

Información General:

En el siguiente documento se encuentran las instrucciones para hacer uso de los siguientes ejecutables entregados para la consolidación y normalización de las bases de datos de eventos.

La arquitectura de los archivos sigue el siguiente formato:

```
C:\.
  analisis_y_arreglo.exe
  analisis_y_arreglo.py
  concatenador_de_excel.exe
  concatenador_de_excel.py
  LEEME.docx

  --config_analisis
    administrador_codigos_invalidos.exe
    administrador_codigos_invalidos.py
    administrador_diccionario_codigos.exe
    administrador_diccionario_codigos.py
    administrador_mapeo_nombres.exe
    administrador_mapeo_nombres.py
    codigos_invalidos.py
    cols_arreglo_base.txt
    cols_arreglo_poblacion_riesgo.txt
    diccionario_codigos.py
    errores_cols_salida.py
    Estrato_Municipio.xlsx
    mapeo_nombres.py

  --config_concatenador
    cols_poblacion_riesgo.txt
    cols_quitar.txt
    cols_renombrar.txt
    _logs_columnas.txt
```

Se pueden encontrar tanto los archivos ejecutables (.EXE) como sus códigos fuente (.PY); en caso de alguna modificación, se debe de recompilar el código fuente para poder generar de nuevo el ejecutable. El código fuente se puede correr usando 'python <nombre_del_archivo>.py' en caso de que se tenga Python y las dependencias necesarias. La compilación del código fuente se puede realizar mediante la librería 'auto-py-to-exe' que puede ser descargada para Python.

Los dos ejecutables principales son concatenador_de_excel y analisis_y_arreglo:

- Concatenador_de_excel, posee unos archivos de configuración, los cuales se encuentran en la carpeta config_concatenador. Mediante estos archivos, se puede modificar el funcionamiento del aplicativo.
- Analisis_y_arreglo, también posee archivos de configuración, que se encuentran en la carpeta 'config_analisis'. En esta carpeta se encuentran archivos para el funcionamiento del aplicativo y archivos base, los cuales permiten realizar algunas de las operaciones que realiza el ejecutable.

En la carpeta 'config_analisis' se encuentran los ejecutables que permiten interactuar con los archivos base que poseen información acerca de cómo mapear la información de las bases de datos.

- Administrador_codigos_invalidos
- Administrador_diccionario_codigos
- Administrador_mapeo_nombres

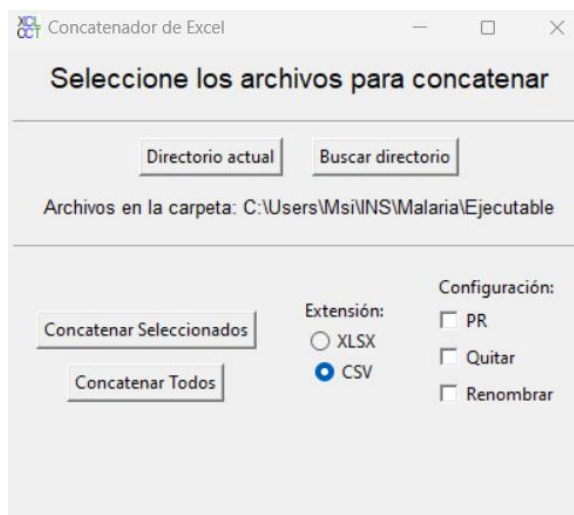
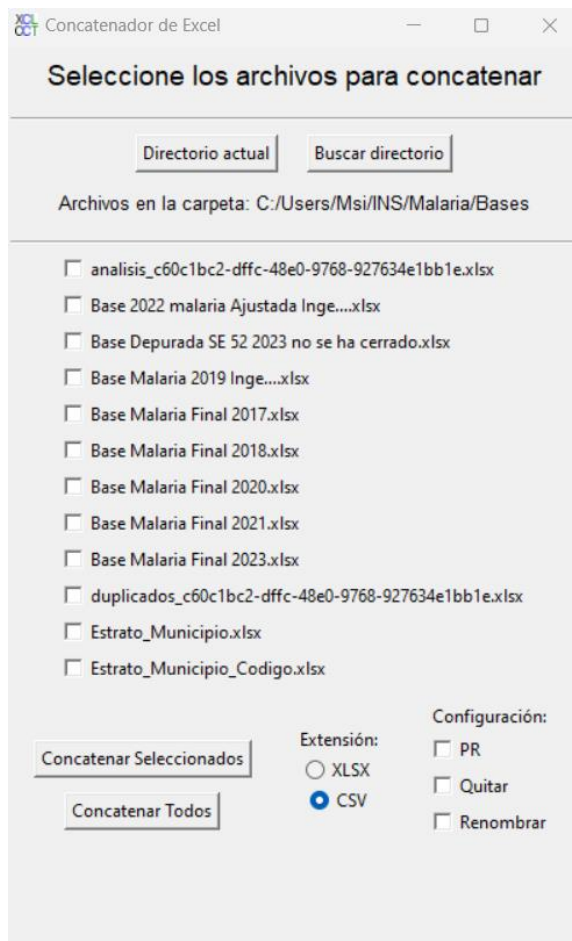
Tabla de Contenido:

Información General:	1
Tabla de Contenido:	2
Ejecutables:	3
Concatenador de Excel:.....	3
Uso:	3
Archivos de Configuración:	4
Salidas:.....	5
Análisis y Arreglo:	6
Uso:	6
Archivos de Configuración:	7
Salidas:.....	9
Administrador Códigos Invalidos:	9
Uso:	9
Administrador Diccionario Códigos.....	10
Uso:	10
Administrador Mapeo Nombres:	11
Uso:	11
Recomendaciones Generales:.....	12

Ejecutables:

Concatenador de Excel:

Uso:



El ejecutable permite seleccionar la carpeta en donde se encuentran las bases de datos. Sea la carpeta actual o una carpeta en otro lugar del disco del computador. Al seleccionar la carpeta, se muestran todos los archivos de Excel en la carpeta seleccionada.

*Se optó por solo archivos de Excel, ya que en caso de archivos de tipo .CSV, pueden tener símbolos de separación diferentes a ','. Por ejemplo: ';' o '|'.

Una vez se tienen los archivos de Excel, se pueden seleccionar usando los Checkbox a la izquierda de cada uno, o usando el botón de 'concatenar todos' el cual toma por defecto todos los archivos de Excel presentes en la carpeta.

El ejecutable permite seleccionar el formato de salida de resultado final en .XLSX o .CSV.

*Se recomienda hacer uso de la salida en .CSV ya que para los computadores es más rápido abrir este tipo de archivos. En un caso posterior, se puede importar la salida a Excel.

Para la configuración, posee una lista de Checkboxes para el funcionamiento del mismo. Este concatenador permite en su versión actual realizar concatenados de archivos de Excel por hojas en formato de bases de datos (fila 1 = títulos) o en formatos de poblaciones de riesgo (títulos en celdas combinadas).

El concatenador junta todas las hojas de los archivos de Excel seleccionados, y une los contenidos usando un set de columnas obtenidas de todas las hojas de los archivos. Esto, con el fin de juntar la información relevante en una sola columna. Esto se realiza consolidando y normalizando los nombres de las columnas y limpiando guiones, comas, apostrofes y otros símbolos que no se encuentren en valores numéricos o en las 26 letras básicas del alfabeto. También reemplaza todos los espacios de la columnas por ' _ '.

Archivos de Configuración:

Este ejecutable posee varios archivos de configuración para su operación, sea para base de datos o poblaciones de riesgo.

El concatenador en su versión actual permite seleccionar 'quitar' y 'renombrar' para el uso de estas opciones para las bases de datos. y 'PR' para las poblaciones de riesgo.

*Al seleccionar 'PR' esta toma prioridad sobre los archivos de 'quitar' y 'renombrar' por las operaciones que puede realizar.

