Forging the Future: Pioneering Approaches in Business, Management and Economics Engineering to Overcome Emerging Global Challenges - 2024

https://doi.org/10.24867/FUTURE-BME-2024-051 Original scientific paper

USE OF EXPLORATORY DATA ANALYSIS AND VISUALIZATION TECHNIQUES TO MEET ORGANIZATION STRATEGIC GOALS

Sunaina Paul[1], Rudra Singh[2], Minja Bolesnikov[3] [0000-0003-3735-602X], Ana Jovičić Vuković[4] [0000-0003-0022-674X]

Abstract

In the era of frequently changing technology and requirements, a competitive team is much needed for the product development. One of the challenging tasks for the Managers are how to effectively analyze the clustered data and use it to visualize the pattern and find the gap areas. To address these gap areas, department needs to first perform Exploratory Data Analysis (EDA) on the available datasets to detect obvious errors, identify outliers in datasets, understand relationships, unearth important factors, find patterns within data, and provide new insights. Secondly, after data has been collected, processed, and modeled, it must be visualized for conclusions to be made. This research will provide an insight on how the existing data could be leveraged to improve departments sustainable growth. To have a better overview on the phenomenon at hand, this research includes a survey from different project teams in the department which investigates the individual competencies and skill set, motivational factors, satisfaction and dissatisfaction, performance level, roles and responsibilities, attrition rate, training needs and effectiveness etc. There are challenges that managers and leaders can mitigate or provide preventing solutions. The analysis section sheds light on them (Ex: attrition) by finding right pattern between problem and solution using EDA and Visualization. These findings are useful for managers who have a task to improve their department competency but are unaware on how to process and visualize the existing data sets effectively. It will also help departments in the organization to retain and improve the quality of their talent pool.

Key words*: Exploratory Data Analysis, Data Visualization, individual competencies, skill set, motivational factors, satisfaction*

[1] Swiss School of Business and Management, Switzerland, sunaina@ssbm.ch
[2] Collins Aerospace, rudra.singh@collins.com
[3] Swiss School of Business and Management, Switzerland, minja@ssbm.ch
[4] Novi Sad School of Business, Serbia, dr.ana.jovicic@gmail.com

1. Introduction

1.1 Broad Area of Work

According to Tukey (1979) exploratory data analysis "emphasizes looking at data from multiple angles, identifying patterns, and using various visualization tools to reveal insights that traditional methods may overlook" Data Visualization is one of the steps of the data science process, which states that after data has been collected, processed and modeled, it must be visualized for conclusions to be made. "Data visualization refers to the graphical representation of information and data, using visual elements like charts, graphs, and maps", (TechTarget, n.d.).

EDA is a critical first step in data analysis that allows for the identification of patterns, outliers, and trends, providing a foundation for further statistical analysis (Smith-Miles, 2011). This research will provide an insight on how the existing data could be leveraged to improve department competency by utilizing the data that has been collected, processed, and modeled from the Exploratory Data Analysis and Visualization. It also helps in identifying challenges by finding patterns from the analysis and its visualization that helps managers and leaders to mitigate or provide preventing solution to overcome these challenges.

1.2 Background

One of the challenging tasks for the Managers are how to effectively analyze the clustered data and use it to visualize the pattern and find the gap areas. Every department in an organization always have these three basic questions to answer:

1. Where do we exist today in terms of our team assets?
2. What is our strategic goal for near future?
3. How to meet the strategic goal with current team assets?

To answer these questions, department needs to perform Exploratory Data Analysis on the available datasets and then visualize those meaningful data to find the pattern and create solution for it.

Exploratory data analysis (EDA) is a critical and important step of analysis in research. The first most objective with exploratory analysis is to examine the data for distribution, outliers, and anomalies to direct specific testing of your hypothesis. It also provides tools for hypothesis generation by visualizing and understanding the data usually through graphical representation. EDA aims to assist the natural patterns recognition of the analyst. Finally, feature selection techniques often fall into EDA. EDA is a fundamental early step after data collection and pre-processing, where the data is simply visualized, plotted, manipulated, without any assumptions, to help assessing the quality of the data and building models.

1.3 Objective

The objective behind this research is:

1. Maximize insight from the dataset by finding the meaningful data out of the clustered data.

2. Visualize the processed data and find out the potential relationships between variables.
3. Develop a predictive or explanatory model/observation that helps to mitigate known challenges.

