```python
invalid_titles = [
    "current openings",
    "openings",
    "careers",
    "jobs",
    "the safetywing digital nomad residency"
]

df["job_title_clean"] = df["job_title"].str.lower().str.strip()

df["is_valid_job_title"] = ~df["job_title_clean"].isin(invalid_titles)
```

```python
df["is_valid_job_title"].value_counts()
```

```
is_valid_job_title
True     4355
False      43
Name: count, dtype: int64
```

Internship Mini Project Report

Remote Ok Jobs Dataset Analysis

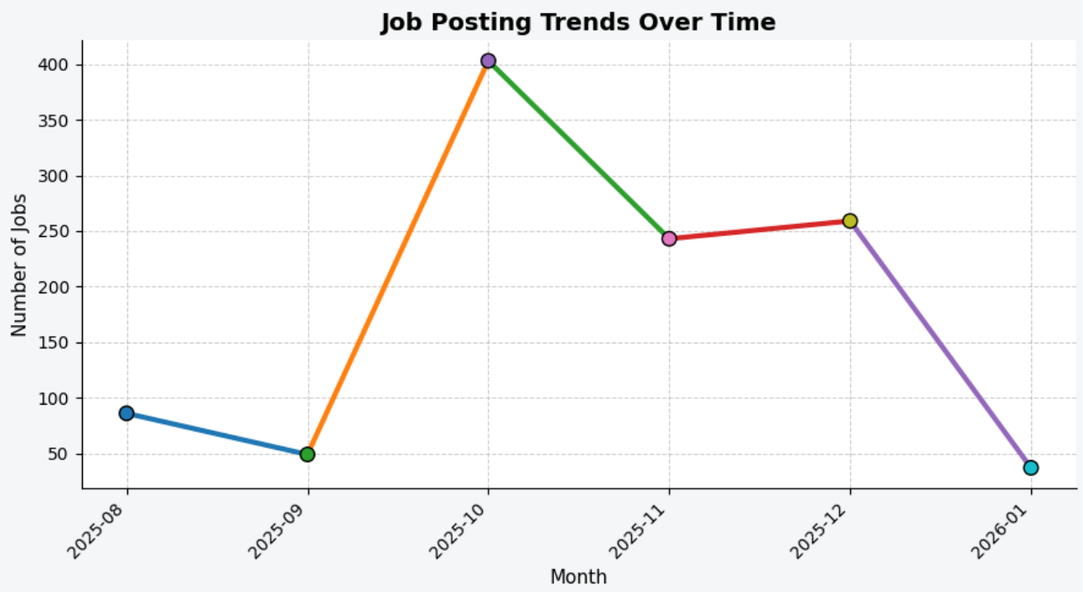
Presented by: Evoastra Limited Intern

PROJECT OVERVIEW

Unveiling Insights from the Remote Jobs Dataset

This project aimed to dive deep into the world of remote work by cleaning and thoroughly analyzing a comprehensive remote jobs dataset. Our primary objective was to transform raw, unstructured data into actionable intelligence, providing clarity on the remote job market landscape.

- **Objective:** Clean and analyze a remote jobs dataset to uncover key trends.
- **Tools Used:** Python's robust libraries: **NumPy**, **Pandas**, **Seaborn**, and **Matplotlib**.
- **Goal:** Extract meaningful insights through systematic data cleaning and impactful visualizations.



