Required Software:

- Visual Studio Code (recommended) or any other software that can run Python and the Pandas library.

Pipeline:

(*Ensure that all three Python files are in the same folder*)

1. **Download the Python files**
   You can download the necessary Python files from my GitHub repository or use Git control flow:
   [GitHub Repository Link]
2. **Building a Single Audience Group**
   - Run the `single_model.py` file:
     - Update the `file_path` variable to point to the dataset you need to analyze.
     - Locate the following line:
       ```
       filtered_df, total_count, default_threshold] =
       building_audience(df,
       "amazonmakesworldbetter_tunnl_7_24",
       "amazonmakesworldworse_tunnl_7_24")
       ```
     - Edit the model names to those you intend to use.
     - You can modify the default settings (`Voter = True`, `default_threshold = None`, `Joint = False`) by adding corresponding values within the parentheses.
   - After running the entire file, you will obtain a result called `combined_dataframe.xlsx`. You can rename this file in the final code block if desired.
3. **Building Multiple Audience Groups**
   - Run the `Outcome.py` file:
     - Update the `name` variable with the audience names.
     - Input the `model_names` pairs that you would like to use for building the audience.
     - Update the `file_path` variable.
     - You can choose to revise the default patterns (these patterns represent the common patterns in the model names).
   - All the exported Excel files will be saved in the folder.