1.4 Scope of Work

This research scope covers gathering of raw clustered data through surveys and using the pre available department data, Perform Exploratory Data Analysis (EDA) on the available datasets to detect obvious errors, identify outliers in datasets, understand relationships, unearth important factors, find patterns within data, and provide new insights. After data has been collected, processed, and modeled, it must be visualized for conclusions to be made.

This research will provide an insight on how the existing data could be leveraged to improve department competency for aerospace. To have a better overview on the phenomenon at hand, this research includes a survey from different project teams from aerospace department which investigates the individual competencies and skill set, motivational factors, satisfaction and dissatisfaction, performance level, roles and responsibilities, attrition rate, training needs and effectiveness etc. Exploratory Data Analysis (EDA) plays a central role in forming hypotheses, defining models, and exploring data characteristics (Wongsuphasawat et al., 2019).

2. Understanding of Exploratory Data Analysis (EDA) and the Visualization Concept

2.1 Literature Review

Exploratory data analysis could be used before the formal analysis as the important data investigation process to test hypotheses with summary statistics and visualizations for finding patterns, outliers, and anomalies, discover trends. “Exploratory Data Analysis (EDA) is a crucial step in understanding the underlying patterns in data before applying any statistical models" (Camelot ITLab, 2019). It provides an idea regarding the data which we will be analyzing in detail and Visualization helps to create meaningful insights easily from massive datasets.

For this research the previously available sample data from public sources was collected, reviewed, and looked for the information’s that we seek. It was observed that the current data was scattered, less informative and had scope for standardization to provide the meaningful detailed information at a department level. There was a case for change to perform Exploratory data analysis on these data, visualize it, summarize the main characteristics, and find out meaningful information for the managers that will help in understanding the overall organization health status. "Exploratory data analysis focuses on discovering the unexpected and can be instrumental in hypothesis generation rather than just confirmation" (Emerson & Hoaglin, 1983).

2.2 What is Exploratory Data Analysis?

Exploratory Data Analysis (EDA) is an approach to extract the information enfolded in the data and summarize the main characteristics of the data (Camelot ITLab, 2019).

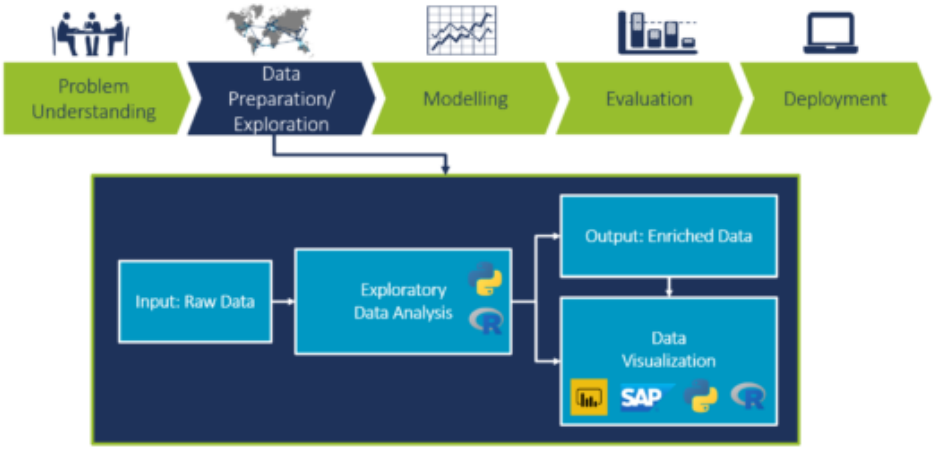

Figure 1: Project Flow with exploratory data analysis (EDA)

As defined in Figure 1, the main pillars of EDA are data cleaning, data preparation, data exploration, and data visualization. There are various exploratory tools (Python and R), and enterprise applications (Power BI, SAP Cloud Analytics, Tableau, etc.) to perform EDA, each of them offering a unique set of tools (Camelot ITLab, 2019). In this research we have mostly used Excel and Tableau for data visualization.

3. Results

This research has considered attrition attribute for the analysis.