DATASET DETAILS

Understanding the Remote Jobs Landscape

Search Rol	Job Title	Company I	Technical S	Job Attribu	Location	Date Poste	Job URL
Support	DeFi Analy	Decentrali	Edu, Teaching, Crypto		?? Worldw	2025-08-2	https://remoteok.com/remote-jobs/remote-defi-an
Engineer	Full Stack I	Jiga	Senior, JavaScript, Rea		???? Unite	2026-01-0	https://remoteok.com/remote-jobs/remote-full-sta
Engineer	Lead Omic	Axle	Software, Assistant, L		?? Upgrade	2026-01-0	https://remoteok.com/remote-jobs/remote-lead-or
Engineer	Copy of Se	Trase Syste	Senior		USA	2026-01-0	https://remoteok.com/remote-jobs/remote-copy-o
Engineer	Senior Soft	Trase Syste	Software, Senior, Eng		?? Upgrade	2026-01-0	https://remoteok.com/remote-jobs/remote-senior-
Engineer	Senior Soft	Trase Syste	Front End, Software, S		?? Upgrade	2026-01-0	https://remoteok.com/remote-jobs/remote-senior-
Engineer	Software E	OpenAI	Software, Design, Syst		?? Probabl	2026-01-0	https://remoteok.com/remote-jobs/remote-softwa
Engineer	Software E	Tech Solut	Software, Web		San Franci	2026-01-0	https://remoteok.com/remote-jobs/remote-softwa
Engineer	Senior Bac	Cloudbeds	Software, SaaS, Amaz		Minnesota	2026-01-0	https://remoteok.com/remote-jobs/remote-senior-
Engineer	Senior Bac	Cloudbeds	Software, SaaS, Amaz		Michigan	2026-01-0	https://remoteok.com/remote-jobs/remote-senior-
Engineer	Senior Bac	Cloudbeds	Software, SaaS, Amaz		Massachus	2026-01-0	https://remoteok.com/remote-jobs/remote-senior-
Engineer	Site Reliab	Bobsled	Technical, Operationa		US/Canada	2026-01-0	https://remoteok.com/remote-jobs/remote-site-rel
Engineer	Full Stack S	Memberup.com			???? Unite	2026-01-0	https://remoteok.com/remote-jobs/remote-full-sta
Engineer	Associate	Recharge	Technical, Support		?? Upgrade	2026-01-0	https://remoteok.com/remote-jobs/remote-associa
Engineer	Senior Soft	Bugcrowd	Software, Design, Sec		India	2026-01-0	https://remoteok.com/remote-jobs/remote-senior-
Engineer	Staff Softv	Malt	Software, Design, Tec		France	2026-01-0	https://remoteok.com/remote-jobs/remote-staff-s
Engineer	Staff Softv	Galileo Fin	Software, Finance, Fir		New York	2026-01-0	https://remoteok.com/remote-jobs/remote-staff-s
Engineer	Senior Clo	ClickHouse	Security, Architect, C,		United Sta	2026-01-0	https://remoteok.com/remote-jobs/remote-senior-
Engineer	Senior Stat	Aegis Vent	Software, Python, Tec		Remote, U	2026-01-0	https://remoteok.com/remote-jobs/remote-senior-
Engineer	Member o	Inflection .	Technical, Support, Er		Palo Alto	2025-12-3	https://remoteok.com/remote-jobs/remote-membe
Engineer	Staff Softv	Muon Spac	Software, Design, Syst		San Jose	2025-12-3	https://remoteok.com/remote-jobs/remote-staff-s
Engineer	Software E	xAI	Software, Support, Tr		?? Upgrade	2025-12-3	https://remoteok.com/remote-jobs/remote-softwa
Engineer	Senior Soft	Chainguar	Software, Design, Clo		?? Upgrade	2025-12-3	https://remoteok.com/remote-jobs/remote-senior-

The dataset is a rich collection of job listings specifically tagged as remote opportunities. It includes vital features that help paint a picture of each job:

- **Job Title:** The specific role offered (e.g., "Software Engineer," "Data Analyst").
- **Company:** The organization offering the remote position.
- **Location:** Geographic specifics, even for remote roles (e.g., "US-based," "Europe").
- **Date Posted:** Details of posting the appliation
- **Job Skills:** Classification such as requirements of core programing, soft skills etc.

Initially, the dataset presented common real-world challenges, including missing values, duplicate entries, and inconsistent data formats, necessitating a robust cleaning phase.

DATA PREPARATION

Streamlining Data for Accuracy and Reliability

Handling Missing Data

Pandas was instrumental in identifying and addressing missing values, employing strategies like imputation or removal based on data context.

Removing Duplicates

Duplicate records were identified and eliminated using Pandas, ensuring each job listing was unique and preventing skewing of analysis.

Numerical Operations

NumPy provided efficient tools for numerical computations, essential for transforming and preparing quantitative data fields for analysis.

Standardizing Variables

Categorical variables like 'job types' and 'locations' were standardized, consolidating similar entries for consistent analysis.

