

MATCHING EMPLOYEES TO PROJECTS

By
Saad Wali Muhammad Joiya

Submitted to Central European University - Private University
Department of Economics

In partial fulfilment of the requirements for the degree of Master of Science in Business Analytics

Supervisor: Eduardo Arino de la Rubia

Vienna, Austria
2025

COPYRIGHT NOTICE

Copyright © Saad Joiya, 2025 (year of submission). Matching Employees to Projects - This work is licensed under [Creative Commons Attribution-NonCommercial-NoDerivatives \(CC BY-NC-ND\) 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.

For bibliographic and reference purposes this thesis/dissertation should be referred to as: Joiya, Saad. 2025. Matching Employees to Projects. MSC Thesis, Business Analytics, Central European University, Vienna.

AUTHOR'S DECLARATION

I, the undersigned, **Saad Wali Muhammad Joiya**, candidate for MSc degree in Business Analytics declare herewith that the present thesis titled “Matching Employees to Projects” is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright.

I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Vienna, 08 June 2025

Saad Wali Muhammad Joiya

PUBLIC SUMMARY

This project focuses on designing a smart, data-driven employee–project matching system to support staffing decisions at scale. Currently, the client uses a manual process to match suitable employees for incoming projects. This process is not optimal considering time, effort, scalability and human errors. Moreover, an organized framework for projects and employee data collection is also currently absent. There is a need for creating a proof of concept that streamlines end to end data collection from feature list to feature data types, utilizes this data to create a framework of scoring employees against projects and interactively presents these results to the user.

In large organizations, traditional project staffing often relies on intuition, making it difficult to objectively assess candidate suitability for projects — especially when evaluating large teams or highly specialized roles. To address this, our system uses structured and unstructured data to score employees based on their alignment with project requirements. In organizations where projects vary greatly in scope and skill requirements, optimally matching resources to projects is vital for maximizing efficiency and enhancing delivery. The goal is to assist project managers and HR teams in identifying the most suitable employees for a given project by evaluating multiple dimensions of fit, thereby reducing manual effort, bias, and inefficiency in the resource allocation process. Moreover, we can enhance the quality of employee allocation by making more accurate and consistent recommendations based on matching features. This system can be improved over time with innovations as there has been less prior work done in this regard at the client’s end.

We simulated realistic employee and project datasets using thematically aligned roles, summaries, and domain-specific pools for skills, products, certifications, and languages. These were based on industry research and job descriptions inspired by our client's organizational structure. This thematic structure helped ensure that the matching logic made practical sense — for example, you would not see a legal advisor matched to a cloud infrastructure rollout. Each employee and project record includes detailed features such as job descriptions, role summary, skills with proficiency levels, language fluency, certifications, availability, and soft skills like cultural awareness and leadership experience and more. We then define corresponding features from both tables for scoring purposes.

Matching is performed using a multi-criteria scoring framework. This includes similarity scoring between text-based features, fuzzy matching for fields with human-written inputs (e.g., skills, locations, languages), keyword overlap for list-based features multi-step statistical scoring algorithms for numeric fields and some additional bonus scores for employee values and characteristics. We also incorporated logic for filtering out employees who are not available within a project's timeline. All these scores are weighted — meaning decision-makers can prioritize what matters most (skills, availability, certifications, etc.) and rolled into a final score that ranks the top best-fit employees per project. The result is a flexible, interpretable, and fair system that offers a scalable foundation for intelligent workforce planning.

For demonstration purposes we implement the complete process in a StreamLit dashboard for interactively displaying top employees for each project based on feature weightages and a detailed

comparison and explanation for each proposed employee ranking is provided. This dashboard was deployed on a VM using dockers for easy reproducibility and accessibility.

For easy reproducibility and collaboration, we generated a Capstone Project Report, Jupyter Notebook, StreamLit App and a GitHub repository. The capstone project report contains project details, complete overview of data, matching and scoring mechanisms and dashboard deployment. The Jupyter Notebook contains a proof of concept of the complete process from data creation to final scoring containing all dependencies and explanations. The StreamLit App is deployed on a VM and is publicly accessible for demonstration purposes. Finally, all these artifacts including reports, codes and requirements are uploaded in a public GitHub repository.

The approach implemented is highly interpretable. It is easy to grasp even for people with non-technical backgrounds. Weights were used as needed to control how the employees are ranked, and which features contribute more than others to the final score. The scores were validated using arena approach in collaboration with the client. This method does not have the usual limitations of a Machine Learning model, making it highly dynamic and extendable. Moreover, it provided the leverage to tweak the model based on client expectations in the future. For example, more features can be added over time, fuzzy thresholds can be tweaked and scoring functions can be easily updated. Furthermore, the approach empowers the client to unlock the full potential of their human capital, fostering both project success and employee growth in an increasingly dynamic business environment. This Capstone Project not only addresses a tangible business problem but also contributes to the growing understanding of how AI can positively transform core HR functions.

BIBLIOGRAPHY

Sources Referenced

Raghunandanan, Arjun. 2025. “How to Generate Synthetic Business Dummy Data with Gemini: using Gemini Library” <https://arjunraghunandanan.medium.com/generate-sythetic-data-with-google-gemini-07e2179b448e>

Dhungana, Kamal. 2023. “An Overview of ChromaDB: The Vector Database” <https://medium.com/@kbdhunga/an-overview-of-chromadb-the-vector-database-206437541bdd>

Nidhiworah. 2024. “ChromaDB – Introduction”. <https://medium.com/@nidhiworah02/chroma-db-introduction-25718915bae6>

Belagatti, Pavan. 2025. “Semantic Search: What is it + How Does it Work?” <https://www.singlestore.com/blog/a-complete-guide-to-semantic-search-for-beginners/>

Espejel, Omar. 2022. “Getting Started with Embeddings” <https://huggingface.co/blog/getting-started-with-embeddings>

Karabiber, Faith. “Cosine Similarity” Accessed May 15, 2025
<https://www.learndatasci.com/glossary/cosine-similarity/>

Sundriyal, Harshir. 2024. “Rapid Development of Streamlit App in Cloud (Virtual Machine)
<https://medium.com/@harshitsundriyal/rapid-deployment-of-streamlit-app-in-cloud-virtual-machine-7193f629be58>

Snowflake Inc. “Deploy Streamlit using Docker” Accessed May 25, 2025
<https://docs.streamlit.io/deploy/tutorials/docker>

AI DISCLAIMER

This project may incorporate content—such as ideas, analyses, text, code, or supporting material—generated or assisted by large language models (LLMs) and other artificial intelligence (AI) tools. These technologies were used to support research, drafting, coding, or source gathering processes. All AI-assisted content was critically reviewed and validated by the author(s). Responsibility for the methodology, analysis, and conclusions presented herein lies solely with the author(s).