3.1 Data Exploration and Visualization for Attrition:

Under this section we performed Univariate graphical (Histogram) and multivariate graphical EDA (Bar chart) using the Tableau Tool for the attrition data.

The goal of exploratory data analysis is to discover hidden insights by using a combination of visual and statistical methods (Patil, 2018).

3.1.1 Collection and Cleaning of data

The raw clustered data is gathered and created after discussion with industry experts and by using the available public source. The data includes respective Employee ID, Age, Attrition information, Gender, Environment Satisfaction, Job Involvement, Job Level, Job Role, Job Satisfaction, Marital Status, Total working

years, Years at company, Number of companies worked, Salary hike (%), Work life balance, Years since last promotion, Years with current manager, Performance rating, Relationship satisfaction etc. Finally, the data was cleaned for any misses, empty field, duplicacy and overlaps.

3.1.2 Visualization and Observation from data

Tableau tool used for extracting appropriate information out of the data set and for its visualization to understand the current attrition rate against various factors through analysis and observation. "Data visualization allows patterns, trends, and correlations to be revealed that might not be obvious in raw data", (Tableau, n.d.). Extracted observations and analysis report would be shared to Managers to take necessary actions.

Development of retention strategies is important for the health of organization. Employee attrition has many consequences such as low morale, work life balance, Job dissatisfaction etc. The goal of this analysis is to discover why employee leave and provide the observations to the Managers to help alleviate attrition.

Example: Age v/s Attrition trend

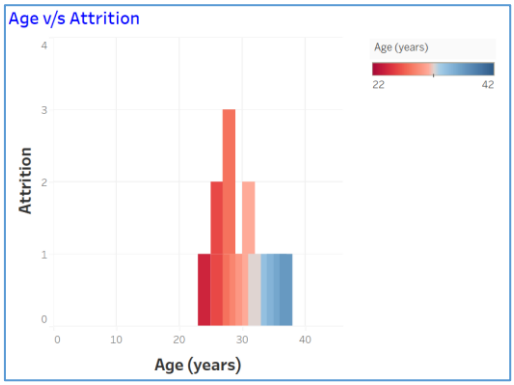

Figure 2: Age v/s Attrition Chart

Observations: To understand the impact of the age factor on attrition rate, we need to divide age into different age group. The entire workforce of any workplace has multiple generations, and they have their crucial role to play in the successful working of the organization. As per gathered data, the workforce belongs to the age in between 22 years to 42 years.

As per Figure 2, it is observed that the employees belonging to the age in between 25 years to 35 years, tends to change their job more frequently. They change their job position at least once. The percentage of changing their job often keeps reducing with the increasing age group. This means with increasing age; employee prefer more stability and like to stay in their comfort zones. Therefore, we could conclude as Age is inversely proportional to Attrition.

Example: Satisfaction v/s Attrition trend

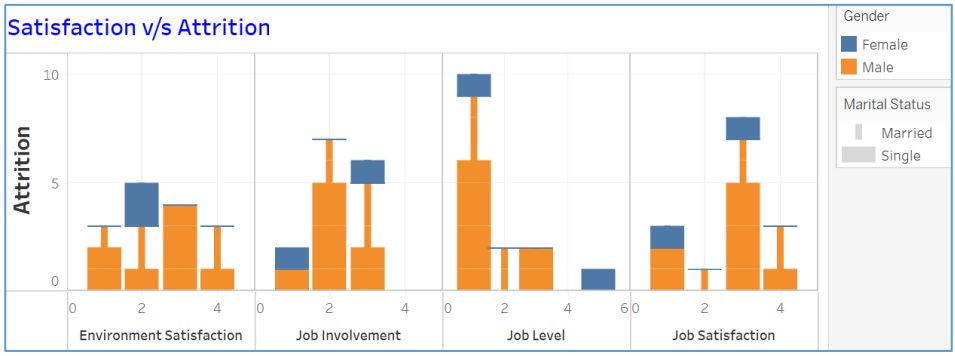

Figure 3: Satisfaction v/s Attrition Chart

Observations: As per Figure 3, observations are drawn below for each subcategory:

