HR Analytics:

Likelihood a Data Scientist is Looking for a Job Change

Context:

- Company X is interested in potentially hiring a select few data scientists out of all candidates who successfully complete some training courses offered by the company.
- To ensure they don’t waste time and resources, Company X has requested help predicting which candidates are open to a job change, with an emphasis on ***reducing false positives***.

Goals:

1) Build a classification model that successfully identifies candidates open to a job change.
2) Identify the characteristics associated with those open to working for Company X.

Key Results:

- City location (and CDI) are main predictors.
- Gender and college major not very influential.
- Current university students and those with 1 or less years of experience more open to working for Company X.
- **Ideal Target**: University students, recent graduates and/or currently unemployed from City_21 and City_11.

Raw Data:

- Data was sourced from Möbius on kaggle.
  - Each row refers to a candidate in a data science training course.
  - Features include information on demographics, education, level of experience, current job.
  - **Target Feature:** Looking for job change (Yes or No)

Transforming Data:

- Rows missing 4 or more values (*n* = 878) were dropped.
- Company Size and Company Type were both about 30% null.
  - Imputed as ‘N/A’ for candidates with no previous employment.
  - Imputed as ‘Unknown’ for all others.
- Missing values of Major were filled in with ‘N/A’ for candidates who did not have a college-level or higher education.

Final Dataset:

- After encoding categorical features and imputing missing values, the final dataset contained
  - 18,280 rows
  - 44 columns

Model Selection:

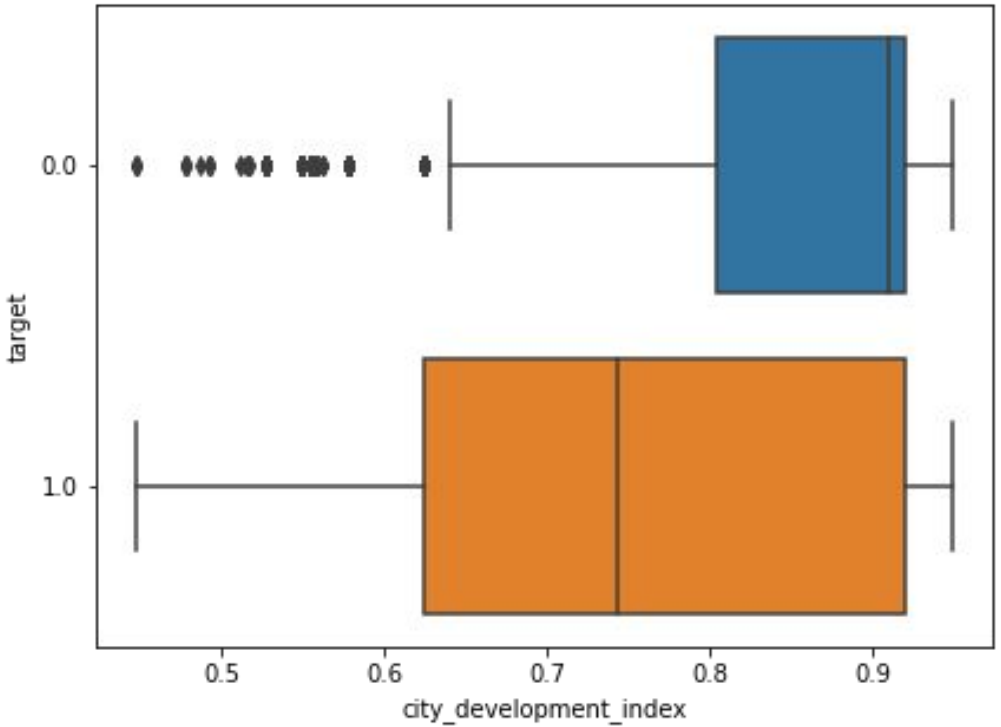
- Four types of classification models were trained and tested, optimized to prioritize a reduction in Type II errors (**F 0.5** metric).
- **Best Model:** XGBoost Classifier

Precision	Recall	Accuracy	F 0.5	AUC
60.10	59.92	80.24%	60.04	.796

Influential Features:

