MINOR PROJECT OF PYTHON

Indian Institute of Technology Mandi

AWESOME HR DATA VISUALIZATION

submitted by

PARTH SHYAM SHUKLA (MB24019)

ABHISHEK MOHAN (MB24041)

V. AISHWARYA (MB24049)

TO

PROF. MANOJ THAKUR

ABOUT OUR DATABASE

A company which is active in Big Data and Data Science wants to hire data scientists among people who successfully pass some courses which conduct by the company. Many people signup for their training. Company wants to know which of these candidates are really wants to work for the company after training or looking for a new employment because it helps to reduce the cost and time as well as the quality of training or planning the courses and categorization of candidates. Information related to demographics, education, experience are in hands from candidates signup and enrollment

PROBLEM STATEMENT

Predict the probability of a candidate looking for a new job.

Note - The main focus of this analysis is on data visualization so the predictive models can almost certainly be improved upon with some basic tweaks.

SOME INPUTS FROM THE DATA SETS

Number of Candidates (Rows): Around 19,000 records.

Number of Features: Over 15 features are used for predicting conversion, such as job-seeking status, experience, and company size.

Key Features:
City Development Index
Experience (years)
Training Hours
Major Discipline
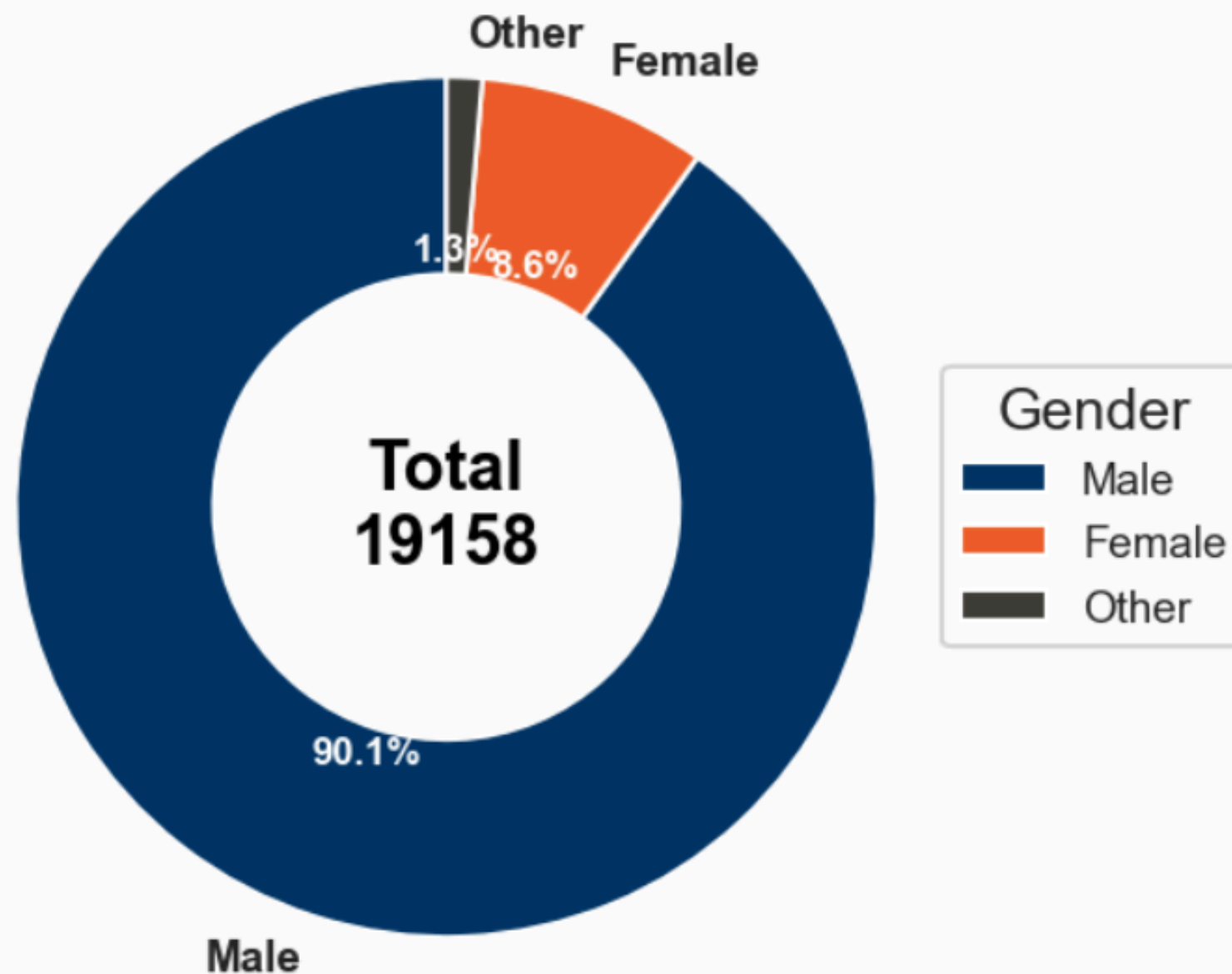Company Size & Type
Gender & Age

Training Data: 19158

Test Data: 2129

Null Value Analysis: Missing values are present in features like gender, education, and company size, which need to be handled for effective prediction.

DATA VISUALIZATION

Gender Distribution
We see majority of participants are Male

Other Female

Total
19158

1.3% 8.6%

90.1%

Male

Gender
- Male
- Female
- Other

The gender distribution chart shows the proportion of males, females, and others among 19,158 participants. Here are the key insights:

1. Male Dominance: Males make up 90.1% of the total participants, indicating a strong male majority in this dataset.

2. Female Representation: Females account for only 8.6% of the participants, suggesting significantly lower representation compared to males.

3. Minimal 'Other' Representation: The 'Other' gender category makes up just 1.3% of the total, indicating a very small presence.

Overall, the chart highlights a substantial gender imbalance, with males forming the overwhelming majority and both females and 'Other' genders having limited representation.

