File structure

This file aims to guide the reader through the critical contents of each file, its location on the reservoir and its relationship within the Jupiter notebooks and with the research projects developed.

Folder (main): Mental_health_cdmx_2020

Inside this folder, you will find the following files and folders:

- Readme: Shows a brief explanation of the files with the research projects and articles.
- License: Mentions the License attribution.
- Mental_health_analysis_main_file.ipynb: This file complements the article "Unveiling the 72-hour psychosocial response at first-level mental health support systems during emergency contexts, Mexico City case" submitted to the IJDRR. It explores the effects of different significant events on the mental health response in Mexico City.
- Database_machine_learning_applications.ipynb: Is part of the Data Article "Machine learning-ready mental health datasets for evaluating psychological effects and system needs in Mexico City during the first year of the COVID-19 pandemic" submitted to the Data in Brief journal prepared to contribute to the understanding of the databases and their fitness for machine-learning applications.
- File_structure_and_central_contents: Works a guide to the contents and internal relationships of the files.
- Databases and coding: This folder contains the Datasets and the coding for its analysis which is summarized in the Jupyter notebooks mentioned before.

Folder: Mental_health_cdmx_2020/Databases and coding

In this folder you will find three subfolders:

- Databases: Contains the primary data files used for parsing in both Jupyter notebooks corresponding to the studies mentioned in the Readme file.
- Coding_Mental_health_analysis: Contains the coding files referred in the coding cells inside the Jupyter notebook "Mental_health_analysis_main_file.ipynb"
- Coding_Database_machine_learning_applications: Contains the coding files referred in the coding cells inside the Jupyter notebook "Database_machine_learning_applications.ipynb"

Folder: Mental_health_cdmx_2020/Databases and coding/Databases

Inside this folder you will find the next files:

- Analysis_clean.csv: This file contains 33,234 applications from the online "Questionnaire for the detection of risks to mental health COVID-19". This file provides socioeconomic information as well as health conditions (prevalent and chronic), loss of loved ones, lockdown status, and screen for mental health symptoms assessing acute stress, avoidance and sadness, distancing, and anger, generalized anxiety, and extra factors: Binge alcohol consumption and Experience of abuse. It also contains composite variables referring to fulfilling the criteria for PTSD, depression, and anxiety disorders.
- Codificacion tamizaje.xlsx: Shows the possible responses and coding (in its original language, Spanish) for every question of the "Questionnaire for the detection of risks to mental health COVID-19".
- Translation of questions, answers and encoding.pdf: This file contains the English translation for every question and its possible answers in the questionnaire as well as the column location in the file: Analysis_clean.csv.
- Registros por dia.csv: This file contains the reports from the Lifeline Emergency Phone Call Service, consisting of 349,202 calls, grouped daily between April 24, 2020, and December 31, 2020, consecutively, and then +-72 hour periods surrounding earthquakes with early warnings, in Mexico City during 2021, and 2022 up to 6/04/2023. The vectors within this data set represent the number of calls (enterqueue), connected calls, number of abandoned attempts, average time (in seconds) before quitting an attempt, average waiting time for connection, and average time of completed calls.
- Defunciones cdmx.xlsx: This file contains the epidemiological reports of daily numbers of confirmed COVID-19 cases, suspected cases, and death tolls between April 13, 2020, and December 7, 2020.
- Mobility_google_cdmx_2020.xlsx: This file provides the mobility results via the percent change from baseline for various types of locations in Mexico City between April 1, 2020, and December 31, 2020. These locations include residences, workplaces, transit stations, parks, grocery and pharmacies, and retail and recreation places.
- GT_Busqueda_Web.csv: This file provides the popularity trends, derived from web searches retrieved from Google Trends, covering the period between April 13, 2020, and December 7, 2020. These trends focus on five key concepts: COVID-19, quarantine, epidemiological traffic light (which establishes appropriate health security measures for work, educational

activities, and public spaces), pandemic, and mental health. Popularity is measured daily on a 0-100 scale relative to its maximum value during the specified period.

- GT_Noticias.csv: This file provides the popularity trends, derived from news appearances retrieved from Google Trends covering the period between April 13, 2020, and December 7, 2020. These trends focus on five key concepts: COVID-19, quarantine, epidemiological traffic light (which establishes appropriate health security measures for work, educational activities, and public spaces), pandemic, and mental health. Popularity is measured daily on a 0-100 scale relative to its maximum value during the specified period.

