Job Posting Insights: Trends and Opportunities

INDEED TEAM 1

Jayesh Bhadane

Ridhi Bandaru

Harsha Karri

Jimmy Bui

Agenda

- Objective
- Overview of model performances
- Our winning model
- Clustering
- Data Summary Recap: what we started with
- Data Summary Recap: insights from the data
- Business Impact

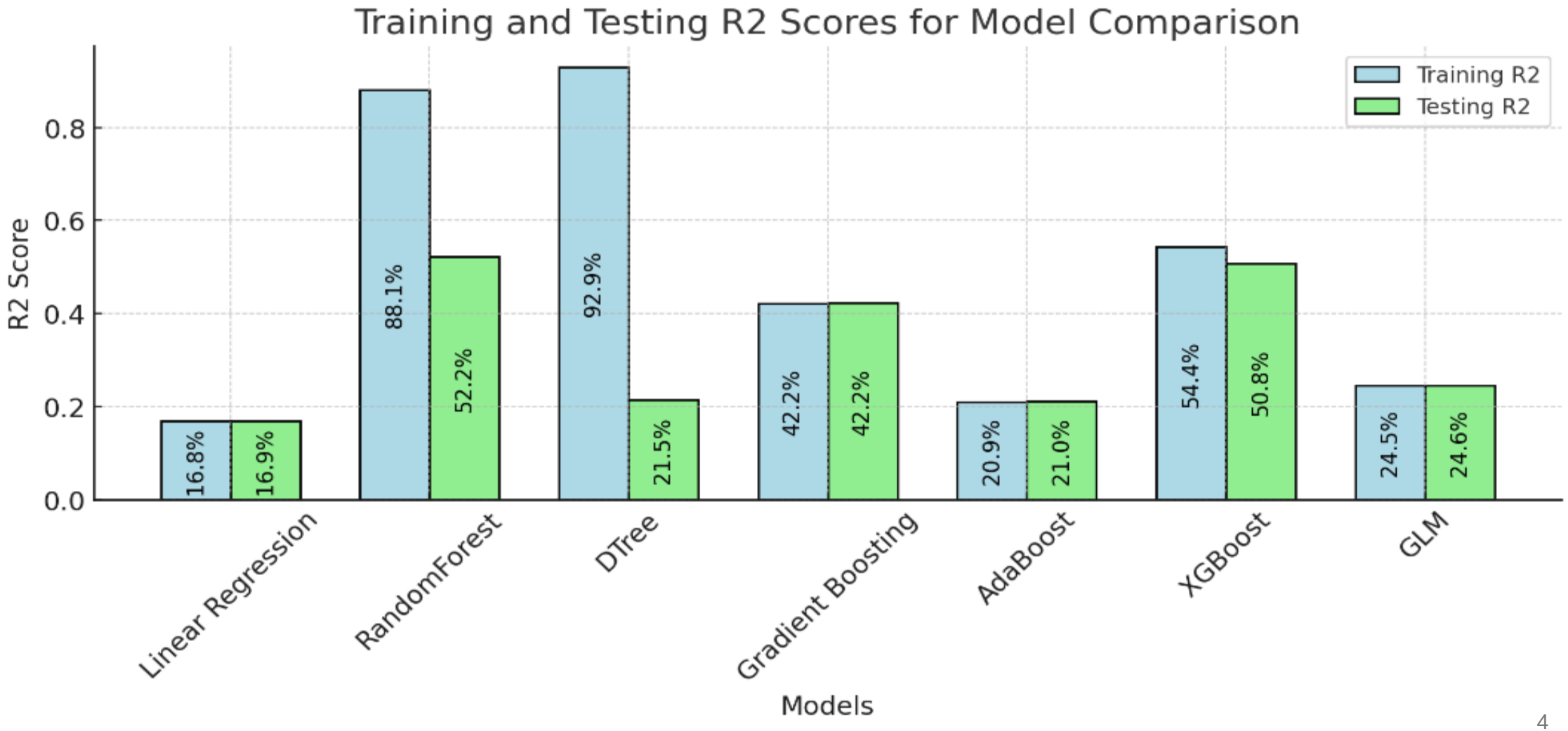
Project objectives

Purpose: Enhance investment decisions for employers by predicting the performance of sponsored job postings on Indeed.

