

Exploratory Data Analysis of RECRUITMENT DATASET (November 2024)

Sudip Bhattarai

Abstract—This paper conducts an Exploratory Data Analysis (EDA) on a recruitment dataset of 3,000 applicants. The dataset includes demographic details, education level, years of experience, desired salary, and application status, providing insights into recruitment patterns and applicant profiles. The analysis focuses on key metrics, such as education level distributions, salary expectations by experience, and application success rates.

Impact Statement—The insights derived from this analysis offer valuable guidance for refining hiring strategies, allowing recruitment teams to make more informed decisions. Understanding trends in applicant profiles enables organizations to align recruitment efforts with industry standards, ensuring that hiring practices are effective and aligned with candidate expectations.

Index Terms—Exploratory Data Analysis(EDA),Recruitment,

I. INTRODUCTION

RECRUITMENT data analysis is essential for identifying trends among applicants and enhancing hiring strategies. This analysis provides a comprehensive understanding of the distribution of education levels, experience, and salary expectations among applicants. It also evaluates how these factors correlate with application statuses, offering insights into potential factors that impact recruitment outcomes.

II. DATASET INFORMATION

The dataset comprises 3,000 records with 18 columns, including applicant ID, demographic details (Gender, Date of Birth), education level, years of experience, desired salary, job title, and application status. Key attributes analyzed include:

- **Education Level:** The highest qualification of each applicant.
- **Years of Experience:** Total professional experience.
- **Desired Salary:** Expected salary specified by applicants.
- **Status:** Current status of each application, categorized as Applied, In Review, Interviewing, Offered, and Rejected.

III. EXPLORING THE DATASET

The recruitment dataset contains information on 3,000 applicants, each identified by a unique Applicant ID. The dataset includes details such as demographic information, education level, years of experience, desired salary, and application status. The primary attributes and their statistical summaries are as follows:

- **Years of Experience:** The average years of experience among applicants is approximately 9.96 years, with values ranging from 0 to 20 years.

- **Desired Salary:** The mean desired salary is \$65,079.10, with a minimum of \$30,047.20 and a maximum of \$99,992.70.
- **Application Dates:** Applications were submitted within a 3-month period, from May 6, 2023, to August 5, 2023.
- **Date of Birth:** Applicant ages vary, with dates of birth ranging from July 8, 1962, to December 2, 2005.
- **Gender:** Gender distribution is balanced, with no missing values.
- **Education Level:** Applicants possess a variety of qualifications, with the most common being a Bachelor's degree.
- **Application Status:** The most frequent status is "Applied," indicating initial stages of the recruitment process.

All columns are complete, with 0% missing data.

	Mean	Min	Median	Dispersion	Min	Max	Missing
Applicant ID	2500.50	1001	2500.50	0.95	1001	4000	0 (0.0%)
Zip Code	51095.09	5408	51018	0.56	541	95897	0 (0.0%)
Years of Experience	9.96	10	10	0.61	0	20	0 (0.0%)
Desired Salary	65079.1	30047.2	64834.9	0.30392	30047.2	99992.7	0 (0.0%)
Application Date	2023-06-21	2023-05-06	2023-06-22	2023-06-06	2023-05-06	2023-06-06	0 (0.0%)
Date of Birth	1983-08-04	1962-07-08	1983-08-05	-43 years	1962-07-08	2005-12-02	0 (0.0%)
Gender		Male		1.1			0 (0.0%)
State		NO		4.07			0 (0.0%)
Education Level		Bachelor's Degree		1.30			0 (0.0%)
Status		Applied		1.61			0 (0.0%)

Fig. 1. Feature Statistics Table of the Recruitment Dataset

A. Box Plot

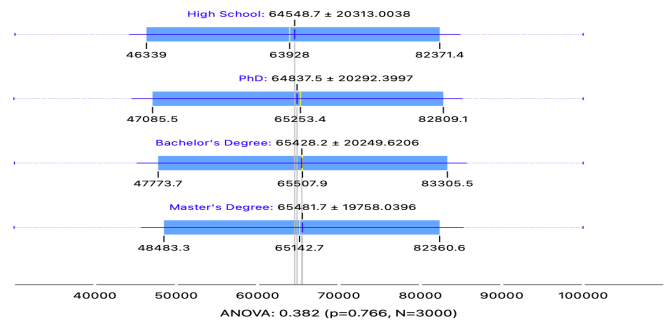


Fig. 2. Box Plot of Desired Salary and Education level

This box plot shows income spread across different education levels (High School, PhD, Bachelor's, Master's). Each box shows the median income and the range for each group. The ANOVA test result ($p = 0.766$) means there's no big difference in income between the education levels, so higher education doesn't really seem to link up with higher income in this sample.