Job Seekers by their course type
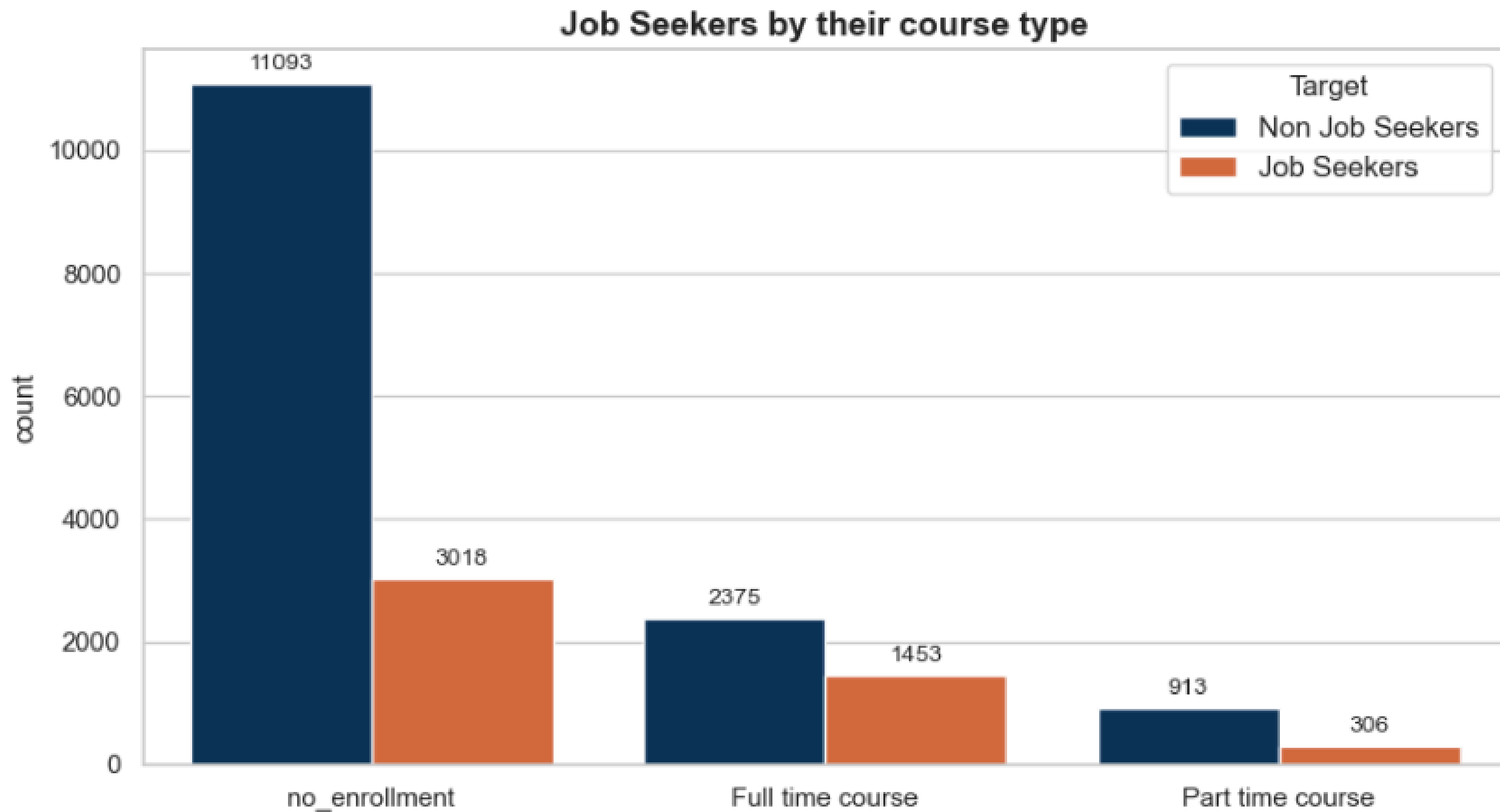
The chart shows that most non-job seekers have no course enrollment (11,087), while the job seekers count is 3,021. Full-time courses have 2,381 non-job seekers and 1,451 job seekers. Part-time courses have the lowest numbers for both categories.