Quitar columnas:

Al seleccionar 'quitar', el concatenador hace uso del archivo 'cols_quitar.txt' que se encuentra en la carpeta 'config_concatenador', el cual posee una lista de columnas en mayúscula, normalizadas (es decir, haciendo uso de solo las 26 letras básicas del alfabeto). El concatenador, al realizar la consolidación de los nombres de las columnas, revisa que el valor consolidado de cada columna no este en el archivo de columnas a quitar. Por lo que estas no se toman en cuenta en el resultado final.

E.j.

nacionalidad_	->	NACIONALIDAD
Complicación	->	COMPLICACION
mues_riñon	->	MUES_RINON

Renombrar columnas:

Al seleccionar 'renombrar', el concatenador hace uso del archivo 'cols_renombrar.txt' que se encuentra en la carpeta 'config_concatenador', el cual posee una lista de ordenes que el concatenador realiza el momento de consolidar y normalizar cada uno de los nombres de las columnas.

E.j.

s NACIONAL	NACIONALIDAD
s CONFIRMADO	CONFIRMADO
= MUESRION	MUESRINON

Cada línea del archivo, posee 3 partes: la orden, la columna que ingresa, y la columna que sale. En el ejemplo anterior, la orden 's' indica que la operación aplica a todas las columnas que comiencen

con la columna que ingresa. La orden '=' indica que la operación solo aplica a las columnas que sea iguales a la columna que ingresa.

Se entiende como:

- Todas las columnas que comiencen con NACIONAL van a ser renombradas a NACIONALIDAD.
- Todas las columnas que comiencen con CONFIRMADO van a ser renombradas a CONFIRMADO.
- Todas las columnas que sea iguales a MUESRION van a ser renombradas a MUESRINON.

PR:

El archivo de configuración de poblaciones de riesgo permite realizar varias operaciones sobre el concatenado de los archivos. Este archivo de configuración tiene prioridad sobre los archivos de 'quitar' y 'renombrar'.

*Se recomienda hacer uso de los archivos 'quitar' y 'renombrar' en conjunto para las bases de datos, y solo el archivo de 'PR' para las poblaciones de riesgo.

E.j.

```
len 5 COD_MUN
z COD_MUN COD_MUN
c TOTAL TOTAL
c CABECERA CABECERA
= RESTO_DE_LA_POBLACION RESTO
c RIESGO,MALARIA POBLACION_RIESGO_MALARIA_LEISHMANIASIS
+ COD_MUN,SHEETNAME ID_POBLACION
```

Estas ordenes para el concatenado de poblaciones de riesgo, se leen como:

- Renombrar a COD_MUN la columna cuyo promedio de caracteres por fila sea más cercano a 5.
- Añadir ceros a la izquierda de la columna COD_MUN para que siempre tenga 5 caracteres y renombrar a COD_MUN.
- Renombrar todas las columnas que contengan la palabra TOTAL a TOTAL.
- Renombrar todas las columnas que contengan la palabra CABECERA a CABECERA.
- Renombrar todas las columnas que sean iguales a RESTO_DE_LA_POBLACION a RESTO.
- Renombrar todas las columnas que contengan las palabras RIESGO y MALARIA a POBLACION_RIESGO_MALARIA_LEISHMANIASIS.
- Añadir la columna ID_POBLACION uniendo las columnas COD_MUN y el nombre de la hoja.

*Al usar el archivo de configuración de PR, solo se guardan las columnas que están en el archivo de configuración.

Salidas:

El ejecutable genera un archivo de '_log_columnas.txt', el cual posee la lista consolidada y normalizada de todas las columnas encontradas en todas las hojas de los archivos de Excel.

Usando este archivo, posteriormente se pueden modificar los archivos de quitar y renombrar columnas para asegurarse que el concatenado de archivos final posea la información necesaria en

las columnas relevantes, sin importar que nuevas columnas sean añadidas en bases de datos posteriores.

El concatenador genera el concatenado en un archivo tomando la extensión deseada por el usuario sea .XLSX o .CSV. El archivo de salida posee un nombre de tipo IDUU (identificador único universal) con el fin de que el concatenado nunca sobrescriba un archivo generado anteriormente.

