Phase II Outer Workflow

Input: Food_Inspections.csv

S0: Data Profiling & Exploratory Analysis using Python notebook
1. Check for Missing values
2. Check for Duplicate values

S1: General Data Cleaning uisng OpenRefine
1. Removed leading & trailing spaces
2. Collapsed Consecutive white spaces
3. Format to same case (uppercase/titlecase)
4. Data type conversions

Output: Chicago_Food_Inspection_EDA.ipynb

Output: Food_Inspections_general_cleaning.csv

S2: Clustering Columns using OpenRefine

Performed mass edits to cluster following cloumns
1. Inspection Type
2. Facility Type
3. DBA Name
4. Address
5. City

Output: Food_Inspections_clustered.csv

S3: Repair Missing Location Data using Python Notebook
1. Use Geocoding API to backfill missing
   Location, Latitude, & Longitude values
2. Use Reverse Geocoding API to backfill missing Zip, City & State values

**Output: Food_Inspections_location.csv,
Backfill_Location_Info.ipynb**

S4: Load data to SQLite using Python
1. Create SQL Constraints for Unique FoodEstablishment & EstablishmentLocation
2. Parse & extract Violation Columns using Python & Regex
3. Load Data into SQLite Database
4. Perform data integrity checks via SQL
5. Extract Cleaned Data from SQLite to CSV flat file

Input: multiple history.json

**Output: foodinspections_cleaned.sqlite,
food_inspections_db_cleaned.csv**

S5: Create Workflow Model
1. Create Outer Workflow Diagram using Draw.io
2. Collect OpenRefine histories for Inner Workflow
3. Generate Inner Workflow model diagrams using OR2YW Tool

S6: Perform Data Visualization & U1 Analysis
2. Load Data into Tableau
3. Create & Publish Dashboard
to Tableau Public
4. Perform U1 analysis

**Output: inner_workflow.yw,
inner_workflow.png
inner_workflow.gv,
outer_workflow.drawio,
outer_workflow.png**

Output: ChicagoFoodInspectionDashboard.twb

Phase II Report
1. Description of Data Cleaning procedure
2. Document Data quality improvements
3. Conclusions & Summary
4. Upload Supplementary Materials

Output: CS513-Phase-2-report-final.pdf