Job Seekers by their major discipline
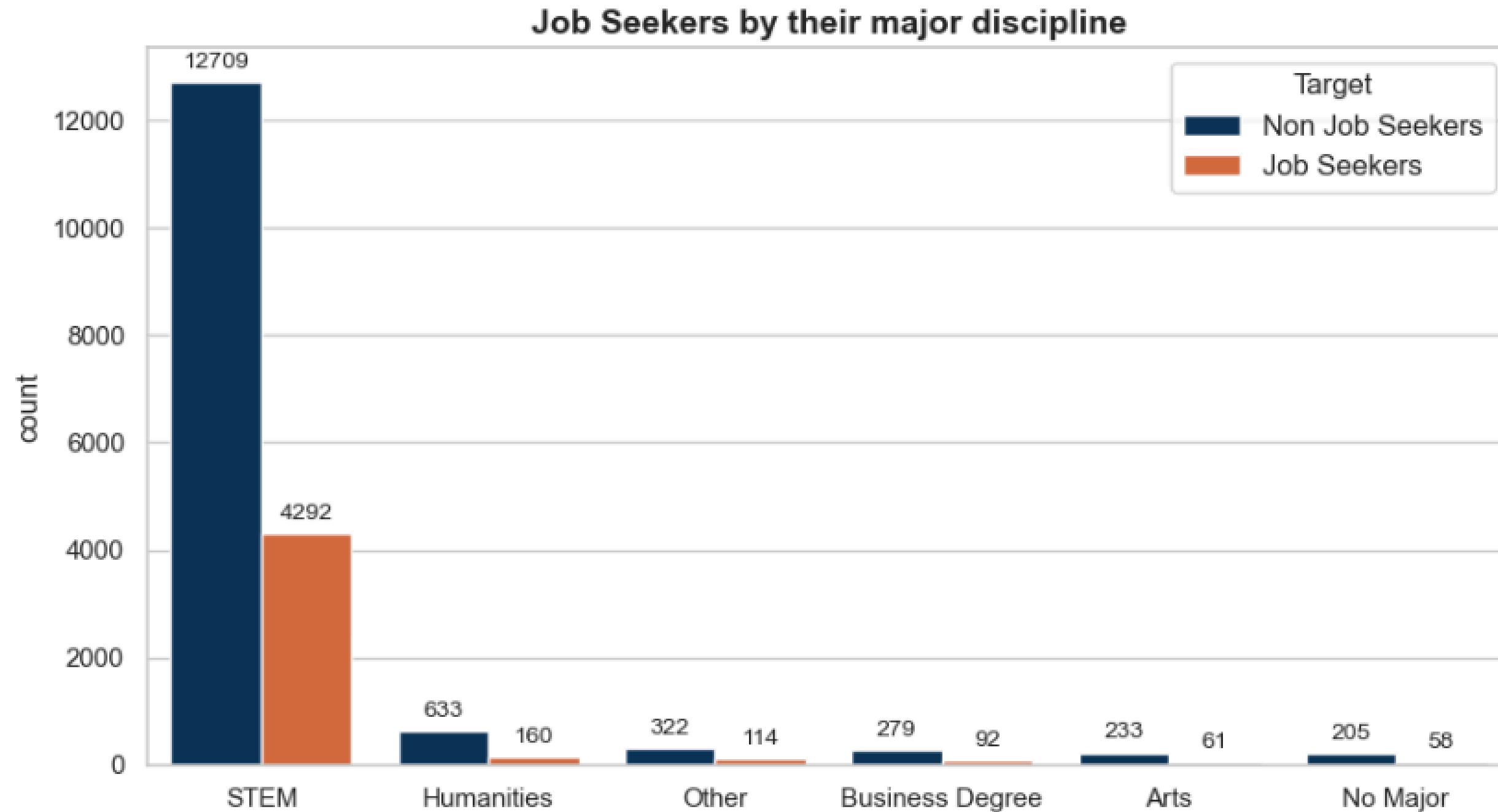
The chart highlights the dominance of STEM majors, with 12,709 non-job seekers and 4,292 job seekers, vastly outnumbering other fields. Non-STEM disciplines like humanities, business, and the arts show much smaller counts. This suggests better job prospects or demand for STEM professionals compared to other majors.
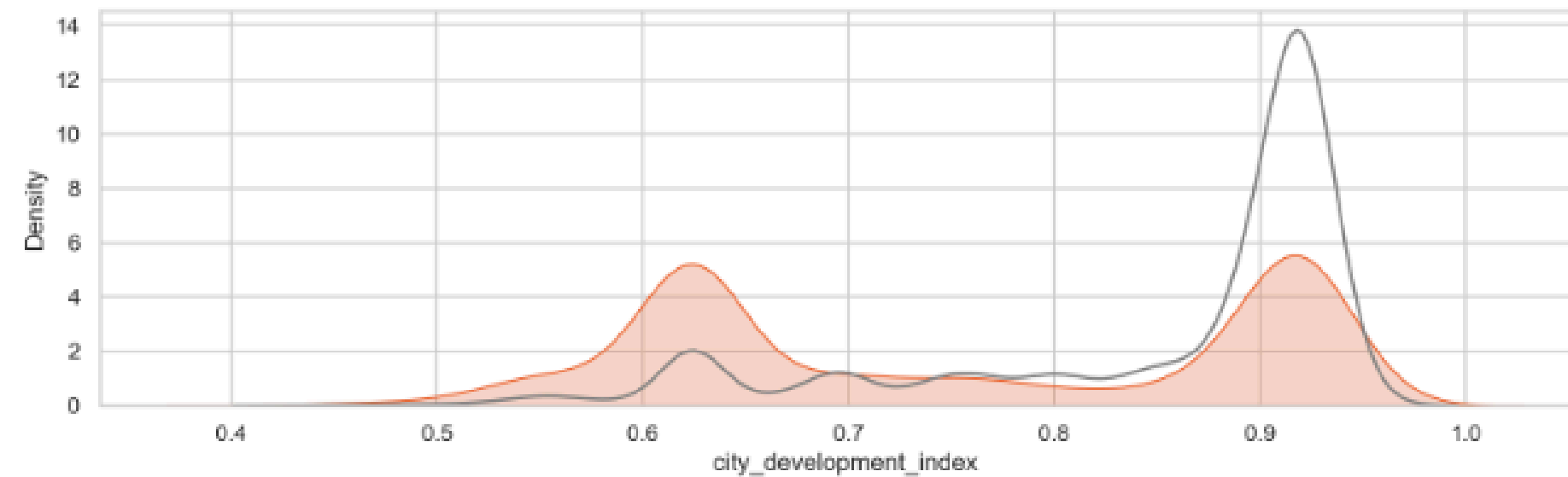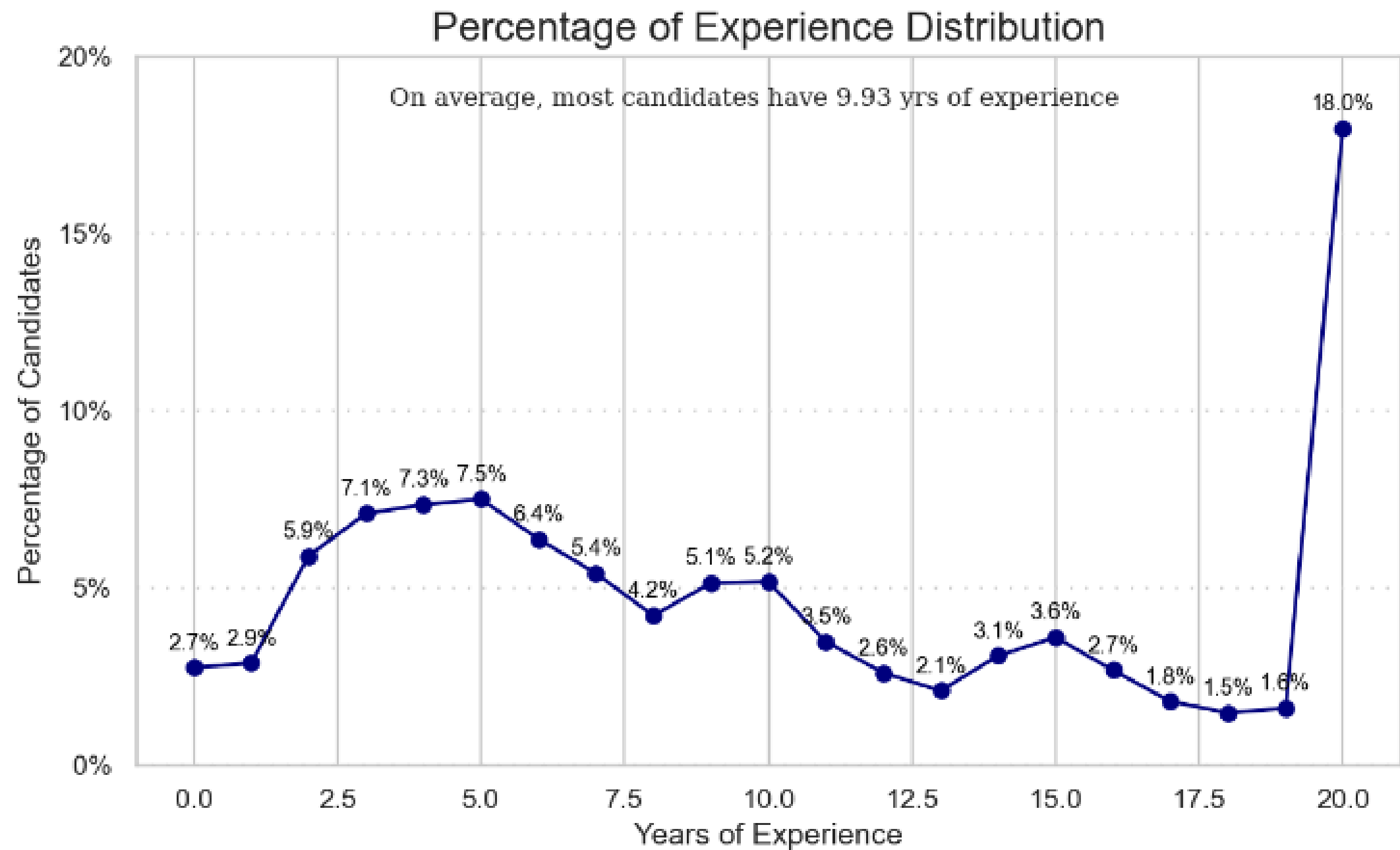
Location based insight of Candidates

