**IE 5374: Foundations of Data analytics - Fall 2022**

**Assignment 1**

1. **Computation of inter-event time (30 points)**

Activities like human communications, neuron spike trains, and seismic signals, consist of high-activity bursty intervals alternating with long low-activity periods. Multiple studies have pointed out that a fat-tailed inter-event time distribution characterizes such bursty behaviors. In the article [**Universal bursty behaviour in human violent conflicts**](https://drive.google.com/file/d/15Jom92SW3TsqZ2vETphHqESAZz46FzLi/view?usp=sharing)**,** researchers find a simple scaling law that governs the inter-event time between two conflicts.

* Read the article titled [**Universal bursty behaviour in human violent conflicts**](https://drive.google.com/file/d/15Jom92SW3TsqZ2vETphHqESAZz46FzLi/view?usp=sharing). The methodology is contained in the Results section.
* Generate a custom function to compute inter-event times
* The function should accept two inputs from the user.1) A list or an array (1 dimension) representing an evenly-spaced time series, 2) A threshold indicating minimum event size.
* The function should return an array representing the inter-event time between successive events that are defined by the threshold.
* Demonstrate that your function works as intended using a threshold of 5 and an array of these events: [10,1,6,4,3,2,7,0,8,8,2,0,7,7,1,0]
* If you are interested in an additional reference on Bursty behavior, you may read [**Universal features of correlated bursty behaviour**](https://drive.google.com/file/d/1pLBa_2DIRhErbrJqvLPRCYjhTgaI39lk/view?usp=sharing) (this is purely optional).

1. **Analyzing the CMS Nursing Home Dataset (70 points)**

The Nursing Home COVID-19 Public File includes data reported by nursing homes to the CDC’s National Healthcare Safety Network (NHSN) Long Term Care Facility (LTCF) COVID-19 Module: Surveillance Reporting Pathways and COVID-19 Vaccinations.

* Use the data provided by [CMS](https://data.cms.gov/Special-Programs-Initiatives-COVID-19-Nursing-Home/COVID-19-Nursing-Home-Dataset/s2uc-8wxp).
* Click on the download button
* Select From = September, 11, 2022
* Select To = September, 11, 2022
* Scroll down the CMS page to obtain the data dictionary for the data. The data dictionary contains details of all the variables included in the dataset.
* *Note: This is going to be a pretty large dataset, don’t be alarmed if it takes some time to load*

**Tasks**

1. Generate a table showing the distribution of the number of nursing home facilities by state and county. (**10 points**)
2. List the top 5 states by the number of nursing home facilities (**10 points**)
3. Conduct a facility analysis on shortage of aides, clinical staff, nursing staff and other staff. Define the following metric for this analysis: **number of weeks the shortage was experienced**. The metric should be categorized into 3 categories, 0-10, 11-20 and >20 weeks. Sample output table is shown below (**20 points**)

|  |  |
| --- | --- |
| **Facility** | N**umber of Shortage weeks** |
| ABC | 0-10 |
| XYZ | 11-20 |
| LMN | >20 |

1. For Middlesex county in Massachusetts compute the hospital bed occupancy in percentage for each facility. How many facilities had 100 % of hospital bed occupancy for 3 weeks or more? (**10 points**)
2. Generate a table showing the distribution of the number of nursing home facilities by division. Use the US Census classification scheme for division as your reference for which states belong to which division: (<https://www.census.gov/geographies/reference-maps/2010/geo/2010-census-regions-and-divisions-of-the-united-states.html>) (**20 points**)

**Submission Format**

1. Submit .ipynb file with the codes and output
2. Name the file as *Assignment1\_FirstName\_LastName.ipynb* (e.g. *Assignment1\_Matthew\_Richards.ipynb)*
3. Upload the file on Canvas
4. Submit the assignment by Monday October 3 at noon (12 PM) PT.