ALY 6120: Leadership in Analytics

Module_3 – Assignment_3 (Group 3)

CRISP-DM Cycle: Data Understanding

Student Names: Parita Gala, Josephine Agbedoawu, Jiansheng Shentu, Godfred Akoto

Professor: Abeba N. Turi

Date: 16th November 2025

1.0 Explain why you have chosen that data

The IBM Employee Attrition dataset was our choice because it provides a genuine and properly organized ground for understanding the determinants of employee turnover. The dataset consists of a great variety of demographics, job-related, and organizational variables, which makes it appropriate for both predictive modelling and insight generation. Moreover, its size is not too large, so it can be analysed quickly and from many angles. At the same time, its extensive application in HR analytics research gives a great number of benchmarks for results validation. All in all, it encourages the in-depth analysis of attrition patterns and the use of data for planning the workforce.

2.0 Source for the data

The IBM HR Analytics Employee Attrition & Performance dataset is a synthetic dataset developed by IBM's data scientists, featuring details on 1,470 employees. In the dataset chosen, we observed that there were 35 variables that included demographic information, job-related attributes, satisfaction indicators, and compensation metrics such as Age, Department, Job Satisfaction, Monthly Income, Performance Rating, Work-Life Balance, and number of Years at the Company. This dataset allows that we can analyze elements that can lead to employee turnover and also help us in developing predictive models to determine those who might be prone to departing. This enables HR teams to implement proactive actions. The dataset is available on Kaggle, making it ideal for machine learning projects and applications centered on predicting attrition.

Dataset Link: (https://www.kaggle.com/datasets/hannycharlotte/employee-attrition-ibm-dataset)

3.0 Key Data Elements to Address the Business Problem and Why

To effectively grasp and forecast employee turnover, seven crucial factors from the IBM HR dataset offer detailed insights into the reasons employees depart and facilitate focused retention efforts. Job Satisfaction evaluates how happy employees are in their positions, aiding in the identification of morale concerns that may forecast premature exits. Monthly Income reflects existing salary levels, enabling evaluation of whether pay disparities contribute to turnover and uncovering possible compensation inequalities based on experience or market rates.

The Years at Company reveals tenure trends - brief tenures might indicate onboarding or engagement challenges, while analyzing tenure patterns aids in recognizing vital retention periods.

Also, Work Life Balance assesses how employees perceive the balance between work and personal life, determining if extended hours or limited flexibility lead to discontent and exhaustion. The Environment Satisfaction is closely tied to workplace culture and management support elements, showing that employees perceive themselves as appreciated and backed in their positions. The Age offers an understanding of generational contrasts and variations in career stages regarding attrition, revealing patterns linked to career advancement expectations or retirement strategies. Potential quitting parameters are over time workload strain on workers, working long hours.

When these factors are examined, they enable the company to pinpoint the underlying reasons for turnover and in turn implement targeted measures to prevent that.

While low pay relative to experience indicates wage discrepancy, high overtime combined with low satisfaction rates indicates insufficient workload management.

When we pay attention to these factors, we can spot when someone is struggling before they quit. Then we can help whether that is flexible hours, fair pay, or just listening. The goal is not sophisticated analytics, it is using what we learn to create a workplace where people genuinely want to stay, and when employees feel valued, everyone wins.

Reference

Hausknecht, J. P., & Holwerda, J. A. (2013). When does employee turnover matter? Dynamic member configurations, productive capacity, and collective performance. Organization Science, 24(1), 210–225. https://doi.org/10.1287/orsc.1110.0720

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. Journal of Applied Psychology, 102(3), 530–545. https://doi.org/10.1037/apl0000103

IBM. (2019). IBM HR Analytics Employee Attrition & Performance Dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/hannycharlotte/employee-attrition-ibm-dataset

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. Academy of Management Journal, 33(4), 692–724. https://doi.org/10.2307/256287

Kumar, R., & Garg, N. (2018). Predicting employee turnover intention: A data mining approach. International Journal of Human Capital and Information Technology Professionals, 9(2), 1–15. https://doi.org/10.4018/IJHCITP.2018040101