Overall
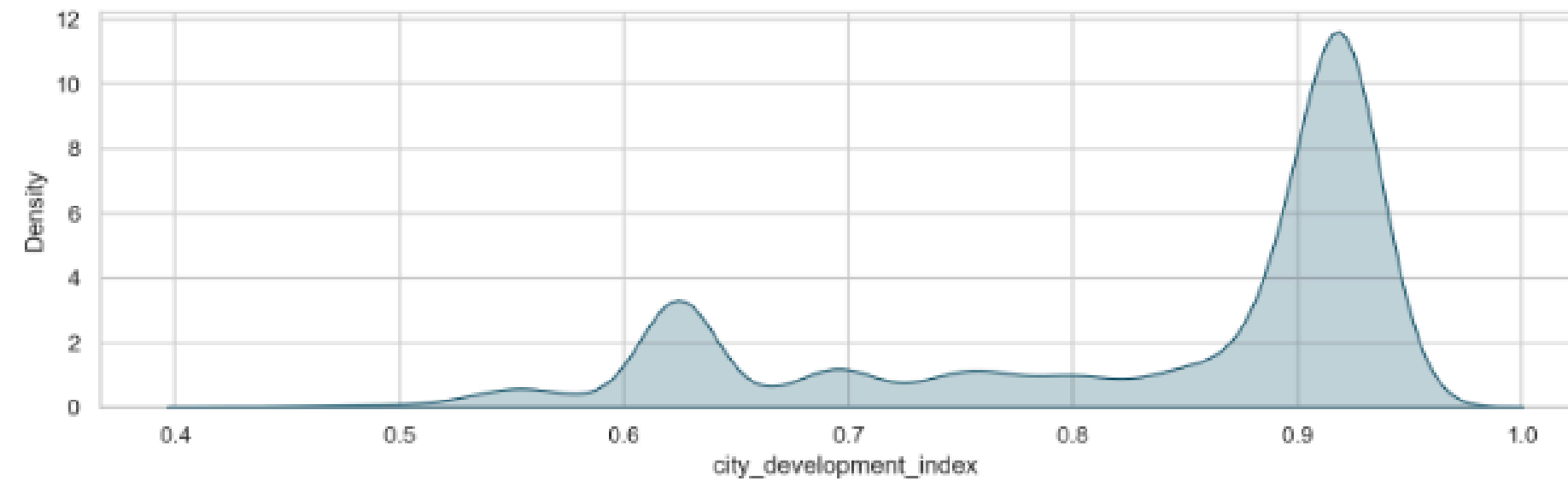
Job Seeker / Non-Job Seeker

It is intresting that most of the job seeker are equally from less developed cities

The plots show most candidates are from highly developed cities (city development index around 0.9–1.0). Non-job seekers are concentrated in these cities, while job seekers are more spread out, including both medium-developed and highly developed cities. This suggests job seekers are more likely to come from less developed areas compared to non-job seekers.

Percentage of Experience Distribution

On average, most candidates have 9.93 yrs of experience

18.0%

7.5%
7.3%
7.1%
6.4%
5.9%
5.4%
5.1% 5.2%
4.2%
3.6%
3.5%
3.1%
2.9%
2.7%
2.7%
2.6%
2.1%
1.8%
1.5% 1.6%

Percentage of Candidates — *Years of Experience*

The chart shows 18% of candidates have 20 years of experience, the largest group by far. Most candidates have 5-10 years of experience, with smaller peaks at 5 years (7.5%) and 10 years (5.2%). The average experience is 9.93 years, with fewer candidates beyond 10 years except for the sharp rise at 20 years.
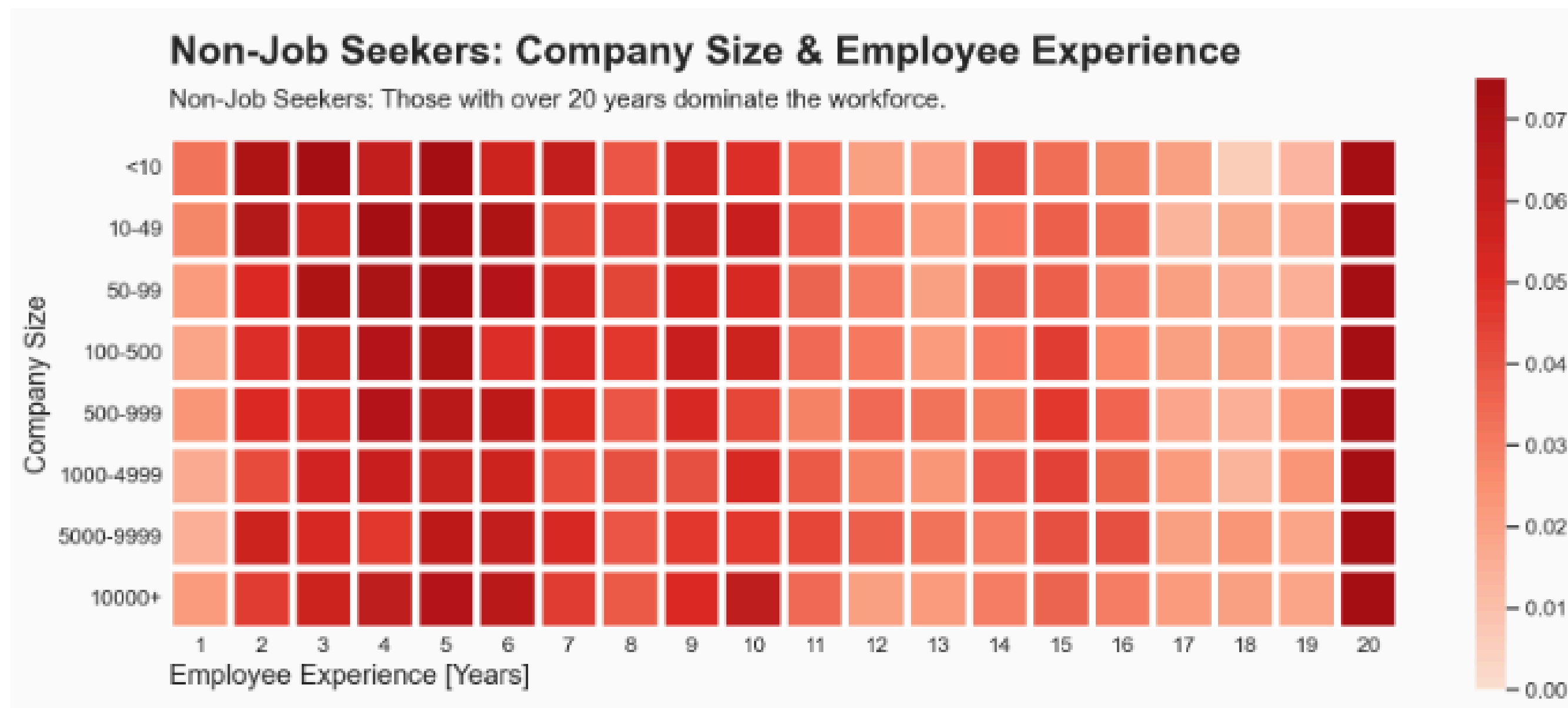
Box Plot of Years of Experience
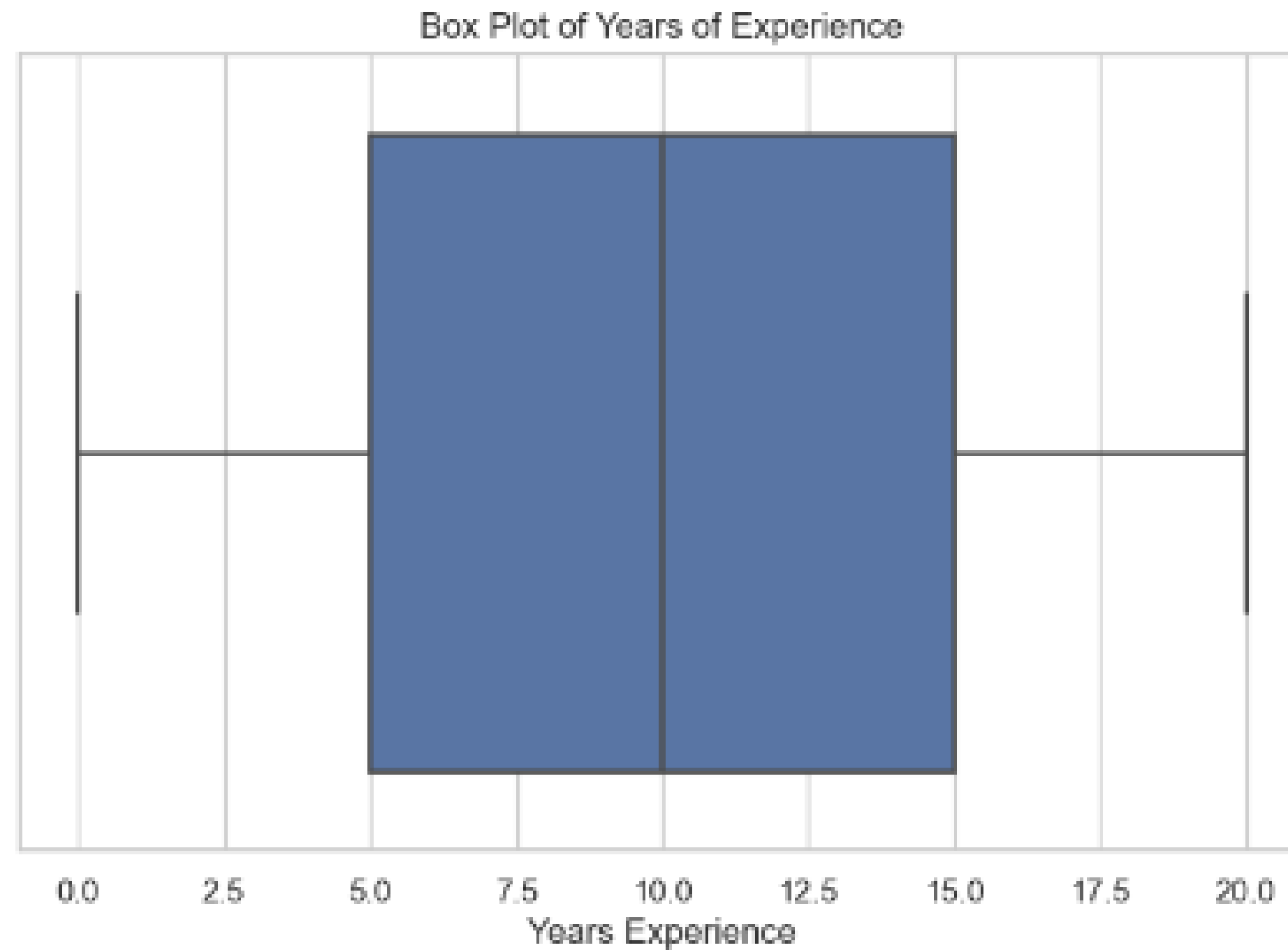
The box plot shows most candidates have 5 to 15 years of experience, with a median of 10 years. Experience ranges from 0 to 20 years, covering early to mid-career professionals.

