# ADS-507 – Initial Final Team Project Proposal

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

Team Number: \_\_\_\_\_\_\_\_\_\_\_\_\_

Team Leader/Representative: \_\_\_\_\_\_\_\_DARREN CHEN \_\_\_\_\_\_\_

Full names of team members:

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Title of your Production-Ready Data Pipeline project: **Data-Driven Strategies to Reduce Illicit Drug Demand Through Public Awareness and Predictive Analytics**

Short description of your project and objectives:

This project aims to develop a **data pipeline** that integrates multiple public health datasets to **analyze illicit drug usage trends, track treatment effectiveness, and assess public sentiment towards drug use**. Using SQL-based transformations and ETL pipelines, we will create an **interactive dashboard** that:

* **Monitors treatment admissions** and rehabilitation success rates.
* **Identifies public sentiment trends** using national survey data.
* **Predicts high-risk demographic groups** using data analytics.
* **Provides insights for policymakers** to develop prevention strategies.

By leveraging **relational databases and SQL queries**, this project will provide actionable insights to help drive **public perception away from illicit drug use** and promote **rehabilitation efforts**.

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

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

Name of your selected datasets:

1. **Treatment Episode Data Set ( TEDS)**\_

2. National Drug Treatment Monitoring System (NDTMS)\_

3. **National Survey on Drug Use and Health (NSDUH)**

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset Name** | **Source** | **Format** | **Size** | **Description** |
| **Treatment Episode Data Set (TEDS)** | **SAMHSA (Substance Abuse and Mental Health Services Administration, USA)** | CSV, SQL (via API) | **Millions of treatment records** | Tracks **treatment admissions**, **patient demographics**, **substance type**, and **rehabilitation outcomes**. Used to analyze trends in addiction treatment. |
| **National Drug Treatment Monitoring System (NDTMS)** | **UK Government (Public Health England)** | SQL Database, CSV | **Thousands of records per year** | Provides **patient-level treatment data** from UK rehab centers, including **admission rates**, **patient characteristics**, and **long-term recovery rates**. |
| **National Survey on Drug Use and Health (NSDUH)** | **SAMHSA (USA)** | CSV, SAS | **Tens of thousands of respondents annually** | Tracks **public drug use trends**, **perceptions of drug risks**, **availability of illicit drugs**, and **effectiveness of drug prevention efforts**. |

Please provide the link for your GitHub repository here: <https://github.com/darrencheninfo/practical-data-engineering-pipeline>

How many times have your members met in the last two weeks? \_\_\_\_2x\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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 1 (Darren) |
| --- |
| 1. everything. |

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

I struggled to find a project purpose that was effective in reducing illicit drug use in the homeland and was ethical. When I finally realized that shifting my strategy to affect domestic demand rather than supply and distribution networks, I realized that this is the ideal way to meet the project requirements and ethical requirements of the school.

**Data Format Variability**: The dataset NSDUH requires conversion from **SAS to CSV before SQL ingestion**.