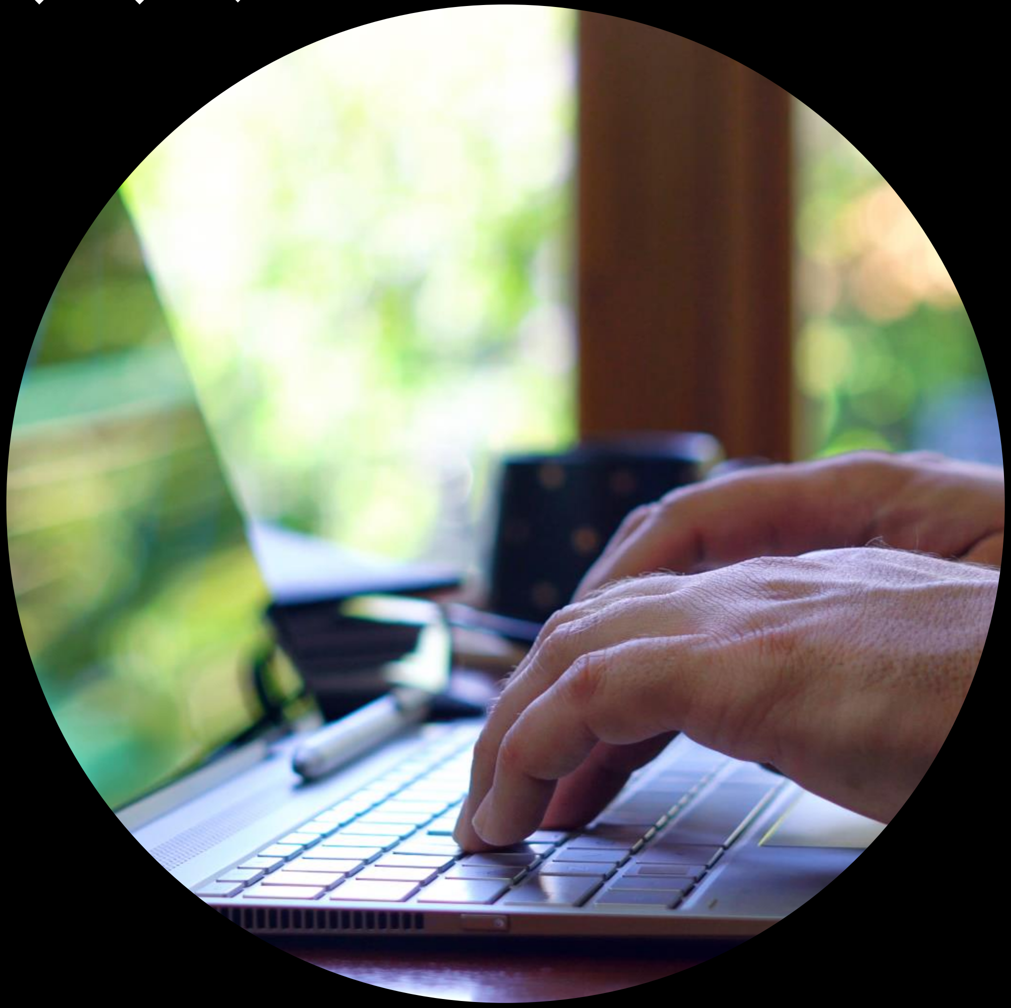
JOB RECOMMENDATION MatchWise: Intelligent Job Matching and Application Trend Prediction

Capstone Project by:
Pheminah Wambui
Isaac Munyaka
Caroline Gesaka
Otiende Ogada
Anne Njoroge
Joan Maina

Business Understanding



Introduction

A 2023 survey by the Federation of Kenya Employers (FKE), titled ***Industries Find it Difficult to Fill Positions***, revealed that employers are struggling to find qualified talent despite a large talent pool and limited job opportunities.

Jacqueline Mugo, FKE CEO, highlighted that rapid technological changes and evolving market dynamics have exacerbated the mismatch between workers' skills and job requirements, making it a critical policy concern.

The survey, involving 521 enterprises from various sectors, found that 20% had hard-to-fill vacancies, with some lowering their qualifications (9.6%) to fill roles. The manufacturing industry faced the most challenges.

Source: Abura, 2023

Background

The gap of job searching/workforce seeking remains weakly bridged in the employment sector.

Challenges include **j**ob seekers often experience challenges in getting jobs that match their skills and employers struggling to attract qualified candidates.

The opportunity is to leverage Advanced Data Analytics & Machine Learning to enhance job matching efficiency and effectiveness for both job seekers and employers.

Business Understanding

Recommending jobs by analyzing preferences, experience, and skills, using algorithms to match seekers with relevant positions.

Helping job seekers find jobs faster and allowing employers to quickly identify top candidates.

Key stakeholders:

I. Job seekers
II. Employers
III. Recruitment Agencies

Problem Statement

Job seekers often find it challenging to identify the most relevant opportunities. Conversely, employers struggle to understand what factors contribute to the attractiveness of their job postings, **leading to a mismatch between job offers and applications.**

Aim:

1. **Personalized Job Recommendations:** Match job seekers with the most suitable opportunities.
2. **Predictive Model for Employers:** Use machine learning to forecast application volume, enabling better job postings.

Objectives

- **Enhance Job Posting Quality:** Improve job descriptions to attract qualified candidates.
- **Predict Candidate Interest:** Develop a model to predict job seekers' likelihood to apply.
- **Optimize Job Recommendations:** Provide personalized job recommendations.
- **Increase Application Rates:** Boost application rates to ensure better applications.

Success Metrics

- **Application Rate:** Percentage of job postings receiving applications.
- **Qualified Application Rate:** Percentage of applications meeting job requirements.
- **Precision:** Measure the proportion of recommended jobs that are relevant to the job seeker's input skills.
- **Recall:** Measure the proportion of relevant jobs that are successfully recommended to the job seeker.
- **F1 Score:** Balance precision and recall to provide a single metric that evaluates the model's performance.
- **AUC (Area Under the Curve):** Measure the model's ability to distinguish between high and low application likelihoods.
- **User Satisfaction:** Evaluate the job matching system based on user feedback and satisfaction with the recommendations provided.

Data Understanding

Data Collection: *LinkedIn Job Postings (2023 - 2024) (kaggle.com)*

Types of Data

Job Data: Titles, descriptions, locations, experience level, salaries, and companies.

User Data: Aggregated from user resumes, skills, interactions, and historical applications.

Data Insights

Job Titles: Standardizing titles helps in matching similar job roles.

Location: Geographical data is standardized to ensure consistency in job location matching.

User Behavior: Historical application data provides insights into user preferences, interactions, and behavior patterns.