The meticulous data cleaning process using Python's powerful libraries resulted in a high-quality dataset, ready to yield reliable insights.

```
df["job_title"].value_counts().head(50)
```

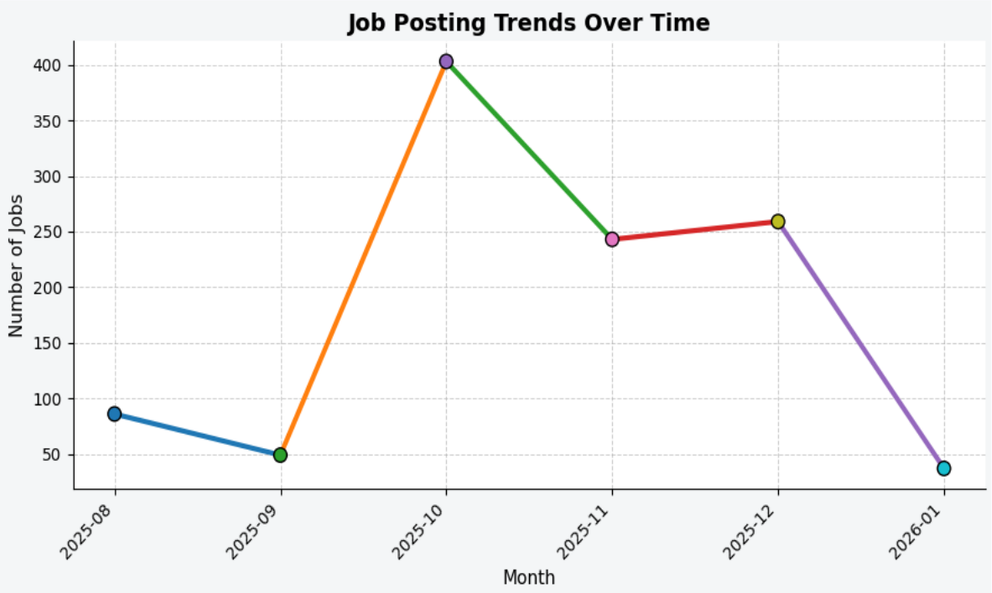
```
job_title
Software Engineer                                  85
Senior Software Engineer                           49
Senior Data Engineer                               40
Senior DevOps Engineer                             37
Data Engineer                                      34
Account Executive                                  32
Current openings                                   27
Product Manager                                    23
DevOps Engineer                                    22
Staff Software Engineer                            20
Enterprise Account Executive                       19
Customer Success Manager                           18
Sales Development Representative                   17
Senior Product Manager                             17
Sales Engineer                                     15
Senior Backend PHP Software Engineer Inventory     15
Growth Marketing Manager                           15
The SafetyWing Digital Nomad Residency             14
Product Designer                                   14
Backend Engineer                                   14
Senior Full Stack Engineer                         14
Project Manager                                    14
Full Stack Software Engineer                       13
Account Manager                                    13
...
Customer Service Representative                     9
Senior Backend Software Engineer                    9
```