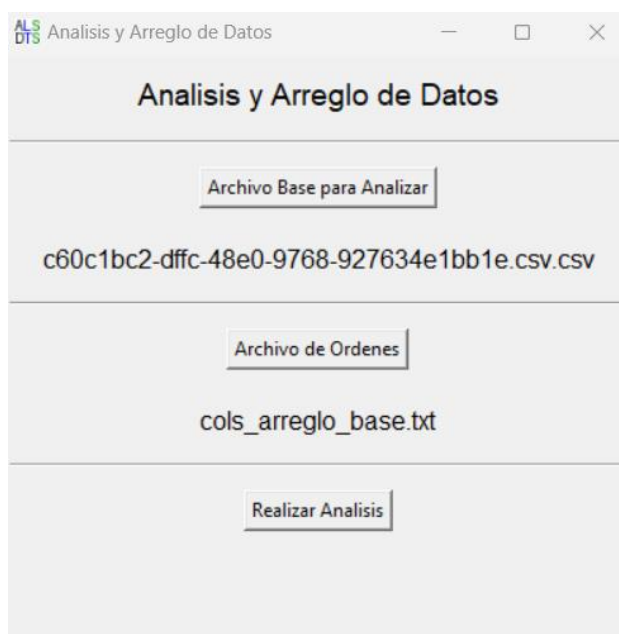
Ej. c60c1bc2-dffc-48e0-9768-927634e1bb1e.CSV.

Una vez el concatenador realiza la unión de la información teniendo en cuenta los archivos de configuración, este revisa cada fila, para eliminar duplicados manteniendo el primer caso de estos. Los duplicados que encuentra se colocan en un archivo de Excel.

E.j. duplicados_c60c1bc2-dffc-48e0-9768-927634e1bb1e.XLSX.

Analisis y Arreglo:

Uso:



El ejecutable permite seleccionar el archivo consolidado de bases de datos o de poblaciones de riesgo generado anteriormente por el aplicativo de concatenador_de_excel.

El ejecutable permite seleccionar el archivo de órdenes para realizar las correcciones necesarias al archivo consolidado. Estos archivos de correcciones se tienen en la carpeta 'config_analisis', y poseen las acciones que se deben de realizar sobre las columnas especificadas para poder realizar un buen análisis del archivo y asegurarse de que los códigos y nombres estén correctos.

*El concatenador permite tener diferentes archivos de órdenes para diferentes análisis.

Al seleccionar el botón de Realizar Análisis, se verá una barra de progreso para ver el avance del proceso. En caso de que se encuentren errores o inconsistencias en los códigos o en los nombres,

se generará un archivo de errores con el fin de que el usuario utilice los ejecutables de 'Administracion_X' para resolver estos errores, dependiendo de lo que se requiera hacer.

En caso de que las correcciones se realicen con éxito, el ejecutable sobre escribe el archivo de entrada.

*El ejecutable puede leer tanto archivos .XLSX como .CSV.

Archivos de Configuración:

Para los archivos de órdenes, se han generado dos archivos base.

Uno para bases de datos: 'cols_arreglo_base.txt', y uno para las poblaciones de riesgo: 'cols_arreglo_poblacion_riesgo.txt'.

También utiliza varios archivos de mapeo para corregir tanto nombres como códigos.

Cols_arreglo_base:

Este archivo posee las ordenes que se deben de realizar para corregir el archivo consolidado de las bases de datos.

