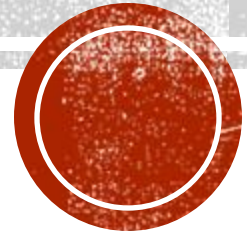
KEY RESUME ATTRIBUTES IMPACTING JOB CALLBACKS



Presenter: Kejing Yan

Brown University Data Science Institute

Oct 21st, 2024

https://github.com/anyfruit/Key-Resume-Attributes-Impacting-Job-Callbacks.git

INTRO

Problem to Solve

- **Objective**: Attributes influential in determining *job application callback rates*
- **Key Question**: Can we predict factors increase the likelihood of receiving callback?

Why This is Important

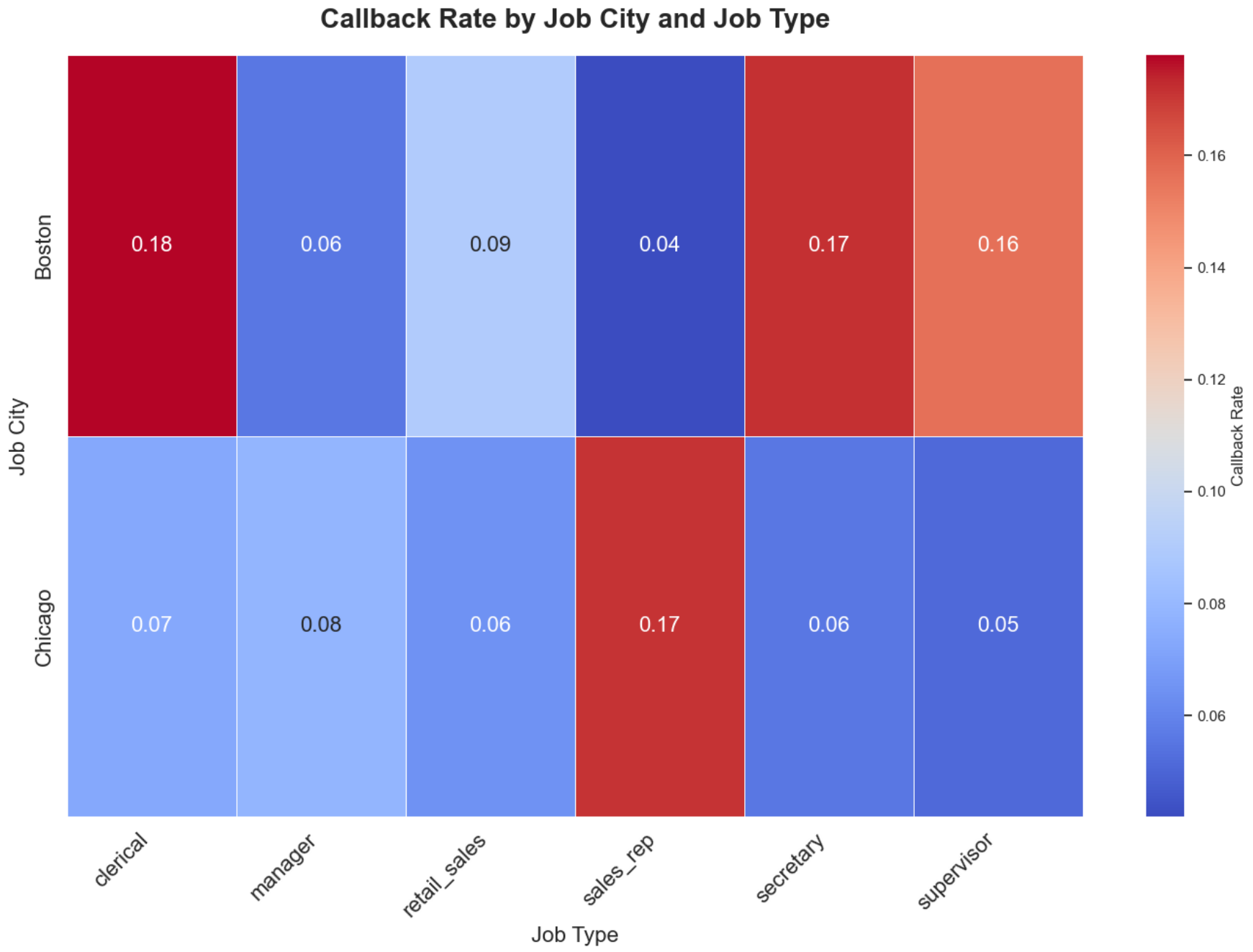
- Guiding applicants in optimizing their resumes
- Beneficial to help addressing bias in hiring practices
- Improving fairness and transparency in hiring processes

