Employee Attrition Case Study

Section 01: Business Objective understanding and High-level Approach

Lay down using a few slides - *refer Industry template.*

- **Overall Business objective** – put in simple words what's the scope of the analysis and what business problem is being solved *(this ensures that you've gone through the client's requirement).*
- **Understanding of the problem** in your own words *(this helps align your understanding of the business problem with client's requirement).*
- **Approach** – Very high-level view of the approach you plan to use to address the problem *(this helps the client see that you have a plan in place to attack the problem).*

Section 02: Data Health Review

Report the below results in a few slides *(ensure slide formatting and grammatical correctness).*

- Do the variables get read in Python in the right format (Integer, Float, Boolean, Date, Object)? ***List down what corrective steps were taken (if any) for the affected variables?***
- Do any variables have missing values? ***Create a list of variables for which you found missing values and what % values in these variables were affected?***
- Do any variables have outliers? ***Use suitable plots to show the outliers for all these affected variables.***
- Are there any variables that require cleaning (extra spaces, special characters, unexpected values, etc.) or replacement of values (e.g., Yes/No to 1/0)? ***List down all such variables and the kind of cleaning required.***
- Are there any duplicate records in the data? ***If yes, how many.***

1. ***Generate Extended Data Dictionary (EDD) of the provided dataset to help comment on data quality*** *– refer Industry template.*

Section 03: Exploratory Data Analysis

Perform Exploratory Data Analysis to comment on what information does the dataset conveys and if it's complete/suitable to solve the given business problem. Also explore for any patterns/ insights that might guide in addressing the overall business problem. To be concrete, generate the below:

A. Observe Univariate distributions on both Object and Numeric variables.

a. Object variables – *Use suitable plots or visuals to show these distributions. Provide suitable commentary on what you observe for each variable.*

b. Numeric variables – *Use suitable plots to show these distributions. Provide suitable commentary on what you observe for each variable.*

B. Observe Bi-variate distributions

a. Scatter plots to show relationship between relevant Numeric variables, and Cross-tabulation for relevant categorical variables, etc.

Section 04: KPI/ Metric based questions – These questions have a specific ask (pin-pointed) and getting to the required outcome is quite straightforward.

1. How does Attrition rate vary with Salary (or for that matter any other Numeric variable)? Use binning to create equal frequency bins and observe the below summaries for different numeric variables:
   I. Average of variable for each bin
   II. Min and Max of variable for each bin
   III. Attrition rate for each bin

Use suitable plots to show these results and report your findings. For discrete variables (*numeric in nature but with low number of unique values*) applying a different binning such that bins have sufficient and equitable number of data points.

2. Observe the Attrition rate for different object variables (for their respective labels) and report your findings. Create these summaries only for relevant object variables (decide based on your understanding of data).
3. While analyzing Attrition rate for a combination of Two variables, which of the below segments show high Attrition:
    I. Employees who got a promotion recently and spent <1 year with their current manager versus employees who have spent >=1 year.
    II. Employees who got a Promotion, but salary increment was low versus employee who got a higher increment.
    III. Employees who had spent a long tenure with the company and did not get a promotion versus employee who got a promotion.
    IV. Employees who had spent a long tenure with the company and got a low salary increment versus employee who got a decent hike.
    V. Other such combinations…

Section 05: Open-ended questions and recommendations – These are business-oriented questions which do not tell much about the kind of expected outcome, rather they require you to check if a certain phenomenon is occurring or not, or whether there is plausibility of a certain pattern. Often these questions need to be asked on your own and to answer them one needs to think through in terms of: 'what kind of output is expected', 'how to get it – which variables and by doing what' and 'whether the achieved outcome helps answer the question'. Since this can be iterative, it requires a lot of brainstorming and asking the right questions (as per the business objective).

1. Which of the below are strong drivers of Attrition:
    - Low salary increments
    - No promotion
    - Current manager
    - Need for a change (after having served the company for long)
    - Shorter Avg. stint of the employees with their earlier employers
    - Low salary compared to Peers (with same work exp.)
    - Lesser training time spent in last year
    - Stock Options not being given
2. Based on your analysis (a quantitative result which provides some guidance/evidence), which of the below areas can be looked at more closely in terms of taking actions to reduce attrition:
    - Re-looking at hiring strategy *(Hint: relevant only when the data shows that many employees who left were the ones who were recently hired)*.
    - Re-looking at overall budget for better salary increments.
    - Organize Manager training to train them on people handling & management and provide mentorship.
    - Have more training programs that can help employees to up-skill themselves.
    - Anything else…

End output expected is a PPT. Keep Python codes, analysis excel files, etc. as a back-up for the final presentation.