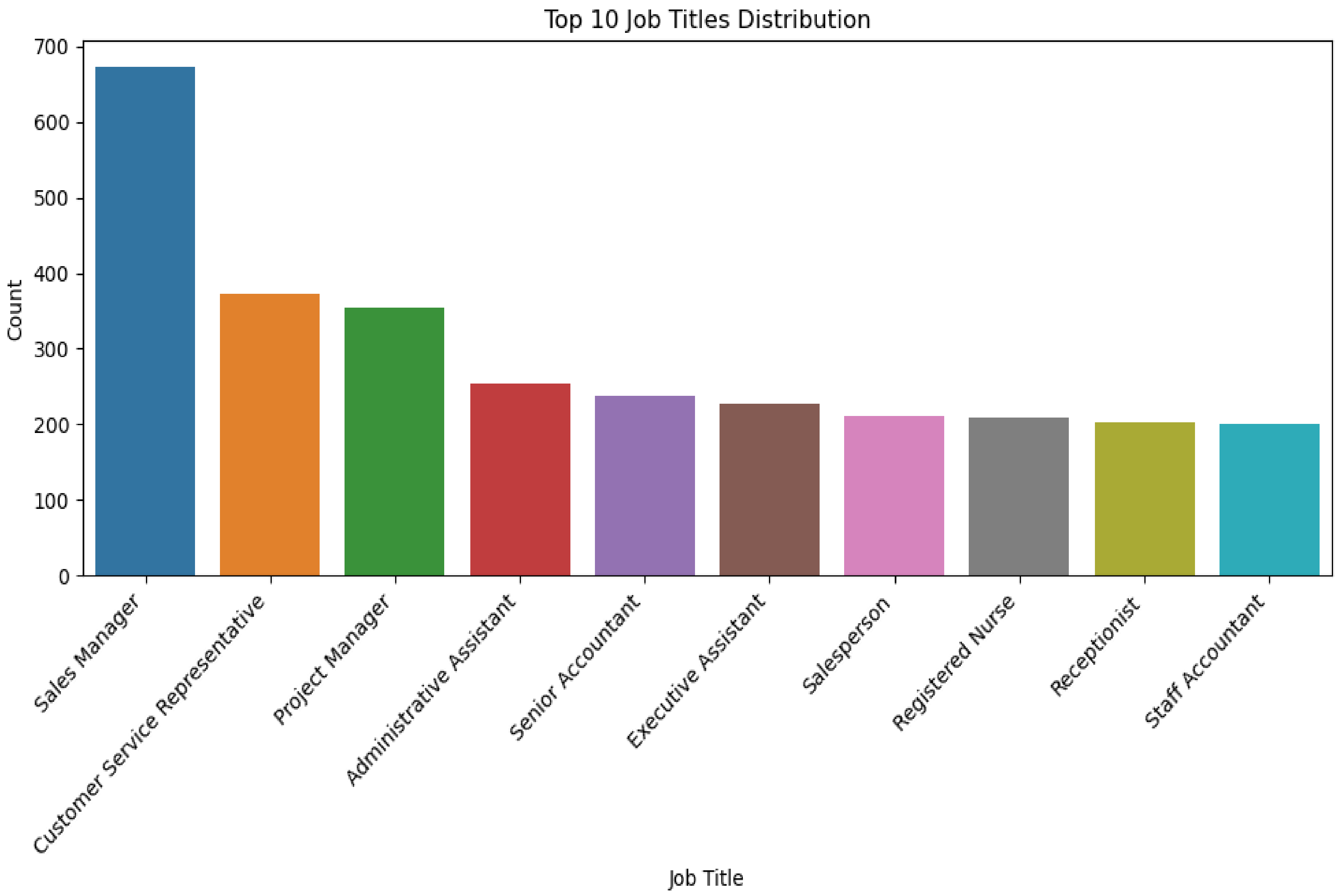
Content: The data contained 123,849 records.

Exploratory Data Analysis (EDA)

Job Title Distribution

Sales managerial roles are in high demand, with more job postings than other positions, reflecting strong business interest.

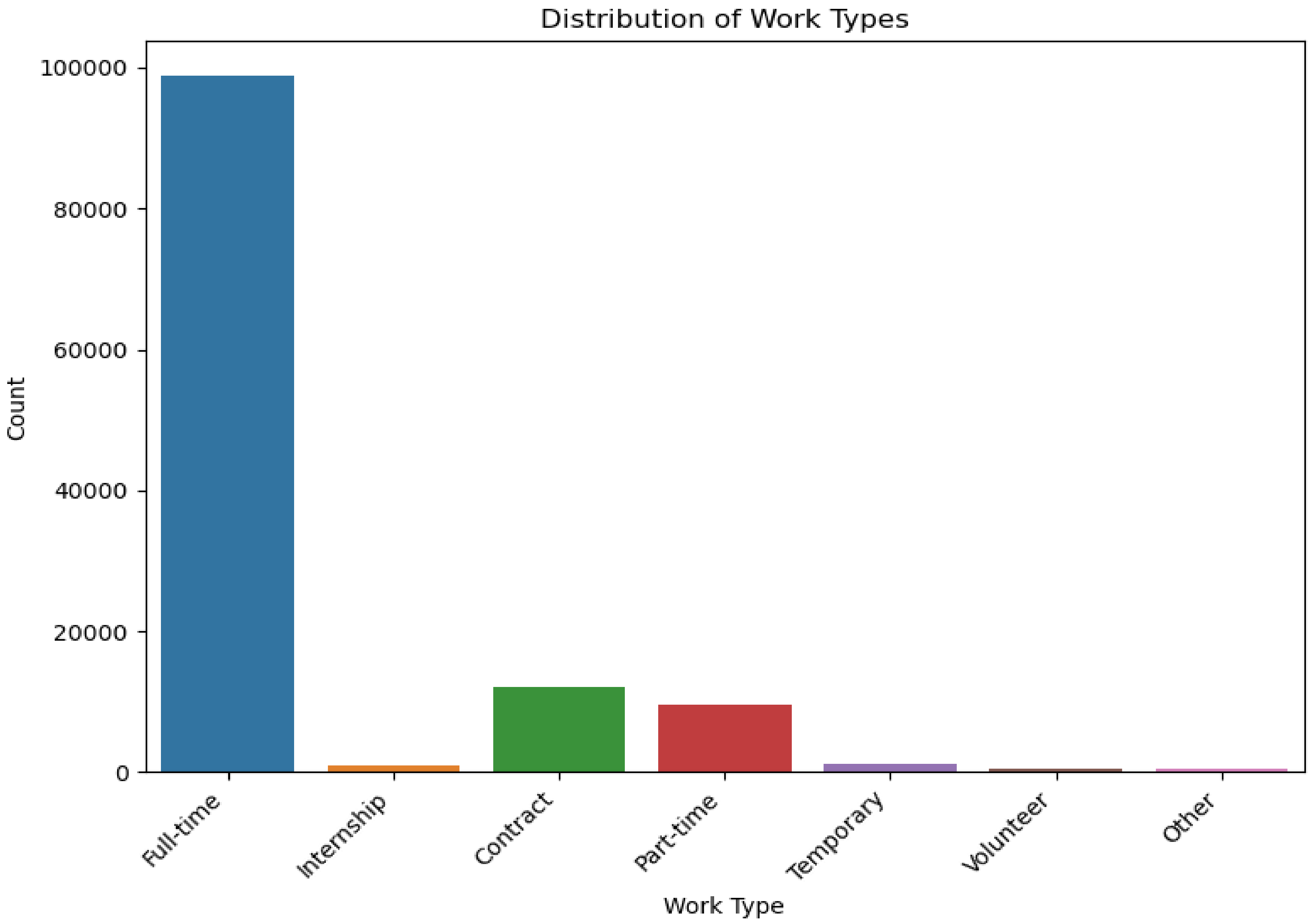
Customer service representatives and project managers are also sought after.



Work type Distribution

Nearly all posted jobs are Full-time, indicating a preference for permanent employees.