1. Environment Satisfaction v/s Attrition: Organization environment plays crucial role to embrace diversity. After visualizing above chart, it shows that when environment satisfaction value is lower then, female employee tends to change their job. But on the other hand, the Male employee have lesser impact on attrition due to Environmental satisfaction. The male attrition rate (either single or married) does not depend on Environmental satisfaction factor.
2. Job Involvement v/s Attrition: The job involvement factor has more impact on attrition for male employees compared to female employee. The single female employee tends to change their job if they are not feeling more inclusive in the work which they are performing in the organization. The employee (male or female) attrition rate (either single or married) is zero when they are more involved in their current job.
3. Job Level v/s Attrition: Job level is directly proportional to the attrition rate. Employee with lower job level have more turnover compared to higher job level employees.
4. Job satisfaction v/s Attrition: Job satisfaction factor has less impact on the male attrition rate. Either the job satisfaction level is high or low, it does not contribute to retaining the male employees within the organization, that is, attrition still happens. In contrast, for married female employees job satisfaction has no impact on attrition rate. They still prefer to stay in the organization irrespective of job satisfaction level.

4. Conclusion

This report set out to understand how the existing data from the respective data against various factors could be leveraged to improve department health in terms to alleviating attrition rate by utilizing the observation from the collected and processed data. These observations will help managers and leaders in identifying

challenges by finding patterns from the analysis and its visualization that help them to mitigate or take necessary action to overcome the challenges. Learned data exploratory and visualization concept could be utilized for other areas like individual competencies and skill set, motivational factors, satisfaction and dissatisfaction, performance level, roles and responsibilities, training needs and its effectiveness etc for improvement management.

Acknowledgments

The making of this research has been a wonderful learning experience when we had an opportunity to learn several details about attrition and its impact, that we were previously blissfully unaware of. We are sure that just the writing of this paper will help us improve as a professional. We would like to first and foremost thank and appreciate the opportunity provided to us by the FUTURE-BME for granting us to submit full paper. We are really grateful to our mentor, Dr. Minja Bolesnikov for his valuable feedback and time throughout the course of this paper submission. Lastly, we would like to thank our mother Seema Singh and our son Advay Singh without whose love and continued support none of this would have been possible.

REFERENCES

The state art is the base for any successful research project. In current project, the literature inclined towards the new domain of conversational information retrieval is considered. The following are referred journals from the preliminary literature review.

[1] Camelot ITLab. (2019, 27 March). *Exploratory Data Analysis – An Important Step in Data Science*. https://blog.camelot-group.com/2019/03/exploratory-data-analysis-an-important-step-in-data-science/

[2] Tableau. (n.d.). *What Is Data Visualization? Definition, Examples, And Learning Resources*. https://www.tableau.com/learn/articles/data-visualization

[3] Tech Target. (n.d.). *Data visualization*. https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization.

[4] Data, M. C., Komorowski, M., Marshall, D. C., Salciccioli, J. D., & Crutain, Y. (2016). Exploratory data analysis. *Secondary analysis of electronic health records*, 185-203.

[5] Singh, G. (2021, 14 October). *A Step-by-Step Guide for Data Visualization using Tableau', analyticsvidhya blog.* Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/10/step-by-step-guide-data-visualization-tableau/

[6] Smith-Miles, K. (2011). Exploratory Data Analysis. In M. Lovric, (ed.), *International Encyclopedia of Statistical Science*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_242

[7] IBM. (n.d.). *What is Exploratory Data analysis?* https://www.ibm.com/ph-en/topics/exploratory-data-analysis

[8] Patil, P. (2018, March 23). *What is Exploratory Data Analysis?*. Towards data science. https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15
[9] Wongsuphasawat, K., Liu, Y., & Heer, J. (2019). Goals, process, and challenges of exploratory data analysis: An interview study. *arXiv preprint arXiv:1911.00568*.
[10] Shmueli, G., & Lichtendahl, K. C. (2010). Exploratory data analysis. In *Practical Time Series Forecasting with R* (pp. 47-65). Springer. https://link.springer.com/chapter/10.1007/978-1-4614-4818-7_4
[11] Tukey, J. W. (1977). *Exploratory data analysis*. Reading/Addison-Wesley.
[12] Emerson, J. D., & Hoaglin, D. C. (1983). Resistant Lines for Y versus X. In D. C. Hoaglin, F. Mosteller, & J. L. Tukey (Eds.), *Understanding Robust and Exploratory Data Analysis* (pp. 129-163). Wiley.