

# Manual de Usuario librería DANNE.EMMET

2023-09-13

## Introducción

Esta librería realiza las tareas de procesamiento mensual de la Encuesta Mensual Manufacturera con Enfoque Territorial EMMET, la cual fue construida en lenguaje de programación R, desarrollando scripts que ayuden a la optimización, flujo y entendimiento de los procesos.

Esta librería es diseñada a partir de la caracterización de la operación estadística donde se evidencia el uso de múltiples herramientas analíticas, como: python, excel, macros de excel, visores, SAS y word. En el siguiente gráfico se muestra el uso de herramientas para cada una de las tareas definidas en el proceso estadísticos.

Para la creación de este piloto se hizo una recopilación de bases de datos de interés, junto con un diccionario de las variables que contienen cada base. Apoyados en el diccionario, se construyó la guía para el tratamiento de estas bases, su uso e importancia. Con esto, se definió qué variables se repetían en las bases, cuáles son las llaves para enlazar las diferentes bases y cuáles de éstas permanecían por la importancia en el proceso y cuales se descartaban.

Revisando los procesos de la EMMET, se concluyó que varias tareas se podían automatizar y estandarizar en un único software libre como R o Python y de acuerdo a las capacidades del equipo técnico se priorizó el uso de R. Con base en esto y teniendo en cuenta la caracterización ya mencionada, se definió la ruta a seguir; esto implicaba realizar el acople entre la base denominada “Logística”; esta base recopila el histórico de datos hasta el mes anterior a tratar, con la base original, es decir la base del mes de interés con los datos que entregaron las fuentes.

Usando todos los recursos disponibles, se transcribió un código que estandarizaba las variables numéricas y de tipo carácter. Este código estaba realizado en SAS, y se transcribió a lenguaje de R, para implementarlo dentro de la ruta que se estableció. Adicionalmente, se realiza un ajuste de fechas para las fuentes que no rindieron esta información, colocando el periodo en el que normalmente reportan.

Después de caracterizada la operación se define la construcción de las siguientes funciones:

- Integración de bases: Se crea una base panel, que contiene toda la información necesaria para ejecutar los diferentes procesos, y se estandarizan las variables numéricas cambiando los datos faltantes o NA por 0.
- Detección de alertas: Analiza los datos entregados en el mes de interés y detecta comportamientos anómalos en comparación a la serie histórica de cada empresa.
- Imputación: De acuerdo al comportamiento de cada empresa identifica el valor por el cual se corrigen las alertas encontradas anteriormente.
- Creación base Temática: Realiza la creación de las nuevas variables, a partir de las encontradas en la base original, que contiene la base temática.
- Anexo nacional: Crea los cuadros de salida del anexo nacional a partir de la base temática.
- Anexo territorial: Crea los cuadros de salida del anexo territorial a partir de la base temática.
- Cuadros de dominios: Crea cuadros de salida en donde se desagrega la información de producción, ventas y empleados por dominios.

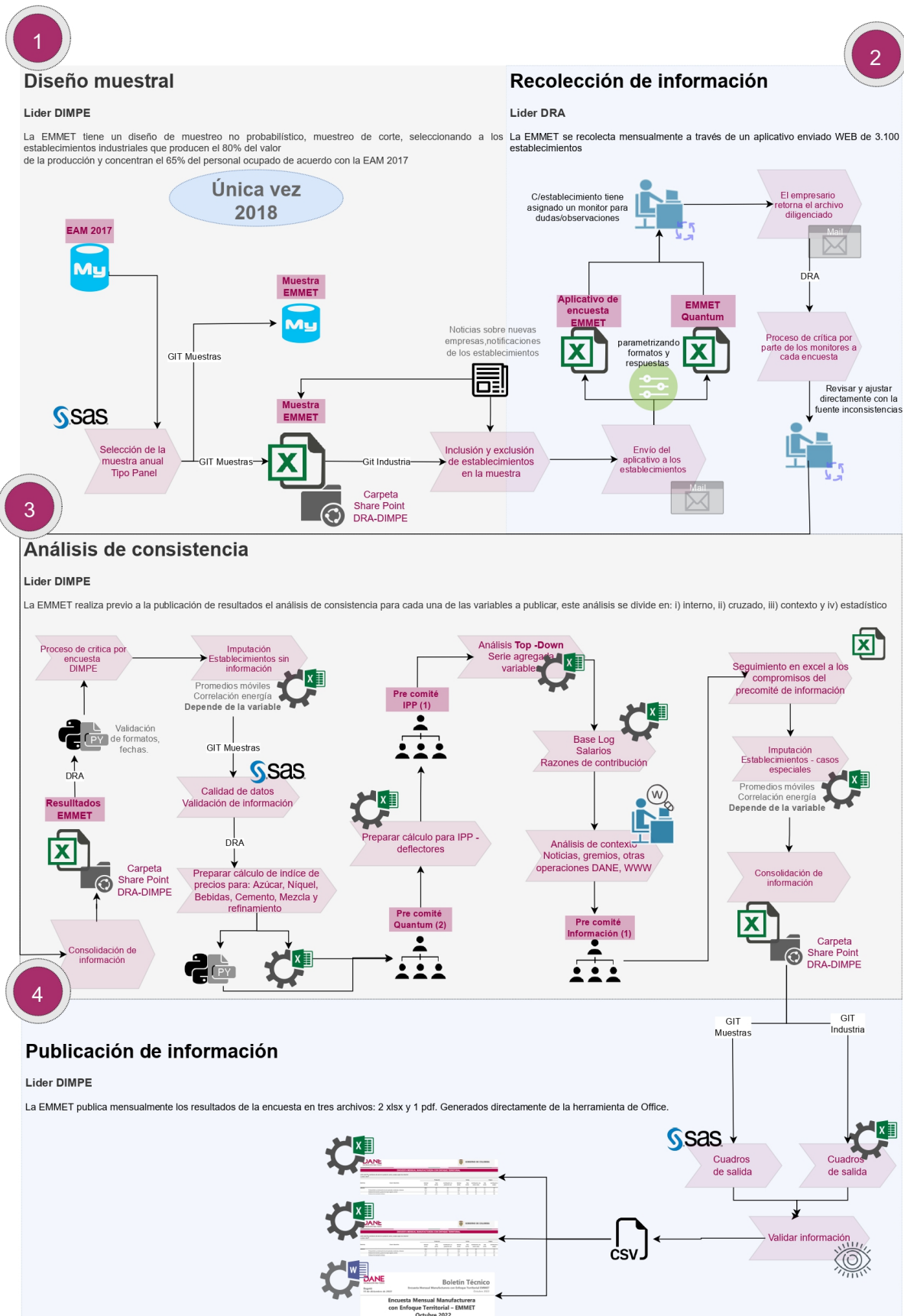


Figure 1: Flujo\_libreria

- Cuadros de regiones: Crea cuadros de salida en donde se desagrega la información de producción, ventas y empleados por áreas metropolitanas, ciudades.
- Boletín: Crea el boletín de publicación, a partir de la base temática.