Training Hours

Most candidates have completed 50-100 hours of training
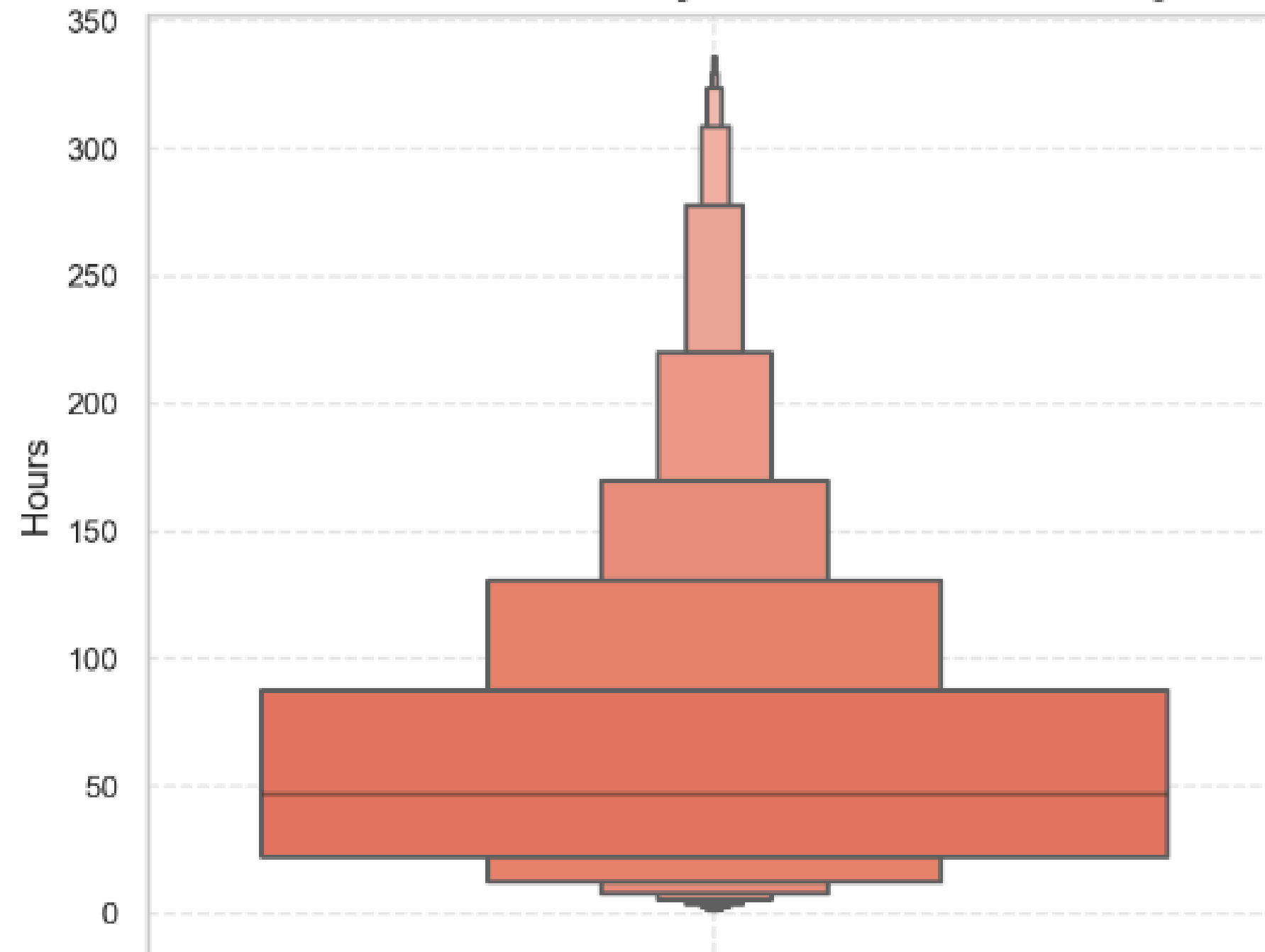
The box plot illustrates the distribution of training hours completed by candidates. Most candidates have completed between 50-100 hours of training, with the median around 50-60 hours. However, the distribution is positively skewed, with some candidates logging up to 350 hours. There's significant variability in training time, ranging from 0 to 350 hours, with several outliers at the upper end. This wide range suggests considerable differences in candidate preparation or training requirements, while the concentration around 50-100 hours indicates a typical training duration for most participants.

