KEY RESUME ATTRIBUTES IMPACTING JOB CALLBACKS

Kejing Yan

Brown University Data Science Institute

GitHub Repo Link

INTRO

Problem to Solve

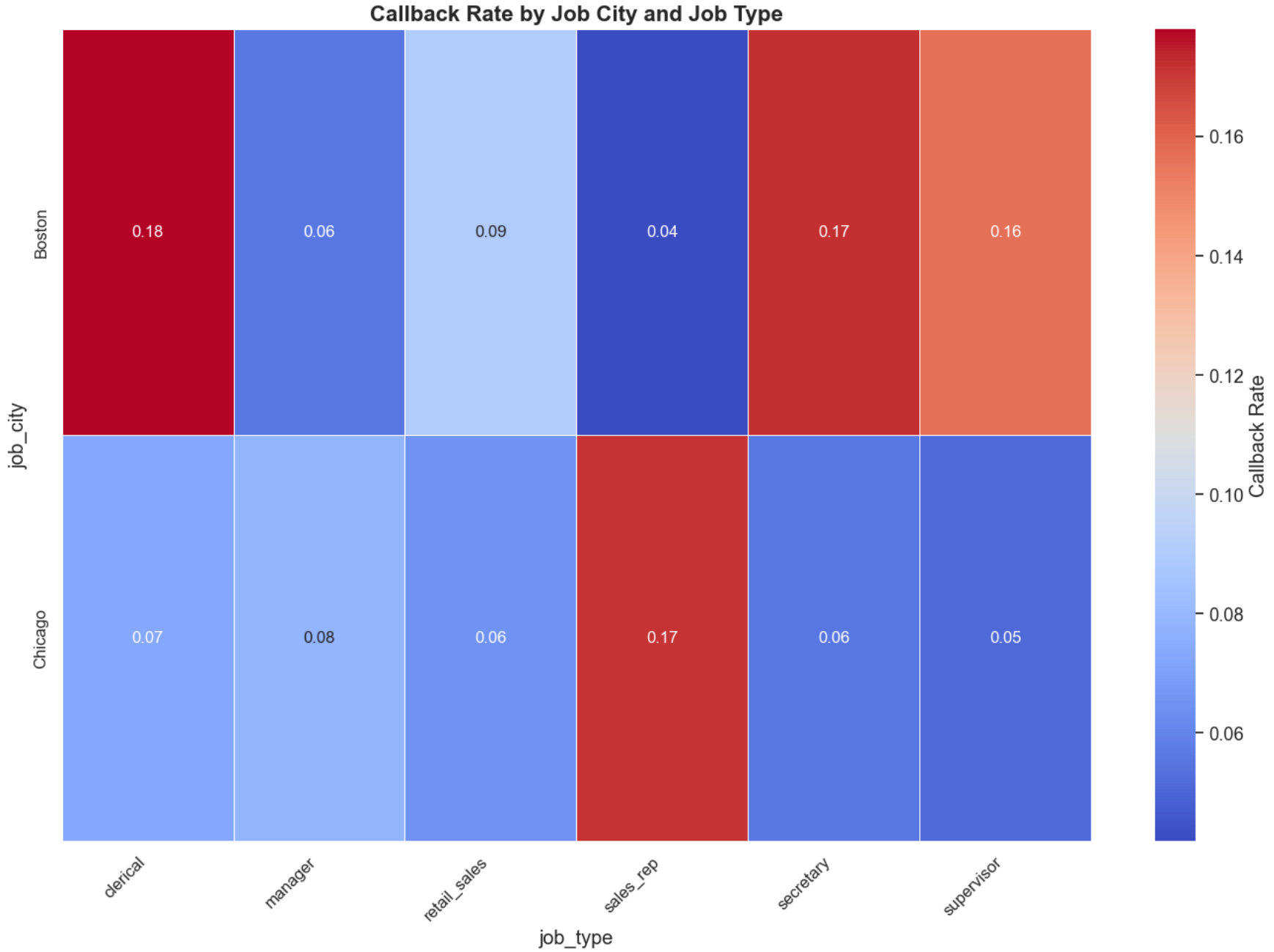
- **Objective**: Attributes influential in determining *job application callback rates*.
- **Key Question**: Can we predict factors increase the likelihood of receiving callback?