Adicionalmente se creó un tablero de control en Shiny que permite a los usuarios interactuar con los datos de manera fluida y dinámica, brindando una visión clara y en tiempo real de los resultados de la encuesta.

Esta librería representa un avance significativo en la automatización y optimización de la Encuesta Mensual Manufacturera. Está diseñada para ahorrar tiempo, reducir errores y brindar una mayor comprensión de los procesos involucrados.

## Caracterización:

Las funciones de la librería desarrollan partes claves que actúan como engranajes para el desarrollo de esta. A continuación, explicamos en detalle cada una de las funciones mencionadas anteriormente.

Las funciones y características clave de esta librería incluyen:

- Inicial: Esta función instala todos los paquetes necesarios para el correcto funcionamiento de la Librería. Además, busca en el directorio suministrado (argumento de la función, que mas adelante se detallará cómo detallarlo), si existen las carpetas en donde se almacenarán los archivos de salida de cada una de las siguientes funciones.
- Integración de Bases: Esta función permite la consolidación y unificación de diversas fuentes de datos en una sola base de datos, para que, a partir de esta ejecuten los demás procesos, adicionalmente en los campos numéricos que posean espacios nulos, se reemplazaran por 0, estandarizando todos las variables numéricas de la base.
- Detección de Alertas: La función de alertas se encarga de analizar los comportamientos de las empresas, registradas en la encuesta, con el fin de alertar cambios inusuales o bruscos en su patrón de respuesta, marcando las alertas de la siguiente manera: Si la novedad de la empresa es 5 se marcará como imputación\_deuda y si el comportamiento es inusual o anómalo se marcará como imputación\_caso\_especial.
- Imputación: La función de imputación reconoce los casillas marcadas en la función de alertas y con diferentes tipos de análisis, como lo son: KNN, variación y TSO outlier, imputa el dato de acuerdo a lo analizado. Con esto se puede concluir que la librería brinda capacidades de imputación automáticas para completar o modificar la información de manera confiable.
- Creación de la base Temática: Esta función genera, la base temática. Esta contiene la información de los datos procesados, de acuerdo a la metodología de la operación, en este sentido se presentan los datos reales a partir de los nominales y a su vez la información ponderada y agrega variables de identificación de los dominios por los cuales se publica.
- Anexo Nacional y Territorial: La librería facilita la generación de anexos nacionales y territoriales de manera automatizada, simplificando la generación de informes detallados. Estos anexos o cuadros de salida muestran información complementaria a la registrada en el boletín de prensa con el fin de brindar la información a un nivel más desagregado tanto total nacional como desagregado a nivel de departamentos, áreas metropolitanas y principales ciudades del país.
- Cuadros de dominios: La librería crea un archivo de excel en donde se genera una hoja por cada dominio que muestra la variación y contribución anual de las variables de producción, ventas, empleados, sueldos y horas.

- Cuadros de regiones: La librería crea un archivo de excel en donde se genera una hoja por cada desagregación de región que muestra la variación y contribución anual de las variables de producción, ventas, empleados, sueldos y horas.
- Boletín: Esta es la etapa final de la Encuesta Mensual Manufacturera, la creación del boletín. Esta función genera un archivo pdf o Word, según se especifique, (esto es un argumento de la función, más adelante se detallará como detalla), en donde el contenido de este es un informe que genera un resumen ejecutivo en el que se presentan los principales resultados con el uso de diferentes tipos de herramientas visuales para mostrar información histórica y lograr hacer comparaciones de variaciones entre las distintas actividades o desagregación regional.

Con la idea de optimizar el proceso de computación y ejecución de las funciones se crearon dos macros, que corren por bloque, cada una, cierta cantidad de funciones. En ese sentido, se optimiza aún más el proceso de automatización de resultados de la EMMET. Cada macro se puede ejecutar con facilidad y están repartidas de la siguiente manera:

- macro1: Ejecuta las funciones de Integración e Identificación\_alertas.
- macro2: Ejecuta en orden las siguientes funciones; Imputación, Temática, territorial, nacional y boletín. Los argumentos que solicita la macro 2 es la unión de los argumentos de las funciones que corre, más adelante se explica detalladamente cómo especificar la función y los argumentos

Esta partición de las funciones con la macro 1 se da para facilitar la revisión de alertas que genera la automatización y con ello, realizar la verificación por parte de los analistas. Una vez hecho esto, se ejecuta en cadena las demás funciones para generar los cuadros de salida, ahorrando tiempo y minimizando los errores en escritura de los argumentos de las funciones.

## Diseño de la librería

Para comprender completamente su funcionalidad, es esencial explorar sus componentes clave, su flujo de trabajo y los parámetros que permiten su ejecución precisa.

Bases Insumo:

La librería se alimenta de cuatro bases de datos fundamentales:

- Base Original (mensual): Esta base se crea cada mes para registrar el reporte que entregan las empresas que rinden información para la Encuesta mensual manufacturera.
- Base Logística: Esta base proporciona información importante para la ejecución de la encuesta, pues recopila el histórico de los datos ya tratados; es decir, con los procesos de crítica e imputación previamente realizados; desde el año 2018 hasta el mes anterior al que desea trabajar.
- Base Paramétrica: Esta base contiene la información asignada a cada establecimiento en cada una de las variables de identificación y publicación. Lo que permite una personalización y adaptación óptima a diferentes contextos.
- Base Deflactor: Contiene los datos de IPP \_\_ PyC, IPP \_\_ EXP, IPC. Estos están designados según el año, mes y código CIU4. Se asocian a cada empresa según corresponda.