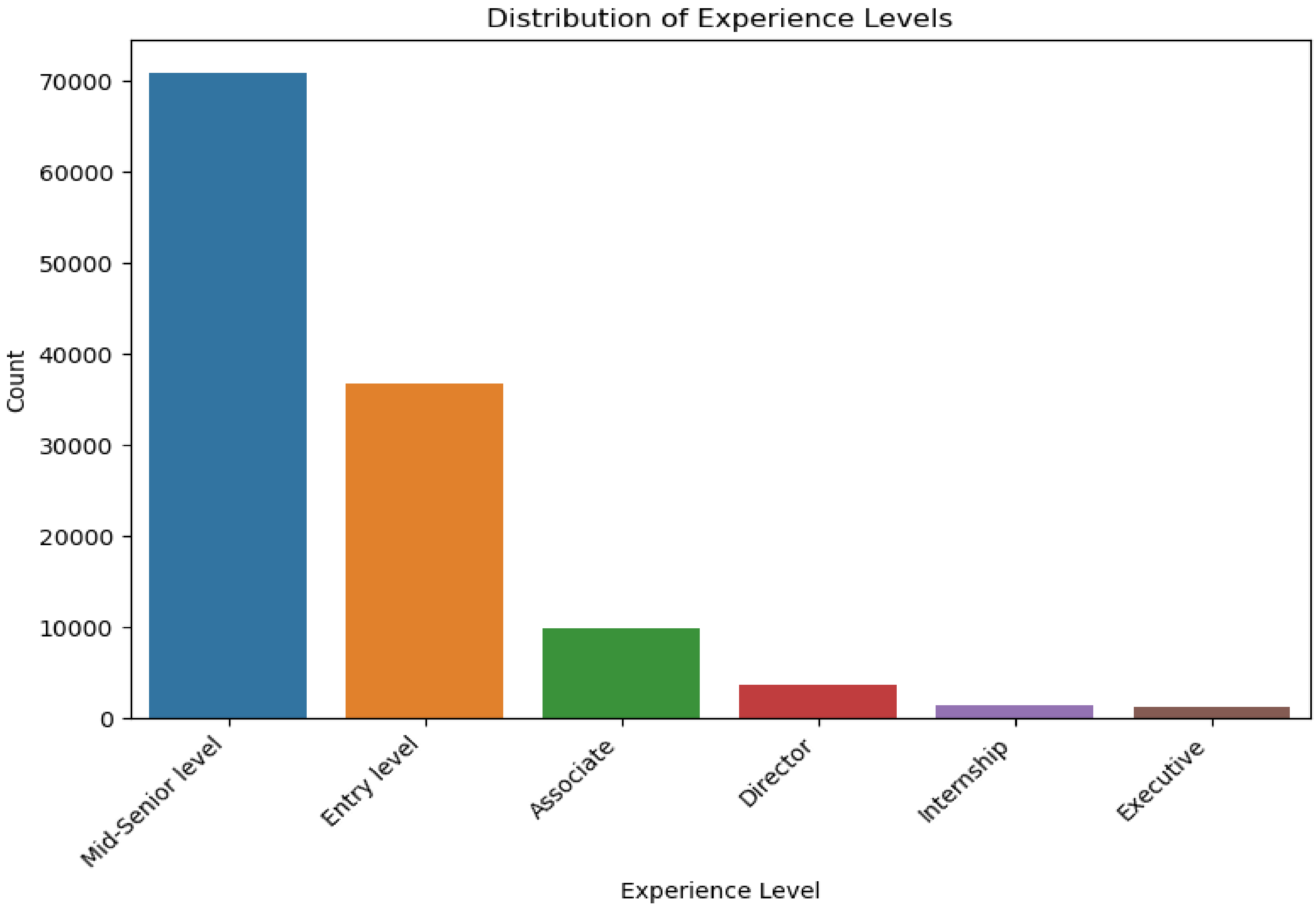
Few postings for temporary jobs or internships suggest lower demand.



Experience Levels

Most postings target mid-senior professionals, followed by entry-level roles.

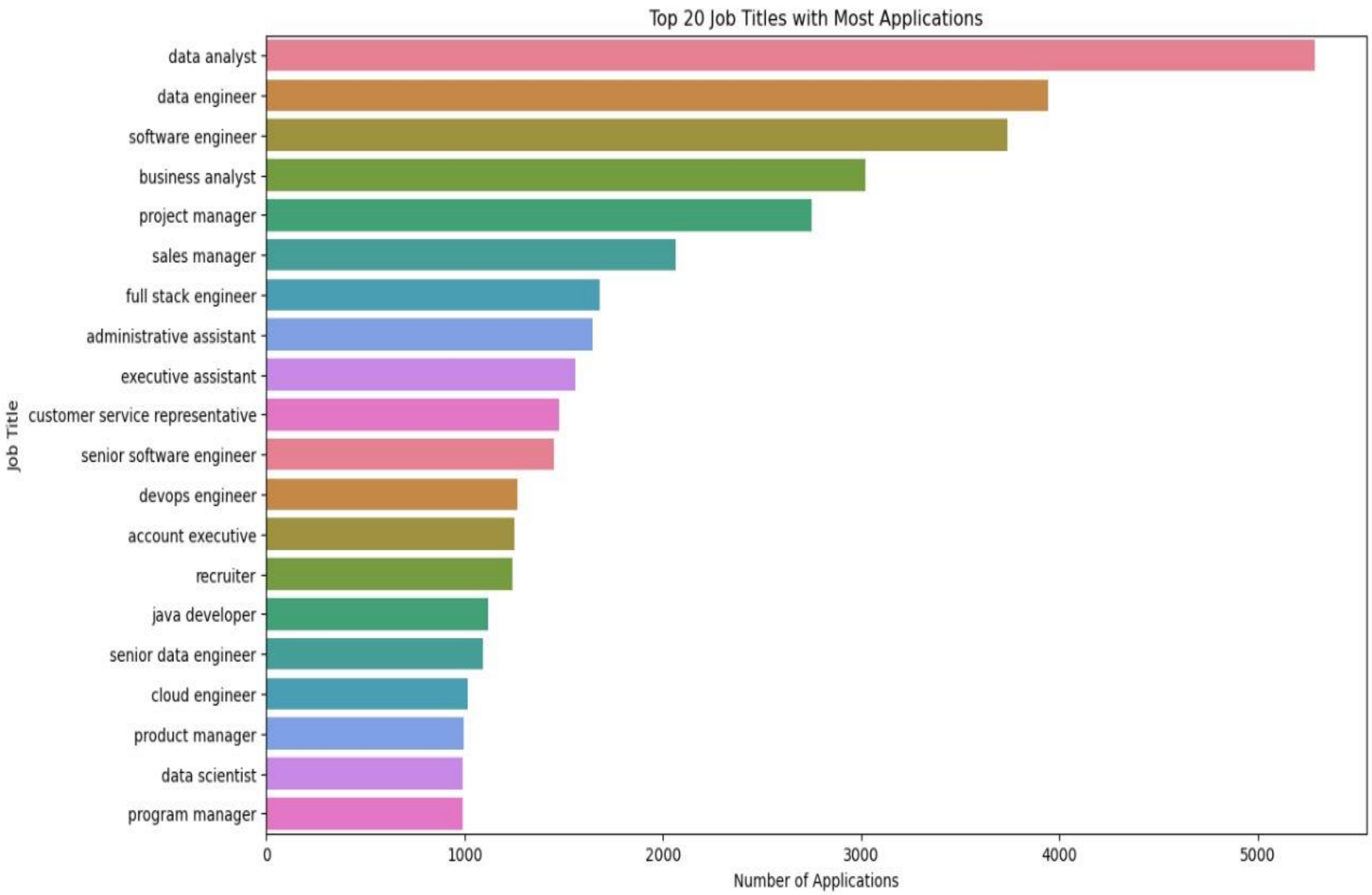
Fewer postings exist for internships and executive positions, highlighting a focus on experienced hires.



Top Applied Jobs

Job seekers' great interest in data analysis, engineering, and software development is evident from the large number of applications tech jobs receive.