Folder: Mental_health_cdmx_2020/Databases and coding/Coding_Mental_health_analysis

Inside this folder you will find the files corresponding to the different analysis and visualizations developed inside the Jupyter notebook called "Mental_health_analysis_main_file.ipynb":

- 3d_plot.py: Constructs a 3d plotting between abandoned call attempts, call duration and waiting time before connection inside the dataset of the lifeline service records.
- analysis_anova_lifeline.py: Performs One-way ANOVA tests to evaluate if there were statistically significant changes in the timeframes (pre and post) for every vector of the dataset of the lifeline service.
- analysis_n_uses.py: Evaluates the percentual change in the number of uses of the questionnaire on pre and post-time windows performed for all events in the study.
- analysis_percentual_change_means_lifeline.py: Evaluates the percentual change in the number of calls per event over the lifeline dataset.
- anova_2_analysis.py: Performs One-way ANOVA tests on the symptom questions to evaluate if significant changes happened between pre and post-event groups for each selected event in the "Questionnaire for the detection of risks to mental health COVID-19".
- contingency_analysis.py: Performs contingency analysis tests on the symptom questions to evaluate if a significant change in the distribution of the responses was present for any of the selected events.
- correlation_matrix_lifeline.py: Performs a correlation matrix between the Lifeline Service Database Vectors.
- correlations_lifeline.py: Performs a correlation between waiting time before connection and abandoned attempts in the lifeline dataset.
- distribution_anx.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by class of symptom (anxiety).

- distribution_avo.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by class of symptom (avoidance).
- distribution_dis.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by class of symptom (distancing).
- distribution_ll.py: Performs general analysis using histograms of the behavior of the lifeline service regarding the calls received, attended, waiting times and call duration.
- distribution_mix.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by information seeking and somatization.
- distribution_str.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by class of symptom (stress).
- gen_res.py: Performs a plot of the mean of percentual changes of the different events across time.
- lifeline_scatter.py: Performs a scatter plot of the number of calls in the lifeline service.
- modules.py: Contains methods needed for performing certain operations in the coding.
- parsing.py: Performs the parsing of the objects needed for the further operations and analysis of the datasets.
- pca_analysis.py: Performs a principal component analysis on the 23 vectors corresponding to the selected questions of the questionnaire due to their relationship with the PTSD criteria supported in the PCL-5 evaluation tool.
- pca_lifeline.py: Performs a principal component analysis to the lifeline vectors.
- percentual_change_max_correct.py: Delivers the Percentual change of the number of calls per event in the lifeline service.
- tamizaje_hist.py: Delivers a histogram of the usage distribution along time for the “Questionnaire for the detection of risks to mental health COVID-19”.
- Time_ranges.py: Create time slices accordingly to the defined times per event as needed for the study.

Folder: Mental_health_cdmx_2020/Databases and coding/Coding_Database_machine_learning_applications

Inside this folder you will find the files corresponding to the different analysis and visualizations developed inside the Jupyter notebook called “Database_machine_learning_applications.ipynb”:

- ann.py: Performs a classification task using K neighbors classification algorithm to visualize the decision surface for high ptsd risk present when evaluating anxiety and acute stress.
- correlations_lifeline.py: Performs a correlation between waiting time before connection and abandoned attempts in the lifeline dataset.
- distribution_anx.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by class of symptom (anxiety).
- distribution_avo.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by class of symptom (avoidance).
- distribution_dis.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by class of symptom (distancing).
- distribution_ll.py: Performs general analysis using histograms of the behavior of the lifeline service regarding the calls received, attended, waiting times and call duration.
- distribution_mix.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by information seeking and somatization.
- distribution_str.py: Performs general analysis using histograms of the behavior of the type of responses for all items grouped by class of symptom (stress).
- lifeline_scatter.py: Performs a scatter plot of the number of calls in the lifeline service.
- mlp_analysis.py: Deliver the results of four classification tasks running 1000 iterations with random test samples of 0.3 size using MLP algorithm with 10,000 iterations of the neural network to evaluate the performance between maintaining the data as prepared and after some optimization strategies.
- modules.py: Contains methods needed for performing certain operations in the coding.
- parsing.py: Performs the parsing of the objects needed for the further operations and analysis of the datasets.
- pca_analysis.py: Performs a principal component analysis on the 23 vectors corresponding to the selected questions of the questionnaire due to their relationship with the PTSD criteria supported in the PCL-5 evaluation tool.
- pca_lifeline.py: Performs a principal component analysis to the lifeline vectors.
- preprocessing_opt.py: Performs different preprocessing operations on the parsed data (scaling standard, scaling with robust scaling algorithm, and then mapping of the distributions after the different scaling algorithms)
- svm_lifeline.py: Delivers the results between the scatter plot of the number of calls from the lifeline service and the predicted results after training a machine using a support vector machine algorithm.

- tamizaje_hist.py: Delivers a histogram of the usage distribution along time for the “Questionnaire for the detection of risks to mental health COVID-19”.