Las funciones que posee la librería se deben ejecutar en orden, ya que depende de los archivos de salida generados por la función anterior. Cada función posee unos argumentos de entrada, que permiten realizar la ejecución de acuerdo con el periodo de interés a tratar.

Los parámetros que se establecieron para las funciones son:

Directorio: debe otorgar la dirección de la carpeta en donde desea que se almacenen los archivos de salida

Mes: Corresponde al mes que desea analizar

Anio: Corresponde al año que desea analizar

Estos deben ser iguales en los argumentos de todas las funciones.

El proceso general se presenta de la siguiente manera:

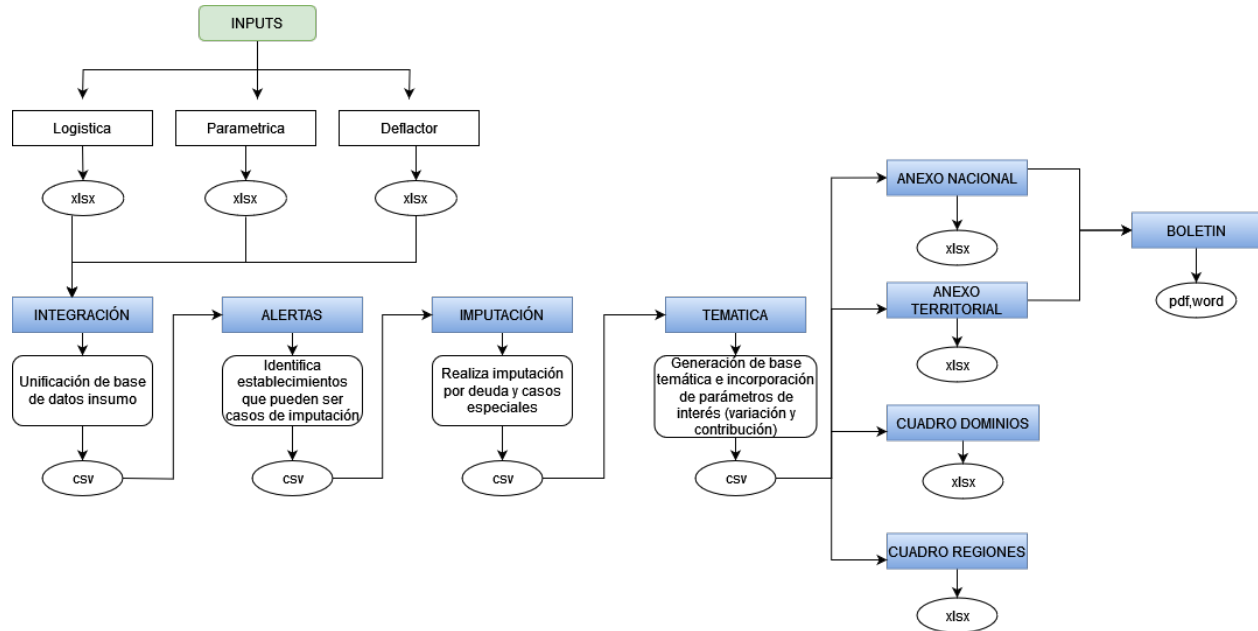


Figure 2: Flujo\_libreria

Ejemplo de cómo llamar la librería, especificar los parámetros en cada función y ejecutarlas:

- Instalar librería en el computador:

```
remotes::install_github("NataliArteaga/DANE.EMMET")
```

- cargar la librería:  
`library(DANE.EMMET)`
- correr función inicial:  
`f0_inicial(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET")`
- correr función integracion:  
`f1_integracion(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- correr función identificacion\_alertas:  
`f2_identificacion_outliers(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- correr función imputación:  
`f3_imputacion_outliers(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`

- correr función temática:  
`f4_tematica(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- correr función anacional:  
`f5_anacional(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- correr función aterritorial:  
`f6_aterritorial(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- correr función boletín:  
`f7_boletin(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022,tipo="word")`
- correr función cdominio:  
`f8_cdominios(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- correr función cregiones:  
`f9_cregiones(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- correr función shiny:  
`f10_shiny(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- correr funcion macro1: `macro1(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- verificar y/o modificar el archivo de alertas para continuar con el proceso
- verificar si se requiere la aplicación shiny `f9_shiny(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022)`
- correr funcion macro2 `macro2(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET",mes=11,anio=2022,tipo="pdf")`

## Documentación

### Integración

#### Descripción

Con esta función se integran los archivos de entrada necesarios en el proceso EMMET para trabajar con una base llamada "base panel.csv".

**Uso** `f1_integracion(directorio, mes, anio)`

#### Argumentos

**directorio:** Define el directorio donde se encuentran ubicados los datos de entrada.

**mes:** Define el mes a ejecutar, por ejemplo, 11.

**anio:** Define el año a ejecutar, por ejemplo, 2022.

## Valor de Retorno

El valor de retorno es un archivo CSV.

## Detalles

En esta función se leen las siguientes bases:

- Base logística: Es la base con todos los datos desde enero de 2018 hasta el mes indicado en el parámetro mes.
- Base paramétrica: Base con las características de cada establecimiento.

El procedimiento para unificar las bases es el siguiente:

Se pegan las filas de la base original debajo de la base logística y se nombra base panel.

Se agregan las columnas de la base paramétrica a la base panel por medio del ID\_NUMORD.

Finalmente se estandarizan los nombres de los departamentos y municipios, uniendo las columnas de la base divipola por el ID\_municipio.

Además, en las variables numéricas, los datos faltantes se cambian por cero. A estas mismas variables se les establece un formato numérico para su posterior manipulación. Finalmente, se exporta un archivo de tipo .csv

```
# f1_integracion(  
#   directorio = "Documents/DANE/Procesos DIMPE /PilotoEMMET",  
#   mes = 11,  
#   anio = 2022  
#)
```

## Ejemplos

## Identificación de alertas

### Descripción

Esta función tiene como objetivo identificar las variables dentro de cada establecimiento que pueden llegar a ser casos de imputación, ya sea imputación por deuda o por casos especiales.