Type of Problem

- *Binary classification* problem; target variable is either 1 or 0

Data Collection

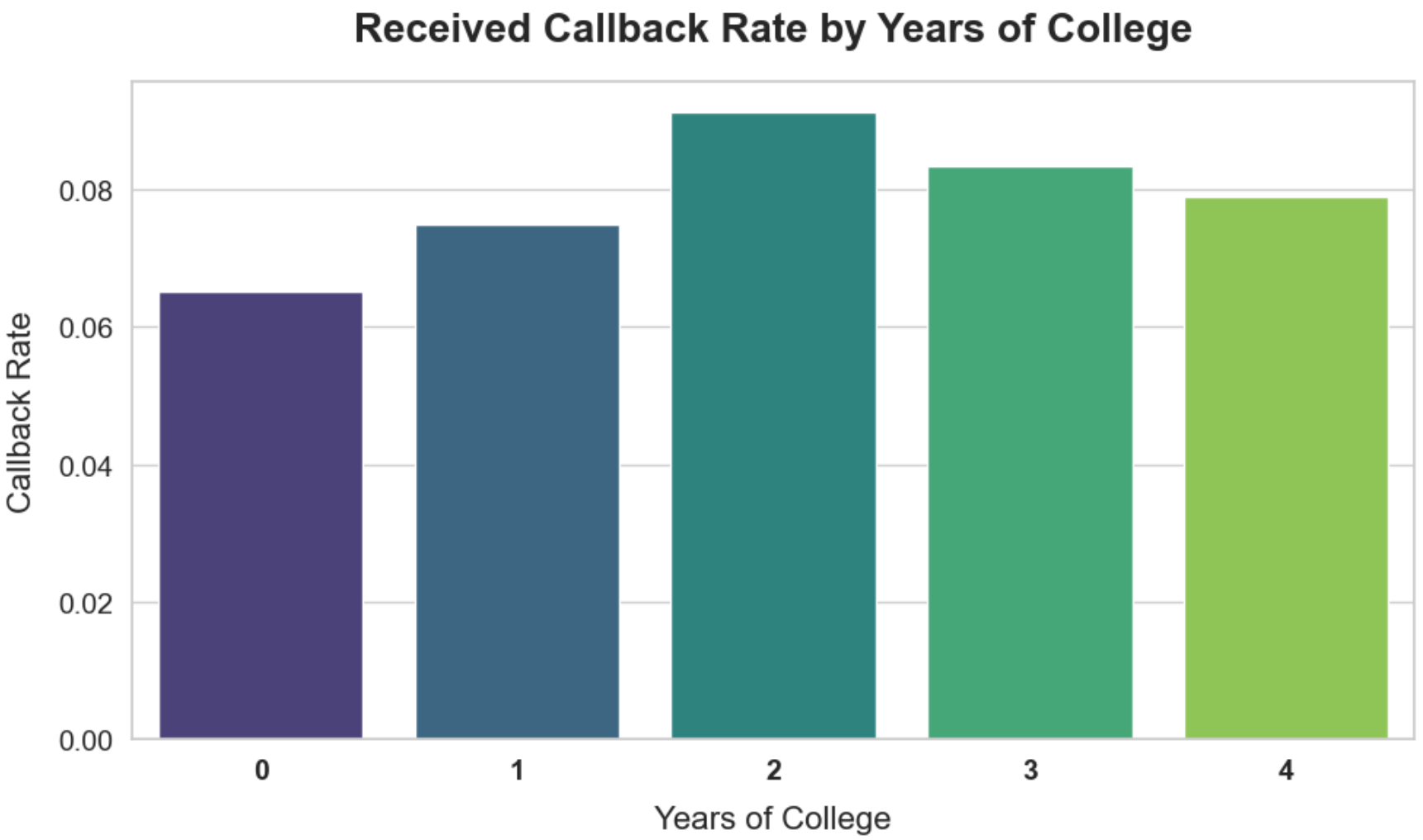
- Kaggle Site: https://www.kaggle.com/datasets/utkarshx27/which-resume-attributes-drive-job-callbacks
- Monitored job postings in Boston and Chicago



