**Jessica Allred, Preethika Gajendran**

Data Analytics Bootcamp

May 18, 2020

**ETL Project – World Happiness and Suicide Statistics**

The aim of this project is to successfully extract the required data from the datasets, transform them to understandable form and load it in to a database, thereby performing a ETL operation.

**Data sets used -**

World Suicide Statistics - WHO (<https://www.kaggle.com/szamil/who-suicide-statistics>)

World Happiness Report - Gallup World Poll (<https://www.kaggle.com/unsdsn/world-happiness>)

**Extraction –**

Read the csv files of the two data sets (World suicide statistics,World Happiness Report)

In to dataframes using pandas.

**Data Cleaning –**

1. Initial inspection of data to understand the dataset and to identify columns needed
2. Renaming of columns to match database table design
3. Replacing/Removing Null values in the dataset
4. Grouping/Filtering of data as required
5. Merging of data frames

**Loading –**

Export the cleaned data in to a PostgreSQL database using SQL Alchemy.