E.j.

```
f FEC_AJU FEC_AJU
f FEC_RESULT FEC_RESULT
cod (COD_DPTO_O,COD_MUN_O) COD_JOIN_O
name_cod (NDEP_PROCE,NMUN_PROCE);(DEPARTAMENTO_OCURRENCIA,MUNICIPIO_OCURRENCIA) COD_JOIN_O
cod_name COD_JOIN_O NOM_DEPT_O,NOM_MUN_O
+ COD_JOIN_O,ANO ID_POBLACION
&Estrato_Municipio.xlsx&ESTRATO&name_cod (ENTIDAD_TERRITORIAL,MUNICIPIO) COD_JOIN_O
```

En este archivo se realizan las siguientes ordenes:

- Aplicar arreglo de fecha para la columna FEC_AJU y renombrar como FEC_AJU.
- Aplicar arreglo de fecha para la columna FEC_RESULT y renombrar como FEC_RESULT.
- Aplicar arreglo de código usando las columnas COD_DEPTO_O y COD_MUN_O y colocar en la columna COD_JOIN_O. (Intenta conseguir un código valido de 5 caracteres usando columnas de código)
- Aplicar arreglo de código usando las columnas (NDEP_PROCE,NMUN_PROCE) y (DEPARTAMENTO_OCURRENCIA,MUNICIPIO_OCURRENCIA) y colocar en la columna COD_JOIN_O. (Intenta conseguir un código valido de 5 caracteres usando columnas de nombres).
- Aplicar arreglo de nombre usando la columna COD_JOIN_O y colocar en las columnas NOM_DEPT_O y NOM_MUN_O. (Intenta conseguir el nombre del departamento y municipio usando la columna de código).
- Añadir la columna ID_POBLACION usando las columnas COD_JOIN_O y ANO.
- Realizar una unión con el archivo 'Estrato_Municipio.xlsx', columna 'ESTRATO', aplicando arreglo de código usando las columnas ENTIDAD_TERRITORIAL y MUNICIPIO y unir usando la columna COD_JOIN_O.

*Se debe mencionar que el archivo 'Estrato_Municipio.xlsx' debe de estar en la carpeta 'config_analisis'.

Cols_arreglo_poblacion_riesgo:

Este archivo posee las ordenes que se deben de realizar para corregir el archivo consolidado de las poblaciones de riesgo.

E.j.

```
cod_name COD_MUN NOMBRE_DEPARTAMENTO,NOMBRE_MUNICIPIO
```

En este archivo se realizan las siguientes ordenes:

- Aplicar arreglo de nombre usando la columna COD_MUN y colocar en las columnas NOMBRE_DEPARTAMENTO y NOMBRE_MUNICIPIO. (Intenta conseguir el nombre del departamento y municipio usando la columna de código).

Codigos_invalidos:

Aunque no permite interactuar directamente con este archivo, se requiere en la carpeta 'config_analisis' para su funcionamiento.

Este archivo es un modulo de Python que contiene una lista de todos los códigos inválidos. En caso de que una fila de la base de datos posea un código invalido que este en esta lista, se intenta obtener el código usando los nombres de departamento y municipio.

Diccionario_codigos:

Aunque no permite interactuar directamente con este archivo, se requiere en la carpeta 'config_analisis' para su funcionamiento.

Este archivo es un módulo de Python que posee un diccionario limpio (la cual fue generada tomando el valor estadístico más repetido para cada uno de los códigos) con los códigos, y sus valores normalizados y consolidados de nombres de departamentos y municipios.

Mapeo_nombres:

Aunque no permite interactuar directamente con este archivo, se requiere en la carpeta 'config_analisis' para su funcionamiento.

Este archivo es un modulo de Python que posee un diccionario limpio de nombres y sus valores correctos en casos de errores ortográficos, los cuales luego de realizar la consolidación, no llegan a coincidir con los valores correctos.

Errores_cols_salida:

Aunque no permite interactuar directamente con este archivo, se requiere en la carpeta 'config_analisis' para su funcionamiento.

Este archivo posee las columnas que se colocan en el archivo de errores, cuando el ejecutable encuentra inconsistencias.