Why This is Important

- **Importance**: Beneficial to help addressing bias in hiring practices.
- **Impact**
  - Improving fairness and transparency in hiring processes
  - Guiding applicants in optimizing their resumes

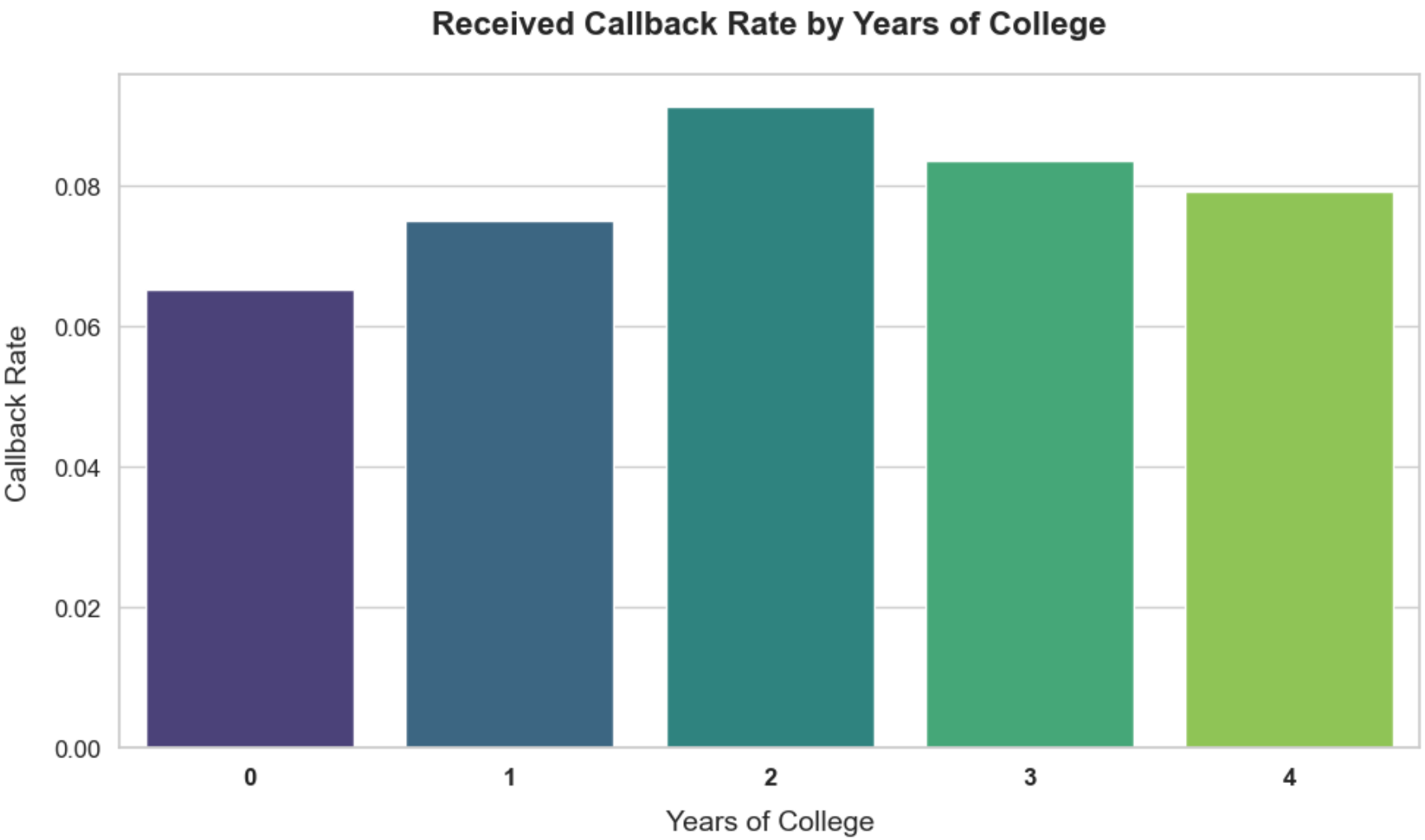
