Week 1- 2 updates:

I made some updates and all files are in this folder: <https://drive.google.com/file/d/1k3geSVTHNF1MrLMGX1be1tUglWtOgWJ2/view?usp=sharing>

For question 1, I used Natural Language Processing (TF-IDF) to compare solution names across two files (compared whether solutions match based on names and description). I updated the [python codebook](https://drive.google.com/file/d/1k3geSVTHNF1MrLMGX1be1tUglWtOgWJ2/view?usp=sharing) here and here's the [updated question 1](https://docs.google.com/document/d/1T42BOz5qUnwfDfyRUSDNPuc2Tn5rYQEUOzZMjqMZJWg/edit?tab=t.0) results doc.

Also, for question 2 - previously, I searched for unique solutions in the Toronto plan ( it seems I should have searched for all solutions in the plan i.e. check if they appear in the plan or not).

I invited you to the Zotero project (attached screenshot). I did a manual query search (checked if a solution appears in the plan or not) - I searched for every solution from the first run in the Toronto plan, annotated the solutions in the toronto pdf file and tagged them with the Solution name (zotero allows annotating pdf and assigning a tag).

Here is the [spreadsheet with solutions](https://docs.google.com/spreadsheets/d/17kgqqoUBBffHkB765rq22Vho_XuaTfo_lu-EAYXvBBY/edit?usp=sharing) and assigned rankings and page numbers (if they appear or not in the plan).

I did the first run of searching for solutions using some python script/AI tool, but it was not always assigning pages correctly, so I manually checked the solutions, whether they match and pages are correct.

Here's a [video recording](https://drive.google.com/file/d/1NVDnloxBUxtZ3xuzv3wn9W_dVvCTC08y/view?usp=sharing) of the updates in case it helps, thanks a lot!

Todo:

Week 3:

1. Splitting descriptions into ngrams - remove excessive ones (not important) - clean up the descriptions
2. Do topic modeling on the remaining descriptions

TODO:

Batch process

Topic modeling - 2 approaches - one raw (from toronto plan) and another descriptions -

Write down short-term and long-term goals

Create web page wih visualizations of topic modeling - ipynb

Visualization - dot for cities and

Github: create repo

Technology enabled climate solutions, cities

Raw text of solutions

Diagrams with tfidf - urban heat, urban flooding, etc, – clustering- problems and solutions