ArcGIS/Arcpy Dataset Generation: (Any time an Arcpy function is used, the same process can be done in ArcGIS Pro but may not be automated to the same degree).

1. Extract the individual raster layer from all .md files(or most other multidimensional raster file types) by using Arcpy's ***auto_multiD_subsetting*** function.
   A script that automatically extracts a given layer from ***.md files*** associated with a given list of dates from a given directory, named ***'MDR_layer_extracterV2'***, can be found in the **GitHub directory**.

2. Create a grid shapefile that encompasses the entire study area.
   a. Open ArcGIS Pro and create a new project or open an existing one.
   b. Open the "Create Fishnet" Tool:
      i. Go to the "Analysis" tab and select "**Tools**".
      ii. Search for "**Create Fishnet**" in the Geoprocessing pane and open the tool.
   c. Configure the "Create Fishnet" Tools:
      i. ***Output Feature Class:*** Choose a location and name for the output fishnet shapefile.
      ii. ***Template Extent:*** Define the extent of the fishnet grid by choosing an existing layer or specifying coordinates manually.
      iii. ***Number of Rows and Columns:*** Specify the number of rows and columns you want for your grid.
      iv. ***Cell Size Width and Height:*** Alternatively, you can specify the cell size for the grid.
      v. ***Geometry Type:*** Set this to "**POLYGON**".
   d. Run the Fishnet Tool:
      i. Click "**Run**" to create the fishnet grid shapefile.

3. Generate and Index the Grid with "Grid Index Features".
   a. Open the "Grid Index Features" Tools:
      i. Go to the "**Analysis**" tab and click on "**Tools**".
      ii. Search for "**Grid Index Features**" in the Geoprocessing pane and open the tool.
   b. Configure the "Grid Index Features" Tool:
      i. ***Input Features:*** Select the fishnet grid shapefile created in **Step 1**.
      ii. ***Output Feature Class:*** Choose a location and name for the output grid index shapefile.
      iii. ***Use Page Numbering:*** Check this option to automatically generate page names like A1, A2, etc.
      iv. ***Number Rows and Columns:*** Specify the number of rows and columns for the grid index, matching those of your fishnet.
      v. ***Generate Polygon Grid that intersects input feature layers or dataset:*** Ensure this option is checked. This ensures that the grid cells are only created where they intersect the input features.
      vi. ***Reference System:*** Choose the coordinate system that matches your input data.
      vii. ***Optional Fields:*** Configure additional parameters such as rotation angle, and polygon width/height, if needed.
   c. Run the Tool:
      i. Click "**Run**" to generate the grid index features.
   d. Verify and Customize Index Labels:

i. Open the attribute table of the generated grid shapefile.
ii. Verify that the "**PageName**" field contains the grid index labels (e.g., A1, A2, etc).
iii. If further customization is required, you can use the Field Calculator to adjust the labels as needed.

4. *Example Configuration:*
    a. Create Fishnet Grid
        i. ***Output Feature Class:*** '*C:\YourProject\Fishnet.shp*'
        ii. ***Template Extent:*** Define manually or select from an existing layer.
        iii. ***Number of Rows and Columns:*** For example, 10 rows and 10 columns.
        iv. ***Cell Size Width and Height:*** Alternatively, specify cell size.
        v. ***Geometry Type:*** Set to "**POLYGON**".
        vi. Click "**Run**".
    b. Generate and Index Grid
        i. ***Input Features:*** Select '*C:\YourProject\Fishnet.shp*'.
        ii. ***Output Feature Class:*** '*C:\YourProject\GridIndex.shp*'.
        iii. ***Use Page Numbering:*** Checked.
        iv. ***Number Rows and Columns:*** Specify the same number of rows and columns as the fishnet.
        v. ***Generate Polygon Grid that intersects input feature layers or datasets:*** Checked.
        vi. ***Reference System:*** Choose the appropriate coordinate system.
        vii. ***Additional Options:*** Set any additional parameters if necessary.
        viii. Click "**Run**".

    By following these steps, you will create a fishnet grid and then generate a grid index with labeled cells, ready for use in mapping and analysis.

5. To organize the data extracted from the rasters the **auto_zonal_stats** program is employed. The **auto_zonal_stats** program allows users to resample files, enabling raster file size conversion. The *Auto Zonal Statistics* program calculates statistics on raster values within zones of another dataset. The tool generates a raster output list, but only one statistic is calculated.

6. Use the ArcGIS function, Aggregate points, **in the "Aggregate_MultiPoints" notebook,** where the point input is the point-shape file from the previous step, and the grid input is the grid-shape file generated in ***step 3***.
    a. Then, select all the columns that need to be aggregated and select the 'Mean' option for all the columns selected. The output is a raster with an attribute table with all data needed. Also, ensure that the 'PageName' column is included when exporting the table.

Note: The next steps do not require Arcpy or ArcGIS.

1. The static variables should be combined in one CSV file and the *index variable* column. In the GitHub directory, a script called ***'static_var_gen'*** trims off all the non-static variable columns. The ***' static_gen'*** function is used to find the recent data for the date and variables needed. The ***static_gen*** function can be found in the ***primary_dataset_gen*** notebook under the ***Dataset_Synthesis*** folder in the GitHub directory.

2. To create the ***primary dataset*** for generating test and train sets, divide all datasets into individual CSV files by date. Each CSV file should include the data column and the PageName column. For the **'primary_dataset_gen'** to function properly, all CSV files, including the static CSV file, should be placed in the same folder.
    a. Verification data, such as MOISST AirMOSS, __ init __ , Insitu, or NDVI data, should have the data names "verify" and "verify2"**.**
3. If partial verification data exists, it will move all rows with verification data into the testing set. If there is no verification or complete verification data, the dataset will be divided up randomly.