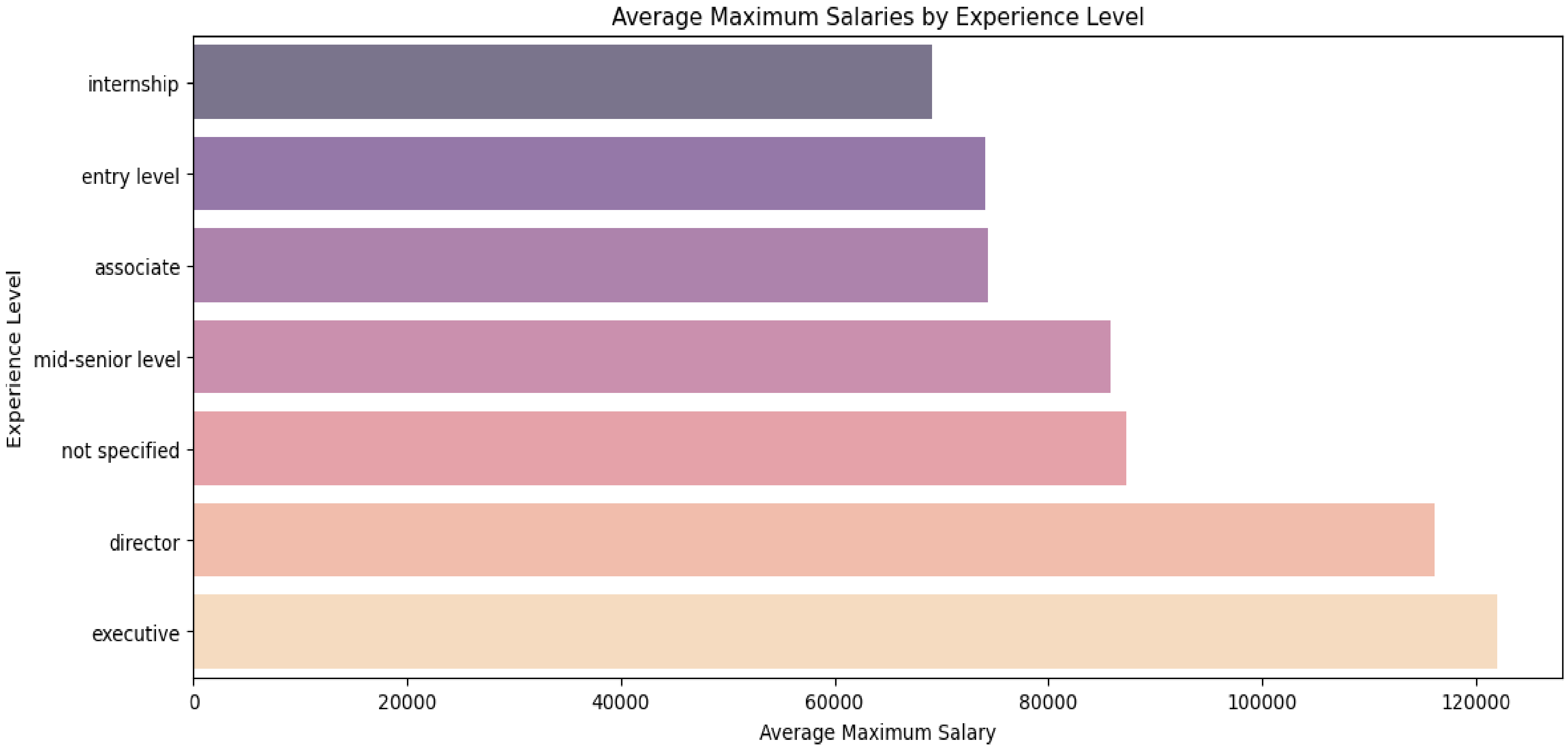
This indicates that digital transformation is driving a shift in the market towards IT roles.



Salary by Experience Level

There is a natural progression where higher experience levels are associated with higher salary amounts.

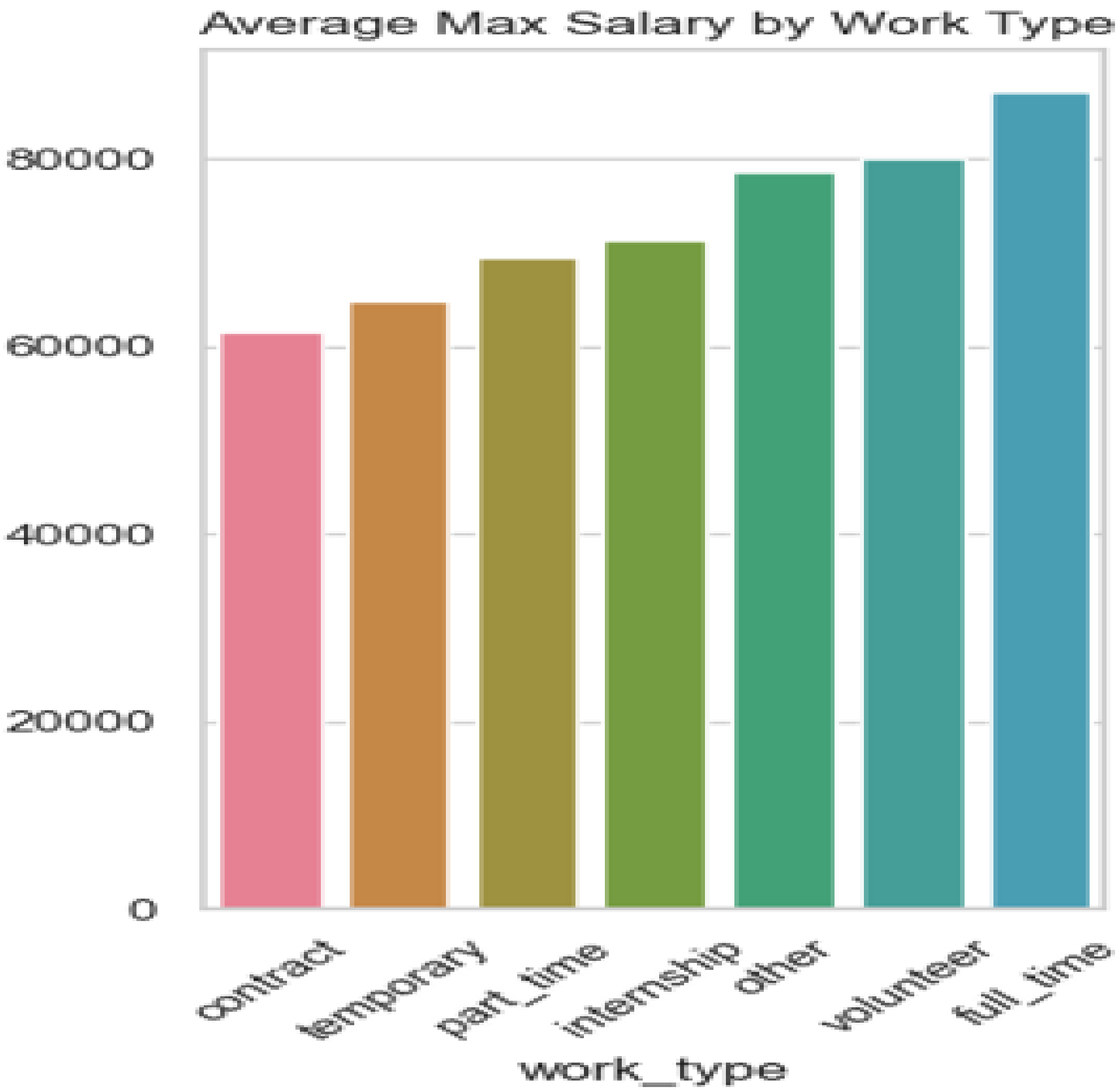
This suggests that experience is a significant factor in determining salary.



Salary by Work type

Jobs with higher salaries tend to offer full-time positions, while contract roles generally offer lower salaries.

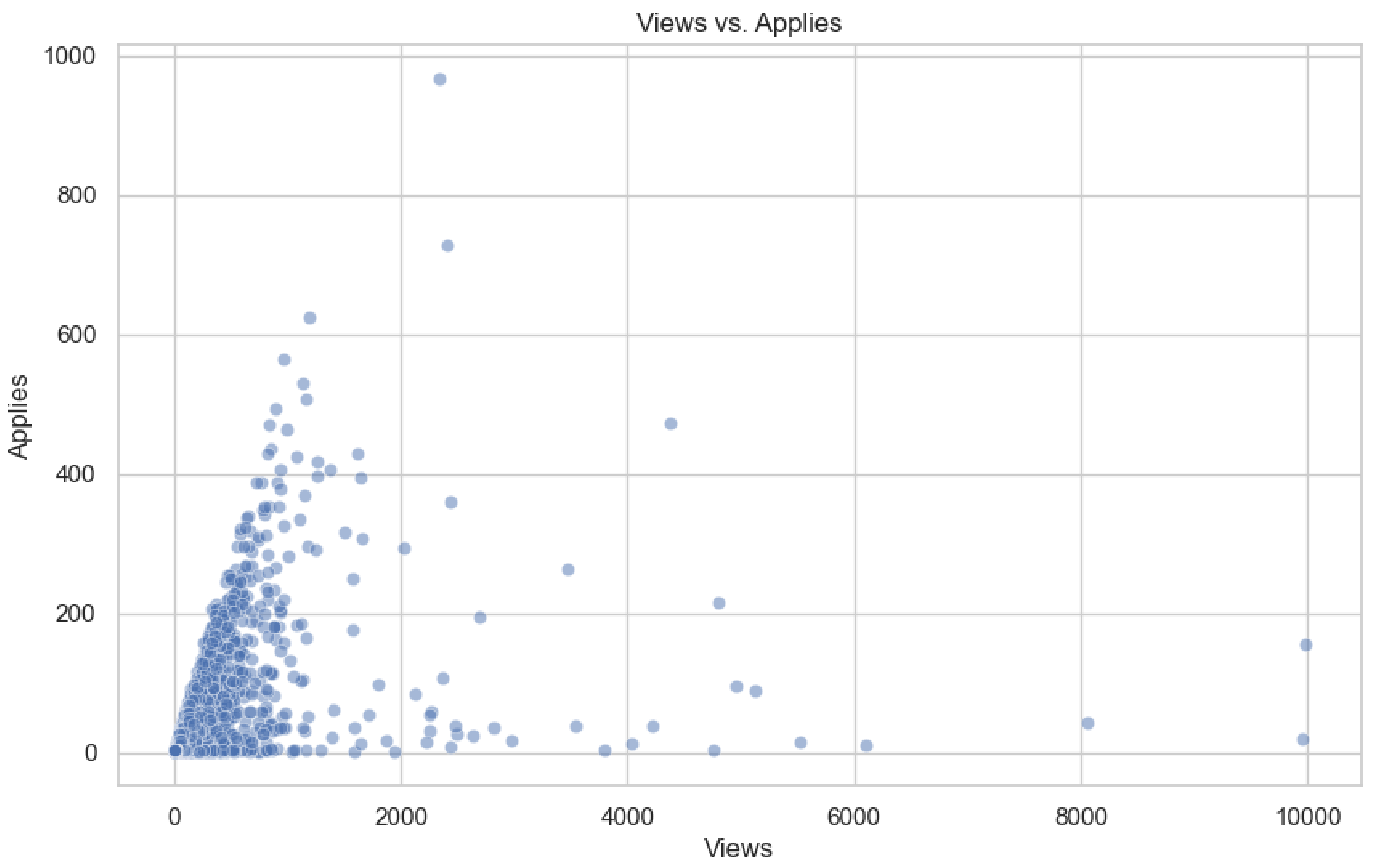
This suggests that full-time roles are valued more highly or are more financially rewarding.