EDA STRATEGY

Visualizing Data for Deeper Understanding

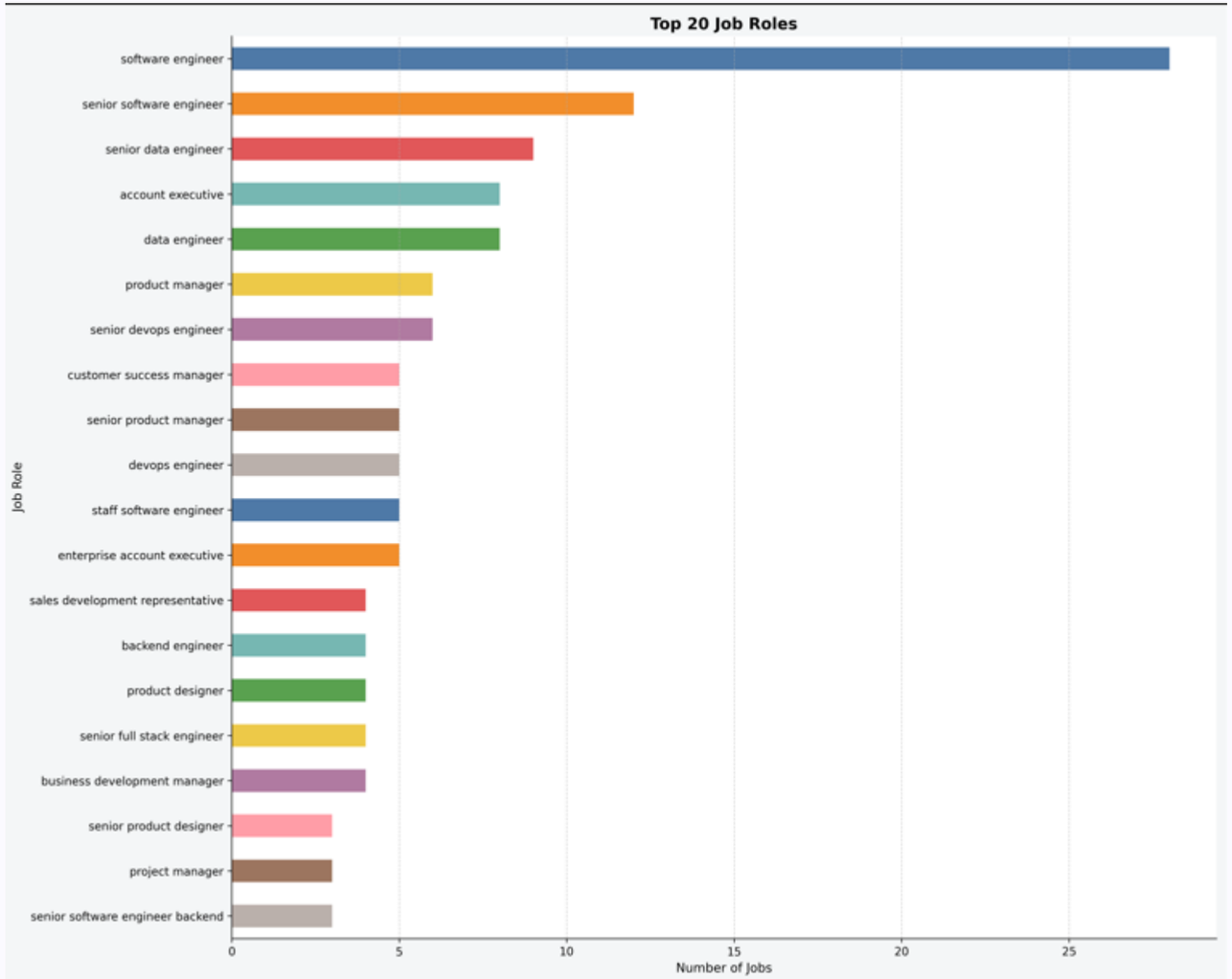
Our Exploratory Data Analysis (EDA) phase leveraged the visualization capabilities of **Matplotlib** and **Seaborn**. This allowed us to visually inspect the dataset, identify key characteristics, and formulate hypotheses before formal modeling.

- **Focus Areas:** We concentrated on understanding data distributions, uncovering correlations between different features, and identifying overarching trends within remote job listings.
- **Plotting Techniques:** Various types of plots, including histograms, scatter plots, and box plots, were generated to highlight patterns, outliers, and potential issues within the data.
- **Objective:** The goal was to gain an intuitive understanding of the dataset's structure and content, paving the way for more targeted analysis.