**Uso** f2\_identificacion\_alertas(directorio, mes, anio, avance = 100)

### Argumentos

**directorio:** Define el directorio donde se encuentran ubicados los datos de entrada.

**mes:** Define el mes a ejecutar, por ejemplo, 11.

**anio:** Define el año a ejecutar, por ejemplo, 2022.

**avance (opcional):** Porcentaje de avance en la identificación de alertas (valor predeterminado: 100).

## Archivo de Retorno

El archivo de retorno es de tipo CSV.

## Detalles

Primero se dividen las variables en capítulo 2 y capítulo 3. Variables capítulo 2

```
II_PA_PP__NPERS_EP
AJU_II_PA_PP_SUELD_EP
II_PA_TD__NPERS_ET
AJU_II_PA_TD_SUELD_ET
II_PA_TI__NPERS_ETA
AJU_II_PA_TI_SUELD_ETA
II_PA_AP__AAEP
AJU_II_PA_AP_AAS_AP
II_PP_PP__NPERS_OP
AJU_II_PP_PP_SUELD_OP
II_PP_TD__NPERS_OT
AJU_II_PP_TD_SUELD_OT
II_PP_TI__NPERS_OTA
AJU_II_PP_TI_SUELD_OTA
II_PP_AP__APEP
AJU_II_PP_AP_AAS_PP
AJU_II_HORAS_HORDI_T
AJU_II_HORAS_HEXTR_T
```

Variables capítulo 3

```
AJU_III_PE_PRODUCCION
AJU_III_PE_VENTASIN
AJU_III_PE_VENTASEX
III_EX__VEXIS
```

El procedimiento para identificar posibles establecimientos a imputar es:

1. Si la novedad es igual 5 sabemos que el individuo se va a imputar por deuda en todas sus variables o si el valor reportado en la variable de interés es cero y el mes inmediatamente anterior fue diferente de cero también se identifica como imputación deuda
2. Si la novedad es diferente a 5 se observa la variación con respecto al mes anterior; si es menor o igual a 20% no se imputa.
3. Si la variación es mayor a 20% y su valor es diferente de cero, entonces, primero se observa si el dato es igual a alguno de los datos reportados por el establecimiento en meses anteriores; en caso de que esté presente, se decide no imputar, si el dato no estuvo anteriormente, entonces se procede a identificar si la variable es referente a capítulo 2, si es así, se realiza una carta de control con los últimos 24 meses, si esta identifica el valor como atípico se imputa por caso especial; si la variable es referente a capítulo 3, entonces, se realiza una prueba mezclando el comando (`locate.outliers.iloop`) y la carta de control de 24 meses, el comando `locate.outliers.iloop` identifica los valores que pueden ser outliers en series de tiempo, si el valor fue identificado como un valor atípico se imputa por caso especial, en caso de que no fuera identificado como un valor atípico no se imputa.



Los límites de la carta control son calculados de la siguiente manera:

$$(\bar{x} + 1.96 + +, \bar{x} + 1.96 + +)$$

Donde  $\bar{x}$  es el promedio de la serie de los últimos 24 meses de la variable y establecimiento de interés y  $S$  es la desviación estándar de la serie de los últimos 24 meses de la variable y establecimiento de interés.

**Guía para actualizar archivo de alertas:** Luego de ejecutar esta función se creará un archivo de excel tipo csv, este contiene variables de identificación de los establecimientos, los valores en cada una de las variables de interés (capítulo 2 y capítulo 3), y si son posibles casos de imputación, las variables en las que tengan un valor diferente a continua son los que en la función de imputación pasaran por ese proceso

Para modificar el archivo debe modificar el valor en la variable que desee y busque la columna cuyo nombre es nombrevariable\_caso\_de\_imputacion y modifique el valor por “continua\_critica”, para tener registro de que valores se modificaron por una critica, si desea, en la última columna puede realizar un comentario

Ejemplo, queremos modificar el valor de la producción del establecimiento con ID\_numord 23; por lo tanto primero se busca la fila cuyo valor de id\_numord es 23, en la columna “AJU\_III\_PE\_PRODUCCION” cambiaremos el valor numerico por el valor que deseamos (evite usar decimales), luego proceda a buscar la columna “AJU\_III\_PE\_PRODUCCION\_caso\_de\_imputacion”, ahí modifique el valor de la casilla por “continua\_critica”, sin importar si el valor anterior era “continua”, “imputacion\_deuda” o “imputacion\_caso\_especial”.

```
# f2_identificacion_alertas(
#   directorio = "Documents/DANE/Procesos DIMPE /PilotoEMMET",
#   mes = 11,
#   anio = 2022
# )
```

## Ejemplos

## Imputación outliers

### Descripción

Esta función realiza la imputación por deuda y por casos especiales en variables de capítulo 2 y capítulo 3.

**Uso** f3\_imputacion(directorio, mes, anio)

### Argumentos

**directorio:** Define el directorio donde se encuentran ubicados los datos de entrada.

**mes:** Define las tres primeras letras del mes a ejecutar, por ejemplo, "11" para noviembre.

**anio:** Define el año a ejecutar, por ejemplo, 2022.

### Valor de Retorno

El valor de retorno es un archivo CSV.

## Detalles

La metodología de imputación es la siguiente:

1. Para las variables de capítulo 2 la metodología a usar es la imputación por el mes anterior.
2. Para las variables de capítulo 3 la metodología a usar para imputación por casos especiales es el KNN combinado con variación mes anterior. Para la imputación por deuda en variables pertenecientes a capítulo 3 se realiza un proceso similar.

El KNN combinado consta de los siguientes pasos:

1. Realizar un primer KNN imputando los individuos atípicos en las variables de interés (KNN1).
2. Calcular la variación con respecto al mes anterior.
3. Realizar un segundo KNN para la variación con respecto al mes anterior de los establecimientos del mismo dominio (KNN2).
4. Calcular el valor final de la siguiente manera:

$$\hat{y} = KNN_1 * (1 + KNN_2) \quad (1)$$

Donde  $\hat{y}$  es el valor resultante con el que se imputara la variable de interés.

Para el caso de imputación deuda en capítulo 3 se sigue la siguiente formula:

$$\hat{y} = MA * (1 + KNN_2) \quad (2)$$

Esto es que el valor resultante con el que se imputara la variable de interés es igual al valor del mes anterior, multiplicado por 1 más la variación con respecto al mes anterior de los establecimientos del mismo dominio.

