Who will leave the company?

Main goal

- Let's find out what causes people to quit your current job through data analysis.
- Also predict the possibility of applicants looking for new jobs or working for a company based on the model that uses current qualifications, demographics, and experience data.

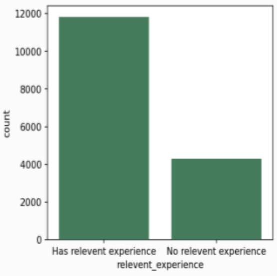
Features

- enrollee_id : Unique ID for candidate
- city: City code
- city_ development _index : Developement index of the city (scaled)
- gender: Gender of candidate
- relevent_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

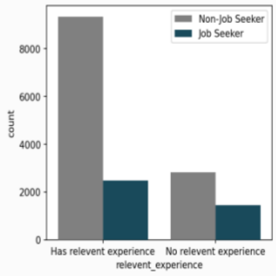
Who is looking for a new job?

About 35% of people with no relevant experience are looking for another job.

Overall



Job searching by relevent experience



Insight

- When classified by the presence or absence of relevant experience, the relevant experience were about 2.5 times more people with no relevent experience people.

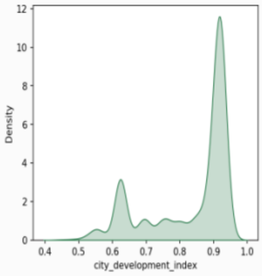
- Considering the proportion of job seekers according to whether they have related experience, the probability of finding another job is about 20% for those with related experience, and about 34% for those without related experience

- From the perspective of companies, in order to prevent the outflow of manpower selecting people who have relevent experience than those who have no relevent experience.

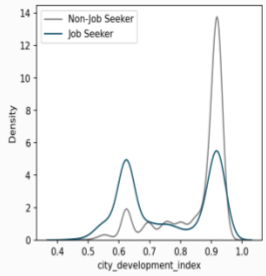
Does the City Development Index play a role?

Interestingly, we see Job Seekers are frequently from cities with a lower CDI score

Overall



Job Seeker / Non-Job Seeker



- What is more interesting though is the City Development Index(CDI) chart. There we see that there are two peaks for job-seekers. The peaks are at high and low CDI scores.

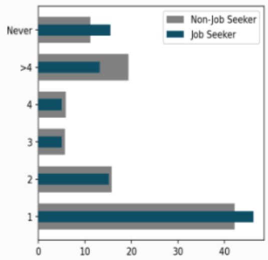
- However, in proportion, the lower the CDI, the more people looking for other jobs, and the higher the CDI, the fewer people looking for other jobs.

- From the perspective of companies, rather than unconditional welfare, it is necessary to identify appropriate problems. according to the CDI and propose solutions to solve them, in order to prevent manpower outflow.

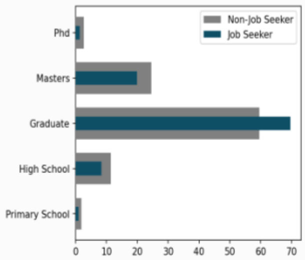
Are there other differences?

We see broadly similar patterns, but notable areas of difference

Last job change (yrs)



Education level



- looking at the last job experience chart, The lower the current job experience, the higher the probability of finding another job.

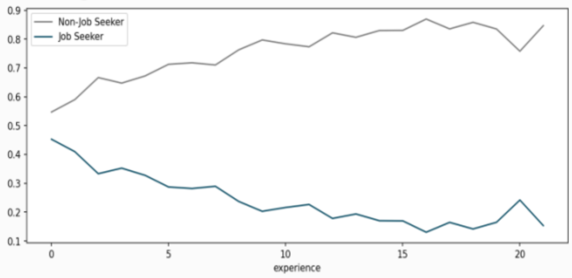
Therefore, from a company's point of view, rather than a new employee or an inexperienced person, It will be necessary to select a person with adequate experience.

- In addition, looking at the education level chart, it can be seen that university graduates are relatively more likely to find other jobs.

Is there a difference according to the total experience?

The higher the total experience, the fewer people look for other jobs.

total experience



- According to the experience chart, the lower the total experience, the more likely you are to find another job. In addition, in the case of people with more than 20 years of experience, the rate of finding other jobs increases.

From the perspective of companies, it seems necessary to manage lower-year employees to prevent manpower outflow.

How many people are looking for new jobs?

Looking at the chart below, the dependent variable is an unbalanced dataset. most people are non-job-seeking

Model Performance

LightGBM is the highest performing model. This model has the highest Roc auc score and recall score.

	Accuracy	Recall	Precision	ROC AUC Score	F1 Score
RandomForest Score	79.0%	51.9%	61.1%	70.2%	56.2%
Logistic Score	74.8%	13.1%	55.3%	54.7%	21.2%
LightGBM Score	79.8%	64.9%	60.3%	75.0%	62.5%
KNN Score	75.8%	20.7%	59.1%	57.8%	30.6%
SVC Score	75.5%	24.3%	56.2%	58.9%	34.0%
Catboost Score	79.2%	54.2%	61.0%	71.1%	57.4%

Dataset distribution with oversampling

We applied oversampling to eliminate the dataset imbalance.

Model performance after oversampling(SMOTE)

The Catboost model performs best. This model has the highest Roc auc score and recall score.

	Accuracy	Recall	Precision	ROC AUC Score	F1 Score
RandomForest Score	76.7%	65.3%	54.2%	73.0%	59.2%
Logistic Score	69.5%	66.0%	44.1%	68.4%	52.9%
LightGBM Score	76.9%	70.5%	54.2%	74.8%	61.2%
KNN Score	65.0%	39.8%	34.7%	56.8%	37.1%
SVC Score	73.0%	11.6%	42.6%	53.1%	18.2%
Catboost Score	77.6%	69.5%	55.5%	75.0%	61.7%

Result

There is not much difference before using SMOTE. However, the recall score and ROC AUC score of the Catboost classification model using SMOTE show the best performance at 69.5% and 75%, respectively.

With the highest recall score and ROC AUC score, this model can be used to predict who is looking for another job in HR.
It will be possible to reduce the company's losses by predicting employees looking for other jobs in advance.