EDA

VISUALIZATION 1 (HEATMAP)

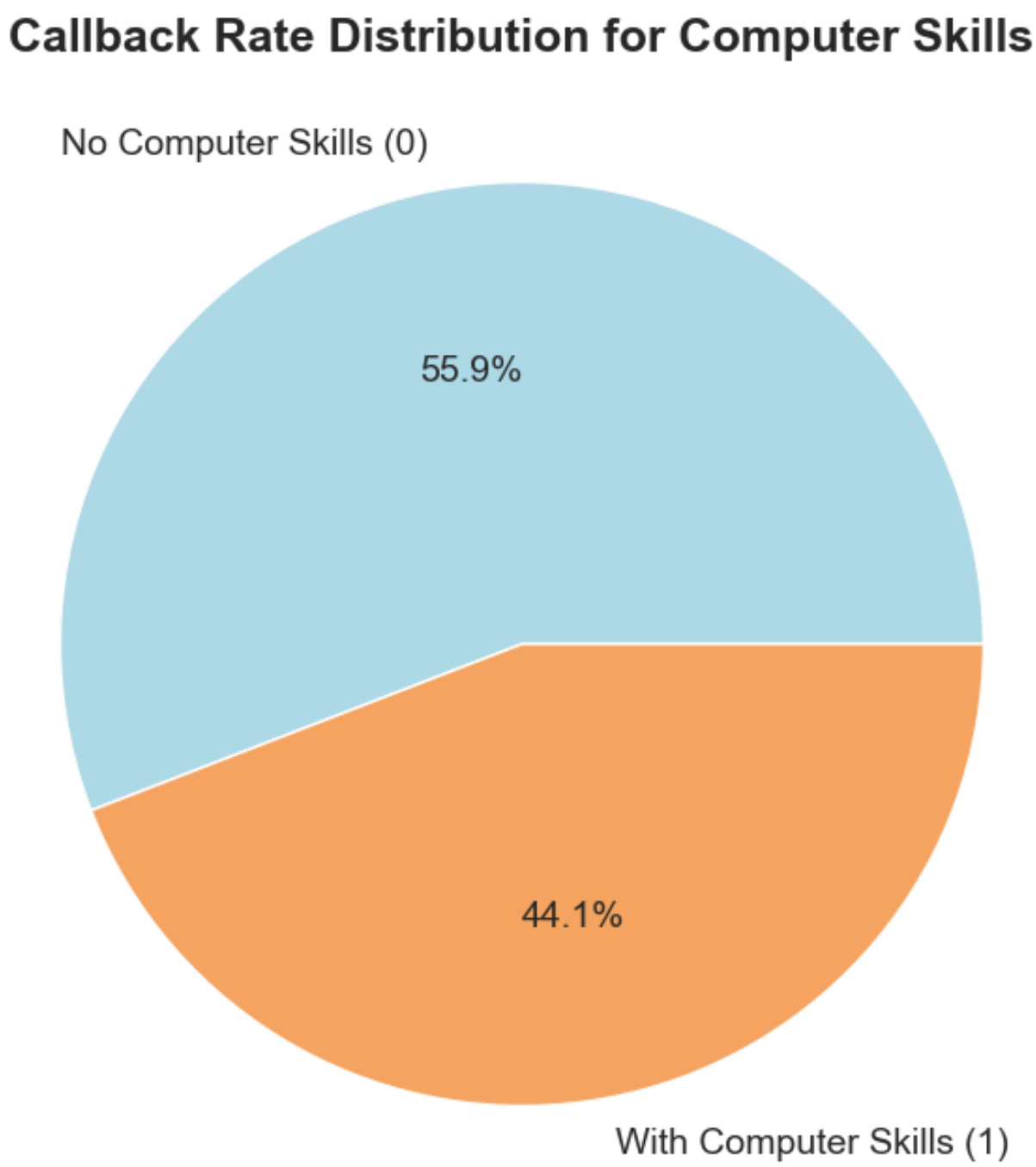
- **Significant geographic and job-specific disparities**
- Boston shows higher callback rates across most jobs