Para conocer más sobre la imputación por el método k-Nearest Neighbors ver (KNN\_VIM)

## Valor de Retorno

El valor de retorno es un archivo CSV.

## Ejemplos

```
#
# f3_imputacion(
#   directorio = "Documents/DANE/Procesos DIMPE /PilotoEMMET",
#   mes = "nov",
#   anio = 2022
# )
```

## Temática

### Descripción

Esta función construye la base temática. La base temática expone los datos procesados de acuerdo con la metodología de la operación, presentando los datos reales a partir de los nominales. También incluye la información ponderada y agrega variables de la identificación de los dominios por los cuales se publica. Con esto, se puede presentar la base temática como la base final de cuadros. Esta función toma como insumo la base Panel con los casos de imputación aplicados y exporta un archivo con extensión .xlsx.

**Uso** f4\_tematica(directorio, mes, anio)

### Argumentos

**directorio:** Define el directorio donde se encuentran ubicados los datos de entrada.

**mes:** Define las tres primeras letras del mes a ejecutar, por ejemplo, "11" para noviembre.

**anio:** Define el año a ejecutar, por ejemplo, 2022.

### Valor de Retorno

El valor de retorno es un archivo CSV.

### Detalles

En esta función se realiza el cruce con el valor de CIIU4, mes y año, entre las variables de la base imputada y la base deflactor, questa base contiene los datos de IPP e IPC para cada CIIU4.

Esta función crea las variables de la base temática de la siguiente manera:

- VentasReales: Suma de la división entre ventas en el interior e IPP\_PYC y la división entre ventas en el exterior e IPP\_PYC.
- TotalPersonas: Suma de las variables que contabilizan el número de trabajadores en las diferentes categorías.
- TotalSueldosNominal: Suma de las variables de sueldos nominales.
- TotalSueldosReales: División entre TotalSueldosNominal e IPC.
- SueldosPermanentesNominal: Suma de las variables de sueldos nominales del personal contratado por la categoría permanente.
- SueldosPermanentesReales: División entre SueldosPermanentesNominal e IPC.
- SueldosTemporalesNominal: Suma de las variables de sueldos nominales del personal contratado por la categoría temporales.
- SueldosTemporalesReales: División entre SueldosTemporalesNominal e IPC.
- SueldosAdmonNominal: Suma de las variables de sueldo nominal del personal administrativo nominal.
- SueldosAdmonReal: División entre SueldosAdmonNominal e IPC.
- SueldosAdmonPermReal: División entre AJU\_II\_PA\_PP\_SUELD\_EP e IPC.
- SueldosAdmonTempNomin: Suma de las variables de sueldo del personal administrativo temporales.
- SueldosAdmonTempReal: División entre SueldosAdmonTempNomin e IPC.
- SueldosProducNominal: Suma de las variables del personal de producción nominal.
- SueldosProducReal: División entre SueldosProducNominal e IPC.
- SueldosProducPermReal: División entre AJU\_II\_PP\_PP\_SUELD\_OP e IPC.
- SueldosProducTempNomin: Suma de las variables del personal de producción temporal.
- SueldosProducTempReal: División entre SueldosProducTempNomin e IPC.
- TotalHoras: Suma de variables de horas extras y ordinarias.
- TotalEmpleoPermanente: Suma de variables del personal de administración y producción.
- TotalEmpleoTemporal: Suma de variables del personal temporal.
- TotalEmpleoAdmon: Suma de variables del personal administrativo.
- TotalEmpleoProduc: Suma de variables del personal de producción.
- EmpleoProducTempor: Suma de variables del personal de producción temporal.
- TOTAL\_VENTAS: Suma de variables en el interior y exterior del país.
- DEFLACTOR\_NAL: División entre TOTAL\_VENTAS y VentasReales.

```
# f4_tematica(
#   directorio = "Documents/DANE/Procesos DIMPE /PilotoEMMET",
#   mes = "nov",
#   anio = 2022
# )
```

## Ejemplos

## Anexo Nacional

### Descripción

Esta función crea el archivo de Anexo Nacional. Utiliza la base de datos Tematica como insumo y genera cada una de las hojas del reporte de Anexo Nacional. Los cuadros de salida o anexos estadísticos de EMMET muestran información complementaria a la registrada en el boletín de prensa con el fin de brindar información a un nivel más desagregado tanto a nivel nacional como desagregado a nivel de departamentos, áreas metropolitanas y principales ciudades del país. Los resultados se presentan con variaciones anuales, año corrido y doce meses, junto con sus respectivas contribuciones, según dominios, por las principales variables que se recolectan en el proceso: Producción (real y nominal), ventas (real y nominal) y empleo, desagregando a su vez esta última variable por área funcional y tipo de contrato. De igual manera se presentan los resultados de sueldos causados y horas totales trabajadas para los dominios nacionales.

**Uso** f5\_anacional(directorio, mes, anio)

### Argumentos

**directorio:** Define el directorio donde se encuentran ubicados los datos de entrada.

**mes:** Define el mes a ejecutar, por ejemplo, "11" para noviembre.

**anio:** Define el año a ejecutar, por ejemplo, 2022.

### Valor de Retorno

El valor de retorno es un archivo CSV.

### Detalles

En los anexos nacionales se calculan las contribuciones y variaciones en tres diferentes periodos. A continuación se muestran los periodos y las formulas para realizar el cálculo de estos:

Contribución anual:

$$CA_{trj} = \frac{(V_{trj} - V_{(t-12)rj})}{\sum_1^n V_{(t-12)rj}} * 100 \quad (3)$$

Donde:

$t$ : Mes de referencia de la publicación de la operación estadística

$V_{trj}$ : Valor en el periodo  $t$  para el territorio  $r$  en el dominio  $j$ .

$V_{(t-12)rj}$ : Valor en el periodo t-12 o en el año anterior, en el territorio r en el dominio j.

$\sum_1^n V_{(t-12)rj}$ : Sumatoria de los valores en el periodo t-12, en el territorio r y en el dominio j

Esta contribución anual se interpreta como el aporte del dominio j en el territorio r a la variación anual del mes de referencia en el dominio j en el territorio r

Contribución año corrido:

$$CAC_{Trj} = \frac{\sum_i^T (V_{trj} - \sum_b^{T-12} V_{trj})}{\sum_b^{T-12} V_{trj}} * 100 \quad (4)$$

Donde:  $t$ : Mes variando de enero a diciembre

$T$ : Mes de referencia.

