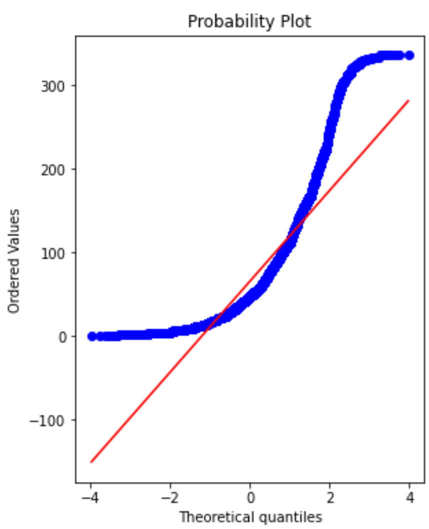
FML LAB

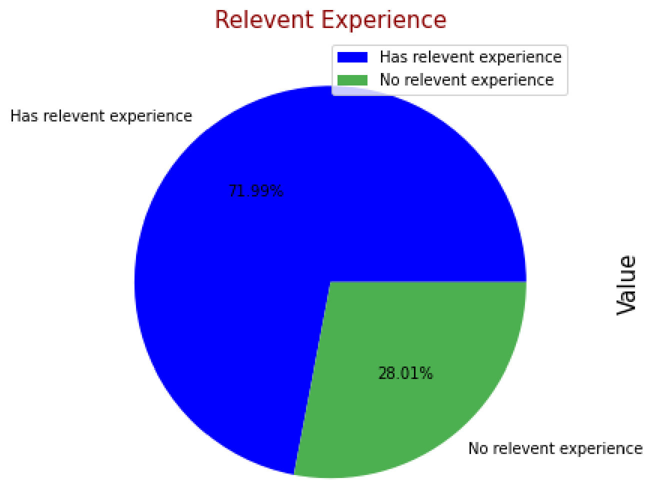
JOB CHANGE PREDICTION

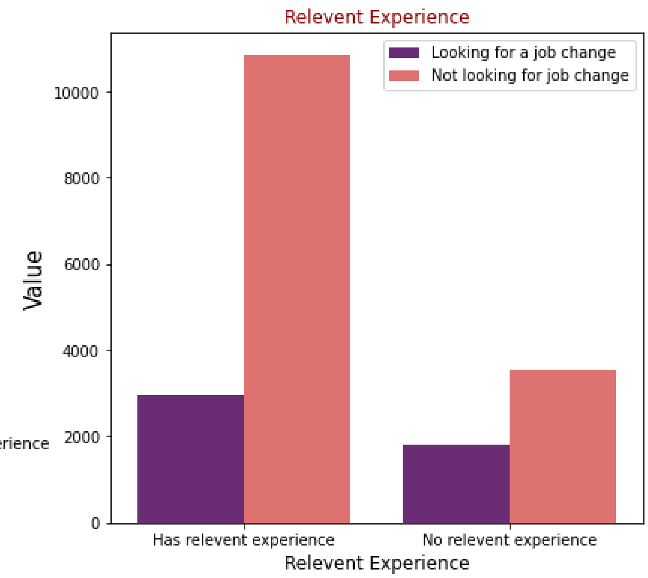
By-Harsh Gupta
Chhavi Lodha
Kartikey Choudhary

Dataset Visualization

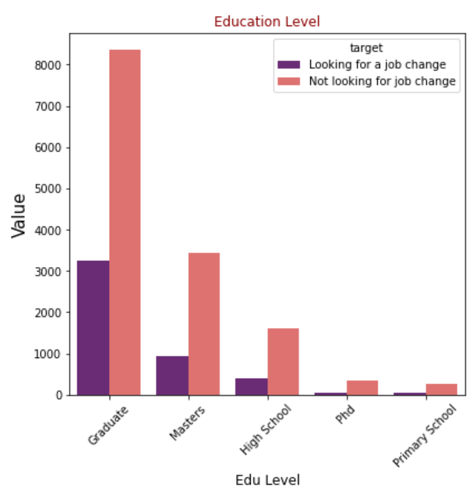


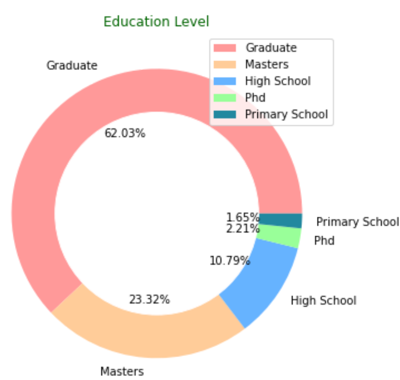
From the probability plot in the above figure, we can infer that the dataset is not normalized.



