**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 focuses on developing a production-ready data pipeline using SQL and MySQL database tools, with the AdventureWorks database as the core data source. The primary objective is to implement an ETL (Extract, Transform, Load) or ELT (Extract, Load, Transform) pipeline to process key datasets from AdventureWorks and produce actionable outputs, such as business analytics dashboards or triggered reports.*

Name of your selected datasets:
*1. SalesOrderHeader*
*2. SalesOrderDetail*
*3. Customer*
*4. Product*
*5. Employee*

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

*PurchaseOrderHeader*
1. **Data Source**: Part of the AdventureWorks transactional data, providing high-level details about sales orders.
2. **Format**: SQL table containing structured fields like OrderID, CustomerID, and TotalDue.
3. **Size**: Approximately 150,000 rows, totaling ~35 MB.

*ProductDescription*
1. **Data Source**: Line-item details for each sales order from the AdventureWorks database.
2. **Format**: SQL table including fields such as ProductID, Quantity, and UnitPrice.
3. **Size**: Approximately 1.2 million rows, totaling ~85 MB.

### *Customer*
1. **Data Source**: Demographic and business data for customers in AdventureWorks.
2. **Format**: SQL table with fields like CustomerID, TerritoryID, and demographics.
3. **Size**: Approximately 19,000 rows, totaling ~1.81 MB.

### *Product*
1. **Data Source**: Product catalog for AdventureWorks, containing data on product names, categories, and pricing.
2. **Format**: SQL table with structured data fields such as ProductID, Name, and StandardCost.
3. **Size**: Approximately 500 rows, totaling ~86.3 KB.

### *Employee*
1. **Data Source**: Employee data for AdventureWorks, detailing staff roles, departments, and salaries.
2. **Format**: SQL table with fields such as EmployeeID, JobTitle, and HireDate.
3. **Size**: Approximately 290 rows, totaling ~93.8KB.

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 |
|---|---|---|
| - Researched and documented the key datasets from AdventureWorks (SalesOrderHeader, SalesOrderDetail, Customer, Product, Employee).<br>- Drafted the SQL schema creation scripts for database setup (create-adventureworks.sql).<br>- Initiated the GitHub repository setup and added the initial folder structure for organizing project files. | - Wrote SQL insert scripts to populate tables with data (insert-data.sql).<br>- Edited the README file to include updated dataset descriptions and setup instructions for AdventureWorks.<br>- Organized team Slack meetings and documented team discussions to keep tasks on track. | - Designed the SQL transformations for cleaning and aggregating data (transformations.sql).<br>- Configured the MySQL database locally and tested loading of schema and data.<br>- Collaborated on refining the pipeline script to ensure it handles data ingestion and storage correctly. |

Comments/Roadblocks: _____

_____

_____