MERCURY GROUP, PROJECT 1

CRISTIAN SALGUEIRO, HARSHA MAHAJAN, SAHIL KUMAR

SEPTEMBER 17, 2020

ECE 59500 SOCIAL NETWORKS W/ MACHINE LEARNING

INDIANA UNIVERSITY PURDUE UNIVERSITY INDIANAPOLIS

1. Introduction

The objective of this project is to compile COVID-19 data and geographic data in order to familiarize students with the process of data collection, and manipulation. Datasets containing, COVID-19, tourism, economic, and geographic data were gathered from various international databases. These datasets were manipulated in order to visualize the top 20 countries for a number of categories.

2. Methodology

The "sortcsv" function takes a csv file, sorts on a specified column, truncates the sorted data to a specified number of rows, and outputs the data to a csv file. The output is a csv file whose name is a concatenation of the input parameters. Excel and python were used to visualize the output data.

3. Results and Analysis

The graphs below depict the top 20 countries globally for COVID-19 cases during the week ending on 9/3, tourism, economic, and geographic data. As one would expect, countries with larger populations and higher international travel volumes such as the United States and Indian tend to be represented near the upper end of the active and confirmed cases. Additionally, countries with large land area but not large population such as Chad, Niger, Peru, and Mongolia are not in the top 20 countries for COVID-19 cases likely due to reduced likelihood of contact.

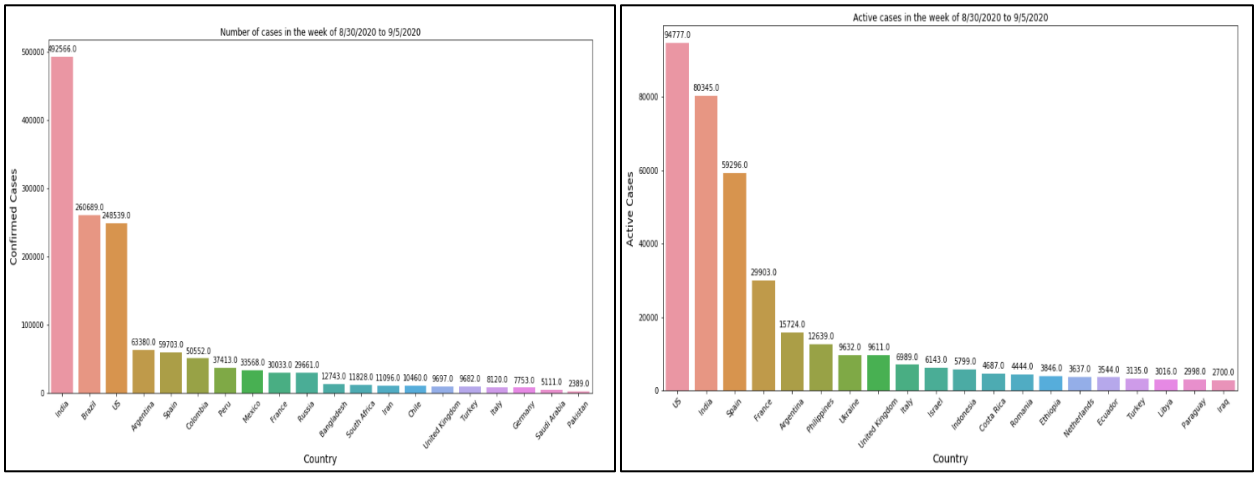

Figure 1. COVID-19 Cases Week Ending On 9/3/2020 - Confirmed (left) Active (right)