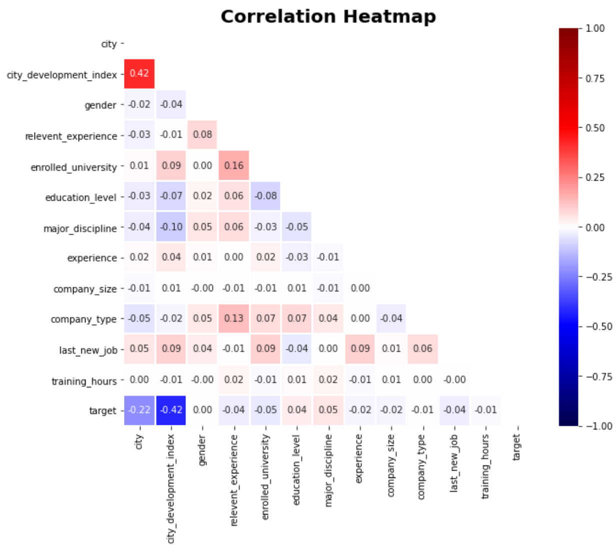


As we can see in the above image, we can infer that people with no-relevant experience are more likely to change jobs rather than relevant experience as most of the experienced people would likely stay in the same job.

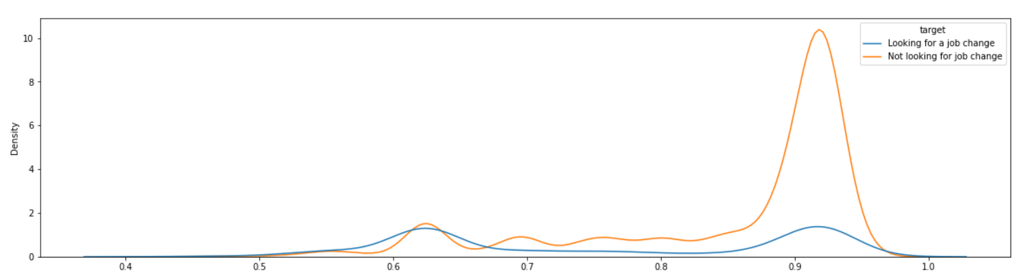




From the above graph we can infer that most of the people in the training dataset are Graduates and masters following them and most of the people that are looking for a change in jobs are from the graduates.



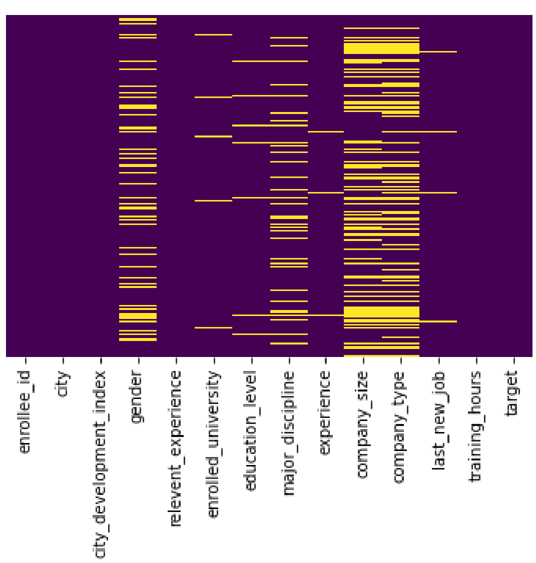
As we can see in the above graphs, we have calculated correlation between features and it can be observed that only highly correlated features are city_devolopment_index and city implying the features are similar and can impact the model towards them.



As we can see in the above graph the people from the well developed cities are most likely to stay in the same job compared to other cities.

Model Preprocessing

We first noticed that most of the values were of a categorical type; most features are categorical (Nominal, Ordinal, Binary), some with high cardinality. To convert the labels in the data to encoded values we used explicit tables so that we can handle them better even after the model is created for the purpose of decoding.
During the pre-processing stage of the model, we saw that there were many missing values in the data, the following graph highlights the missing values:



To handle the following missing values we simply imputed them using linear regression. The best model performance we saw was the Light XGBoost Model with a 91% accuracy and the following confusion matrices.

When we decoded the model we observed that the last_new_job has the most important in predicting the job change of the person.