The Team

Matt Clewis, Nathan Cruz, Clint Jansson

Problem Statement a.k.a Decompose The Ask

Collect data from 2 data sources pull them down and load them to a database.

Process the data into the database in a manner than it can be managed or manipulated.

Join the datasets once up on database

Identify Data Sources

Data source 1 = IMDB database from Kaggle, from 2006 to 2016 movie titles.

Data source 2 = Wikipedia web scrape for list of NC media actors and actresses

Define Strategy and Metrics

Collect CSV from IMDB website.

Collect List of actors from Wikipedia born in North Carolina.

Load to mysql using python.

Merge data and create view on mySQL

Description of Data Analysis Tools You Plan to Use

Pull data from web sources

Clean data using pandas

Upload data to database using SQLAlchemy

Create database structure using database mysql

Join data sets and create a view of just the Movies that incorporate North Carolina Actors and Actresses.

Describe the Data Products Your Project Will Produce

Database will house data in 2 tables that can be merged based upon the actor and actresses name.