$i$ : Siempre es el mes de enero.

$b = i - 12$ : Corresponde a enero del año anterior

$V_{trj}$ : Valor de la variable en el periodo t en el territorio r en el dominio j

Esta contribucion de año corrido se interpreta como el aporte del dominio j en el territorio r a la variación año corrido del mes de referencia en el dominio j en el territorio r.

Contribución año acumulado:

$$CAA_{Trj} = \frac{\sum_{a+1}^T (V_{trj} - \sum_{b+1}^a V_{trj})}{\sum_{b+1}^a V_{trj}} * 100 \quad (5)$$

Donde:

$t$ : Mes variando de enero a diciembre

$T$ : Mes de referencia.

$a=T-12$

$b=a-12$ : Corresponde al mes a del año anterior

$V_{trj}$ : Valor de la variable en el periodo t en el territorio r en el dominio j

Nota: cuando las variables que denotan meses (a, b) son negativas representan el mes del año inmediatamente anterior.

Esta contribución de año acumulado se interpreta como el aporte del dominio j en el territorio r a la variación acumulada anual del mes de referencia en el dominio j en el territorio r.

Variación anual:

Es la relación del índice o valor (para producción y ventas, categoría de contratación, sueldos, horas) en el mes de referencia ( $t_i$ ) con el índice o valor absoluto del mismo mes en el año anterior ( $t_{i-12}$ ), - 1 por 100.

$$VA = \frac{\text{índice o valor del mes de referencia}}{\text{índice o valor del mismo mes del año anterior}} - 1 * 100 \quad (6)$$

Se interpreta como el crecimiento o disminución porcentual, dependiendo de si el resultado es negativo o positivo, de la variable correspondiente en el mes de referencia, en relación al mismo mes del año anterior

Variación Año Corrido:

$$VAC = \frac{\sum \text{índice o valor de enero al mes de referencia del año actual}}{\sum \text{índice o valor de enero al mes de referencia del mismo mes del año anterior}} - 1 * 100 \quad (7)$$

Se interpreta como el crecimiento o disminución porcentual, dependiendo de si el resultado es negativo o positivo, de la variable correspondiente en lo corrido del año hasta el mes de referencia, en relación al mismo periodo del año anterior

Variación Acumulado Anual:

$$VAA = \frac{\sum \text{índice o valor desde } a_{+1} \text{ hasta el mes de referencia}}{\sum \text{índice o valor en el año anterior desde } a_{+1} \text{ hasta el mes de referencia}} - 1 * 100 \quad (8)$$

Donde:

$t$ =mes de referencia

$a=t-12$

Se interpreta como el crecimiento o disminución porcentual, dependiendo de si el resultado es negativo o positivo, de la variable correspondiente en los últimos 12 meses hasta el mes de referencia, en relación al mismo periodo del año anterior.

Contribuciones porcentuales: aporte en puntos porcentuales de las variaciones individuales a la variación de un agregado.

La función escribe, en formato excel, las hojas:

1. Var y cont\_Anuar: Se calcula la Variación anual (%) y contribución, del valor de la producción, ventas, y empleo, según las clase industrial
2. Var Anual\_Emp\_Sueldos\_Horas: Se calculan las Variaciones anuales, según clase industrial, producción, ventas, personal ocupado, sueldos y horas totales trabajadas
3. Var y cont\_año corrido: Se calcula la Variación año corrido (%) y contribución, del valor de la producción, ventas, y empleo, según las clase industrial
4. Var año corr Emp\_Sueldos\_Hor: Se calculan las variaciones año corrido, según clase industrial producción, ventas, personal ocupado, sueldos y horas totales trabajadas
5. Var y cont\_doce meses: Se calcula la aariación doce meses (%) y contribución, del valor de la producción, ventas, y empleo, según las clase industrial
6. Var doce meses Emp\_Sueldos\_H: Se calculan las variaciones doce meses, según clase industrial, producción, ventas, personal ocupado, sueldos y horas totales trabajadas .
7. Indices total por clase: Se calculan los índices de producción nominal y real, ventas nominal y real, empleo, sueldos y horas Totales trabajadas, según clase industrial para cada uno de los dominios que se encuentran en la encuesta mensual manufacturera.
8. Índices Desestacionalizados: Para la contrucción de la desestacionalización se usa la librería: seasonal. Esta libreria es una adaptacion del programa X13 seasonal, para el programa de R.

Para esta función se crea un calendario en donde se ingresan los festivos fijos y los de fecha variables, para asignar la ponderación a los días, en donde finalmente quedan expresadas los días de Lunes a Sábados.

Posteriormente para realizar la desestacionalización en las variables de producción, ventas y empleo, se escoge de manera individual la variable de interés con la columna de las fechas, este nuevo data frame (ej: producción, fecha) se convierte en una serie de tiempo, con frecuencia mensual.

Para la desestacionalización, se usa la serie de tiempo construida anteriormente, y se especifican:

- fecha de inicio y fin de la serie (ej: 2001.1,2021.11).

- variables regresoras: el calendario, construido anteriormente,
- variables dicotomicas de outliers: estas son festivos patrios, años bisiesto, y fechas de alto impacto, como:  
2016.Jul: Paro camionero  
2016.Aug:  
2020.Mar: Inicio de pandemia Covid-19  
2020.Apr: Inicio de la cuarentena  
2021.May: Paro Nacional

Adicionalmente se especifican los parámetros estacionales y la informacion de la serie que se desea observar.

Finalmente se crea una tabla con la fecha y los datos desestacionalizados.

Esto, por cada una de las variables de interés.