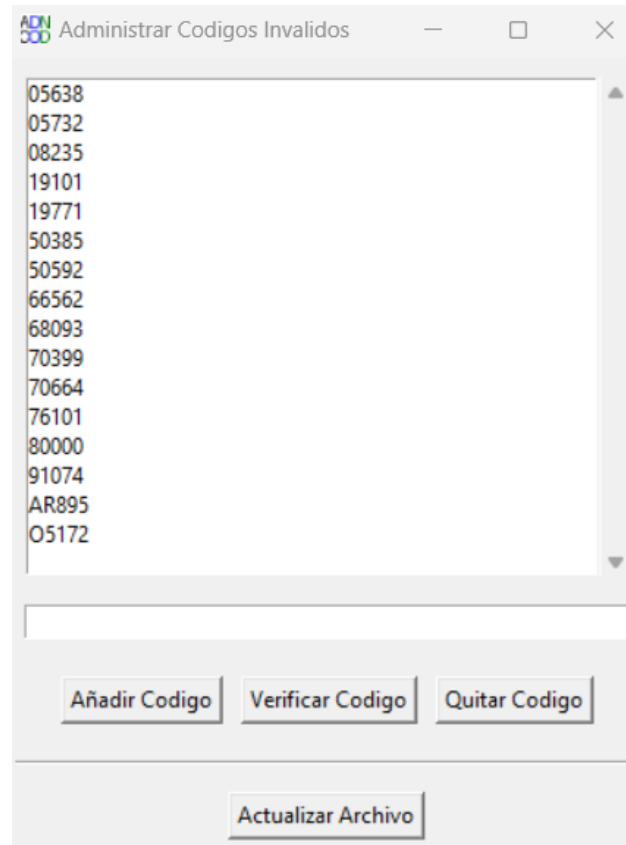
Salidas:

En caso de que se encuentren inconsistencias o errores en la ejecución, se genera el archivo: ERROR_c60c1bc2-dffc-48e0-9768-927634e1bb1e.XLSX, con las columnas que se encuentran en el archivo: 'errores_cols_salida.txt' en la carpeta 'config_analisis'.

En caso de que sea exitoso en realizar las correcciones, se sobre escribe el archivo de entrada.

Administrador Codigos Invalidos:

Uso:



Este ejecutable funciona dentro de la carpeta 'config_analisis' y permite interactuar de manera segura con el módulo codigos_invalidos.

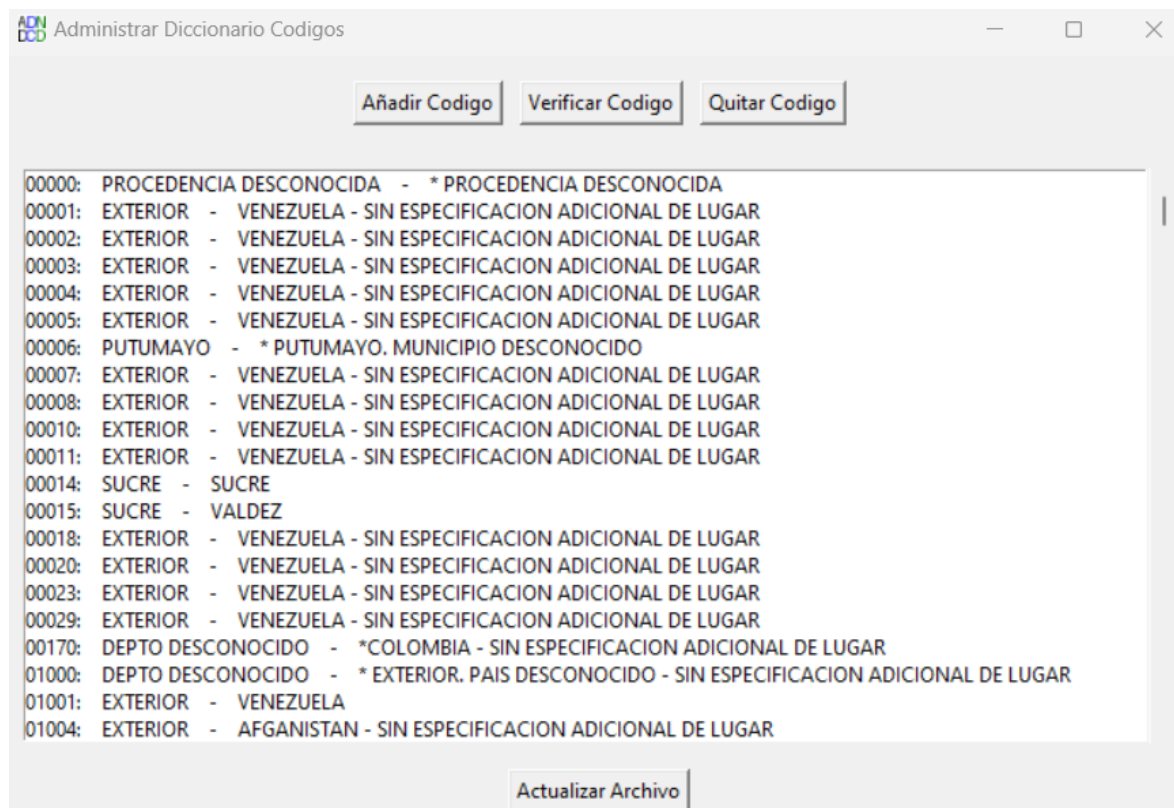
Permite añadir códigos, usando las validaciones necesarias para verificar que sea un código real, antes de ser añadido. También, permite verificar si algún código ya esta dentro de la lista de códigos inválidos. Al seleccionar uno de los códigos de la lista, se puede quitar el código.

