

Career Compass: A Platform Guiding Graduates in Science and Engineering

Junzhe Liao, Fangzhe Gu, Shiyu Ji, Xi Wang

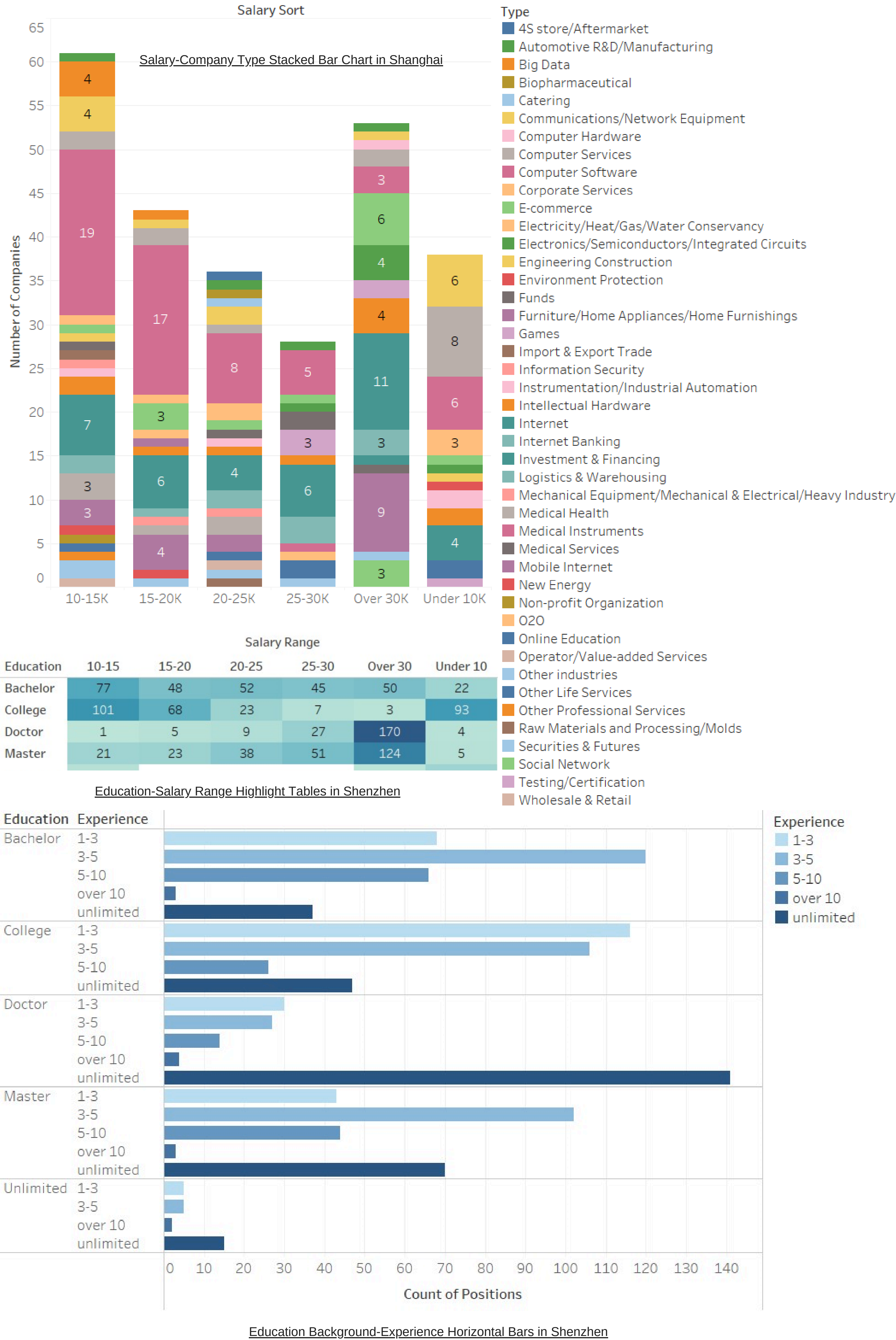
The ranking is not prioritized; All members have equal contributions

METHODOLOGY

- Data Scraping:** Scraped employment-related data from the Boss Zhipin website.
- Data Cleaning:** Thoroughly cleaned the data using OpenRefine to ensure accuracy and consistency.
- Data Visualization:**
 - Visualized the data using Tableau to create insightful representations of employment trends and patterns.
- Frontend Development:**
 - Started a React application using Vite.
 - Enhanced the user interface with Material UI for intuitiveness.
 - Implemented dynamic and interactive chart visualizations with Recharts and Echarts.
- Backend Development:**
 - Handled data requests using Node.js and the Express.js framework.
 - Integrated Flask to facilitate the connection between the predictive model and the website frontend, enabling dynamic data interaction.
- Server and Cloud Connection:**
 - Connected the entire application to a cloud server via Nginx, ensuring stable and responsive performance.
- Predictive Modeling:**
 - Chose the XGBoost algorithm for predictive modeling due to its effectiveness with structured data.
 - Conducted extensive feature engineering before model training to optimize predictions for specific regions and industries.

INTRODUCTION

This project investigates the employment trends for science and engineering graduates in Beijing, Shanghai, Guangzhou, and Shenzhen, highlighting a critical shortage of skilled professionals in tech-related sectors. Our research presents diverse career trajectories: some graduates pursue further academic studies, others secure government roles, and some venture into entrepreneurship. We have developed a website that visually presents current employment data and features an interactive tool allowing visitors to input their personal details to predict salary outcomes. Our analysis, driven by comprehensive employment statistics, educational backgrounds, key skills, and industry trends, aims to provide a detailed overview of the professional landscape for these graduates. Despite potential data gaps, we utilize advanced predictive models to ensure accuracy and utility in our findings, effectively employing data analytics to navigate employment risks.



ANALYSIS

- Salary Ranges and Industry:**
 - Most companies offer salaries between 10-15K, typical for entry-level to mid-level roles.
 - Higher salary ranges are less common but are seen in e-commerce and semiconductor sectors.
 - This distribution shows the diversity of salaries in the tech industry.
- Job Market Orientation:**
 - The job market is heavily oriented towards entry-level positions with no experience required, creating many opportunities for new graduates and early-career individuals.
 - Mid-level roles are significant, indicating a demand for professionals with moderate experience.
 - Senior positions that require over 10 years of experience are limited, suggesting high entry barriers or fewer opportunities at this level.
- Education and Salary Relationship:**
 - Individuals with college degrees typically earn under 10K.
 - Those with bachelor's degrees are most commonly found in the 10-15K salary range.
 - Master's and doctoral degree holders are more likely to earn over 20K.
 - Doctoral degree holders are most prevalent in the highest salary bracket, over 30K.
 - Educational attainment strongly correlates with salary levels: higher education links to higher pay, particularly at the upper end of the salary spectrum.
 - Bachelor's and master's degree recipients generally earn mid-range salaries (15-25K).

RESULTS/FINDINGS

- Model Experimentation:**
 - Experimented with various models including linear regression, neural networks, Support Vector Regression (SVR), and XGBoost.
- Model Selection:**
 - Compared the r-square scores of these models. Found that the XGBoost model significantly outperformed others with an r-square of 0.56.
- Performance Issue:**
 - Despite its relatively high performance, the XGBoost model still exhibited noticeable cognitive errors in predictions.
- Future Plan:**
 - Suggested more data collection to improve accuracy and reliability.