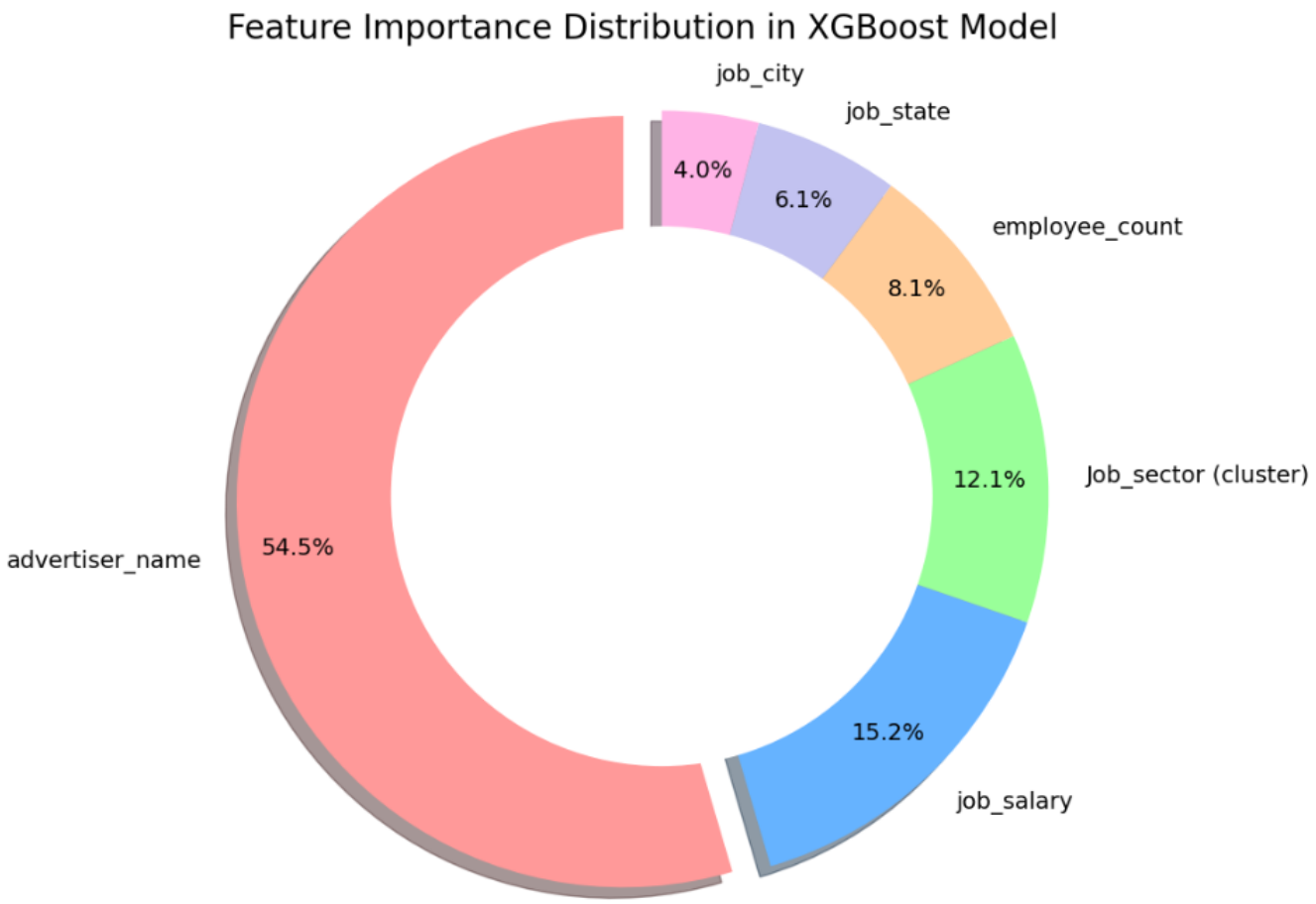
Goal: Provide predictive insights into the number of apply starts a job posting will receive, helping advertisers gauge the potential return on investment before spending.

