

Instructivo de Actualización de validaciones (Malla I)

Mayo, 2020



Buenas cifras,
mejores vidas

Introducción:

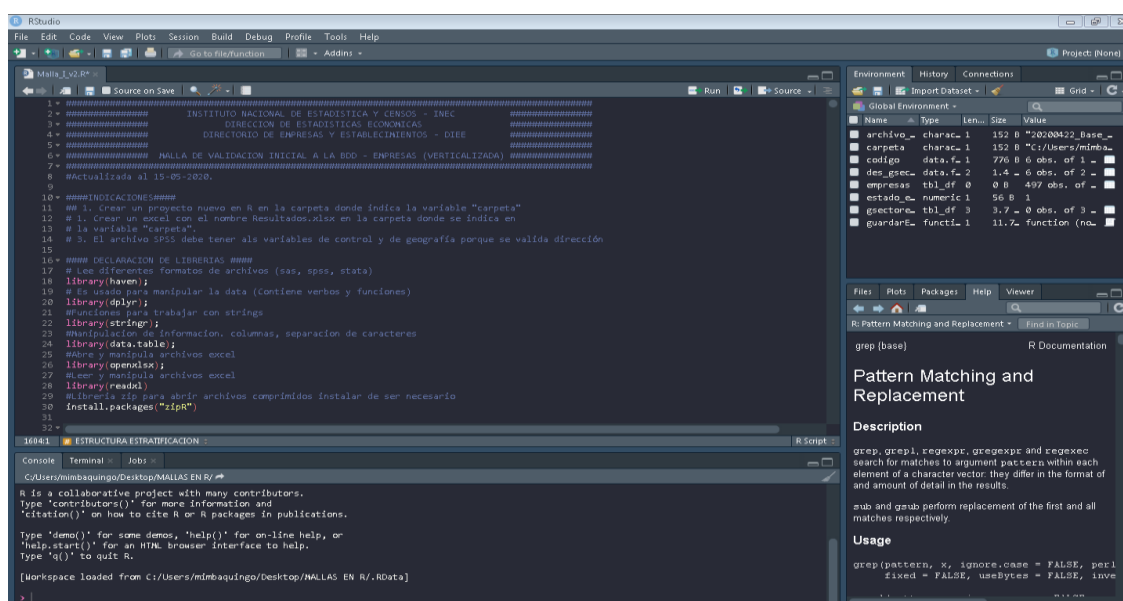
La unidad de gestión de estadísticas económicas en base a registros administrativos y en función al calendario establecido anualmente, calcula y publica los resultados del Directorio de Empresas y Establecimientos DIEEE, operación estadística que se construye a partir de Registros administrativos, que se obtienen de dos entidades técnicas, autónomas, con personería jurídica y derecho público, como lo son: el Servicio de Rentas Internas (SRI) y el Instituto Ecuatoriano de Seguridad Social (IESS).

En la construcción del DIEEE, se toma en cuenta estándares mínimos necesarios de acuerdo a experiencias y recomendaciones internacionales, desde su integración hasta llegar a su difusión, etapa que también cuenta con normas de publicación; adicional a ello también se considera las necesidades de usuarios de la operación estadística.

En GEERA se aplican una serie de validaciones que permiten detectar errores en la información de la base de datos las mismas se realizan en el programa R mediante un script que se ajusta a las necesidades del Directorio de empresas; esta serie de validaciones son realizadas por el personal técnico de la unidad, con ello se corrige errores antes de obtener la base de datos final.

Por ello el presente documento constituye una relatoría del uso de las validaciones aplicadas (Malla I) en la base de datos del Directorio de Empresas y Establecimientos DIEEE. El software que se utilizó para realizar las validaciones (Malla I) es R, se usa para procesamiento de información ejecutar funciones que nos permite detectar errores del Directorio de Empresas y Establecimientos.

Con el fin de que cualquier persona que se integre al equipo de trabajo del DIEEE se desenvuelva sin inconveniente en la etapa del Procesamiento de información, se ha creado el presente instructivo del funcionamiento del script y la elaboración del mismo.



