# IS-Project-2

## Data sources

- Population Growth by Country

- https://www.amerigeo.org/datasets/02a4e8117bbf40bdba63144b1740e507\_0/explore?location=-2.206385%2C0.000000%2C1.47&showTable=true

- Income by Country

- https://www.kaggle.com/frankmollard/income-by-country

- World Happiness Report

- https://www.kaggle.com/unsdsn/world-happiness

## The type of final production database

- Postgres SQL Database

## Description of expectations

- Correlating Population Growth Rate with GDP per capita and happiness score by country from 2015 to 2019.

- We don't believe that if Population Growth Rate and the GDP per capita has relevant relationship.

- We also believe that happiness has weak correlation to population growth rate.

## ETL Process

\*\*Extract\*\*

File 1.

- Title: 'population\_growth\_rate'

- Formatting: CSV File

- Sheets: 'Population\_Growth\_Rate\_by\_country'

- Sheets that we used: 'Population\_Growth\_Rate\_by\_country'

- Columns: 'OBJECTID', 'type', 'name', 'name\_long', 'abbrev', 'formal\_en', 'region\_wb', 'SqMi', 'POP\_GROWTH\_RT', 'YEAR', 'DATE\_START', 'DATE\_END'

- Columns that we used: 'name\_long' as country name, 'POP\_GROWTH\_RT' as popular growth rate.

File 2.

- Title: 'Income\_by\_Country'

- Formatting: CSV File

- Sheets: 'Income Index', 'Labour share of GDP', 'Gross fixed capital formation', 'GDP total', 'GDP per capita', 'GNI per capita', 'Estimated GNI male', 'Estimated GNI female', 'Domestic credits'

- Sheets that we used: 'GDP per capita'

- Columns: 'Country', '1990', '1995', '2000', '2005', '2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017', '2018', 'info'

- Columns that we used: 'Country' as country name, '2014', '2015', '2016', '2017', '2018' as GDP per capita.

File 3.

- Title: 'world\_happiness\_report\_2015', 'world\_happiness\_report\_2016', 'world\_happiness\_report\_2017', 'world\_happiness\_report\_2018', 'world\_happiness\_report\_2019'

- Formatting: CSV File

- Sheets: 'world\_happiness\_report\_2015', 'world\_happiness\_report\_2016', 'world\_happiness\_report\_2017', 'world\_happiness\_report\_2018', 'world\_happiness\_report\_2019'

- Sheets that we used: 'world\_happiness\_report\_2015', 'world\_happiness\_report\_2016', 'world\_happiness\_report\_2017', 'world\_happiness\_report\_2018', 'world\_happiness\_report\_2019'

- Columns: 'Country', 'Region', 'Happiness Rank', 'Happiness Score', 'Standard Error', 'Economy (GDP per Capita)', 'Family', 'Health (Life Expectancy)', 'Freedom', 'Trust (Government Corruption)', 'Generosity', 'Dystopia Residual'

- Columns that we used: 'Country' as country name, 'Happiness Score' as happiness score

\*\*Transform\*\*

File 1.

- Title: 'population\_growth\_rate'

- Copy three columns 'name\_long', 'POP\_GROWTH\_RT', 'YEAR' to the new data frame

- Rename 'name\_long' to 'country', 'POP\_GROWTH\_RT' to 'PGR', 'YEAR' to 'year'

- Drop duplicates items in the column 'country'

- Set the column 'country' as the index column

File 2.

- Title: 'Income\_by\_Country'

- Copy six columns 'Country', '2014', '2015', '2016', '2017', '2018' to the new data frame

- Rename 'Country' to 'country'

- Drop unnecessary data '..' in columns '2014', '2015', '2016', '2017', '2018'

- Set the column 'country' as the index column

- Unified the data type as float in columns '2014', '2015', '2016', '2017', '2018'

- Calculate the mean of GDP per capita for each country from 2014 to 2018

File 3.

- Title: 'world\_happiness\_report\_2015', 'world\_happiness\_report\_2016', 'world\_happiness\_report\_2017', 'world\_happiness\_report\_2018', 'world\_happiness\_report\_2019'

- Copy two columns 'Country'('Country or region'), 'Happiness Score'('Happiness.Score')('Score') to the new data frame

- Rename 'Country', 'Country or region' to 'country', and 'Happiness Score', 'Happiness.Score', 'Score' to 'happiness\_score',

- Drop duplicates items in the column 'country'

- Set the column 'country' as the index column

- Merge 'happiness\_score' columns from 2015 to 2019 only if each country name is matched

- Calculate the mean of happiness score each columns and put the mean data to the new data frame

- Round up the data in columns 'happiness\_score', 'GDP per capita' and 'PGR' to 2 decimal places.

- Merge three data frame into the new data frame 'GDP\_happiness\_population\_df'

\*\*Load\*\*

- Move the data in the data frame 'GDP\_happiness\_population\_df' to PostgreSQL Database 'countries\_db'

- Table name is 'countries' including the columns 'country', 'happiness\_score', 'GDP\_per\_capita', 'PGR'

- The reason the topic was chosen: To figure out the relevance between Happiness Score and GDP per capita, between Happiness Score and Population Growth Rate, between GDP per capita and Population Growth Rate.

## Findings

- Correlation between Happiness Score and Population Growth Rate is 0.0. There are no relevance between both factors.

Chart, scatter chart

Description automatically generated

- Correlation between Happiness Score and GDP per capita is 0.72. There are meaningful relevance between both factors.

Chart, scatter chart

Description automatically generated

- Correlation between Happiness Score and Population Growth Rate is 0.14. There are lack of relevance between both factors.

Chart, scatter chart

Description automatically generated

- Only the relevance between Happiness Score and GDP per capita are valid.