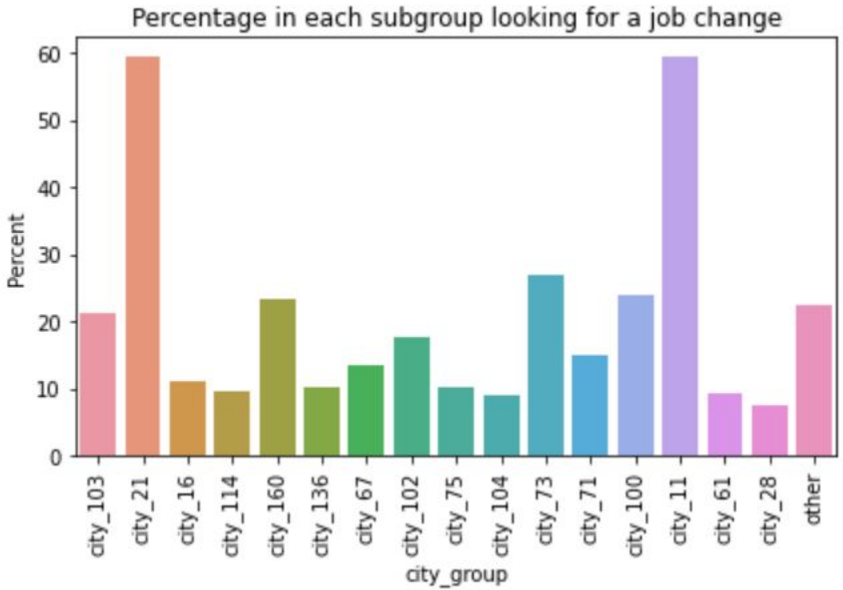
- City Development Index
- Years of Experience
- Current company information unknown
- Lack of relevant experience
- Full-time university student

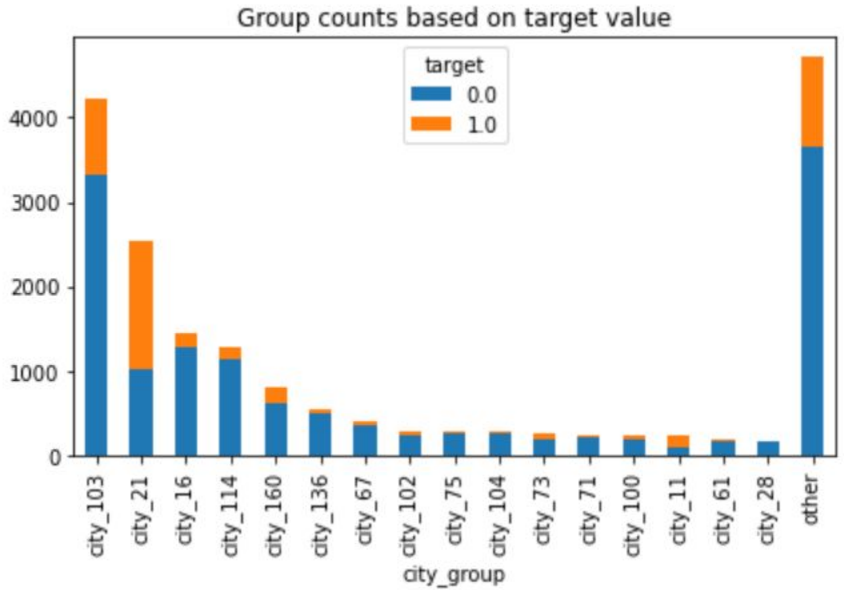
a) City Development Index



- Candidates who are open to working for Company X come from a wider range of cities.
- Median CDI of interested candidates = 0.74
- Median CDI of disinterested candidates = 0.91.

