Avery Twitchell-Heyne

Diamond Williams

Michael Goldman

Tolu Olayemi

ETL Project: Legal trade in wildlife

Data Sources: Kaggle CITES Wildlife Trade Database (csv) IUNC Red List API

Type of transformation needed: Dataset from Kaggle will be cleaned and filtered to provide an easily reader format. Post cleansing, data set will be joined with API using Pandas and/or SQL to create an overall data. Once we have a combined data set from Kaggle and the API it will be loaded and store in to a Mongo or SQL data base.

We are planning to transform the data from the two sources (CSV and API) and recombine it in different ways it based on the final information we want to present to the end user. For example, our CSV source does not include the common name of the animals (only scientific classification), but we would like to incorporate common names into our tables/collections to make them more understandable.

The final tables/collections are as follows:

* Taxonomical
* Countries of Origin
* Countries of Destination
* Endangered Status
* Appendix Classification
* Threats by Species
* Term
* Institution Type
* Common Name