Type of Problem

- **Problem Type**
  - This is a *binary classification* problem
  - Target variable (received callback) is either **1** (callback received) or **0** (no callback received)



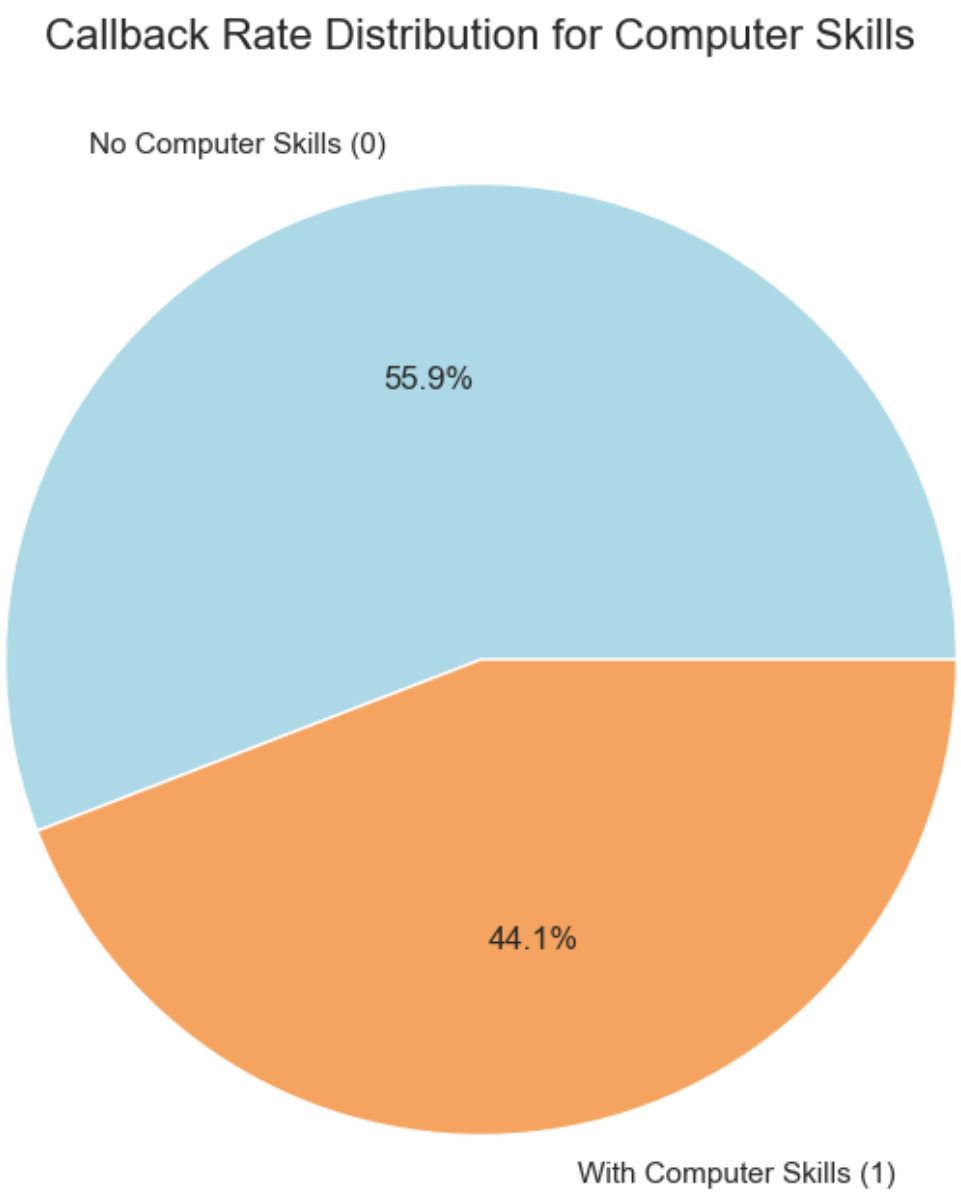
VISUALIZATION 1 (HEATMAP)

