

# Project Proposal

By Team

Unattached and Unhinged

Team member 1: Varad Paradkar (801418318)

Team member 2: Mitali Yadav (801453849)

Team member 3: Sudeepta Bal(801455628)

Team member 4: Ruthwik Dovala(801431661)

---

# Motivation, Analytics & Prediction

- ❖ As upcoming graduates, we are entering a highly dynamic and competitive job market. Through this project, we aim to gain valuable insights into current employment trends and better understand the evolving landscape.
- ❖ Our goal is to build a comprehensive data pipeline to generate insights into the current job market.
- ❖ Using Spark SQL, we will analyze key trends such as hiring companies, high-demand locations, and sought-after roles and skills.
- ❖ Additionally, we will implement real-time monitoring of new job postings and train MLlib models to classify jobs by industry and role specialization.

# Progress so far

1. The project group has been formally established and is prepared to begin collaborative work.
2. The team has decided to work with a job related dataset as the focus, to be compiled through web scraping from LinkedIn and a startup-oriented job platform.
3. An initial sample dataset has been acquired to support demonstration activities and preliminary testing.

Autosave | Home | Insert | Draw | Page Layout | Formulas | Data | Review | View | Automate | Search (Cmd + Ctrl + U) | Comments | Share

Calibri (Body) | 11 | A | A' | General | Conditional Formatting | Cell Styles | Insert | Delete | Sort & Filter | Find & Select | Sensitivity | Add-ins | Analyze Data | Yes | No

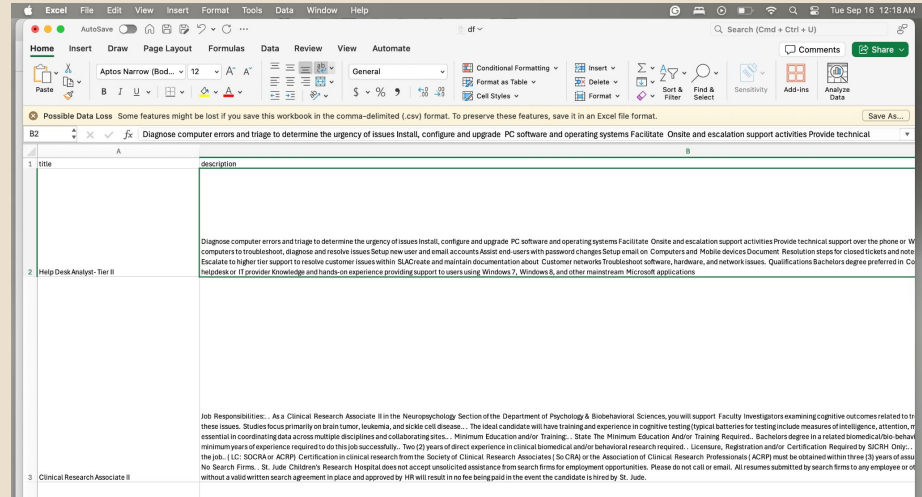
Open recovered workbook? Your recent changes were saved. Do you want to continue working where you left off?

