GitHub Repo Instructions for Team 5, DonorsChoose Domain

Arthur Spirou (aspirou), Akin Joseph (ababujos), & Nikki Chen (yiqichen)
Github Repo Link: https://github.com/aspirou55/DonorsChoose

Directions For Running Notebooks and Importing Libraries:

First, make sure you have the following packages installed:

- numpy
- pandas
- matplotlib
- seaborn
- scikit-learn
- xgboost
- Ipython
- imblearn

If necessary, these can be installed by running the following command in your terminal:

- % pip install numpy pandas matplotlib seaborn scikit-learn xgboost ipython

Next, download Jupyter Notebooks and the data file TIDY.csv

- Save all notebooks and the data files (TIDY.csv, and flagged_projects_top10_xgboost.csv) in the same local directory where you will run them from
    - Notebooks are the following .ipynb files, and can be downloaded directly from the GitHub repo (link above):
        - ClassificationModels_DonorsChoose.ipynb
        - RatioModels_DonorsChoose.ipynb
        - DonorsChoose_EDA_final.ipynb
        - Bias_report.ipynb

- TIDY.csv can be downloaded from the following public link (also found in README file):
    - https://drive.google.com/file/d/1WihsvXdqblYZmjeb7BU2wGsyFOU172tP/view?usp=sharing
- Likewise, flagged_projects_top10_xgboost.csv can also be found either in the README file or here: flagged_projects_top10_xgboost.csv

You are now ready to run all notebooks in any order you choose, as they each only require the TIDY.csv dataframe, or the flagged_projects_top10_xgboost.csv dataframe (in the case of Bias_report.ipynb) in order to operate.

- EDANotebook.ipynb contains the analysis and results of our exploratory data analysis
- ClassificationModels_DonorsChoose.ipynb contains the classification task analysis
- RatioModels_DonorsChoose.ipynb contains the alternative regression task analysis explored
- Bias_report.ipynb contains the bias audit analysis of projects flagged by our model at top 10% precision