Views vs Applies

Although there is some correlation between views and applies, increase in views does not necessarily suggest an increase in applies.

Factors such salary, location, job title and experience level may contribute to rate of views and applies.



Correlation insights

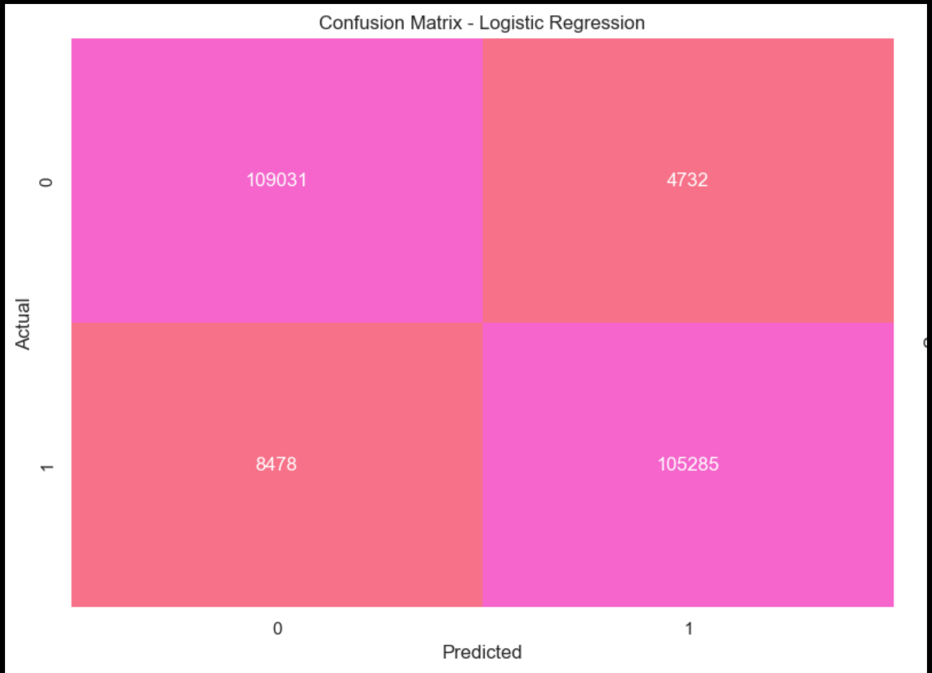
max_salary and min_salary: These are perfectly correlated (correlation of 1), which is expected since the maximum salary usually relates directly to the minimum salary within the same job category or position.

views: No significant correlation with other features, suggesting that the number of views is independent of salary, job type, or description length.

applies: Similarly, no significant correlation with other features, indicating that the number of applications is independent of salary, job type, or description length.

description_length: Correlated only with itself, suggesting that the length of job descriptions does not have a direct impact on other features like salary or views.

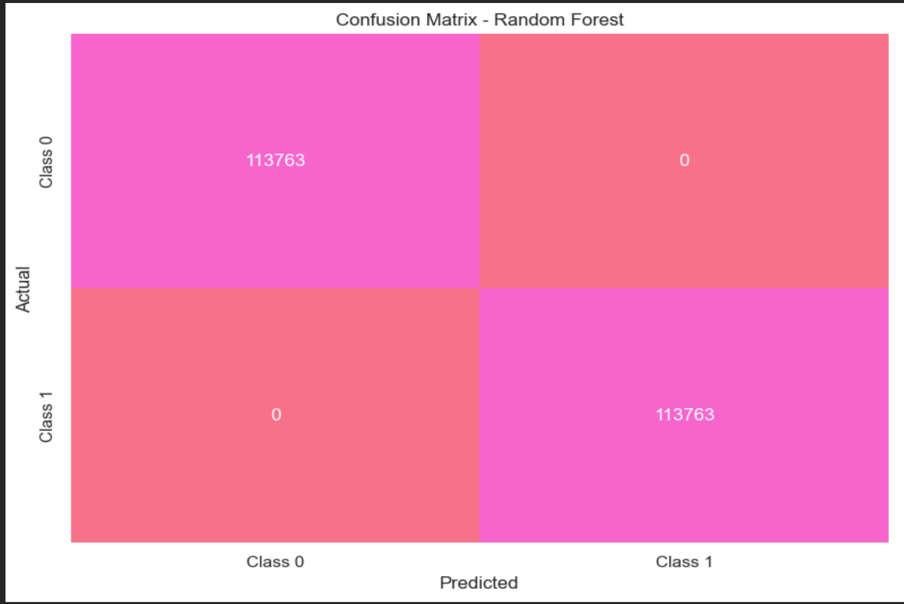
Modeling

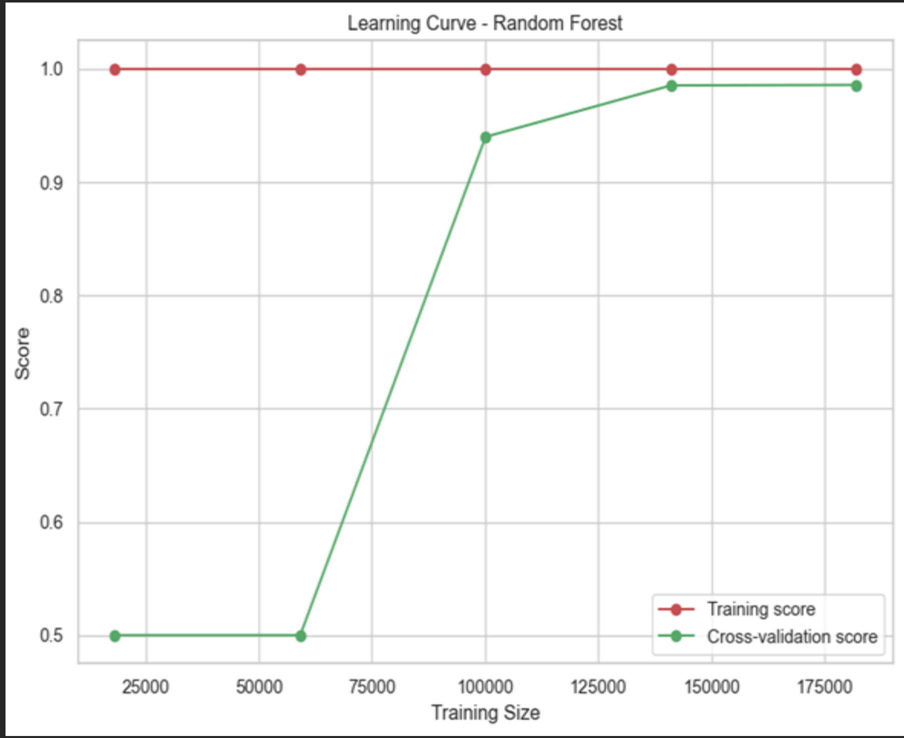


Modeling: Predictor Model – Baseline Modeling

- **F1-Score:** Both classes: 0.94 ROC-AUC Score: 0.9879
- **Precision**: Class 0: 0.93 Class 1: 0.96
- **Recall:** Class 0: 0.96 Class 1: 0.93