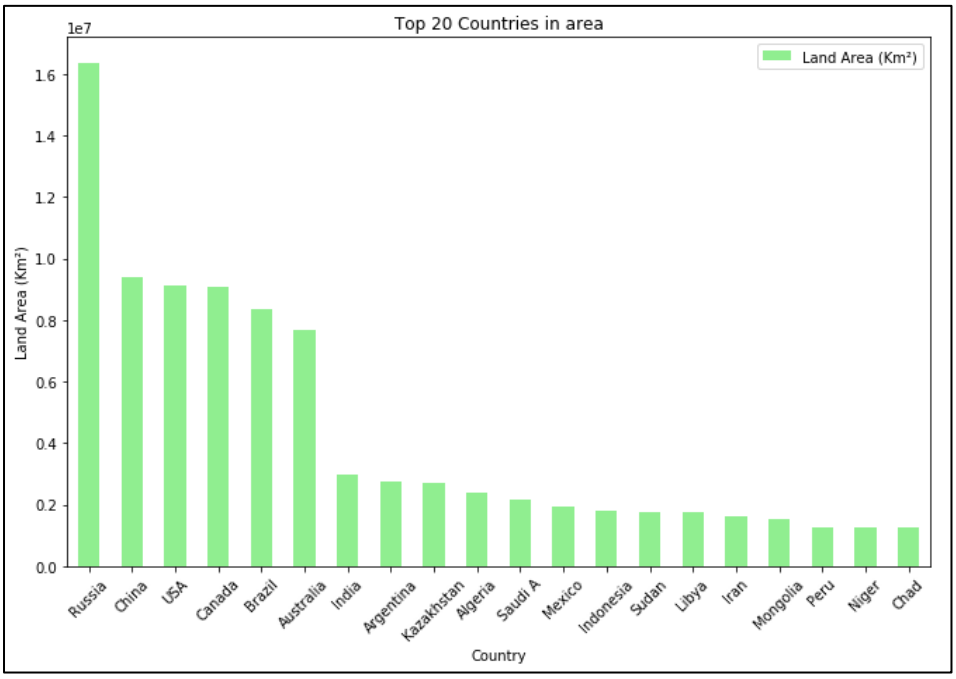

Figure 2 Top 20 countries in the area

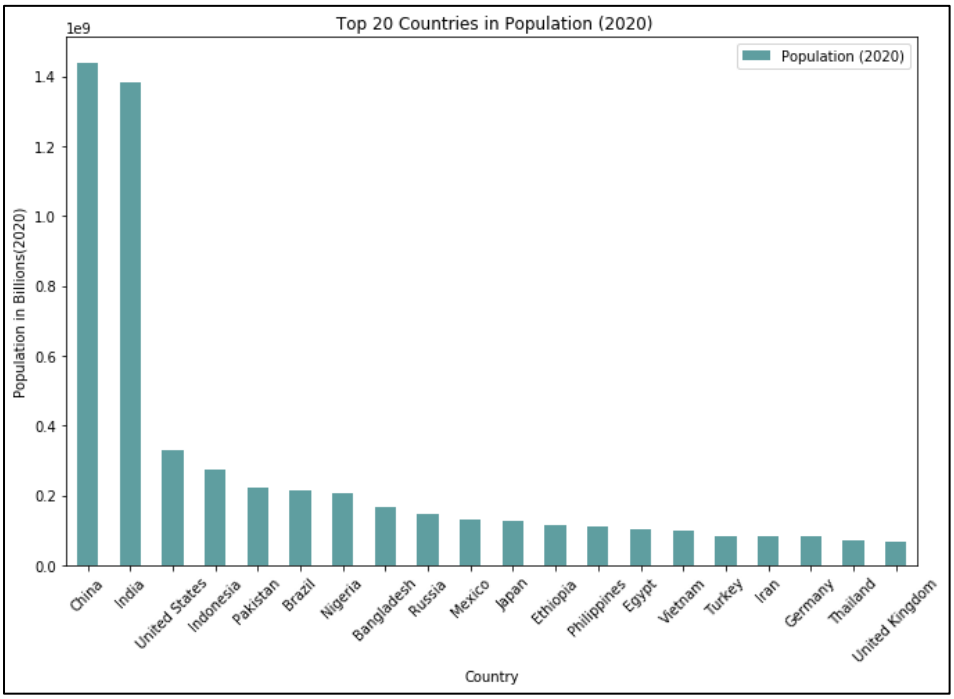

Figure 3 Top 20 Countries in Population

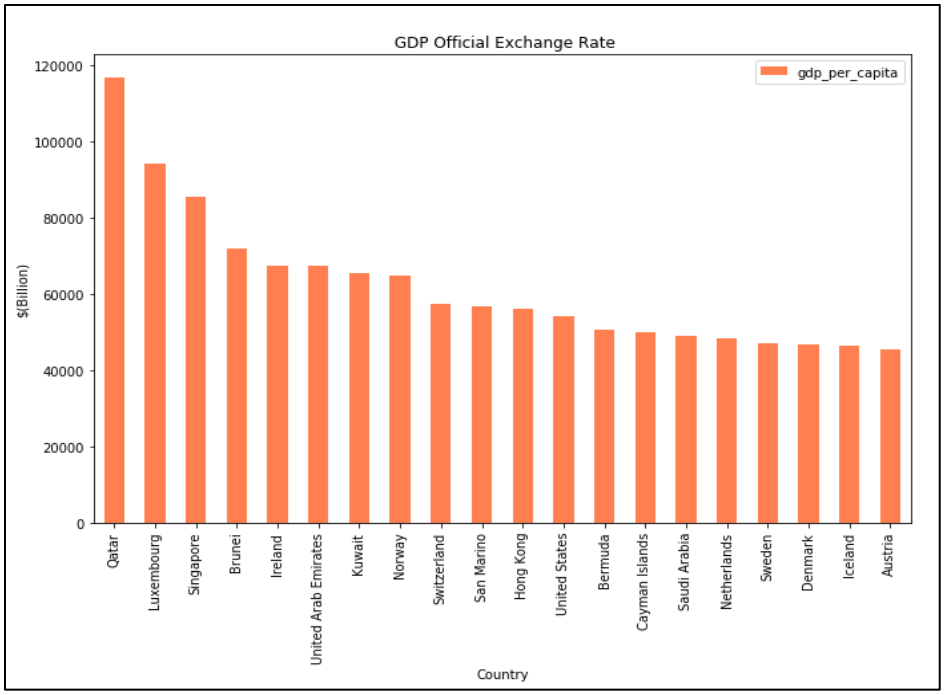

Figure 3 Top 20 countries in per-capita GDP 2019

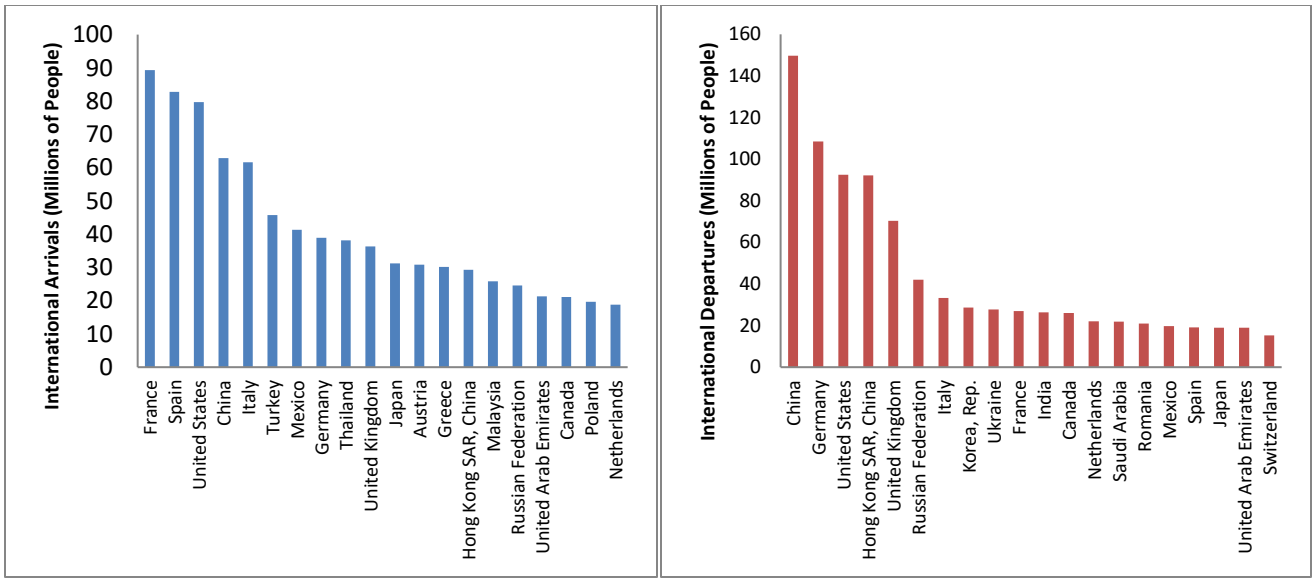

Figure 4 Top 20 International Arrivals and Departures By Country (2018)

References:

1. https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data
2. https://tcdata360.worldbank.org/indicators/ST.INT.DPRT?country=BRA&indicator=1842&viz=line_chart&years=1995,2018
3. https://ourworldindata.org/coronavirus-source-data

Appendix

Source Code:

```
import pandas as pd

def sortcsv(file,column,top_num):
    data = pd.read_csv(file)
    out = data.sort_values(column,axis=0,ascending=False,inplace=False)
    out = out.drop_duplicates("Country")
    outtop = out.head(top_num)
    outtop.to_csv("Top" + " " + str(top_num) + " " + column + ".csv")
```