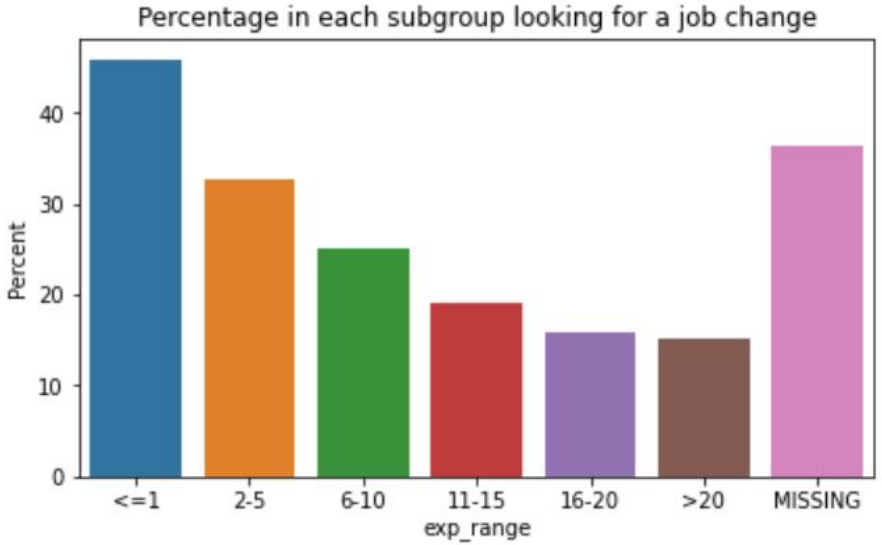
b) City Groups

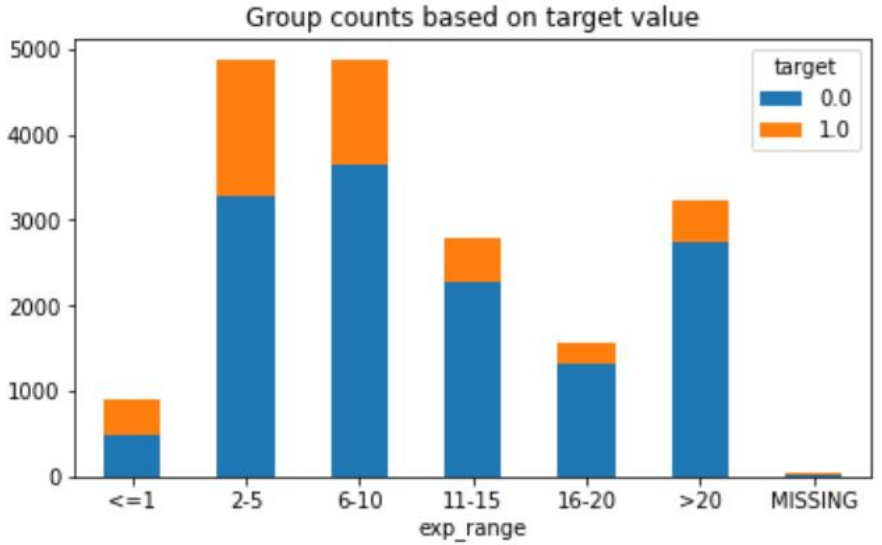




- City_21 represents second largest group of candidates, and is also negatively correlated with CDI.
- Candidates from City_21 (CDI = 0.62) and City_11 (CDI = 0.55) are most likely to be open to a job change.