- The logistic regression model is highly accurate and reliable in predicting the target variable

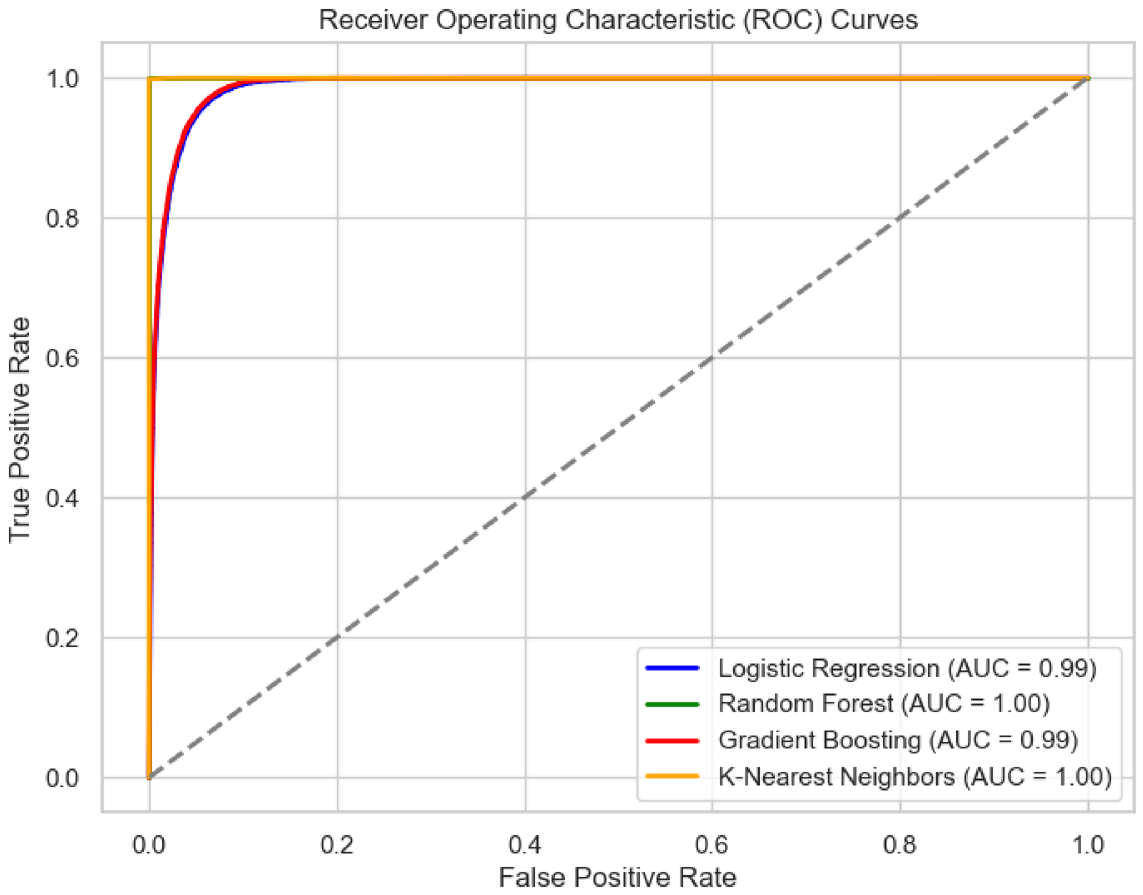


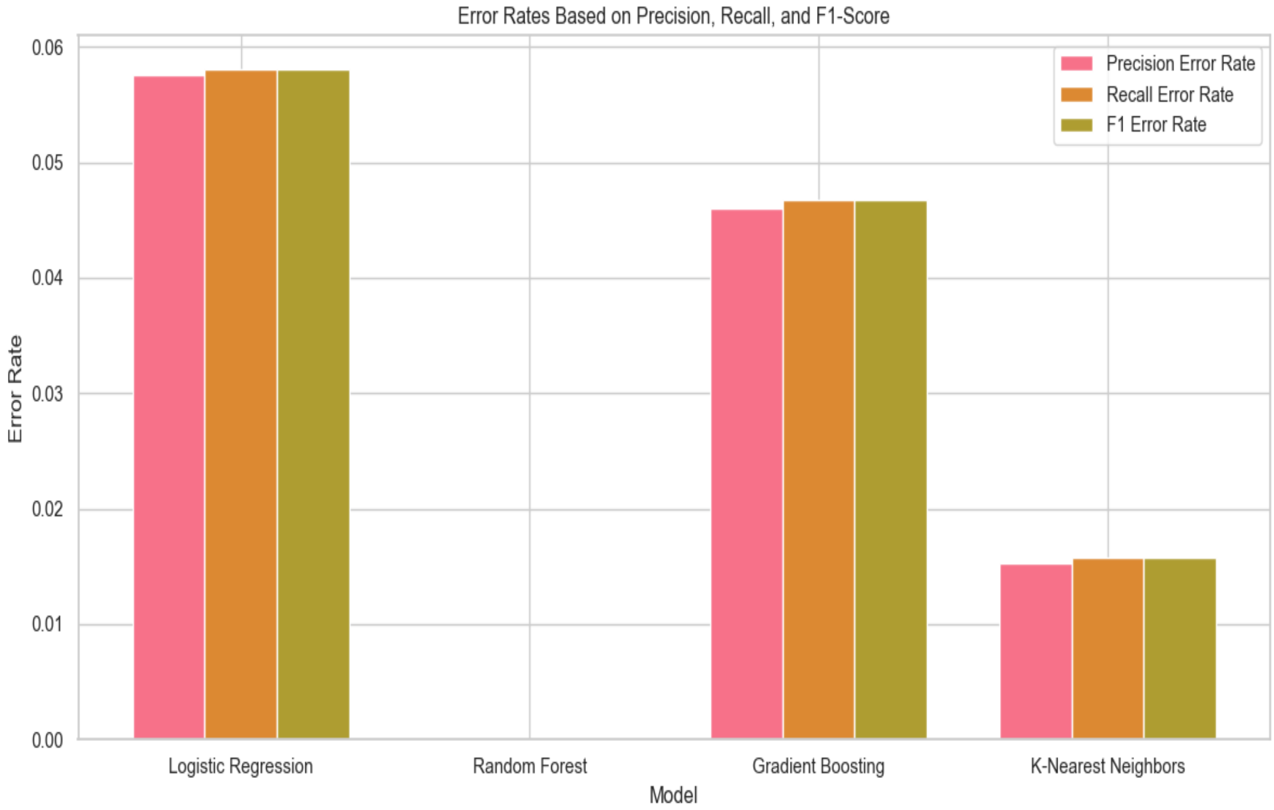


Modeling: RandomForestClassifier

- The **Random Forest model** shows exceptional performance with perfect accuracy, precision, recall, F1-scores, and an almost perfect ROC-AUC score.

ROC Curves & Error Performance for the Models





Random Forest achieves perfect accuracy, precision, recall, and F1-Score on both classes, and an ROC-AUC score of 1.0. It's the best-performing model in terms of classification metrics.

RandomForest model is error-free while K-NearestNeighbors is second with least errors.

Feature Importance

Views: The high importance score suggests that this feature has a strong predictive power in the model.

