**Description**

CharityML is a fictitious charity organization located in the heart of Silicon Valley that was established to provide financial support for people eager to learn machine learning. After nearly 32,000 letters were sent to people in the community, CharityML determined that every donation they received came from someone that was making more than $50,000 annually. To expand their potential donor base, CharityML has decided to send letters to residents of California, but to only those most likely to donate to the charity. With nearly *15 million* working Californians, CharityML has brought you on board to help build an algorithm to best identify potential donors and reduce overhead cost of sending mail. Your goal will be evaluate and optimize several different supervised learners to determine which algorithm will provide the highest donation yield while also reducing the total number of letters being sent.

**Software and Libraries**

This project uses the following software and Python libraries:

* [**Python**](https://www.python.org/downloads/release/python-364/)
* [**NumPy**](http://www.numpy.org/)
* [**pandas**](http://pandas.pydata.org/)
* [**scikit-learn**](http://scikit-learn.org/0.17/install.html) (v0.17)
* [**Matplotlib**](http://matplotlib.org/)

You will also need to have software installed to run and execute a **[Jupyter Notebook](http://ipython.org/notebook.html" \t "_blank)**.

If you do not have Python installed yet, it is highly recommended that you install the [**Anaconda**](http://continuum.io/downloads) distribution of Python, which already has the above packages and more included.

**Data**

The modified census dataset consists of approximately 32,000 data points, with each datapoint having 13 features. This dataset is a modified version of the dataset published in the paper *"Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid",* by Ron Kohavi. You may find this paper [online](https://www.aaai.org/Papers/KDD/1996/KDD96-033.pdf), with the original dataset hosted on [UCI](https://archive.ics.uci.edu/ml/datasets/Census+Income).

**Features**

* age: Age
* workclass: Working Class (Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked)
* education\_level: Level of Education (Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool)
* education-num: Number of educational years completed
* marital-status: Marital status (Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse)
* occupation: Work Occupation (Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces)
* relationship: Relationship Status (Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried)
* race: Race (White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black)
* sex: Sex (Female, Male)
* capital-gain: Monetary Capital Gains
* capital-loss: Monetary Capital Losses
* hours-per-week: Average Hours Per Week Worked
* native-country: Native Country (United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands)

**Target Variable**

* income: Income Class (<=50K, >50K)

**Starting the Project**

For this assignment, you can find the p1\_charityml folder containing the necessary project files on the [**Data Scientist GitHub**](https://github.com/udacity/DSND_Term1), under the projects folder. This project contains three files that you will need to use:

* finding\_donors.ipynb: This is the main file where you will be performing your work on the project.
* census.csv: The project dataset. You'll load this data in the notebook.
* visuals.py: This Python script provides supplementary visualizations for the project. Do not modify.

In the Terminal or Command Prompt, navigate to the folder on your machine where you've put the project files, and then use the command jupyter notebook finding\_donors.ipynb to open up a browser window or tab to work with your notebook. Alternatively, you can use the command jupyter notebook or ipython notebook and navigate to the notebook file in the browser window that opens. Follow the instructions in the notebook and answer each question presented to successfully complete the project. A **README** file has also been provided with the project files which may contain additional necessary information or instruction for the project.

**Submitting the Project**

**Evaluation**

Your project will be reviewed by a Udacity reviewer against the [**Finding Donors for CharityML project rubric**](https://review.udacity.com/#!/rubrics/406/view). Be sure to review this rubric thoroughly and self-evaluate your project before submission. All criteria found in the rubric must be *meeting specifications* for you to pass.

**Submission Files**

Following files would be needed for evaluation:

* The finding\_donors.ipynb notebook file with all questions answered and all code cells executed and displaying output.
* An **HTML** export of the project notebook with the name **report.html**. This file *must* be present for your project to be evaluated.

When you are ready to submit your project, There are three ways in which your project can be submitted for evaluation.

1. If you ran the notebook from your **local machine** collect the above files and compress them into a single archive for upload.
2. You could supply the above files on your **GitHub Repo** in a folder named finding\_donors for ease of access. This would build a good Github profile in parallel.
3. If you worked using the **workspace inside the classroom** you can submit your project directly for review using the submit button at the end of project, just make sure you download the HTML report to local machine and upload it back into workspace BEFORE submitting your report.