- Boston shows significantly higher callback rates across most job
- Chicago outcompete Boston in Sales_rep



VISUALIZATION 2 (HISTOGRAM)

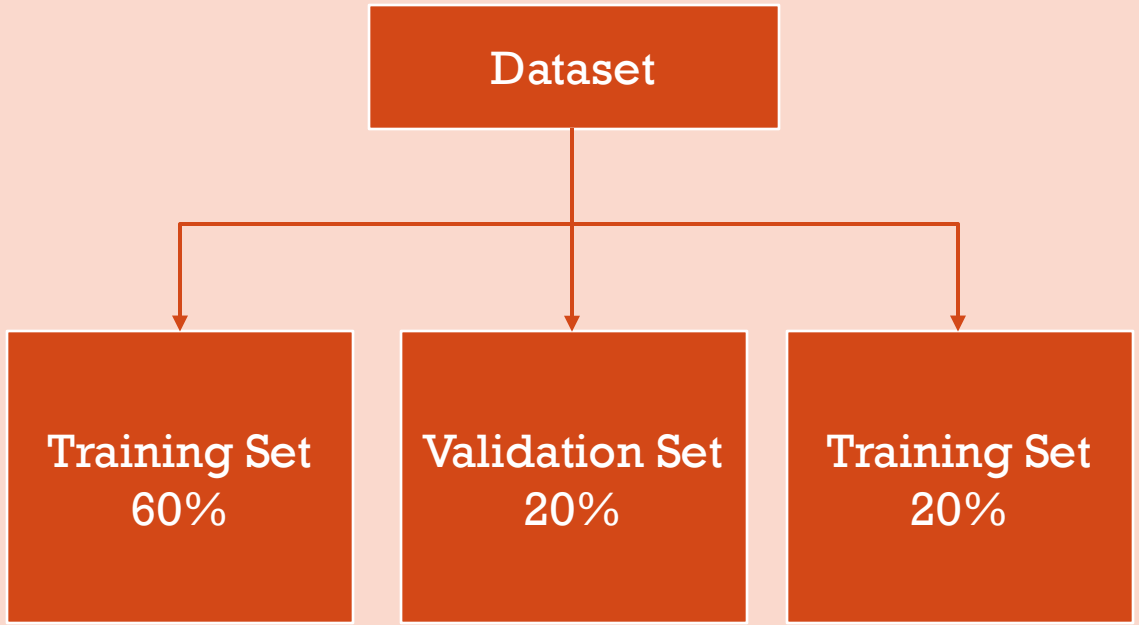
- Candidates with 2 years of college experience have the highest callback rate
- Surprisingly, those with 4 years show lower callback rates.



VISUALIZATION 3 (PIE CHART)

- A surprising 55.9% of callbacks are for applicants without listed computer skills
- While 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	
job_fed_contractor	1768	NA
job_ownership	1992	Unknown
job_req_min_experience	2746	NaN

Renamed Values	
1768	unknown
1992	unknown
2746	unknown

FEATURE ENCODING

Ordinal Encoding

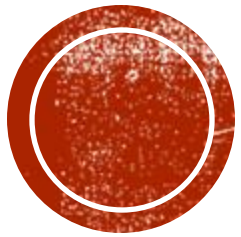
- Applied to features with a natural order
- Example: years_college:
  - 0
  - 1
  - 2
  - 3
  - 4

One-Hot Encoding

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

FEATURE ENCODING

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



Q&A