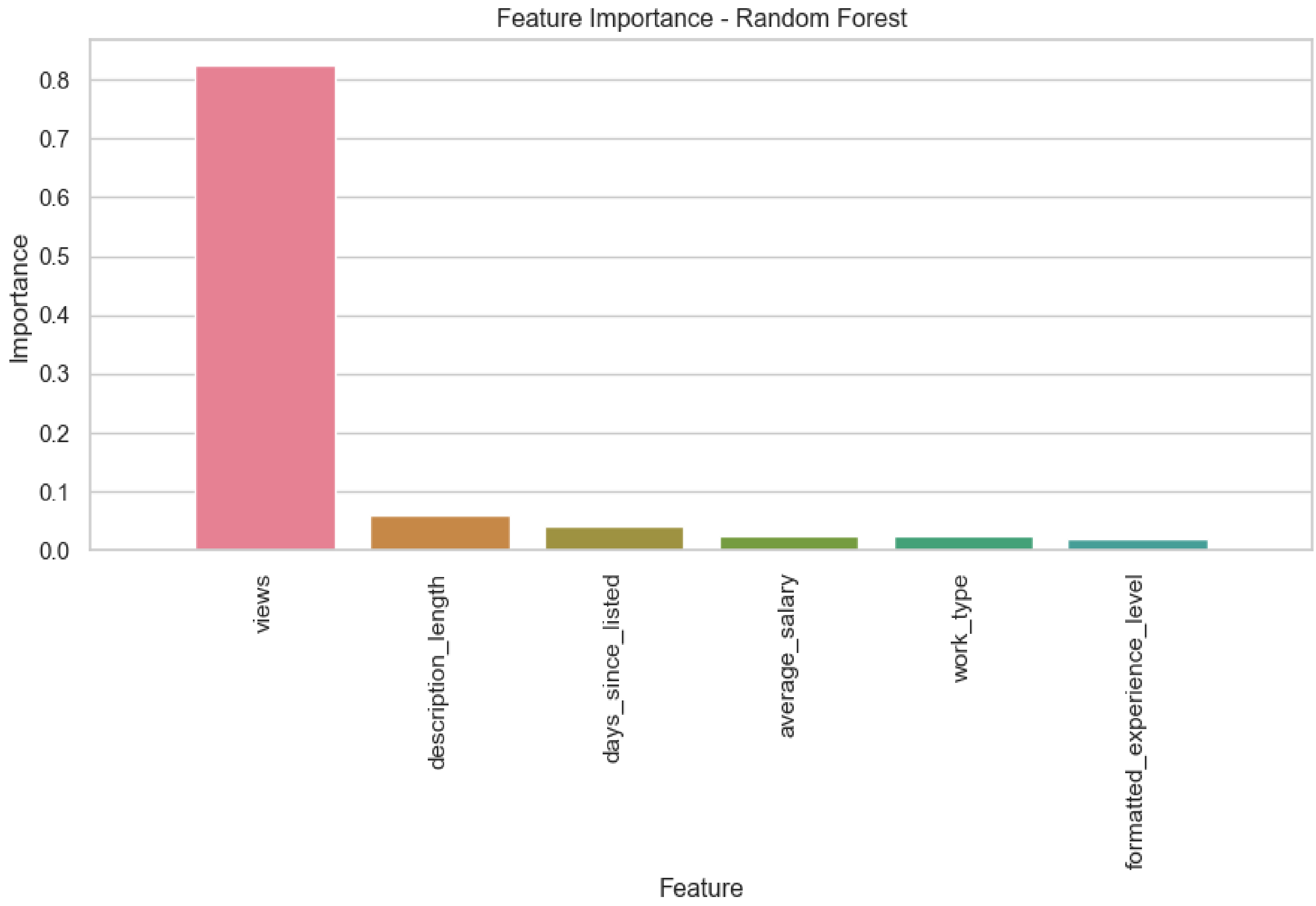
Description: The length or quality of the job description is the second most important feature.

Days_since_listed: This feature might help in understanding the freshness of the job posting and its attractiveness over time.

Average_salary: The average salary offered by the job posting has some impact on the prediction.

Formatted_experience_level: Suggests that while the experience level is a factor, it may not a strong predictor of the number of applications.

Work_type: The type of work (e.g., full-time, part-time, remote) also has minimal importance in the model is a factor but does not significantly influence the prediction.



Model Evaluation

Model Evaluation

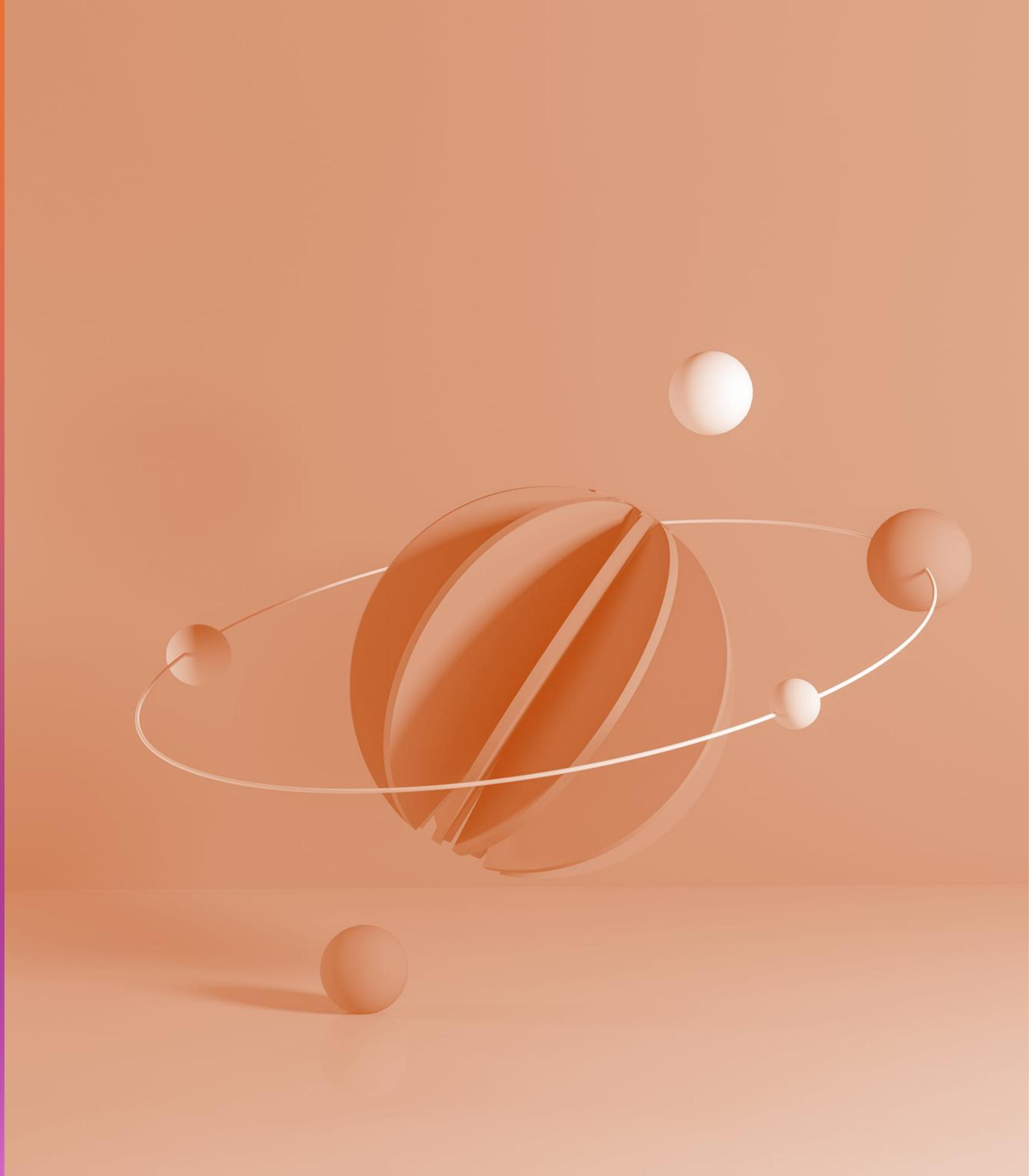
Model	Cross-Validation Accuracy	Test Accuracy	Precision (Class 0)	Precision (Class 1)	Recall (Class 0)	Recall (Class 1)	F1-Score (Both Classes)	ROC-AUC Score
Logistic Regression	0.941	0.942	0.93	0.96	0.96	0.93	0.94	0.988
Random Forest	0.986	1.0	1.00	1.00	1.00	1.00	1.00	1.0
Gradient Boosting	0.952	0.953	0.97	0.94	0.93	0.97	0.95	0.989
K-Nearest Neighbors (KNN)	0.973	0.984	1.00	0.97	0.97	1.00	0.98	0.99998

Conclusion: Random Forest performs best based on test accuracy and F1-score.