KEY VISUALIZATIONS

Job Type Distribution: A Snapshot of Remote Employment

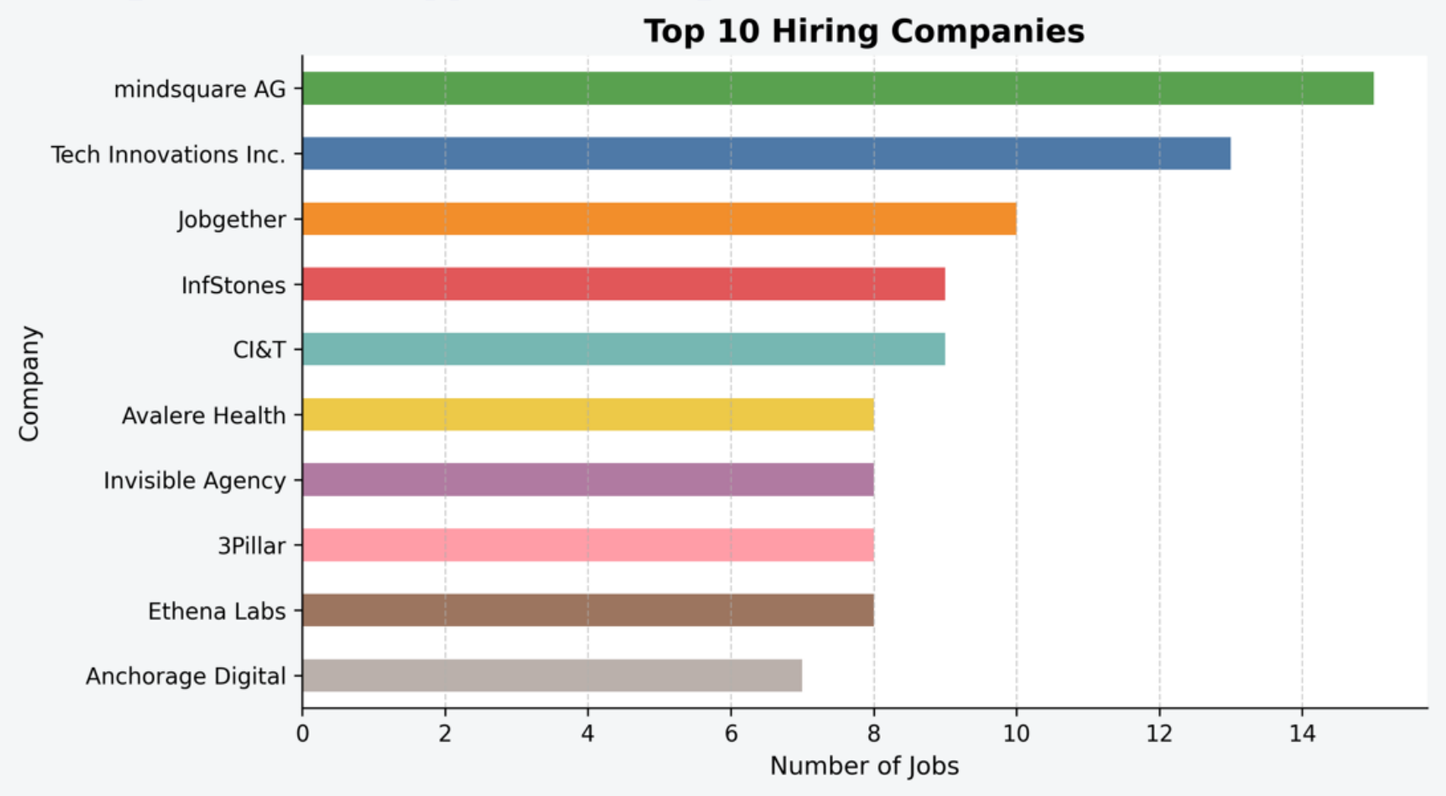


A crucial step in understanding the remote job market was analyzing the distribution of job types. This bar chart clearly illustrates the prevalence of different employment models within the remote sector.

- **Predominance:** The chart reveals that the vast majority of remote opportunities are for **SOFTWARE ENGINEER** positions, indicating a strong demand for IT GRADUATES.
- **Contract Roles:** Contract-based remote jobs also represent a significant portion, highlighting flexibility for both employers and professionals.
- **Market Composition:** This visualization provides a clear understanding of the remote job market's composition, valuable for both job seekers and companies formulating hiring strategies.

