1. **Data Cleaning (R) - Data Cleaning_Binary Transformation.R**
    - It uses ***culvertdata.csv*** dataset
    - It removes the outlier data from the following columns:
    - 1. TMAC 2. TMINC 3. PPTIN 4. YEAR_OF_FUTURE_ADT_115 5. FUTURE_ADT_114 6. DECK_WIDTH_MT_052
    - It converts the non-standard values into NA's and removes the data for the following columns:
    - 1. SCOUR_CRITICAL_113 2. WATERWAY_EVAL_071 3. STRUCTURAL_EVAL_067 4. CHANNEL_COND_061
    - After creating 4 new columns with integer types for the above columns, the 4 non-standard columns are removed
    - The following 3 character/string type columns are removed that will not be considered as input for correlation:
    - 1. FID 2. BRIDGE_CONDITION 3. NAME
    - Cleaned data is taken with 33 observations: ***clean.csv***
    - Cleaned data with names with 36 observations: ***clean_name.csv***
2. **Binary Variable Transformation (R) - Data Cleaning_Binary Transformation.R**
    - 0 ~ 5 – Unsatisfactory Condition is represented by 0
    - > = 6 - Satisfactory Condition is represented by 1
3. **Stratified Random Sampling (R) - Stratified Sampling.R**
    - It uses ***clean_name.csv*** dataset
    - It uses seed value 10% for random sampling
    - After random sampling is done total data = 13280 Observations
    - Two Bar Plots are made
    - 1. Total Values per state
    - 2. Total Values per state for Stratified Data
    - Stratified Data is taken out as: ***strat_data.csv***

- The following 3 columns are removed from the stratified dataset as they will not be considered for correlation
- 1. FID 2. BRIDGE_CONDITION 3. NAME

4. **Correlation Plot (Python) - Model Entropy_Correlation.py**
   - It uses ***strat_data.csv*** as dataset
   - It includes 13280 observations with 33 columns, among which culvert_cond is the output column having a series of 0's and 1's
   - From the correlation plot it is evident that the following columns show highest correlation with the output column:
   - 1. YEAR_BUILT_027 2. APPR_ROAD_EVAL_072 3. LOWEST_RATING 4. Struc_eval 5. Channel_cond
5. **Model Entropy (Python) - Model Entropy_Correlation.py**
   - From the mutual information we find the following giving more information about the culvert condition:
   - 1. YEAR_BUILT_027 2. LOWEST_RATING 3. Struc_eval 4. Channel_cond
6. **Final List of Inputs: (Taken from correlation + model entropy) - Final Input List.csv**
   - YEAR_BUILT_027
   - LOWEST_RATING
   - Struc_eval
   - Channel_cond
7. **State-wise Culvert Location Mapping (R + Python) - Culvert Location Mapping.R + Mapping.py**
   - It uses ***Map Data.csv*** as dataset for Python
   - Few data points of few states are taken to plot the culvert presence using Geopandas
   - It uses ***clean_name.csv*** as dataset for R
   - Both maps are printed as images