ETL Project

JungJe Son and McKenna Miranda

World Happiness Report and Country COVID-19 Data

Our goal : To see relation between covid-19 and happiness and draw graphically using gmap

EXTRACTION:

Our data (from Kaggle and all of our files as CSV files):

* Worldometer COVID-19 database
* 2019 World Happiness Report
* OurWorldInData COVID-19 data.

TRANSFORMATION:

What we did:

* We initialized our tables in pgAdmin and completed them using the engine through pandas and SQL queries on Jupyter Notebook.
* We analyzed the relation between COVID data and World Happiness data and visualize it with gmap api.

Issue 1: We had originally found the Worldometer and Happiness Report data from Kaggle. Though after working with the worldometer data and attempting to merge it with the Happiness data, we ran into value differences and had to find more appropriate as well as more recent COVID-19 data.

Issue 2: The complete COVID-19 files were organized by date, beginning in December of 2019 (when the first global case of COVID-19 was recorded). Because of this, there were roughly twenty-three thousand rows of data.

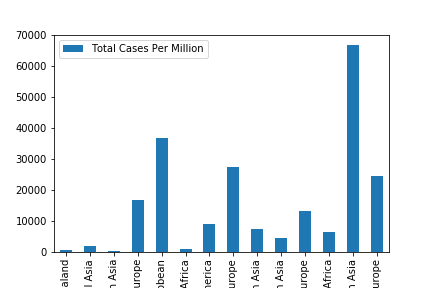
Resolution on issue 1: We found a country code data frame which consisted of known aliases of countries so we could combine the data and use a singular country code to merge the data-frames without error.

Resolution on issue 2: Luckily, the Total Cases per country were recorded daily as a consecutive sum so we took from the absolute latest date (June 14th,2020).

LOAD:

Analysis Process: Once we had finally had a consistent value for our countries, we were able to load the countries through gmaps’ GeoJSON so we could visualize the rankings on a world map. Our completed analysis used pandas, sqlalchemy, matplotlib, and gmaps.

What we analyzed :

* Our final database consisted of our initial columns from both the COVID-19 data and World Happiness report as well as a ranking we created ourselves.
* The Happiness data consisted of a ranking: 1-142 countries ranked by how happy they were found.
* We wanted to emulate this ranking with our COVID-19 data.
* We did this by calculating the rank of each country by its “Total Cases Per Million”.
* We found this to be a representative sample of COVID-19’s effect on the country.
* By doing this we were able to create a rank of each country’s COVID-19 effect on a scale of 0-10; 0 being the least amount of cases per million, 10 being the most amount of cases per million.

A close up of a map

Description automatically generated