Non-Job Seekers: Company Size & Employee Experience

Non-Job Seekers: Those with over 20 years dominate the workforce.

The heatmap shows that most job seekers have 2-6 years of experience. Larger companies prefer more experienced candidates (6-10 years), while smaller firms are open to those with 1-5 years. Mid-career professionals are moderately sought after, and entry-level and highly experienced candidates are less common but still present. This highlights how candidate experience aligns with company size preferences.

Job Seekers: Company Size & Employee Experience
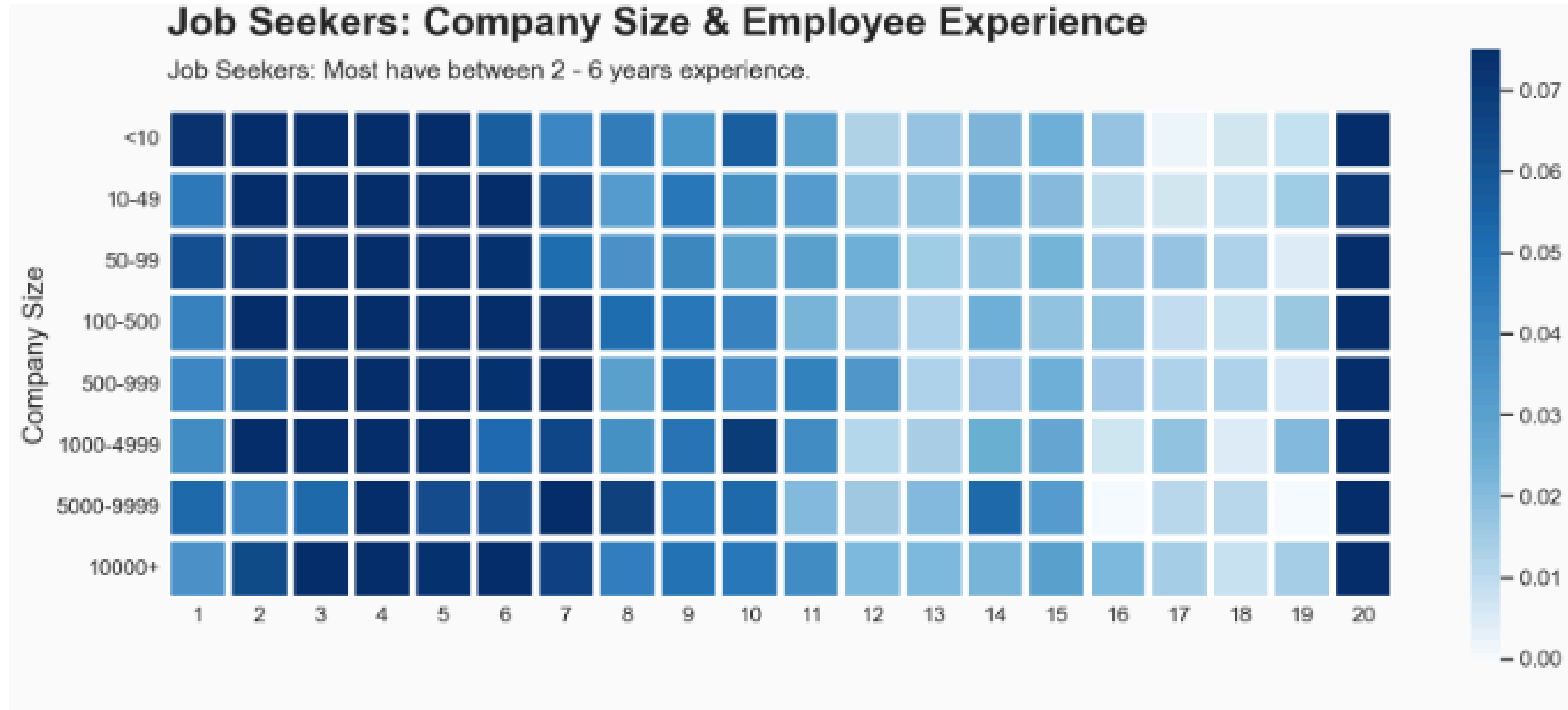
The heatmap shows that employees with 20+ years of experience dominate non-job seekers across all company sizes, indicating strong workforce stability. Mid-career employees (10-15 years) also show stability, while early-career professionals (1-5 years) are more mobile. Retention patterns are consistent regardless of company size, with larger firms slightly better at retaining experienced employees.

