**Analyze People Migration data**

Human migration is the movement by people from one place to another, particularly different countries, with the intention of settling temporarily or permanently in the new location. It typically involves movements over long distances and from one country or region to another.

Migrations are the essence of peopling the world and they are believed to have begun approximately 2 million years ago with the early expansions out of Africa by Homo erectus.

In more modern times, migrations have often been driven by necessity, that could include forced displacement (in various forms such as deportation, slave trade, trafficking in human beings) and flight (war refugees, ethnic cleansing).

This study is focused on analyzing the data to get insights on the migration dynamics, addressing questions such as:

* What are the major flows of migration (meaning from where to where)?
* Where people from the more developed regions are migrating to?
* Where people from the less developed regions are migrating to?
* What are the dynamics by income and geographic region?
* Drill down in one region of your choice, addressing questions similar to the above.

Using the insights from the analysis of the factors, write a report with your interpretation, along with the details you extracted from the files using your code.

The data for this study (from the UN) can be found on Canvas. The dataset is an Excel spreadsheet with 2 tabs: one with the data, one with the metadata. For your convenience, the data tab is also available as csv file.

The data cover the period 1990-2019, with some years missing. For each year you have similar information: you will calculate the cumulative values to address the questions above.

Because you have data by year, an analysis of the differences over the years it is also possible (not required).

The use of the data provided is mandatory, but feel free to use additional data from different sources (to be attached to your submission).

**Analyze research projects**

The Naval Postgraduate School is a graduate university offering master’s and doctoral degrees in more than 70 fields of study to the U.S. Armed Forces, DOD civilians and international partners.

In 2019, they awarded about $100 million in sponsored research funding.

Research material is stored in a centralized archive maned Calhoun, that is the Naval Postgraduate School's digital repository for research materials and institutional publications created by the NPS community. Materials in Calhoun are openly accessible to anyone on the web and will be preserved for future generations. All the publications are available for search/download.

This study is focused on analyzing the distribution in time of the research activities, addressing questions as:

* What are the key areas/topics in time?
* What are the networks of researchers that are more active?
* What are the institutions that are more active?
* What is the collaboration between institutions?

Using the answers to those questions (and to other you may create), write actionable recommendations for future research activities.

The data for this study can be found at the following URL: https://calhoun.nps.edu

You can use "re- usable links", like the following:

"https://calhoun.nps.edu/handle/10945/16/discover?search-result=true&query=author%3Aberry&current-scope=10945%2F16&rpp=20&sort\_by=dc.date.issued\_dt&order=desc”

and then looping in a predefined list of search keywords (like authors). As a “brute force” alternative, a “wild card” instead of the author (basically an “\*”) can be used, generating a list of ~33K thesis and dissertations. Once the list is extracted as return value in a set of web pages, the content can be scraped using a script. Content would be pdf files.

A relevant subset (10-20) of the total documents would be sufficient for the analysis.

**COVID-19**

Coronavirus disease 2019 (COVID-19) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In the US only there have been more than 30 millions of cases, with more than 550,000 casualties (data Apr. '21).

Explore the datasets to find out how the disease propagated.

There are 4 files to be used:

* *COVID-19\_US-counties.csv*

This is a list of cases and deaths by county and date. County FIPS is a unique identifier for the county. For your reference the full list is in the file *US\_FIPS\_Codes.xls*

* *CO-EST2019-ANNHU.xlsx*

Annual Estimates of Housing Units for Counties in the US. Consider only the 2019 values

* *GDP\_Counties.xlsx*

GDP per each county

* *cc-est2019-alldata.csv* and related description at *cc-est2019-alldata.pdf*

This is a list of county-level information from the Census.

This study is focused on analyzing the files provided to get insights on the propagation of the virus and find out any trends or linkage between factors.

Using the insights from the analysis of the factors, write a report with your interpretation, along with the details you extracted from the files using your code.

The data for this study can be found on Canvas.

The use of the data provided is mandatory, but feel free to use additional data from different sources (to be attached to your submission).