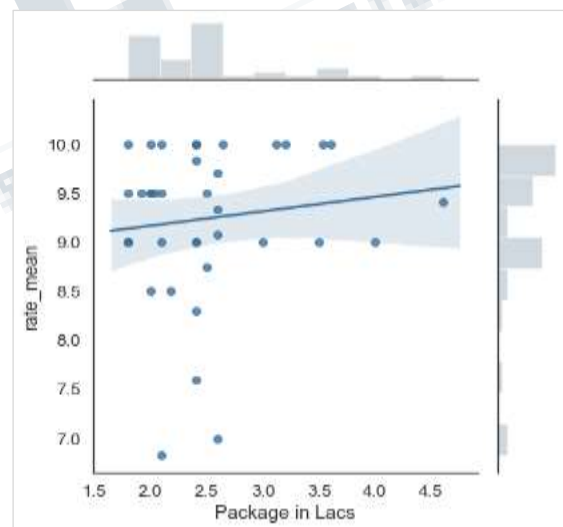


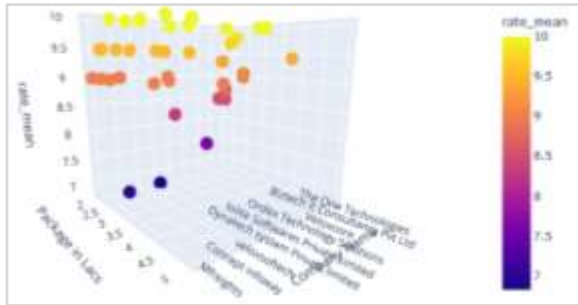
Figure 8 - Correlation between Student Rating and Package

E. 3D Plotting:

From above 2D plotting we can get the overview of the placement analysis but to make the exact summery see the 3D Plot showing results of Company Name v/s Package v/s Student Rating.

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 8, issue 5, May 2021



This 3D plotting provides with the exact analysis of Placements - 2019-20 for Institute at Ahmedabad.

III. CONCLUSION

With the correlation function we can get following conclusions which can be considered for the placements 2020-21 at institute based in Gujarat with 200 students and 75 companies on-campus.

Company specification to be Considered for First Round	Companies with Yellow and Orange Plotting and Package > 2.5
Company specification to be Considered for Second Round	Companies with Pink and Purple Plotting and Package > 2.0
Company specification to be Considered for Remaining Students	Companies with Blue Plotting

IV. FUTURE SCOPE

This process is applied over only the placement-data for MSc students, this method can also apply over MBA Branch. Moreover moving towards the large scale this analysis method can be applied over more than one institute based in Gujarat, which may have data-set for fresher and job-seekers across Gujarat State. To perform the large-scale analysis we can propose the “Apriori Mend Algorithm” to define association rules for Placement Fairs in Gujarat.

REFERENCES

- [1] D. Magdalene Delighta Angeline and I. Samuel Peter James. (2012). Association Rule Generation Using Apriori Mend Algorithm for Student’s Placement, Int. J. Emerg. Sci., 2(1), 78-86, March 2012 : ISSN: 2222-4254
- [2] Correlation(s) in Python (July 2019) <https://raphaelvallat.com/correlation.html>
- [3] Sebastian Norena (April 2018) Finding Correlation Between Many Variables (Multidimensional Dataset)

- with Python - <https://medium.com/@sebastiannorena/finding-correlation-between-many-variables-multidimensional-dataset-with-python-5deb3f39ffb3>
- [4] Jake VanderPlas (November -2016)- Python Data Science Handbook - Essential Tools for Working with Data - O'Reilly Media
- [5] 3D Plotting with plotly - <https://plot.ly/python/ipython-notebook-tutorial/#3d-plotting>