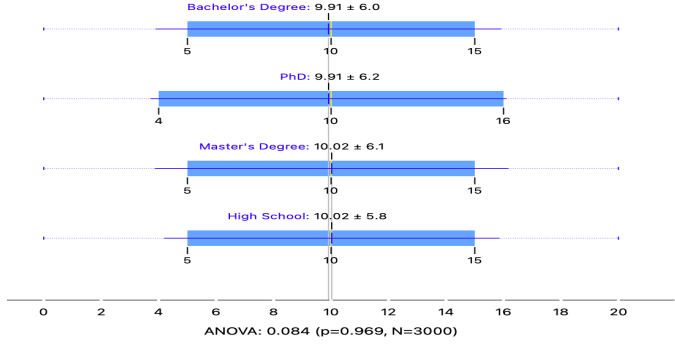


Fig. 3. Box Plot of Experience and Education level

This box plot shows scores across different education levels (Bachelor's, PhD, Master's, High School) related to experience level. Each box shows the spread and middle score for each group, giving an idea of how scores vary. The ANOVA result ($p = 0.969$) means there's no big difference in scores based on education level, so experience doesn't seem to link up with higher scores much in this sample.

B. Distributions

This chart shows how different education levels are linked to salary expectations. Each bar on the x-axis represents a salary range, and the colors in each bar show the education levels: Bachelor's Degree, High School, Master's Degree, and PhD. As the salary range goes up, the bars get taller, meaning more people expect higher salaries. People with a Bachelor's Degree are found in all salary ranges but are mostly in the lower ranges, while those with a Master's or PhD show up more in the higher salary ranges. High School graduates are mainly in the lower salary ranges and are less common as salaries increase. Overall, this chart suggests that higher education levels are connected to higher salary expectations.

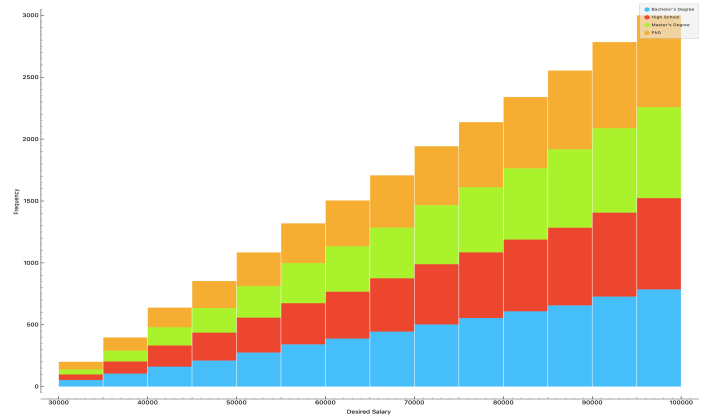


Fig. 4. Desired Salary vs Frequency Graph Split by Education Level

This chart shows how years of experience relate to different education levels. Each bar on the x-axis represents a number of years of experience, and the colors in each bar show the education levels: Bachelor's Degree, High School, Master's Degree, and PhD. As experience increases, the bars get taller, meaning more people have higher years of experience. People with a Bachelor's Degree are spread across all experience levels, especially in the early years. As years go up, more people with Master's and PhD degrees appear, showing that higher degrees are common at higher experience levels. High School graduates are mostly in the lower experience levels and become less common as experience increases. This suggests that people with higher education tend to have more years of experience, while high school graduates are mainly in the early stages of their careers.

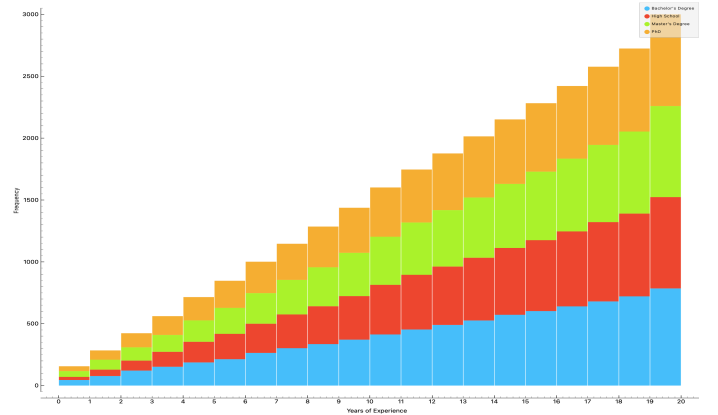


Fig. 5. Year of Experience vs Frequency Graph Split by Education Level

C. Chi-Square Test

The Chi-Square Statistic of 10.54 and the p-value of 0.57 mean that there is no strong evidence of a relationship between Education Level and Application Status in this dataset.

In simple terms, the high p-value (greater than 0.05) tells us that any differences we see in application status across different education levels could just be due to random chance, not because education level actually affects application status.

So, based on this test, we don't find a meaningful link between a person's education level and the outcome of their job application. .

IV. CONCLUSION

This paper looked at a recruitment dataset with 3,000 applicants to find patterns in profiles and outcomes. We analyzed different factors like education level, years of experience, desired salary, and application status. Using visual charts like box plots and distribution graphs, we tried to see if education, experience, and salary expectations have any connections.

The main findings show that applicants with higher degrees, like Master's and PhD, usually have more years of experience and ask for higher salaries. But, the Chi-Square test showed there's no strong link between education level and whether someone gets hired or not, meaning education alone may not really decide if a person gets the job.

In general, this analysis gives useful insights for recruitment teams to improve hiring strategies to better fit applicant expectations. By knowing how education, experience, and salary expectations are spread out, companies can make their recruitment process match up more with what candidates want. In the future, more research could look at other factors, like specific skills or job roles, to understand hiring outcomes better.