Proposal feedback:

A data pipeline diagram might help you better scope out the entire project from end-to-end.

Questions to ask about project proposal in consult:

| Q | A |
| --- | --- |
| What are the requirements of the proposal since it is not stated except 1 page? Is there a format for the proposal? Like requirements for font size, pages, etc | No other stuff |
| Do we need to store the data into the DB and utilise it for downstream usage as an example in the project? | yes |
| What about the data source, is it recommended to scrape/ come up with our own data and store it accordingly (for example transforming into structured data for unstructured data) ourselves? | No necessary - more important to address selected problem, give data engineering solution (must use airflow) -> must solve concrete problem |
| What kind of data engineering project is good/ what type is usually done? | Examples in pdf, e.g. housing data aggregated |
| Is it necessary to develop the downstream application? What downstream applications are needed, if any? | Yes - show that the data collection can be used - simple visualisation will do |
| What is the final product/solution of the data engineering project? | Entire proj - point 3 of guideline |
| What are the recommended processes/stages/milestones for this project? | Identification of data sources & structure  Data retrieval to storing (ETL)  Downstream applications etc |

**Project Proposal: Job Market Analysis**

**1. Project Overview:**

The goal of this project is to analyse the job market in Singapore by collecting, processing, and visualising data from various job portals. The team will focus on extracting insights into job trends, skill demand, popular job titles, and salary ranges based on the different industries. The data will allow downstream applications such as an interactive dashboard to provide valuable information for job seekers, hiring managers and career planners to make decisions.

**2. Methodology:**

We aim to collect job postings from various job portals through web scraping and APIs. Some sample websites that are relevant include LinkedIn Job Portal, InternSg, JobStreet, SkillsFuture and Careers@Gov. Moving forward, we will implement a data pipeline to clean and transform the raw data, ensuring consistency and accuracy throughout the process. For efficient storage and retrieval of the processed data, we will leverage on a suitable database or data warehouse solution. Subsequently, we will implement an analytical engine to identify trends, including industry demand, skill sets requirements, salary ranges, popular job titles, etc. This is to create an interactive dashboard for visualising job market insights with customizable filters and user-friendly features. This will benefit both users, employees and employers, to utitlise it according to their needs.

**3. Resources:**

Open-source libraries: BeautifulSoup, Scrapy, Pandas, and Plotly for efficient data handling and visualisation.

Workflow Management: Apache Airflow for managing and orchestrating the data pipeline.

**4. Conclusion:**

The Job Market Analysis project aims to provide valuable insights for job seekers and career planning by leveraging data engineering techniques. The resulting dashboard will empower users to make informed decisions based on current job market trends and demands. This project aligns with the team's academic learning and will showcase practical skills in data engineering, analysis, and visualisation.

**API/Sources**

<https://developer.ssg-wsg.gov.sg/webapp/docs/product/6Gl44K5M46EuDgn7LCsAs2#>

<https://github.com/pwaaron/jobscrapers>

<https://beta.data.gov.sg/collections?query=job>