Recommender Models

1. **Content-Based:** Recommends jobs similar to the input job description.
2. **Collaborative Filtering (KNN):** Suggests jobs based on similarity of job attributes like views, applies, and salary.
3. **Keyword-Based:** Filters jobs based on specific keywords in the job title.
4. **KNN-Based Similarity:** Provides jobs very similar to a specific job (based on low average distance).
5. **Title-Based:** Recommends jobs with similar titles to the input job title.

Recommendations



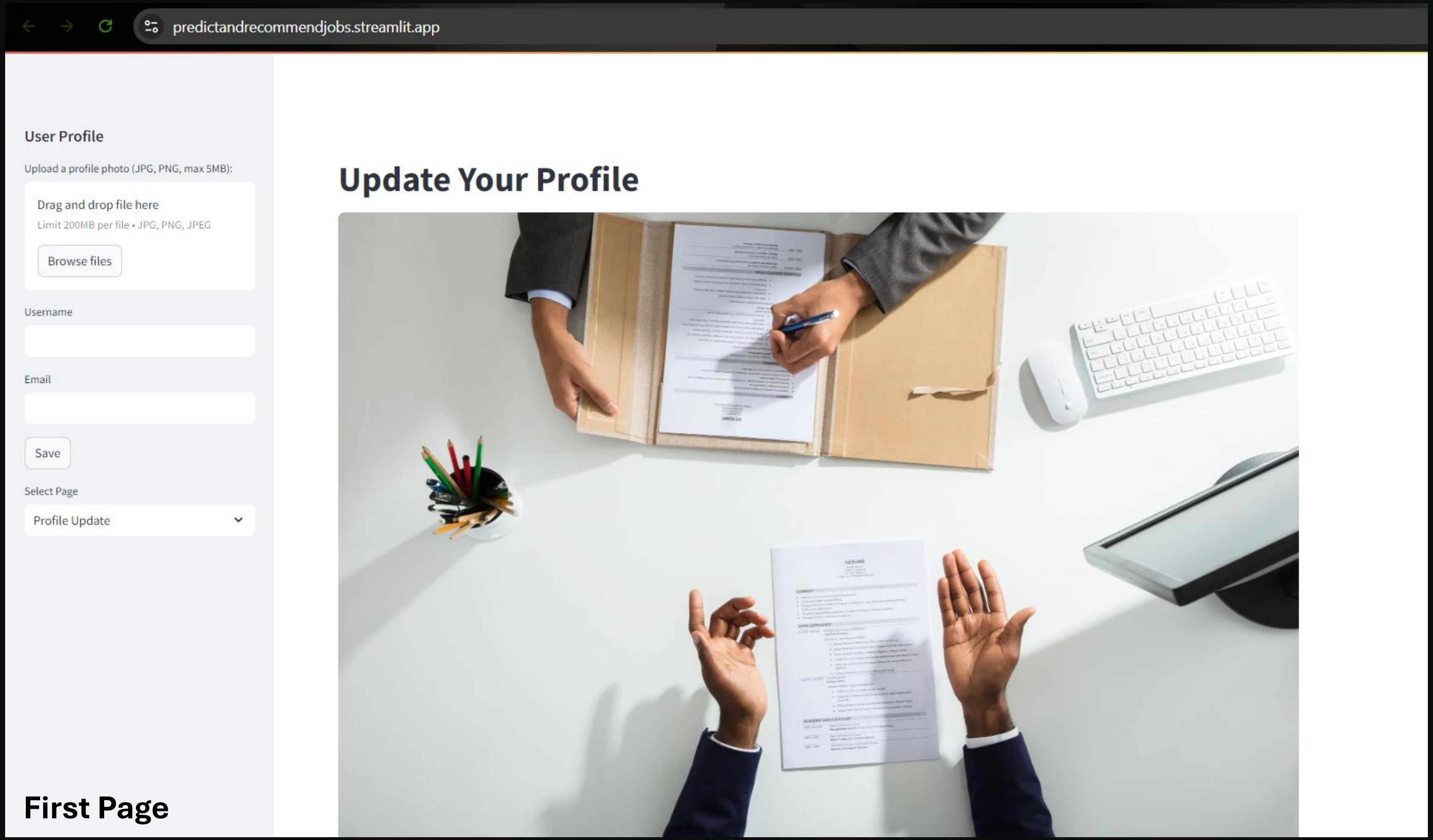
Recommendations

1. Boost Job Visibility: Encourage recruiters to increase job views through strategic promotion and SEO, ensuring postings reach a wider audience.

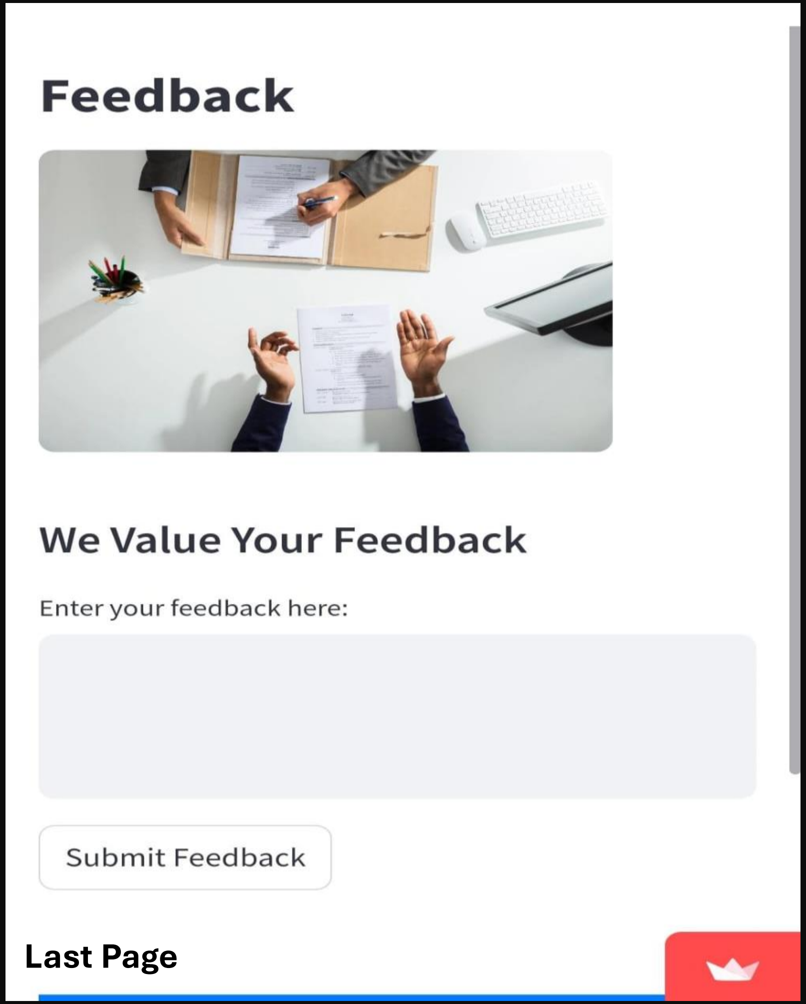
2. Enhance Job Descriptions: Advise recruiters to craft detailed, compelling job descriptions to attract top talent.

3. Promote Fresh Listings: Encourage recruiters to emphasize new job postings to capitalize on their initial attractiveness and draw immediate attention.

Job Recommendations App: Interface Preview



First Page



Last Page

Limitations of the Study

1. Data Access- Inability to get local data.

2. Data Privacy and ethical issues.

3. Inadequate computation resources for running Machine Learning Models

Thank you!

- Job Recommendation System App: https://predictandrecommendjobs.streamlit.app/

- Github: https://github.com/ge-saka/CAPSTONE-Group2