9. Enlace legal hasta 2014: Se realiza una conexión entre la serie de la metodología actual, con las variaciones de los índices de la metodología anterior con el fin de tener una serie más larga de comparación de la información.
10. Enlace legal hasta 2001: Se realiza una conexión entre la serie de la metodología actual, con las variaciones de los índices de la metodología anterior con el fin de tener una serie más larga de comparación de la información.
11. Var y cont\_Trienal: Se calcula la contribución y variación trienal, es decir usando como año base los datos del año 2019, del valor de la producción, ventas, y empleo, según clase industrial

## Ejemplos

```
# f5_anacional(  
#   directorio = "Documents/DANE/Procesos DIMPE /PilotoEMMET",  
#   mes = "nov",  
#   anio = 2022  
# )
```

## Anexo Territorial

### Descripción

Esta función crea el archivo de Anexo Territorial utilizando la base de datos Tematica como insumo. El reporte generado contiene información complementaria a la registrada en el boletín de prensa con el objetivo de proporcionar información a un nivel más desagregado tanto a nivel nacional como a nivel departamental, áreas metropolitanas y principales ciudades del país. Los resultados se presentan con variaciones anuales, año corrido y doce meses, junto con sus respectivas contribuciones, según dominios, por las principales variables que se recolectan en el proceso: Producción (real y nominal), ventas (real y nominal) y empleo. También se incluyen resultados de sueldos causados y horas totales trabajadas para los dominios nacionales.

**Uso** f7\_ateritorial(directorio, mes, anio)

## Argumentos

directorio: Define el directorio donde se encuentran ubicados los datos de entrada.

mes: Define el mes a ejecutar, por ejemplo, "11" para noviembre.

anio: Define el año a ejecutar, por ejemplo, 2022.

**Valor de Retorno** El valor de retorno es un archivo CSV.

## Detalles

En los anexos territoriales se calculan las contribuciones y variaciones en tres diferentes periodos. A continuación se muestran los periodos y las formulas para realizar el calculo de estos:

Contribucion anual:

$$CA_{trj} = \frac{(V_{trj} - V_{(t-12)rj})}{\sum_1^n V_{(t-12)rj}} * 100 \quad (9)$$

Donde:

$t$ : Mes de referencia de la publicación de la Operación Estadística

$V_{trj}$ : Valor en el periodo  $t$  para el territorio  $r$  en el dominio  $j$ .

$V_{(t-12)rj}$ : Valor en el periodo  $t-12$  o en el año anterior, en el territorio  $r$  en el dominio  $j$ .

$\sum_1^n V_{(t-12)rj}$ : Sumatoria de los valores en el periodo  $t-12$ , en el territorio  $r$  y en el dominio  $j$

Esta contribucion anual se interpreta como el aporte del domino  $j$  en el territorio  $r$  a la variación anual del mes de referencia en el domino  $j$  en el territorio  $r$

Contribucion anio corrido:

$$CAC_{Trj} = \frac{\sum_i^T (V_{trj} - \sum_b^{T-12} V_{trj})}{\sum_b^{T-12} V_{trj}} * 100 \quad (10)$$

Donde:  $t$ : Mes variando de enero a diciembre

$T$ : Mes de referencia.

$i$ : Siempre es el mes de enero.

$b=i-12$ : Corresponde a enero del año anterior

$V_{trj}$ : Valor de la variable en el periodo  $t$  en el territorio  $r$  en el dominio  $j$

Esta contribucion de anio corrido se interpreta como el aporte del domino  $j$  en el territorio  $r$  a la variación año corrido del mes de referencia en el domino  $j$  en el territorio  $r$ .

Contribucion anio acumulado:

$$CAA_{Trj} = \frac{\sum_{a+1}^T (V_{trj} - \sum_{b+1}^a V_{trj})}{\sum_{b+1}^a V_{trj}} * 100 \quad (11)$$

Donde:

$t$ : Mes variando de enero a diciembre

$T$ : Mes de referencia.



$a=T-12$

$b=a-12$ : Corresponde al mes a del año anterior

$V_{trj}$ : Valor de la variable en el periodo t en el territorio r en el dominio j

Nota: cuando las variables que denotan meses (a, b) son negativas representan el mes del año inmediatamente anterior.

Esta contribucion de año acumulado se interpreta como el aporte del domino j en el territorio r a la variación acumulada anual del mes de referencia en el domino j en el territorio r.

Variación anual:

Es la relación del índice o valor (para producción y ventas, categoría de contratación, sueldos, horas) en el mes de referencia (ti) con el índice o valor absoluto del mismo mes en el año anterior (t i-12), - 1 por 100.

$$VA = \frac{\text{índice o valor del mes de referencia}}{\text{índice o valor del mismo mes del año anterior}} - 1 * 100 \quad (12)$$

Se interpreta como el crecimiento o disminución porcentual, dependiendo de si el resultado es negativo o positivo, de la variable correspondiente en el mes de referencia, en relación al mismo mes del año anterior

Variación Año Corrido:

$$VAC = \frac{\sum \text{índice o valor de enero al mes de referencia del año actual}}{\sum \text{índice o valor de enero al mes de referencia del mismo mes del año anterior}} - 1 * 100 \quad (13)$$

Se interpreta como el crecimiento o disminución porcentual, dependiendo de si el resultado es negativo o positivo, de la variable correspondiente en lo corrido del año hasta el mes de referencia, en relación al mismo periodo del año anterior

Variación Acumulado Anual:

$$VAA = \frac{\sum \text{índice o valor desde } a_{+1} \text{ enero al mes de referencia}}{\sum \text{índice o valor en el año anterior desde } a_{+1} \text{ hasta el mes de referencia}} - 1 * 100 \quad (14)$$

Donde:

$t$ =mes de referencia

$a=t-12$

Se interpreta como el crecimiento o disminución porcentual, dependiendo de si el resultado es negativo o positivo, de la variable correspondiente en los últimos 12 meses hasta el mes de referencia, en relación al mismo periodo del año anterior. Contribuciones porcentuales: aporte en puntos porcentuales de las variaciones individuales a la variación de un agregado.

