John Lorenz IV
CS 510
Irvine
April 29th, 2022

In class Assignment 4

Initial Discussion Questions:

I don't have an itemized list of responses for this section of the in-class activity since it turned more into a discussion about it rather than each of us independently answering it. Some of the group hadn't had the opportunity to watch the videos and read the slides prior to class so there was some confusion on some of the terminology. We used this opportunity to skim through the slide deck and talk about the problem with data transformation, and the step in which validation occurs as well as the pipeline process.

Some initial responses:

Myself: Sometimes if data is in a rough state, or difficult to extract then forming it into a workable structure that achieves your goals may pose difficulties. I believe transformation occurs before validation, as the process of extraction typically happens before assertions are applied to the data set. There are opportunities where assertions can be applied to well understood data prior to pulling it, but it isn't guaranteed in exploratory or 'bit rotten' data sets.

Adam Shih: Transforming data and normalizing it into a format can be a challenge. Having some steps afterwards for removing incorrect rows is helpful.

Joshua Kasten: *Didnt catch it*

Ajay Gorantla: *Didnt catch it*