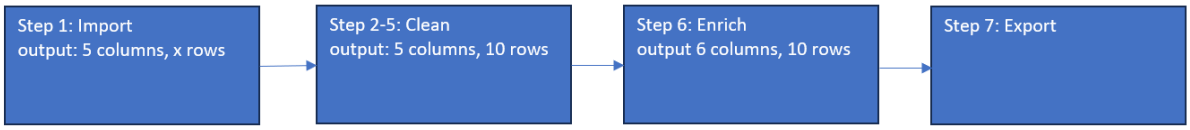
Earthquake Pipeline Luigi



Step 1: Import data from csv

Import the data from the database.csv file. Only use the Data, Time, Latitude, Longitude and Magnitude columns.

Use the pandas read_csv, to_csv functions. To select specific columns, you can check https://www.datacamp.com/tutorial/python-select-columns

Step 2: Clean(Task)

In this clean step we want to get the 10 earthquakes with the highest magnitude from the dataset. You can use sort_values() and head()

Step 3: Enrich(Taks)

```
def getCountry(lati, longi):
    url = 'http://api.geonames.org/countryCodeJSON'

    params = dict( formatted='true',
    lat = lati,
    lng = longi,
    username = 'pbtraining',
    style = 'full')

    resp = requests.get(url=url, params=params)

    if not resp.text.startswith('{"status"'):
```

```
        jsonresp = resp.json()
        return jsonresp['countryName']
    else:
        return 'None'
```

Use the GetCountry() function to get the information from the web API to convert longitude, latitude to a country name.

You can use the following code to add the values to the dataframe:

```
df['Country'] = df.apply(lambda x: getCountry(x['Latitude'], x['Longitude']), axis=1)
```

Step 4: Export(Task)

In this step we print the data, normally here we would export the data to a usable format for further steps.