**ADS-507 – Team Project Status Update Form**

Fill out this form and submit it by the end of Module 4 in Blackboard.

Team Number: 5

Team Leader/Representative: Jack McCullers

Full names of team members:

1. Oscar Gil

2. Christopher Richardson

3. Jack McCullers

Title of your Production-Ready Data Pipeline project: College Scorecard Data Pipeline

Short description of your project and objectives:

The purpose of our project is to take a flat dataset, that we have obtained on college scorecards, and implement batching to mimic the dataset as being continuously updated. The team will then utilize SQL queries to create tables, and models to group the data for analytics purposes. The team will employ techniques learned from the module homeworks for use of systems and querying. We will go in further detail what the dataset is comprised of.

Name of your selected dataset: CollegeScorecardDataDictionary.csv

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

In completing searches online for datasets that could be used for the final project, the team found the College Scorecard dataset from <https://data.ed.gov/dataset/docs/college-scorecard-all-data-files-through-6-2020> where data on college scorecards are pulled from the published from collegescorecard.ed.gov. The data is broken up across 5 tabs of data. Each tab has numerous amounts of variables ranging from 20 to 700+ variables. We have columns such as institution name, city, state, fafsa data, on community colleges vs 4 year universities, zip codes, address, women only schools and others to allow us to create tables for analysis. There are roughly 6700 records with almost 2400 columns.

Please provide the link for your GitHub repository here: <https://github.com/CFRichardson/ADS507_Project>

How many times have your members met in the last two weeks? 1 time over Zoom, but continuous Slack conversations on our topic.

What was the agreed upon method of communication? Are you using any teamwork project management software, such as [Trello](https://trello.com/en-US) or [Asana](https://asana.com/)? Slack, Github, and Google Drive

Comments/Roadblocks: Noting identified now.