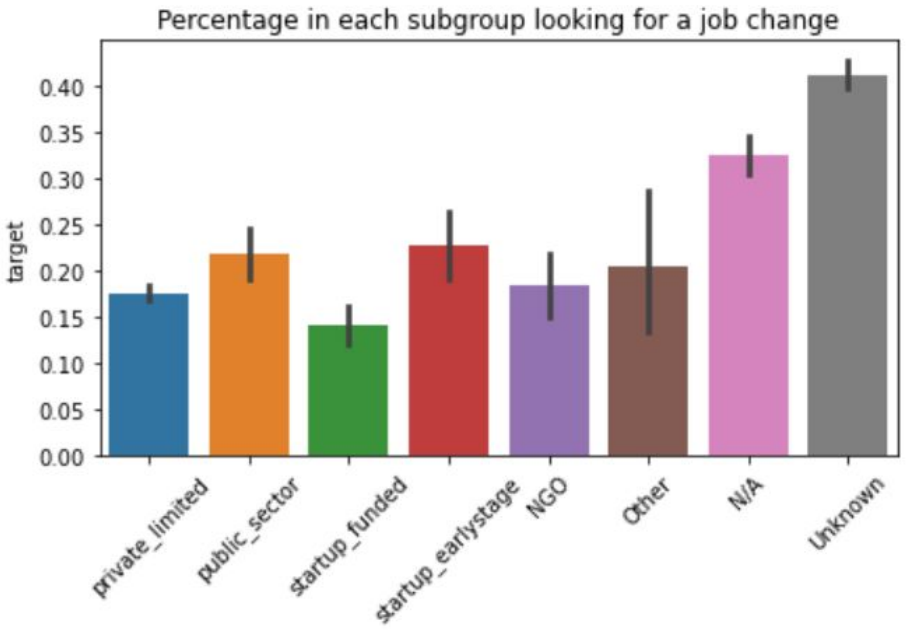
c) Experience Range

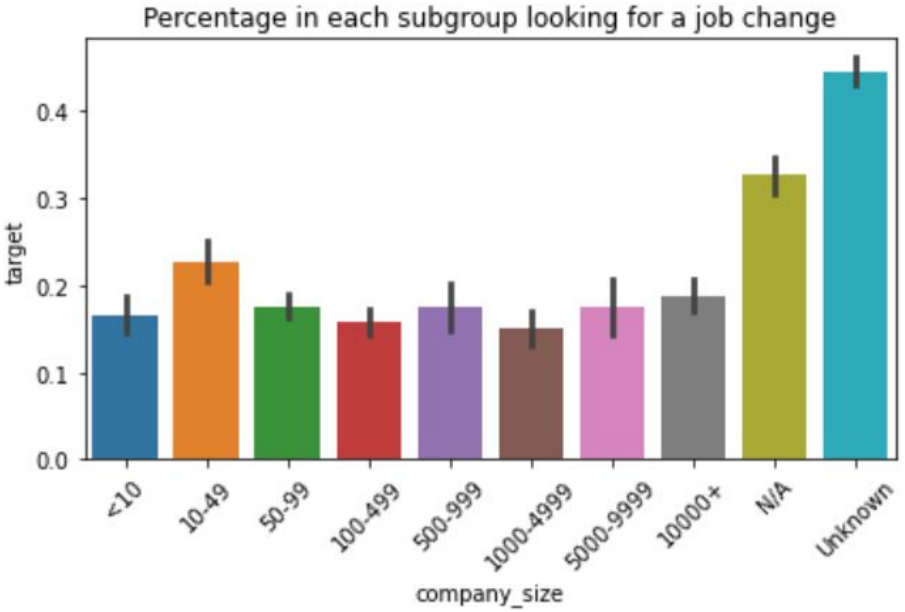




While the training course primarily appeals to those with 2+ years of experience, those with less than 1 year of experience are most likely to be interested in working for Company X.