KEY VISUALIZATIONS

Top Hiring Companies in the Remote Space



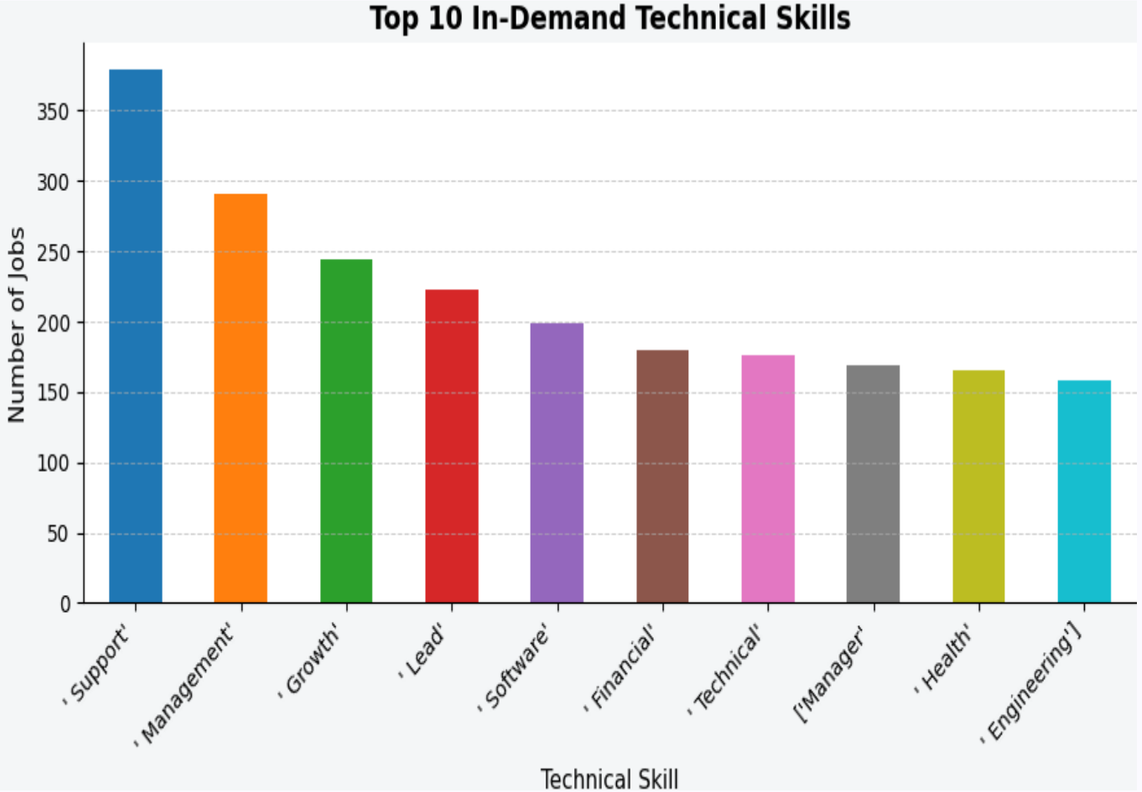
Mindsquare AG-leading recruiter

Expanding rapidly with remote opportunities in IT and digital consultation and services.

This analysis, typically represented by a **Seaborn horizontal barplot**, revealed the companies most actively recruiting remote talent. It provided valuable insight into which industries and organizations are leading the charge in embracing remote work models.

DEEPER DIVE

Skills Analysis: Benchmarking requirements for remote jobs



Understanding skills is paramount in any job market analysis. We utilized **boxplots** to visually compare salary distributions across different job types and locations for remote positions.

- **Identifying Trends:** These visualizations allowed us to discern median salaries, interquartile ranges, and the spread of compensation for various remote roles.
- **Outlier Detection:** Boxplots effectively highlighted salary outliers, indicating exceptionally high or low paying roles that deviate from the norm.
- **Strategic Value:** This insight is incredibly valuable for job seekers to set realistic salary expectations and for recruiters to benchmark competitive compensation packages in the remote sector.

REFLECTIONS

Navigating Challenges and Embracing Learning

Data Inconsistencies

Dealing with missing and inconsistently formatted data proved to be the most significant challenge, requiring iterative cleaning steps and careful validation.

Unexpected Patterns

Visualizations often revealed patterns and relationships within the data that were not immediately apparent, leading to deeper investigative paths.

Skill Reinforcement

This project significantly strengthened my proficiency in Python data libraries and enhanced my ability to translate data into a compelling narrative.

Every challenge encountered served as a valuable learning opportunity, reinforcing the importance of thorough data handling and thoughtful analysis.

Conclusion & Future Directions

Project Success

Successfully cleaned, analyzed, and visualized the remote jobs dataset, extracting valuable insights into the market.

Impactful Insights

Our visual analysis provides a clear understanding of remote job trends, company landscapes, and salary benchmarks.

Next Steps

Future work could involve applying machine learning models for job recommendation systems or predictive trend analysis.

Thank you for your attention and interest in this project. I am happy to answer any questions you may have!