**Question:**

Compare the 2 runs and see how different they are. I'm not sure the most efficient way to do it, but give it some thought. We want to get an initial sense of how consistent the tool is at finding the solutions.

**Steps performed:**

A summary of the steps performed in the scripts for comparing the two runs:

1. Loading and Normalizing Data:

- Both CSV files were loaded into separate dataframes (solutions\_run\_1 and solutions\_run\_2).

- The Technologies and Description fields were normalized by converting the text to lowercase and removing any extra spaces. This ensures consistency for comparison and matching.

solutions\_run\_1['Technologies'] = solutions\_run\_1['Technologies'].str.lower().str.strip()

solutions\_run\_2['Technologies'] = solutions\_run\_2['Technologies'].str.lower().str.strip()

solutions\_run\_1['Description'] = solutions\_run\_1['Description'].str.lower().str.strip()

solutions\_run\_2['Description'] = solutions\_run\_2['Description'].str.lower().str.strip()

2. Identifying Unique Solutions:

- Solutions unique to each run were identified by comparing the "Solution Name" field. Any solution present in one run but not the other was considered unique.

- Two separate dataframes (unique\_to\_run\_1 and unique\_to\_run\_2) were created to store these unique solutions.

unique\_to\_run\_1 = solutions\_run\_1[~solutions\_run\_1['Solution Name'].isin(solutions\_run\_2['Solution Name'])]

unique\_to\_run\_2 = solutions\_run\_2[~solutions\_run\_2['Solution Name'].isin(solutions\_run\_1['Solution Name'])]

3. Preparing Comparisons of Unique Entries:

- For each set of unique solutions, only the Solution Name, Technologies, and Description columns were kept for comparison.

- These were stored in new dataframes (unique\_to\_run\_1\_comparison and unique\_to\_run\_2\_comparison).

unique\_to\_run\_1\_comparison = unique\_to\_run\_1[['Solution Name', 'Technologies', 'Description']]

unique\_to\_run\_2\_comparison = unique\_to\_run\_2[['Solution Name', 'Technologies', 'Description']]

4. Saving Results to CSV Files:

- The results (unique solutions from each run) were saved to two CSV files for further review.

- File paths were generated for each unique set.

unique\_to\_run\_1\_comparison.to\_csv('/mnt/data/unique\_to\_run\_1\_comparison.csv', index=False)

unique\_to\_run\_2\_comparison.to\_csv('/mnt/data/unique\_to\_run\_2\_comparison.csv', index=False)

Resulting Output:

- The files unique\_to\_run\_1\_comparison.csv and unique\_to\_run\_2\_comparison.csv now contain the unique solutions identified in each run, along with the associated technologies and descriptions.

1. Category Differences:

* **Run 1**: Categories are often more specific (e.g., "Water-Related Issues," "Geological Hazards").
* **Run 2**: Some categories combine multiple issues (e.g., "Water-Related Issues; Extreme Weather Events"), indicating broader groupings.

2. Solution Name Variability:

* Some solutions appear in both runs but with slight variations in the names. For example:
  + "Wildfire Risk Assessment Tools" in Run 1 versus "Fire Weather Monitoring System" in Run 2.
* Different names are used for solutions that may refer to similar ideas, leading to potential inconsistencies.

3. Technologies Field:

* Both runs reference similar technologies like machine learning, remote sensing, and satellite imagery.
* In Run 2, the "Technologies" field sometimes lists more technologies together, whereas Run 1 might focus on more specific solutions (e.g., "Big data analytics" in Run 1 vs. "Data analytics systems" in Run 2).

4. Description Differences:

* Descriptions are often similar in content but vary in phrasing. For example:
  + Run 1 might state, "Utilizing satellite imagery and data analytics" while Run 2 says "Satellite imagery and predictive analytics."
* These differences are not always significant in meaning, but they do introduce variability in extraction.

5. Unique Solutions:

* Each run contains unique solutions not found in the other, such as:
  + **Run 1**: "Snow and Ice Management Technologies" and "Community Engagement Platforms."
  + **Run 2**: "Sustainable Urban Mobility Solutions" and "Landslide Monitoring System."
* The unique solutions suggest that the extraction tool captures different nuances in each run.

In summary, while many solutions overlap, there are differences in naming conventions, categorization, and phrasing. These differences indicate some variability in how the tool identifies and labels climate adaptation solutions across runs. This could be refined by harmonizing naming conventions and categories to reduce discrepancies.