

Guidelines for ETL Project

This document contains guidelines, requirements, and suggestions for ETL Project

Project Proposal

Before you start writing any code, remember that you only have one week to complete this project. View this project as a typical assignment from work. Imagine a bunch of data came in and you and your team are tasked with migrating it to a production data base.

Finding Data

Your project must use 2 or more sources of data. We recommend the following sites to use as sources of data:

- data.world
- [Kaggle](https://www.kaggle.com)

You can also use APIs or data scraped from the web. Again, there is only a week to complete this!

Data Cleanup & Analysis

Once you have identified your datasets, perform ETL on the data. Make sure to plan and document the following:

- The sources of data that you will extract from.
- The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).
- The type of final production database to load the data into (relational or non-relational).
- The final tables or collections that will be used in the production database.

You will be required to submit a final technical report with the above information and steps required to reproduce your ETL process.

Project Report

At the end of the week, you will submit a Final Report that describes the following:

- **Extract:** your original data sources and how the data was formatted (CSV, JSON, MySQL, etc).
- **Transform:** what data cleaning or transformation was required.
- **Load:** the final database, tables/collections, and why this was chosen.

Please upload the report (an excel sheet with 1000 Rows) to Github and submit a link to Bootcampspot.