

## Contributions

List of data wrangling/analysis/visualization tasks accomplished:

Daniel:

Scraping/cleaning/merging separated data from primary CSVs

Geolocation (extremely long processing time 2-6 requests/second for about 24 hours)

Matplotlib mapping

Fuzzy Matching (where most of the misnames in final viz happen)

Bokeh mapping/plotting/interactions (extremely time consuming to figure out)

Hoa:

Scraped wikipedia for links and images

scraped google for images

cleaned wikipedia data

visualized data from wikipedia

IPYNB descriptions:

Daniel: daniel\_luft-martinez\_

1 - a first look at the primary data source

2 - color mapping of registered address country codes

3 - geolocation for addresses

4,5 - bokeh testing ground

6 - static pin map of all addresses in primary data set, fuzzy matching to Hoa data, and the final Bokeh interactive map of the "interesting" people that were fuzzy matched

7 - Final presentation ipynb, all visualizations, pull from joblib files to reduce runtime when possible

Hoa: hoa\_nguyen\_

1 - "" scrapes wikipedia for panama and paradise entities and extracts their website link and image, stores data in a json file, contains some visualizations""