- **Why This Matters:**
  - **Transparency**: Empowers employers with data-driven expectations, promoting confidence in the value of sponsored posts.
  - **Optimization**: Enables precise targeting and budget allocation, maximizing the effectiveness of recruitment efforts.
- **Expected Outcomes:**
  - **Decision Support:** Equip advertisers with a predictive tool to estimate apply starts for their job postings.
  - **Strategic Posting:** Guide customers in crafting job advertisements that are more likely to attract applicants, ensuring better use of their recruitment budget.

Comparative Model Performance Analysis

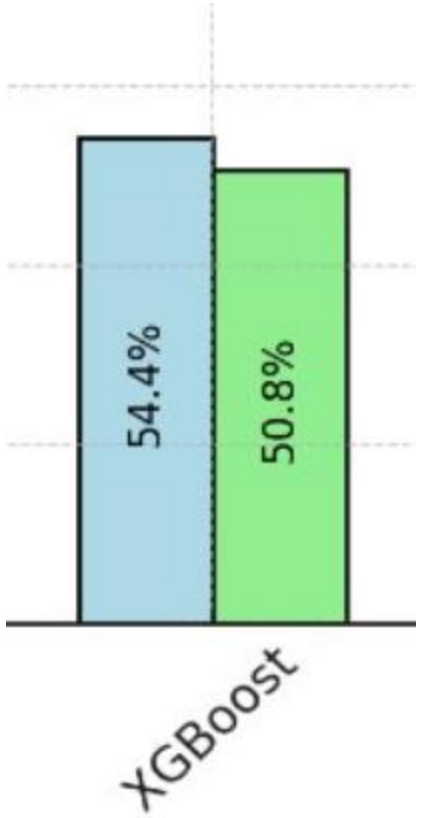


Understanding Feature Importances



- **Advertiser Name (54%)** - Dominates as the top influence, indicating the strong impact of the employer's brand and reputation on attracting applicants.
- **Job Salary (15%) and Job Sector (12%)** - Compensation and industry type are also critical, with competitive salaries and popular sectors drawing more interest.
- **Employee Count (8%), Job State (6%), Job City (4%)** - Company size and location contribute to applicant preferences, though to a lesser extent, highlighting the importance of perceived stability and desirable locales.

XGBoost – Best Model



- **Consistent Performance:** XGBoost demonstrated high R-squared values for both training (54.4%) and testing (50.8%) datasets, indicating strong predictive accuracy.
- **Balanced Model:** The minimal difference between the training and testing R-squared values maintained balanced.
- **No Overfitting:** The comparable R-squared values for training and testing datasets indicate that XGBoost did not overfit in the training dataset.

Understanding XGBoost - Our Top Performing Model

What is XGBoost? :

- **How It Works:** Think of teaching someone to recognize fruits. Each time they mistake an apple for an orange, they learn a little more about the difference. XGBoost improves in a similar way, but with data instead of fruits.

Advantages of XGBoost:

- **Effective and Reliable:** XGBoost performed the best among all the methods we tried, making it the most reliable choice for our needs.
- **Handles Different Situations Well:** It works effectively across a wide range of scenarios, adapting to various types of data.

Performance Explained:

- **How Well Does It Learn?** XGBoost learned to predict outcomes with about **54%** accuracy from the data it was trained on.
- **How Well Does It Predict New Situations?** When faced with new data it hadn't seen before, it predicted outcomes with nearly the same accuracy, about **50%**.
- Our choice to use XGBoost was validated by its strong performance, making it a dependable tool for predicting outcomes based on historical data.