La función escribe, en formato excel, las hojas:

1. Var y Cont Anual Dpto: Se calcula la contribución y variación anual (%) del valor de la producción, ventas, y empleo, según departamento
2. Var y Cont Anual Desagreg Dp: Se calcula la contribución y variación anual (%) del valor de la producción, ventas, y empleo, según clase industrial por departamento
3. Var y Cont Anual Áreas metrop: Se calcula la contribución y variación anual (%) del valor de la producción, ventas, y empleo, según área metropolitana

4. Var y Cont Anual Ciudades: Se calcula la contribución y variación anual (%) del valor de la producción, ventas, y empleo, según ciudad
5. Var y Cont Año corrido Dpto: Se calcula la contribución y variación año corrido (%) del valor de la producción, ventas, y empleo, según departamento
6. Var y Cont Año corrido Desag Dp: Se calcula la contribución y variación año corrido (%) del valor de la producción, ventas, y empleo, según clase industrial por departamento
7. Var y Cont Año corrido Áreas met: Se calcula la contribución y variación a año corrido (%) el valor de la producción, ventas, y empleo, según área metropolitana
8. Var y Cont Año corrido Ciudad: Se calcula la contribución y variación año corrido (%) del valor de la producción, ventas, y empleo, según ciudad
9. Var y Cont doce meses Dpto: Se calcula la contribución y variación doce meses (%) del valor de la producción, ventas, y empleo, según departamento
10. Var y Cont doce meses Desa: Se calcula la contribución y variación doce meses (%) del valor de la producción, ventas, y empleo, según clase industrial por departamento
11. Var y Cont doce meses Áreas: Se calcula la contribución y variación doce meses (%) del valor de la producción, ventas, y empleo, según área metropolitana
12. Var y Cont doce meses Ciu: Se calcula la contribución y variación doce meses (%) del valor de la producción, ventas, y empleo, según ciudad
13. Índices Departamentos: Se calcula los índices, de producción nominal y real, ventas nominal y real, empleo según departamento y clase industrial
14. Índices Áreas Metropolitana: Se calcula los índices de producción nominal y real, ventas nominal y real, empleo según área metropolitana
15. Índices Ciudades: Se calcula los índices, de producción nominal y real, ventas nominal y real, empleo según ciudades
16. Var y Cont Trienal Dpto: Se calcula la contribución y variación trienal (%), es decir; usando como año base los datos del año 2019, del valor de la producción, ventas, y empleo, según departamento
17. Var y Cont Trienal Desa: Se calcula la contribución y variación trienal (%), es decir usando como año base los datos del año 2019, del valor de la producción, ventas, y empleo, según clase industrial por departamento
18. Var y Cont Trienal Áreas: Se calcula la contribución y variación trienal (%), es decir usando como año base los datos del año 2019, agrupando los datos por áreas metropolitanas de la producción, ventas y personal
19. Var y Cont Trienal Ciuda: Se calcula la contribución y variación trienal (%), es decir usando como año base los datos del año 2019, del valor de la producción, ventas, y empleo, según ciudad.

Finalmente se exporta un archivo excel, que contiene la información de las 19 hojas.

###vEjemplos

```
# f6_aterritorial(
#   directorio = "Documents/DANE/Procesos DIMPE /PilotoEMMET",
#   mes = "nov",
#   anio = 2022
# )
```

## f7\_boletin

### Descripción

Esta función crea el boletín, un documento que se puede exportar en pdf o word, según se desea.

Este es un resumen ejecutivo en el que se presentan los principales resultados con el uso de cuadros de líneas para presentar información histórica y graficas de barras para hacer comparaciones de variaciones entre las distintas actividades o desagregación regional.

Además, se presentan tablas con el resumen por actividad para el total nacional, departamentos, áreas metropolitanas y ciudades incluyendo las tres variables principales que permiten hacer un análisis cruzado entre estas.

**Uso** f7\_boletin(directorio, mes, anio, tipo = "pdf")

####Argumentos

directorio: Directorio donde se encuentran ubicados los datos de entrada.

mes: Mes a ejecutar, ej: 11.

anio: Año a ejecutar, ej: 2022.

tipo: Tipo de archivo de salida (pdf, word u html). Por defecto, se usa "pdf".

### Detalles

El boletín generado presenta los principales resultados utilizando cuadros de líneas para información histórica y gráficos de barras para comparaciones de variaciones entre actividades o desagregación regional. Además, se incluyen tablas con el resumen por actividad para el total nacional, departamentos, áreas metropolitanas y ciudades, lo que permite un análisis cruzado.

```
# f7_boletin(
#   directorio = "Documents/DANE/Procesos DIMPE /PilotoEMMET",
#   mes = 11,
#   anio = 2022,
#   tipo = "word"
# )
```

### Ejemplos

## Cuadro dominios

### Descripción

Esta función crea un archivo de excel que contiene las variaciones y contribuciones anuales desagregadas por dominio 39. Por cada dominio se crea una hoja en donde se consigna las variaciones y contribuciones de las variables de producción, ventas, empleados, sueldos y horas.

Adicionalmente se crea una hoja en donde se presentan los valores de defaltores y se calcula la variación anual y mensual por cada dominio.

**Uso** f8\_cdominios(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET", mes=11,anio=2022)

### Argumentos

directorio: Define el directorio donde se encuentran ubicados los datos de entrada.

mes: Define el mes a ejecutar, por ejemplo, 11.

anio: Define el año a ejecutar, por ejemplo, 2022.

### Valor de Retorno

El valor de retorno es un archivo CSV.

### Detalles

Esta funcion se hace con la finalidad de crear cuadros informativos con datos desagregada según el interés, para facilitar el análisis del comportamiento de cada secotr del dominio.

```
# f8_cdominios(  
#   directorio = "Documents/DANE/Procesos DIMPE /PilotoEMMET",  
#   mes = 11,  
#   anio = 2022  
#)
```

### Ejemplos

## Cuadro regiones

### Descripción

Esta funcion crea un archivo de excel que contiene las variaciones y contribuciones anuales desagregadas por cada región. Se crea una hoja por cada desagregación en donde se consigna las variaciones y contribuciones de las variables de produccion, ventas, empleados, sueldos y horas.

**Uso** f9\_cregiones(directorio="Documents/DANE/Procesos DIMPE /PilotoEMMET", mes=11,anio=2022)

### Argumentos

directorio: Define el directorio donde se encuentran ubicados los datos de entrada.

mes: Define el mes a ejecutar, por ejemplo, 11.

anio: Define el año a ejecutar, por ejemplo, 2022.

### Valor de Retorno

El valor de retorno es un archivo CSV.

## Detalles

Esta funcion se hace con la finalidad de crear cuadros informativos con datos desagregada según el interés, para facilitar el análisis del comportamiento de cada región.

```
# f9_cregiones(  
#   directorio = "Documents/DANE/Procesos DIMPE /PilotoEMMET",  
#   mes = 11,  
#   anio = 2022  
#)
```

## Ejemplos