**Data Retrieval Plan – Data to import for main project DataFrame**

Suggested column names for final DataFrame in Red

**World-happiness-report-2021.csv**

Main Columns to include

* Country name – **Use this is as primary list of countries.** (Rename Country)
* Ladder score (Rename Happiness)

Possible Additional Columns we could consider using

* Regional indicator (Rename Region)
* Logged GDP per capita (compare to world bank data?) (Rename GDP PC)
* Healthy life expectancy (definition?) (Rename Life Expectancy)
* Social Support (keep same name Social Support)
* Freedom to make life choices (Rename Freedom)
* Generosity (keep same name Generosity)
* Perceptions of corruption (Rename Corruption)

**World Bank GDP Data**

* **Question/hypothesis:** Does money make you happy? Are countries with higher GDP more ‘happy’
* Is it adequate to use the
* GDP Total (Rename GDP Total)
  + Download dataset from: <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?most_recent_year_desc=true>
  + Use 2019 data (most recent)
* GDP Per Capita (Rename GDP PC)
  + <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?most_recent_year_desc=true>
  + Use 2019 data (most recent)
* Do we want to look at average weekly (or monthly wage) as well?

**Weather**

* **Question/hypothesis:** does better weather make a county’s citizens more happy?
* Need to think about how to compare weather at a country level -
  + The World Bank Climate API looks easier to use than OpenWeatherAPI as it provides averaged values by country (instead of city): <https://datahelpdesk.worldbank.org/knowledgebase/articles/902061-climate-data-api#:~:text=The%20Climate%20Data%20API%20provides,World%20Bank's%20Terms%20of%20Use>
  + Columns to add to dataframe:
    - Annual Average Temperature (Rename Temp)
    - Annual Average Precipitation (Rename Rain)
* **Question/hypothesis:** Does more sunlight make you happy? Are countries with more annual sunlight happier?
  + Average sunlight data pulled from this source (Cities by Sunshine Duration.xlsx link below). Used a pivot table in excel to average the annual sunlight for each city in each country and saved as sunlight.csv (will upload to github): <https://data.world/makeovermonday/2019w44>
  + Column to add to dataframe:
    - Sunlight

**COVID Data**

* **Question/hypothesis:** were countries with less COVID cases/deaths ‘happier’ in 2020/21
* Should compare each country’s *change* in happiness score compared to the previous year’s results, as this will give a fairer assessment of whether COVID has affected ‘happiness’ rather than other, more general, socioeconomic factors
  + Previous years ‘happiness’ score available from **world-happiness-report.csv**
  + Rename 2019 Happiness
  + Change in happiness = (happiness - 2019 happiness) / 2019 happiness \* 100
* Mulesoft COVID data API see: <https://www.mulesoft.com/exchange/68ef9520-24e9-4cf2-b2f5-620025690913/covid19-data-tracking-api/>
* For each country in **World-happiness-report** list:
  + Total COVID cases (Rename COVID cases)
  + Total COVID deaths (Rename COVID deaths)

**Depression Rates**

* **Question:** Are rates of depression higher in less ‘happy’ countries?
* For each country in **World-happiness-report**:
  + Depression rate (% of population) (Rename Depression rate)
    - https://www.who.int/publications/i/item/depression-global-health-estimates
  + Suicide rate
    - (per 100,000 population) (Rename Suicide rate)
    - <https://apps.who.int/gho/data/view.main.MHSUICIDEv>