Understanding R^2 in Regression Models

What is R^2?:

- R^2, or R-squared, measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1, where 0 means the model explains none of the variability and 1 means it explains all the variability.
- In our case, if our model predicts the number of applicants for job postings and has an R^2 of 45%, this means 45% of the variation in applicant numbers is explained by our model, based on the data it has learned from.

Why Training and Testing R^2?:

- Training R^2 tells us how well the model learned from the data we already knew about. It shows if the model is picking up the patterns it should from this data.
- Testing R^2 tells us how well the model can use what it learned on new, unseen job postings. It's important because it shows whether the model can be trusted to work in real-life situations, not just in tests.
- Both training and testing R^2 are crucial for assessing our model's effectiveness. They ensure that our model is not only good at learning but also excellent at predicting new data accurately, making it a reliable tool for forecasting job application starts.

Data Summary

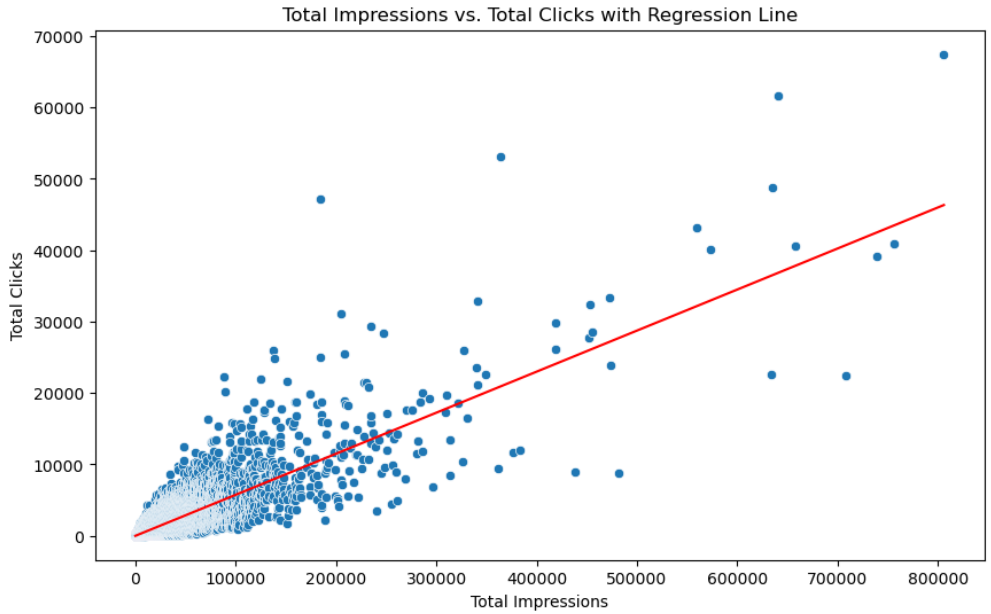
- 3,484,869 Job Listing with 9 variables
- Data type
- Missing percentages