Para añadir el código se digita dentro de la barra de texto.

Al presionar el botón de Actualizar Archivo, el módulo de Python de códigos inválidos es sobrescrito.

Administrador Diccionario Codigos

Uso:



Este ejecutable funciona dentro de la carpeta 'config_analisis' y permite interactuar de manera segura con el módulo diccionario_codigos.

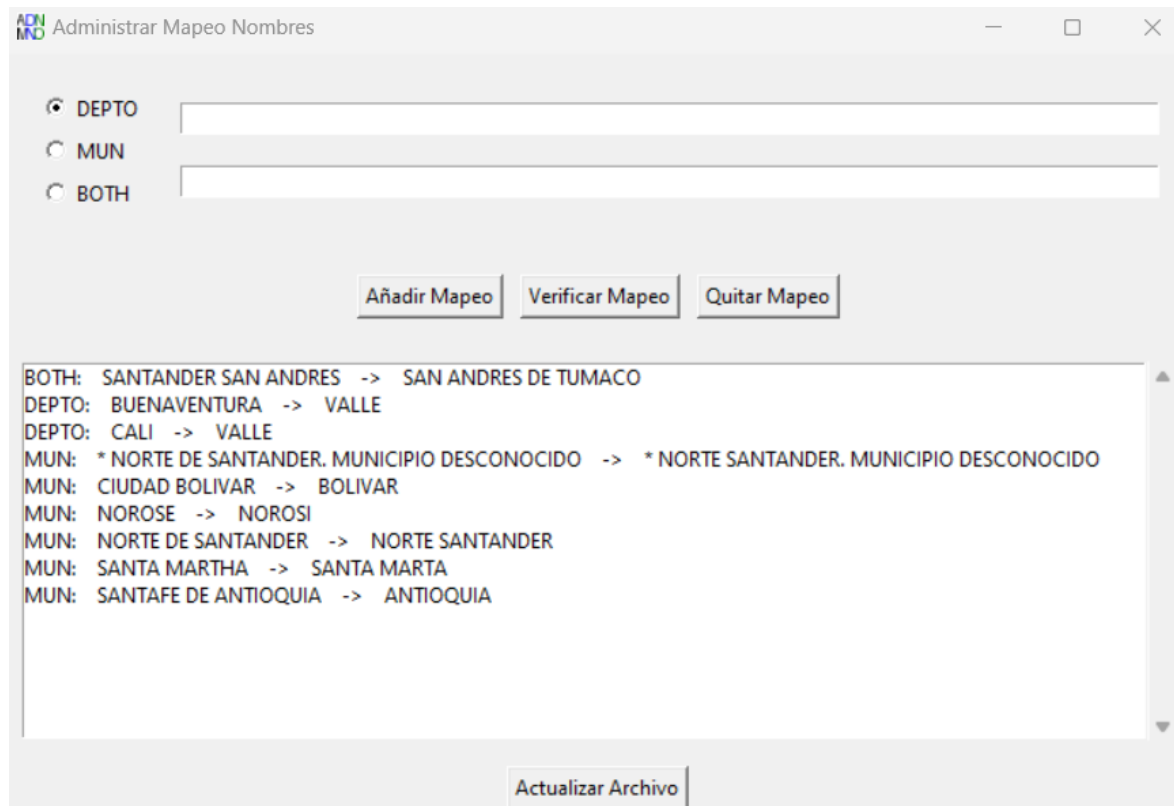
Permite añadir códigos y su respectivo departamento y municipio, usando las validaciones necesarias para verificar que sea un código real, que no esté en la lista, y que el nombre del departamento y municipio no esté en el diccionario bajo otro código, antes de ser añadido. También, permite verificar si algún código ya está dentro del diccionario, y cuáles son los valores de los nombres que posee. Al seleccionar uno de los códigos de la lista, se puede quitar el código con sus respectivos nombres.

