# ADS-507 – Initial Final Team Project Proposal

Fill out this form and submit it by the end of Module 3 in Canvas.

Team Number: *Group 3*

Team Leader/Representative: *Duy Nguyen*

Full names of team members:

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ *Duy Nguyen* \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*Gabriel Mancillas Gallardo*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*Jorge Roldan*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Title of your Production-Ready Data Pipeline project: MySQL Pipeline

Short description of your project and objectives:

*This project aims to create a production-ready data pipeline using MySQL and Python to process and transform data from raw sources into actionable insights. The primary goal is to design an ETL (Extract, Transform, Load) or ELT (Extract, Load, Transform) pipeline that ingests selected datasets, performs SQL-based transformations, and outputs meaningful results such as visualizations or notifications. The project is designed to be fully reproducible and scalable for larger datasets in the future.*

Name of your selected datasets:

1. Pagila Schema

2. Pagila JSON Schema

3. Pagila Insert Data

Description of your selected datasets (data source, format, size of dataset, etc.):

**Pagila Schema**

1. **Data Source**: Adapted from the Sakila database for PostgreSQL.
2. **Format**: SQL file defining the structure of the database (tables, relationships, constraints).
3. **Size**: Approximately 51.2 KB for schema definition.

**Pagila JSON Schema**

1. **Data Source:** A variant of the Pagila schema with JSONB support.
2. **Format:** SQL file including schema and JSONB fields.
3. **Size:** Approximately 1.77 KB.

**Pagila Insert Data**

1. **Data Source**: Sample data for the Pagila schema, including customer, rental, and film information.
2. **Format**: SQL file for inserting records into the Pagila database.
3. **Size**: Approximately 5.09 MB of data, representing a typical medium-sized dataset.

Please provide the link for your GitHub repository here: ***https://github.com/Gabeleo24/ADS-507***

How many times have your members met in the last two weeks? *Our team has been meeting through Slack every other day, which means we’ve met approximately 7 times in the last two weeks.*

List the specific contributions that each team member is providing for the Final Team Project in the table below.

* **NOTE:** ALL students on the team should contribute equally to the Final Team Project.

| Team Member Duy Nguyen | Team Member 2 Gabriel Mancillas Gallardo | Team Member 3 Jorge Roldan |
| --- | --- | --- |
| List of contributions | List of contributions | List of contributions |

Comments/Roadblocks: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_