

Group 6 Project Proposal

Project Title: Recipe Data Exploration and Analysis

Collaborators: Jack Wall, Jorge Santos Jr, Anton LeBeque, Mitchell Lor, Maia Yang and Bukola Fatile

Project Overview: The Recipe Data Exploration and Analysis project aim to leverage the Recipe Search API, enabling us to search through millions of web recipes. As a collaborative effort within the data engineering track, our focus will be on exploring, transforming and analyzing the vast dataset available through this API.

Objectives: Data Exploration, Utilize the Recipe Search API (API URL: <https://api.edamam.com/api/recipes/v2>) to retrieve a diverse set of recipe data, covering various cuisines, ingredients, and cooking methods.

ETL Workflows: Implement Extract, Transform, Load (ETL) workflows to ingest data from the API. The original dataset(s) will undergo transformations to ensure consistency and relevance to our analysis.

Database Integration: Utilize SQLite as the database technology to store and manage the transformed recipe data. SQLite offers simplicity, portability, and ease of use, making it ideal for our data exploration and analysis needs.

Collaborative Analysis: Collaborate with team members to perform exploratory data analysis on the recipe dataset. Analyze patterns, trends, and relationships within the data to derive meaningful insights.

Project Phases:

- **Data Collection and Ingestion:** Develop Python scripts to interact with the Recipe Search API and retrieve a diverse set of recipes. Implement ETL workflows to clean and structure the raw data.
- **Database Setup:** Create an SQLite database to store the transformed recipe data. Define the database schema to accommodate the necessary fields for analysis.
- **Exploratory Data Analysis:** Collaboratively analyze the dataset to uncover patterns and trends. No data visualization tools will be used, focusing solely on data engineering aspects.
- **Documentation and Reporting:** Document the ETL processes, database schema, and analysis methodologies. Generate a comprehensive report summarizing key findings and insights.

Expected Outcomes: A well-structured SQLite database containing cleaned and transformed recipe data. Insights into popular cuisines, trending ingredients, and cooking methods. Collaboratively developed Python scripts and workflows for future data exploration projects.

Project Links:

- API URL: <https://developer.edamam.com/edamam-docs-recipe-api>
- GitHub Repository: <https://github.com/maiayang/project-3-data-engineering.git>