EDA

VISUALIZATION 2 (HISTOGRAM)

- **Unexpected trend with college experience**
- Candidates with 2 years of college experience have highest callback rate
- Surprisingly, those with 4 years show lower callback rates

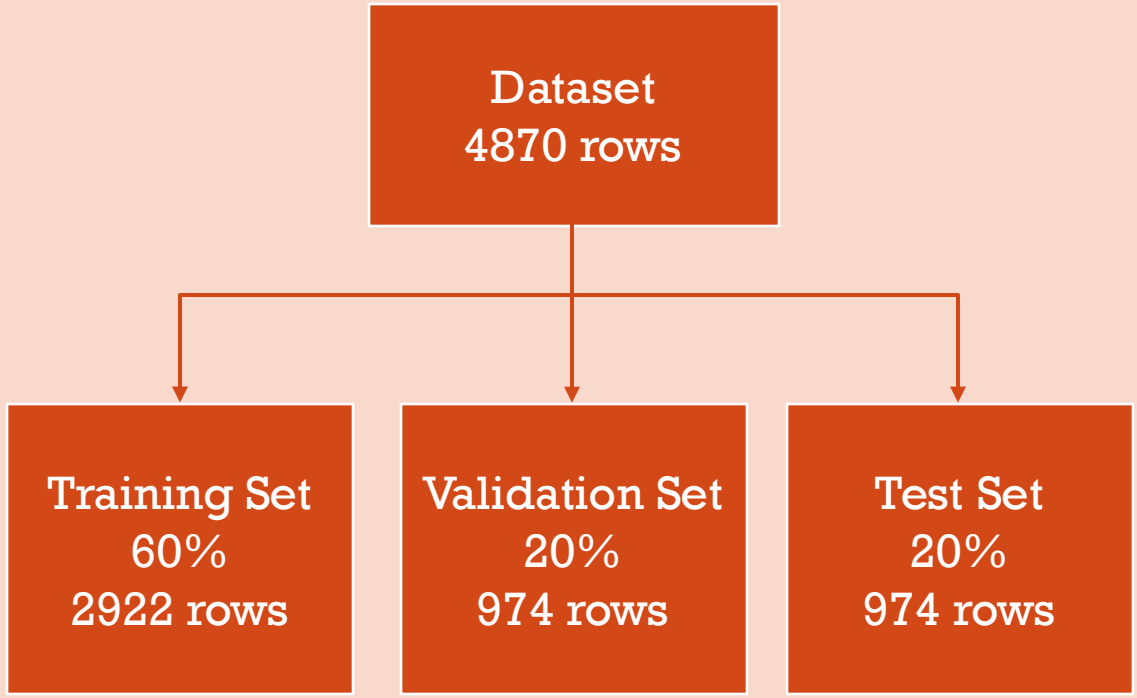


EDA

VISUALIZATION 3 (PIE CHART)

- 55.9% of callbacks are for applicants without listed computer skills
- 44.1% of callbacks are for those with computer skills

SPLITTING



Features excluded:

firstname

*not relevant for predicting callbacks

job_ad_id

*not relevant for predicting callbacks

PREPROCESSING

Handling Missing Data

- Missing values in 3 categorical columns 'UNK', 'Unknown', and NaN were standardized to "unknown" to maintain consistency

Feature Name	Missing Values	Count	%
job_fed_contractor	NA	1768	36.3
job_ownership	Unknown	1992	40.9
job_req_min_experience	NaN	2746	56.4

Renamed Values	
1768	unknown
1992	unknown
2746	unknown

FEATURE ENCODING

Ordinal Encoding

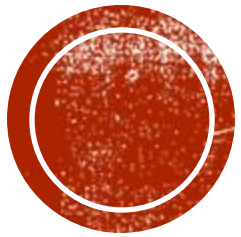
- Applied to features with a natural order
- Example: years_college:
  - 0
  - 1
  - 2
  - 3
  - 4

One-Hot Encoding

- Used for categorical features with no inherent order
- Example: job_ownership
  - unknown
  - nonprofit
  - private
  - public

PREPROCESSING

- **Before Preprocessing**
  - 27 features
- **After Preprocessing**
  - 60 features (one-hot encoding expansion)
- **Final Training Dataset**
  - Contains 2922 data points with 60 features



THANK YOU FOR LISTENING