D2D6 | Be an early applicant

	A	B	C	D	E	F
	job_title	company_name	location	hiring_status	date	seniority_level
1	Store Business Manager - JAMES CHRISTENSEN	M. Inc.	Brisbane, Queensland, Australia	Be an early applicant	2023-04-10 00:00:00	Not Applicable
2	Full-time	Guterman	Chicago, IL	Be an early applicant	2023-03-10 00:00:00	Mid-Senior level
3	Senior Machine Learning Engineer	Robert I. Houch	Adelaide, South Australia, Australia	Be an early applicant	2023-04-01 00:00:00	Mid-Senior level
4	Senior Data Scientist	Rupa	Melbourne, Victoria, Australia	Be an early applicant	2023-04-09 00:00:00	Entry level
5	Solution Architect	Hykon Digital	Chennai, Tamil Nadu, India	Be an early applicant	2022-01-20 00:00:00	Mid-Senior level
6	Lead Data Scientist	Spry	Gurgaon, Haryana, India	Active Hiring	2023-03-10 00:00:00	Mid-Senior level
7	Chief AI Data Scientist & Data Analysts	Endavour Group	Sydney, New South Wales, Australia	Active Hiring	2023-03-26 00:00:00	Mid-Senior level
8	Blockchain Engineer	Sapit	India	Active Hiring	2023-03-10 00:00:00	Mid-Senior level
9	Senior Software Engineer (Fullstack - React/PHP/NodeJS)	Highway International Inc.	Brisbane, Queensland, Australia	Active Hiring	2023-03-10 00:00:00	Mid-Senior level
10	Full Stack Front-End Developer	Digitalart	Bangalore, Karnataka, India	Active Hiring	2023-02-06 00:00:00	Entry level
11	Executive Assistant	Sony Music Publishing	New York, United States	Active Hiring	2023-04-06 00:00:00	Mid-Senior level
12	Senior Developer (Full Stack) Remote	Glassdoor	Germany	Be an early applicant	2023-04-13 00:00:00	Mid-Senior level
13	Executive Assistant - Strategic Software Support	CD Ross Group	London, England, United Kingdom	Active Hiring	2023-04-27 00:00:00	Mid-Senior level
14	Project Manager	SPHinx	Bangalore, Karnataka, India	Active Hiring	2023-03-10 00:00:00	Mid-Senior level
15	Executive Assistant	National Grid Remittance	Bloomington, MN	Be an early applicant	2023-04-10 00:00:00	Mid-Senior level
16	HR Generalist	Chia Bahwan Construction Co., Ltd	Shanghai, Guangdong, China	Be an early applicant	2023-03-10 00:00:00	Not Applicable
17	Personal Assistant to Chief Executive Officer	BrainBlox	London, England, United Kingdom	Active Hiring	2023-03-01 00:00:00	Full-time
18	Senior Project Manager (Full-time/Remote)	Investitech	Hyderabad, Telangana, India	Active Hiring	2023-03-10 00:00:00	Mid-Senior level
19	Senior Marketing Manager	JM Mariotti	Pune, Maharashtra, India	Active Hiring	2023-03-01 00:00:00	Mid-Senior level
20	Senior Software Engineer	Stashbuilder	Palo Alto, CA	Be an early applicant	2023-04-10 00:00:00	Not Applicable
21	Application Developer Full Stack	ETS	Portsmouth, England, United Kingdom	Be an early applicant	2023-03-14 00:00:00	Entry level
22	Senior Software Engineer	Pharmia, Inc.	Pharmia, CA	Be an early applicant	2023-03-10 00:00:00	Mid-Senior level
23	HR Manager	HR Manager	Mumbai, Maharashtra, India	Active Hiring	2023-01-18 00:00:00	Mid-Senior level
24	Senior Data Scientist	Recruiting from Toronto	Ontario, CA	Be an early applicant	2023-02-10 00:00:00	Mid-Senior level
25	Technical Program Manager - Editor	Cuma	Sydney, New South Wales, Australia	Active Hiring	2023-04-10 00:00:00	Entry level
26	Software Engineer - Java/Python	The HCL	Bangalore, Karnataka, India	Be an early applicant	2023-04-10 00:00:00	Entry level
27	Executive Assistant	BBC Coffee Limited	Leamington, England, United Kingdom	Active Hiring	2023-02-27 00:00:00	Full-time
28	Associate/Content writer	Quarant	Hyderabad, Telangana, India	Active Hiring	2023-04-10 00:00:00	Mid-Senior level
29	Associate/Content writer	Quarant	Bangalore, Karnataka, India	Active Hiring	2023-02-27 00:00:00	Associate
30	Associate/Content writer	Quarant	Pune, Maharashtra, India	Active Hiring	2023-04-10 00:00:00	Not Applicable
31	Project Manager	Ori	Los Angeles, CA	Active Hiring	2023-02-14 00:00:00	Associate
32	Program Manager - Operations (India)	Bellini	Bangalore, Karnataka, India	Active Hiring	2023-04-10 00:00:00	Not Applicable
33	Consult - Contracts Solutions GSE - IN	Unistatim	Gurgaon, Haryana, India	Active Hiring	2023-04-10 00:00:00	Full-time
34	Client Services	Geography	Gurgaon, Haryana, India	Active Hiring	2023-04-10 00:00:00	Full-time
35	Human Resource Generalist	Helio Nova Talent Solutions	El Segundo, CA	Active Hiring	2023-04-10 00:00:00	Full-time

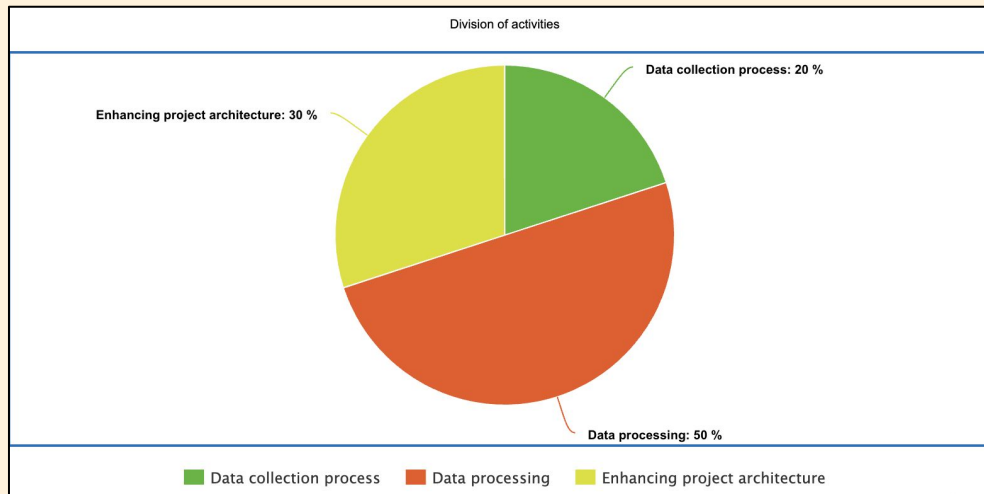
# Current Challenges

- ❖ Ongoing efforts are focused on enhancing the scraper to capture the latest job postings from LinkedIn and Y Combinator.



# Next Week activities

- ❖ To complete the data collection process and finalize job filtering criteria, including country and role specifications.
- ❖ To process the collected data by implementing ingestion and cleaning workflows using Spark.
- ❖ To enhance and document the end-to-end project architecture to validate the design and ensure clarity for each component.



# Dataset acknowledgement (Final

Dataset will be created by us after scraping)

- **Linkedin\_job\_posts\_insights:**  
<https://www.kaggle.com/datasets/sindhumadhurii/linkedin-job-posts-insights-dataset>
- **Linkedin data Job Titles and Postings:**  
<https://www.kaggle.com/datasets/tianyimasf/linkedin-data-job-titles-and-postings>

# Thank You!