

## Problem

Employee turnover poses significant challenges for companies, including increased costs associated with recruitment, training, and lost productivity. By accurately predicting which candidates are more inclined to remain with DataCorp after completing their training program, we aim to optimize our training initiatives, reduce turnover costs, and enhance candidate categorization. Our objective is clear: build a robust machine learning model that predicts the probability of a candidate seeking a new job or staying with DataCorp post-training. Leveraging demographic, educational, and experiential attributes of candidates, this model will enable us to allocate resources effectively, focus on candidates more likely to stay, and ultimately enhance the quality and efficiency of our training programs.

## Data Exploration

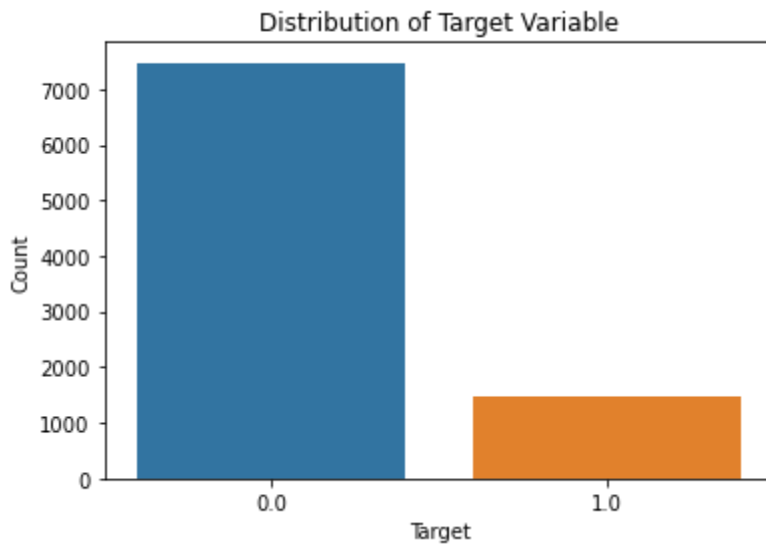
The dataset provided contains anonymized information about candidates' demographics, education, experience, and other relevant features. It includes features such as city development index, gender, relevant experience, education level, major discipline, total experience, company size, company type, last new job, and training hours. The target variable indicates whether a candidate is looking for a job change (1) or not (0).

## Features

- enrollee\_id : Unique ID for candidate.
- city: City code.
- city\_development\_index : Development index of the city (scaled).
- gender: Gender of candidate
- relevant\_experience: Relevant experience of candidate
- enrolled\_university: Type of University course enrolled if any
- education\_level: Education level of candidate
- major\_discipline :Education major discipline of candidate
- experience: Candidate

total experience in years • company\_size: No of employees in current employer's company • company\_type : Type of current employer • last\_new\_job: Difference in years between previous job and current job • training\_hours: training hours completed

- target: 0 – Not looking for job change, 1 – Looking for a job change.



## Modelling Results

Accuracy: 0.8565047459519821

Classification Report:

	precision	recall	f1-score	support
0.0	0.90	0.93	0.92	1511
1.0	0.55	0.44	0.49	280
accuracy			0.86	1791
macro avg	0.73	0.69	0.70	1791
weighted avg	0.85	0.86	0.85	1791

---

## Conclusion & Recommendation

The logistic regression model achieved an accuracy of approximately 85.65% on the test data. The classification report provides insights into the model's performance for each class, including precision, recall, and F1-score. In this case, the model performs better at predicting the negative class (0.0) compared to the positive class (1.0), as evidenced by higher precision, recall, and F1-score for the negative class.

This model can inform DataCorp's decision-making process in optimizing their training programs and resource allocation. By identifying candidates who are more likely to stay, DataCorp can focus their efforts on providing targeted support and incentives to retain these individuals, thereby reducing turnover costs and enhancing the effectiveness of their training programs. Additionally, insights from the model can help DataCorp tailor their recruitment strategies and employee retention initiatives to attract and retain high-potential candidates who are committed to long-term engagement with the company. Overall, leveraging predictive modeling techniques like logistic regression enables DataCorp to make data-driven decisions that support their strategic goals of talent retention and organizational development.