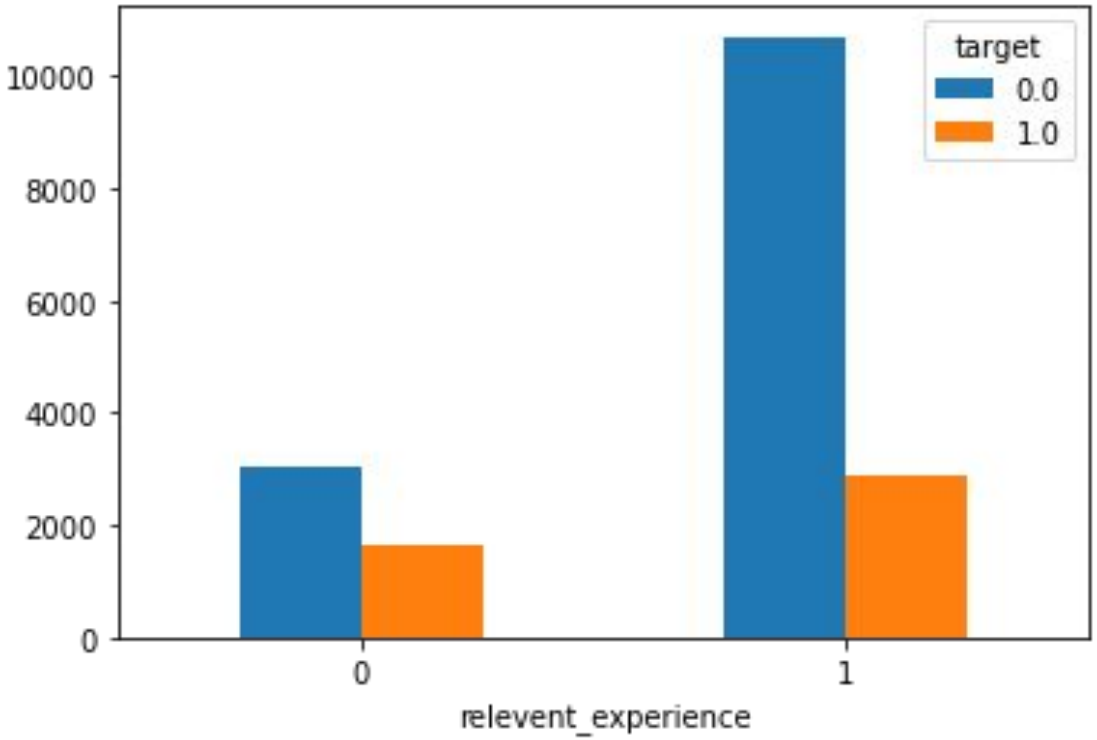
d) No Current Company





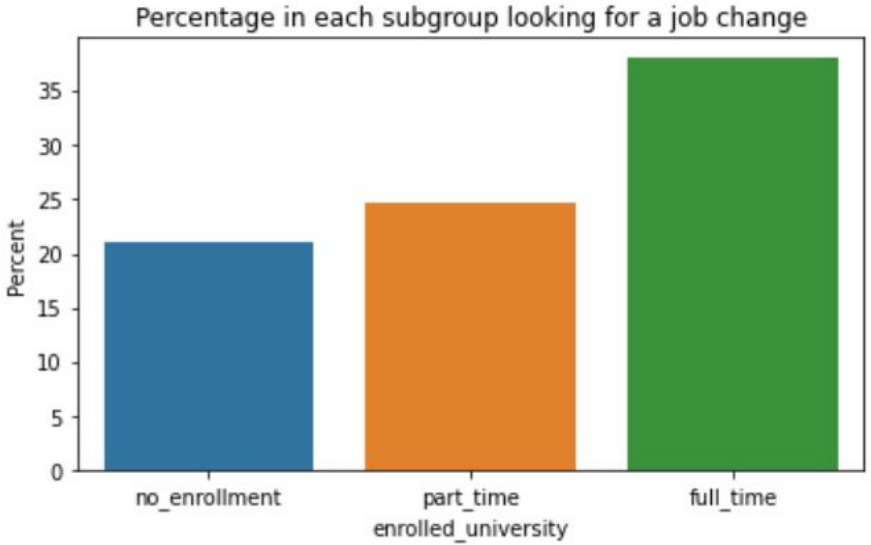
- **N/A** refers to candidates who have never held a previous job, and **Unknown** refers to candidates with missing work information, and are possibly unemployed.

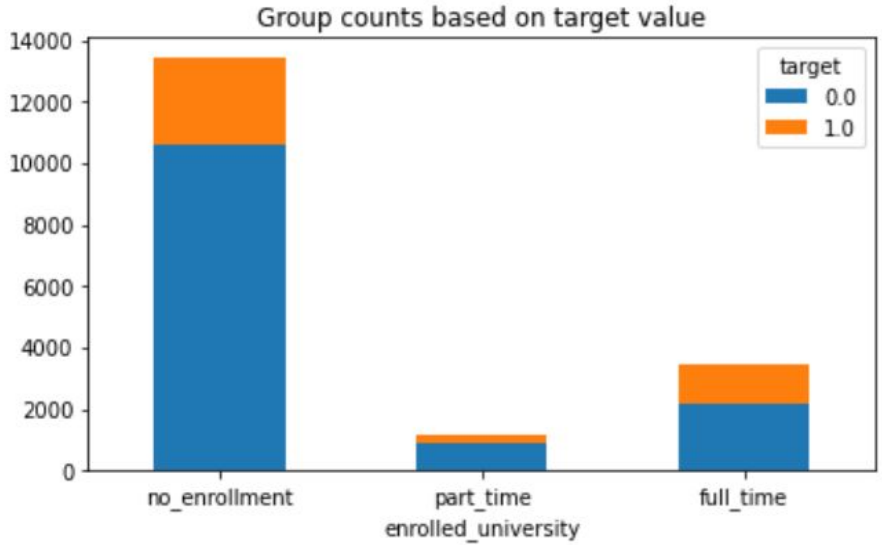
e) Relevant Experience



- Unsurprisingly, candidates without relevant experience were more open to a job opportunity.

f) University Enrollment





- Full-time students are much more likely to be interested in a job change than either non-students or part-time students.

Recommendations:

- When determining which candidates to move forward with, prioritize those from low CDI cities, recent graduates, and career changers.
- Offer internship or entry-level opportunities for candidates currently enrolled in universities or having recently graduated.
- Design a career changer program to streamline the process of being considered for a role after completing certain training courses.

Limitations:

- Missing data on current company size and type is unexplained.
  - Suggestion: Provide a ‘N/A’ option for questions that may not refer to all applicants.
  - Suggestion: Collect more data about candidate employment status and history.
- Since city is so influential, it would be useful to collect more data on each city, such as distance from Company X, population demographics, proximity to universities, etc.

Limitations (cont.):

- Dataset also does not include candidate's current job title (if applicable), and whether their years of experience pertain to data science or a different field.
    - Collecting this info would make it easier to differentiate between recent/soon-to-be grads and career changers.

Thank you!