**ETL Project – Happiness & Life Expectancy**

Group members: Jasjeet Kaur, Saleha Khatun, Erika Hoshino,

**Project Proposal**

Extract 2015 - 2021 World Happiness data from Kaggle.com, Transform csv files into two clean data-frames, Load data-frames directly from pandas into PostgreSQL

**Tools Used:**

* Pandas
* PostgreSQL
* SQL Alchemy

**Extract**

Imported 2015 -2020 world happiness csv files from Kaggle.com into Pandas.

**Transforms**

* Dropped unwanted columns, filtered for:
  + Country Name
  + Happiness Rank
  + Happiness Score
  + Healthy Life Expectancy
  + Generosity
  + Trust
* Renamed columns
* Rounded Numerical values to (2) decimal places
* Merged data frames based on country

**Load**

**PostgreSQL**

1. Created world happiness db in PostgreSQL with seperate tables.
2. Imported data into tables directly from Pandas using SQL Alchemy engine.
3. Used Pymongo to create a connection from Pandas to MongoDB local host.

**Insights**

With the kind of the data-set we fond relational, so SQL best suitable to store our data.