# ETL Application Report

### **By David, Jean and Elvis.**

**Suicide Countries (WHO Statistics) and**  [**Wealth – GDP (WBG data) 1995 - 2014.**](https://datawrapper.dwcdn.net/vXnEw/4/)

**Extraction:**

* Extracted data from two different csv files (sources – last page of this document) using pandas read\_csv
  + - [**https://www.kaggle.com/szamil/who-suicide-statistics/download**](https://www.kaggle.com/szamil/who-suicide-statistics/download)**.**
    - <https://datacatalog.worldbank.org/dataset/wealth-accounting/resource/f3a8d0aa-621c-43a1-a59d-58365a62ae51>

**Transformation:**

***Suicide Data***

* Define and extract columns from suicide data file relevant to our project:  Country, Year, and Suicide Count.
* Drop rows with null values using pandas ‘dropna’.
* Used ‘loc’ method to extract relevant years: 1995, 2000, 2005, 2010, 2014.
* Calculate sum of suicides by country using pandas sum aggregate and groupby functions to list by country.
* Panda merge results from all years to one table.
* Use pandas ‘melt’ method to unpivot table and normalize data.

**GDP Data**

* Used columns from GDP data (Wealth) for all countries with data provided
* Use loc to extract (Wealth) – GDP as indicator from world bank data (open source)
* Used Python/Pandas to extract and transform the data to clean and turn into a usable source to determine if there was a correlation. For example, most of the data from the GDP report was turned into a pivot table and normalize data for use with suicide data.

**Loading**

* Create database connection to postgres using SQLAlchemy.
* 1) We created a database “ETL”
* 2) Following syntax was used to establish the connection between the cleaned file to be loaded on SQL.
* connection\_string = "postgres:postgres@localhost:5432/ETL"
* # engine = create\_engine(f'postgresql://{connection\_string}')
* 3) clean tables from the SQL database and uploaded to github.

# Resources Used:

**Suicide Statistics by Country Source:**

<# https://www.kaggle.com/szamil/who-suicide-statistics/download>

**Suicide Statistics by Country Description:**

Basic aggregate numbers covering 1996 - 2014, by country, year, age groups and sex. There is only one file, with only a few columns. I made this file using the [WHO Mortality Database online tool](http://apps.who.int/healthinfo/statistics/mortality/whodpms/).

## **Wealth – GDP**

<https://datacatalog.worldbank.org/dataset/wealth-accounting/resource/f3a8d0aa-621c-43a1-a59d-58365a62ae51>

<https://datacatalog.worldbank.org/total-wealth-capita-constant-2014-us>

Total wealth is calculated as the sum of produced capital, natural capital, human capital, and net foreign assets. Values are measured at market exchange rates in constant 2014 US dollars, using a country-specific GDP deflator.