Present Company Type: Job Seekers and Non-Job Seekers

Company Type	Job Seekers	Non-Job Seekers
Pvt Ltd	0.75	0.75
Funded Startup	0.063	0.08
Public Sector	0.081	0.071
Early Stage Startup	0.057	0.045
NGO	0.042	0.041
Other	0.0094	0.0094
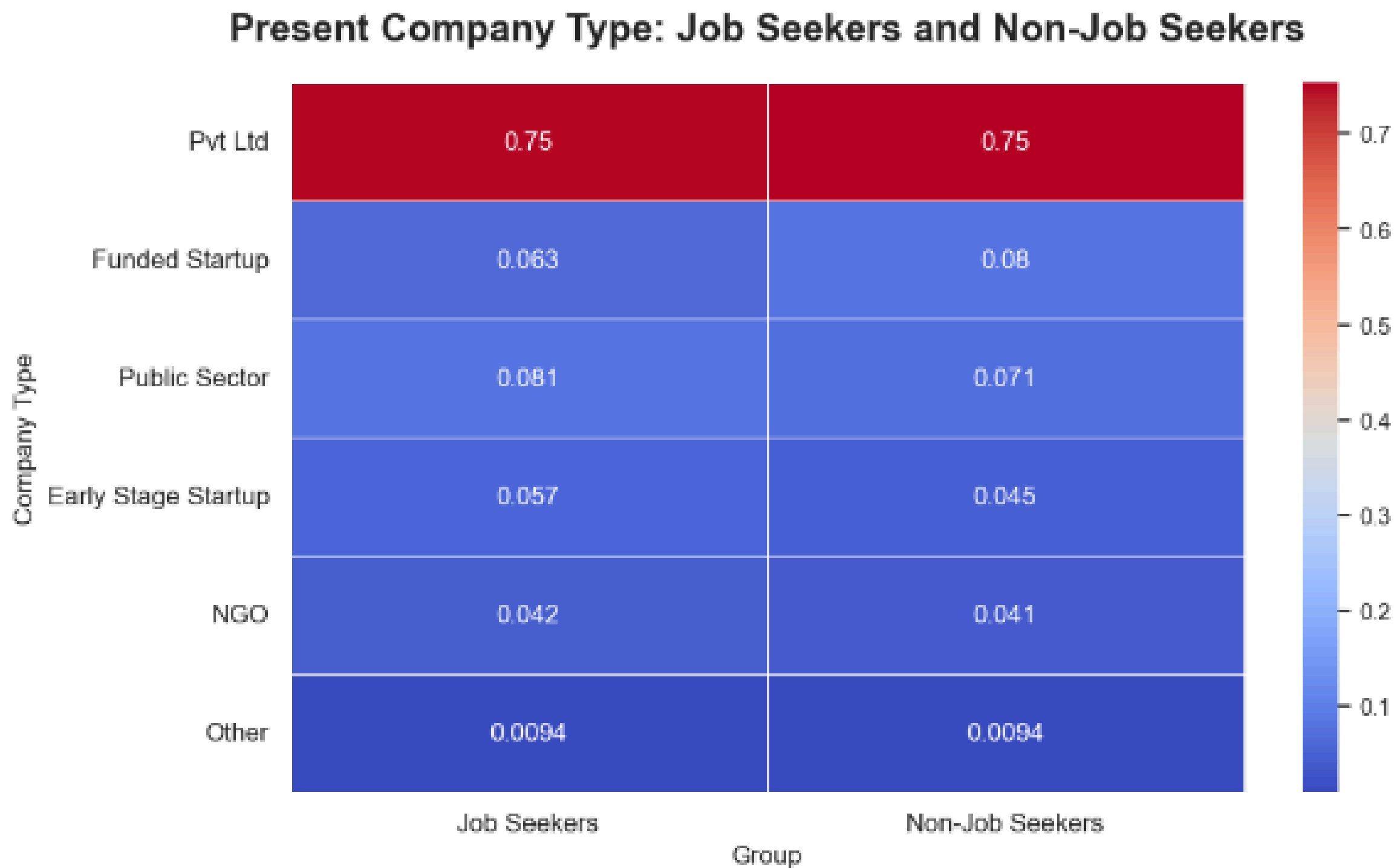
Group

The heatmap reveals that Private Limited companies are the most popular among both job seekers and non-job seekers. Startups, especially funded ones, are favored by job seekers, while the public sector offers a stable option for both groups. NGOs and early-stage startups have limited appeal to both groups. These insights highlight the preferences of job seekers and the current landscape of the Indian job market.

Comparison of Job Seekers and Non-Job Seekers by Company Type
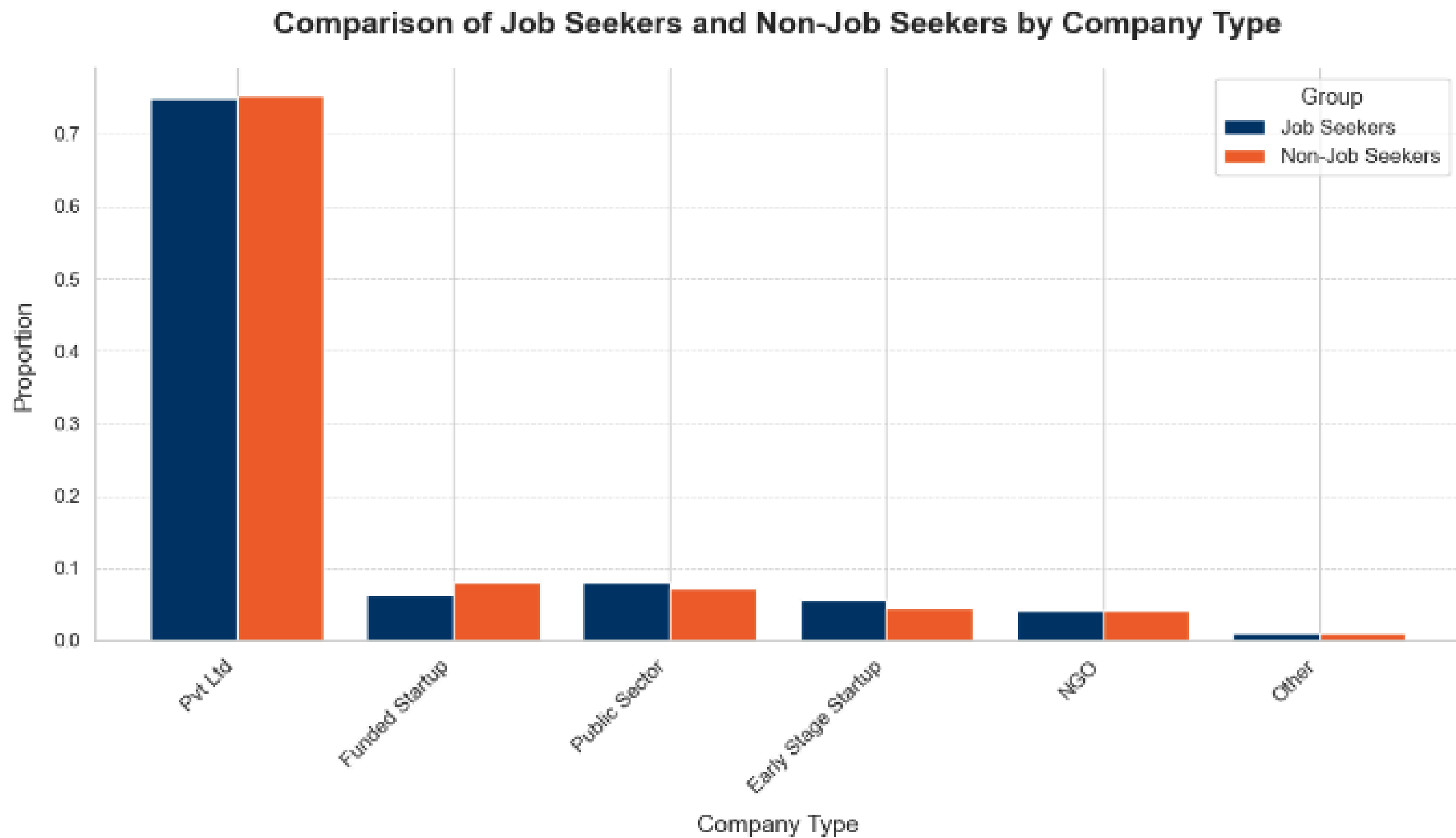
The bar chart shows that Private Limited companies are the most popular among both job seekers and non-job seekers. Startups, especially funded ones, are favored by job seekers, while the public sector offers a stable option. NGOs and early-stage startups have limited appeal to both groups. These insights highlight the preferences of job seekers and the current landscape of the Indian job market.

