TECHNICAL REPORT

|  |  |
| --- | --- |
|  |  |
| Data Visualization | Project 2 |
| Team: | Zac Cheatle  Adam McVey  Amr Alwakeal  Heather Wright  Shelly Fischer |
| Project Name: | World Happiness 2015 v. 2020 |
| Project Proposal: | Engineer SQL database from Pandas dataframe. Survey countries based on their subjective (well-being) happiness levels |
| Data Sourced | csv files from kaggle: https://www.kaggle.com/mathurinache/world-happiness-report?select=2015.csv (and 2020). |
| Data Clean-up & Analysis: | Discrepancies, in the form of column names, appeared in both datasets used. Transformation entailed the following:  **Extract:**  ***2015 World Happiness CSV:***   * Clearly defined Happiness Index, inclusive of Rank & Score * Representation of all country participants present     ***2020 World Happiness CSV*:**   * Happiness Score for 2020 show dissimilarities whereas it was determined the ‘Ladder score’ equates to Happiness Score * Additionally, this most recent data set provided additional statistical values vs. the 2015 dataset     *NOTE:* *There is no happiness rank for 2020 yet as the data is still being collected. There are also some countries that don’t have data for both years. Considering this we decided not to join tables later on in SQL.*  **Transform**:   * Dropped unwanted columns from 2015 & 2020, filtered for:   + Country Name   + Ladder Score   + Happiness Score   + Healthy Life Expectancy   + Generosity   + Perceptions of Corruptions   + Trust * Renamed columns in 2015 * Renamed columns in 2020 * Merged data frames based on country * Rounded Numerical values to (2) decimal places   **Load**:   * Loaded final World Happiness tables 2015 & 2020 data frames into Postgre SQL (relational) * Confirmed data was accurate and present in Postgre & Pandas * SQL (Postgre) was chosen because:   + SQL works better with Pandas data frame   + SQL easier to manage the data base, less coding involved   + SQL works well for smaller datasets   + Portable, can be used on various platform   + SQL supports foreign key   + SQL is Row and Column based which is more compatible with our data |
|  |  |
|  |  |