**📄 Internship Report**

**Name:** Rahul Rai  
**Internship Role:** Data Analyst Intern  
**Company/Platform:** [NullClass]  
**Duration:** [17-07-2025] – [17-08-2025]

**1. Introduction**

This report summarizes my internship experience as a Data Analyst Intern, where I contributed to data-driven decision-making processes and analytics dashboard development. The internship was aimed at enhancing my analytical thinking, programming proficiency, and practical understanding of real-world data workflows.

**2. Background**

With a background in Python programming, data science, and exploratory data analysis (EDA), I joined this internship to gain practical exposure to solving real business problems using data. The company/project focused on digital transformation, and my role involved supporting analytical models and reporting pipelines.

**3. Learning Objectives**

* Apply statistical methods and machine learning models on real datasets.
* Use tools like **Pandas, Matplotlib, Streamlit**, and **SQL** in production-like environments.
* Collaborate with team members using **GitHub** and project boards.
* Translate business problems into data tasks and KPIs.

**4. Activities and Tasks**

* Cleaned and analyzed job description datasets with over 100,000 entries.
* Built a **Job Analytics Dashboard** using **Streamlit** and deployed it with **Render and Netlify**.
* Used **TF-IDF** and **Logistic Regression** to predict job matches from user resumes.
* Implemented time-based filters, company clustering, and sentiment analysis of reviews.
* Created charts (heatmaps, bar charts, timelines) for visual storytelling.
* Handled large CSVs and resolved Git/GitHub push errors related to file size (LFS issues).

**5. Skills and Competencies Gained**

| **Skill** | **Description** |
| --- | --- |
| **Data Cleaning** | Preprocessing unstructured and inconsistent datasets. |
| **Python & Pandas** | Advanced data manipulation and modeling. |
| **Streamlit** | Interactive dashboard and frontend development. |
| **Version Control (Git)** | Working with branches, resolving merge conflicts. |
| **Deployment** | Hosted app using Render; debugged deployment issues. |
| **Communication** | Shared updates via Git commits and documentation. |

**6. Feedback and Evidence**

* Received positive feedback from mentors and peers for:
  + Proactive problem-solving.
  + Creating visually appealing and functional dashboards.
* Maintained a public GitHub repository as evidence:  
  🔗 https://github.com/rahulraimau/job\_analytics\_portal\_project.git

**7. Challenges and Solutions**

| **Challenge** | **Solution** |
| --- | --- |
| GitHub push failed due to file size | Used .gitignore, avoided uploading CSV >100MB, considered Git LFS. |
| Module not found in Render | Updated requirements.txt, ensured transformers, pytz were installed. |
| Time-based logic for India (IST) | Used pytz with datetime to handle timezone accurately. |
| Data imbalance in target variable | Applied stratified sampling and balanced metrics. |

**8. Outcomes and Impact**

* Developed a deployable analytics dashboard for job trends and prediction.
* Helped simulate a job recommendation system using user-uploaded resumes.
* Demonstrated real-world problem-solving with data and full-stack tools.
* Created a solid GitHub portfolio to showcase my work.

**9. Conclusion**

This internship was a transformative experience that enhanced both my technical and soft skills. It bridged the gap between theoretical knowledge and real-world application. The opportunity to work independently on end-to-end data science components—from cleaning to modeling to deployment—has prepared me for future roles in data analytics and software engineering.