Job Change Frequency: Job Seekers vs Non-Job Seekers
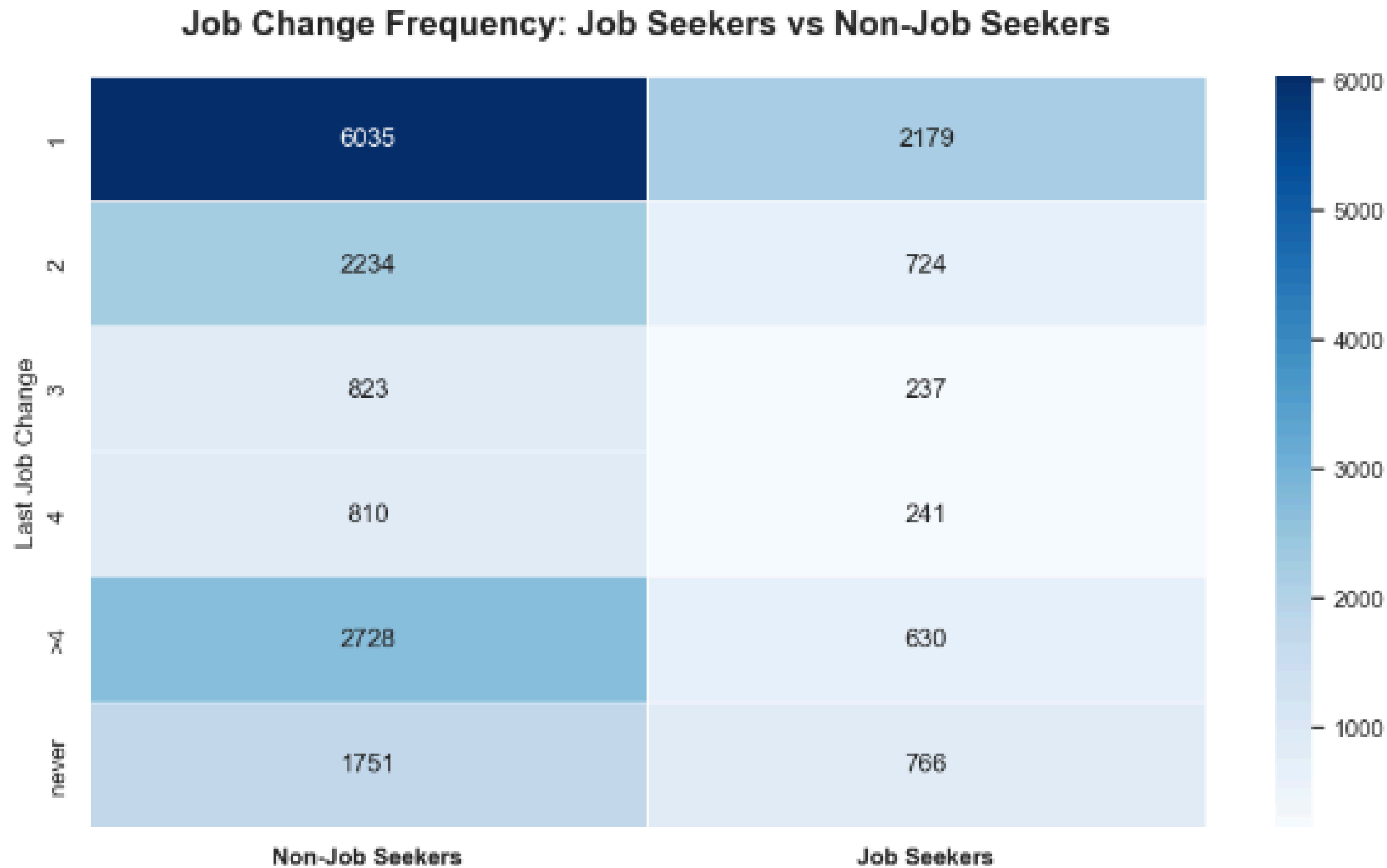
The heatmap shows that job seekers are more likely to change jobs frequently than non-job seekers. As time passes, both groups tend to change jobs less often. The "never" category is larger for non-job seekers, indicating greater career stability. These insights suggest a dynamic job market for job seekers and a trend towards longer job tenures over time.

INFERENCES FROM THE ANALYSIS

The data reveals several key trends across various aspects of the job market:

1. **Gender Imbalance:** Males dominate the workforce, comprising 90.1% of participants, while females (8.6%) and 'Other' (1.3%) have significantly lower representation.

2. **STEM Dominance:** STEM majors overwhelmingly dominate both job seekers and non-job seekers, indicating higher demand or job security in this field. Non-STEM majors like humanities and arts have much lower representation.

3. **City development influence:** Most candidates, especially non-job seekers, are from highly developed cities. Job seekers are more geographically dispersed, including candidates from medium-developed cities.

4. **Experience Distribution:** A significant portion of candidates (18%) have 20 years of experience. The average experience is 9.93 years, with most candidates having 5-10 years of experience.

5. **Training Hours:** Most candidates have completed between 50 and 100 hours of training, though the distribution is skewed with some logging up to 350 hours, indicating a wide range of training commitment.

6. **Company Preferences:** Private limited companies are the most popular among both job seekers and non-job seekers, with funded startups appealing more to job seekers. Larger firms prefer mid-career professionals, while smaller companies are open to early-career candidates.

7. **Job Mobility:** Job seekers tend to change jobs more frequently than non-job seekers, who show greater career stability, especially in larger firms.

8. **Retention and Experience:** Non-job seekers, particularly those with 20+ years of experience, show strong retention across company sizes. Early-career professionals (1–5 years) are more likely to change jobs.

Overall, the data reflects a workforce dominated by experienced STEM professionals, with strong gender imbalances, a preference for larger firms, and stable job retention for seasoned workers.

What Additionally we did in our project

1. Data cleaning (proportional imputation)
2. Strategic approach of data analysis through demographic, education, location and career-based insights

3. We have compared the city development index with the target value and checked whether CDI plays any role or not.
4. How many have relevant experience?
5. What streams most graduates come from?What streams do their education level?

CONTRIBUTIONS

Name	Roll No.	Contribution in analysis (The part of data analysis you did)	Details of use of web resources/ Codes/AI tools etc. (Name the tool and the purpose it was used)	Overall contribution to the work done. (in % after discussion with all team members)
Parth Shyam Shukla	MB24019	Data Clearing Presentation making	Chat GPT, Kaggle	32%
Abhishek Mohan	MB24041	Data Visualization Presentation making	Chat GPT, Kaggle	33%
V. Aishwarya	MB24049	Data Visualization Machine learning Algorithms	Chat GPT, Kaggle	35%

THANK YOU