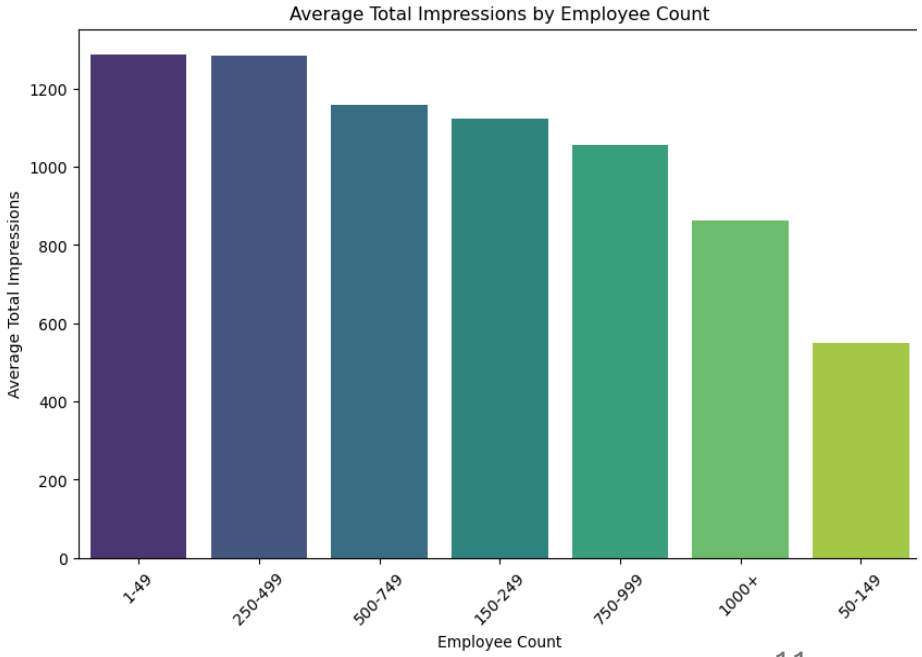
Clustering

- **Data Preparation**
  - **Source Variable:** 'job_title' containing detailed job descriptions and roles.
  - **Cleaning Steps:** Converted text to lowercase, removed punctuation, and applied basic text preprocessing to standardize job titles.
- **Embedding and Clustering**
  - **Embedding Technique:** Utilized Hugging Face's sentence transformers to convert cleaned job titles into numerical data (embeddings).
  - **Clustering Approach:** Employed clustering algorithms to group similar job titles based on their embeddings.
- **Outcome: Job Sector Categorization:**
  - **Sectors Identified:** Successfully categorized all job titles into 6 distinct job sectors, reflecting the diverse range of industries and roles on Indeed.
  - **Purpose:** This categorization aids in targeted analysis and prediction of apply starts for sponsored job posts.

Key Data Insights

- **Interdependence of Impressions, Clicks, and Apply Starts**
  - Clicks-to-Impressions Ratio
- **Higher Impressions in Smaller Companies**
  - Employee count categorized ordinally based on company size
- **Key Industry Sectors**
  - Grouped industries into 6 main clusters after initial analysis and refinement





Other Data Insights

- **Consistent Salaries**: Region not a factor for job posting
  - Average salaries are consistent across regions, indicating that the job's location does not significantly impact salary offers.
- **Smaller Companies Offer Higher Salaries**: Attracting more job seekers
  - Smaller companies tend to offer higher salaries, which draws more job seekers to these opportunities.
- **Location Impact**: Key Locations, Industry Density, and Living Costs Influence Applications
  - Job applications are influenced by key locations, industry density, and the cost of living in different areas.
- **Remote Jobs**: Lead in Apply Starts
  - Remote jobs have the highest number of apply starts, making them the most sought-after positions.

Business impact

- **Strategic Advertising:** Enables precise targeting of job ads, improving the likelihood of attracting qualified applicants and reducing expenditure on less effective postings. This targeted approach boosts ROI by channeling resources into high-performing campaigns.
- **Data-Driven Decisions:** Provides stakeholders with actionable insights to optimize advertising spend, focusing on platforms and positions that yield the highest return on investment. Leveraging predictive analytics ensures that each dollar spent is an investment towards optimal outcomes.
- **Resource Allocation:** Supports smarter allocation of recruitment resources, enhancing overall business operations by improving hiring efficacy. Efficient resource management aligns recruitment strategies with organizational goals, maximizing workforce productivity.
- **Market Insight:** Empowers stakeholders with deeper market understanding, allowing for quicker adaptation to industry trends and applicant behaviors, ensuring job postings resonate better with potential candidates. This strategic foresight positions the organization advantageously in a competitive labor market.