Al seleccionar los botones, aparecen campos que indican que tipo de información se debe de colocar. Sea el código, el nombre del departamento o el nombre del municipio.

Al presionar el botón de Actualizar Archivo, el módulo de Python de diccionario códigos es sobrescrito.

Administrador Mapeo Nombres:

Uso:



Administrador Mapeo Nombres

☒ DEPTO
☐ MUN
☐ BOTH

BOTH: SANTANDER SAN ANDRES -> SAN ANDRES DE TUMACO
DEPTO: BUENAVENTURA -> VALLE
DEPTO: CALI -> VALLE
MUN: * NORTE DE SANTANDER. MUNICIPIO DESCONOCIDO -> * NORTE SANTANDER. MUNICIPIO DESCONOCIDO
MUN: CIUDAD BOLIVAR -> BOLIVAR
MUN: NOROSE -> NOROSI
MUN: NORTE DE SANTANDER -> NORTE SANTANDER
MUN: SANTA MARTHA -> SANTA MARTA
MUN: SANTA FE DE ANTIOQUIA -> ANTIOQUIA

Este ejecutable funciona dentro de la carpeta 'config_analisis' y permite interactuar de manera segura con el módulo mapeo_nombres.

Permite añadir mapas de nombres a los departamentos y a los municipios. Esto usando los botones de DEPTO, MUN y BOTH.

La barra superior es el nombre original y la barra inferior es el nombre al que se quiere cambiar. Se debe de seleccionar que tipo de cambio es: si es un cambio para los nombres de departamento o un cambio para los nombres de municipio. En caso de que existan municipios iguales en departamentos diferentes, se puede usar el botón de BOTH, en el cual se coloca el nombre original de la siguiente manera: DEPARTAMENTO<espacio>MUNICIPIO, y en la barra de inferior, solo el nombre del municipio.

Al presionar el botón de Actualizar Archivo, el módulo de Python de mapeo nombres es sobrescrito.

Recomendaciones Generales:

- Intentar mantener el formato de las bases de datos y de poblaciones de riesgo actual.
- Intentar no realizar cambios innecesarios a los nombres de las columnas en archivos de bases de datos y poblaciones de riesgo posteriores.
- El concatenador de Excel realiza una limpieza de columnas, no debe haber columnas con nombres iguales, o nombres que, al realizar la limpieza y normalización, sean iguales.
- Los nombres de las columnas en los archivos de configuración se usan luego de la normalización y consolidación de las columnas.
- Los archivos que utilizan ordenes de tipo 'contener' para ciertas palabras en los títulos de las columnas, solo funcionan si todas las palabras se encuentran en el título.
- Intentar mantener un formato de títulos para poblaciones de riesgo que no contenga palabras repetidas que se encuentren en los archivos de configuración.