Descripción:

- 1. Características básicas:** el script con el nombre "Malla_I_v2" que fue actualizada el 15 de mayo del 2020, actualizando las validaciones y corrigiendo errores que estaba generando la anterior malla.

- 1.1** En las indicaciones se puede visualizar los documentos previos que se deben crear y tener para el funcionamiento del script:

```
#####
##### INSTITUTO NACIONAL DE ESTADISTICA Y CENSOS - INEC #####
##### DIRECCION DE ESTADISTICAS ECONOMICAS #####
##### DIRECTORIO DE EMPRESAS Y ESTABLECIMIENTOS - DIEE #####
#####
##### MALLA DE VALIDACION INICIAL A LA BDD - EMPRESAS (VERTICALIZADA) #####
#####
##### Actualizada al 15-05-2020. #####

#####INDICACIONES#####
# 1. Crear un proyecto nuevo en R en la carpeta donde indica la variable "carpeta"
# 2. Crear un excel con el nombre Resultados.xlsx en la carpeta donde se indica en la variable "carpeta".
# 3. El archivo SPSS debe tener las variables de control y de geografía porque se valida dirección
```

- 1.2** Se especifica las librerías y paquetes que vamos a usar para correr las validaciones los mismos nos permiten cargar, manipular y procesar los datos, además se especifica para que funciona cada librería, las mismas se puede incrementar dependiendo de las necesidades o funciones que se van a realizar:

```
#### DECLARACION DE LIBRERIAS ####
# Lee diferentes formatos de archivos (sas, spss, stata)
library(haven);
# Es usado para manipular la data (Contiene verbos y funciones)
library(dplyr);
#Funciones para trabajar con strings
library(stringr);
#Manipulacion de informacion. columnas, separacion de caracteres
library(data.table);
#Abre y manipula archivos excel
library(openxlsx);
#Leer y manipula archivos excel
library(readxl)
#Libreria zip para abrir archivos comprimidos instalar de ser necesario
install.packages("zipR")
```

- 1.3** A continuación está la declaración de variables en donde se indica la base de datos a usar, el año y los diferentes valores que pueden tomar ciertas variables:

```
#### DECLARACION DE VARIABLES ####
# Limpiar objetos de ser necesario
rm(list = ls()) #Borra objetos
#rm(list = ls(all.names = TRUE)) #Borra todo
# Carpeta de trabajo
carpeta <- 'C:/Users/mimbaquingo/Desktop/MALLAS EN R';
# Nombrelibrary(readxl) archivo SPSS
#archivo_spss <- '20200422_Base_ejemplo_500_casos.sav';
#archivo_spss <- '20191125_DIEE2018_anio2018_activas_direcciones.sav';
archivo_spss <- '20200417_empresas_marco_2012_2018.sav';
#archivo_spss <- 'BASE_CON_GEOGRAFIA_500.sav';
# Situación de empresas
estado_empresa <- 1;
# Años proceso
anio_proceso <- 2018;
# Setear ubicacion en la carpeta
setwd(carpeta);
# Valor maximo que puede tomar las variable
val_max_plaza <- 33345
val_max_empleo <- 33345
val_max_ventas <- 13861085576
val_max_remun <- 607357888
# Valores X
x0 <- -4.94
x1 <- 1.40
#Valores Y
y0 <- -81.00
y1 <- -75.00
```

- 1.4 Se procede con la carga de la base de datos, en este caso se encuentra en formato SPSS, se seleccionan todos los años y el estado de situación en donde se seleccionan las empresas activas.

```
#Carga de data de empresas activas

system.time(Leer_spss <- read_spss(archivo_spss))
system.time(empresas<-Leer_spss %>% filter(situacion == estado_empresa))
rm(Leer_spss)
```

- 1.5 Para guardar el archivo se creó previamente un Excel con el nombre resultados, que es el documento en donde se van a ir guardando los errores que se encuentren en la base de datos.

```
##### Guardar archivos global #####

guardarExcel <- function(nombreHoja, titleSheet, resultDataSet){
  wb1 <- loadWorkbook(file = "Resultados.xlsx")
  addWorksheet(wb = wb1, sheetName = nombreHoja, visible = TRUE)
  writeData(wb = wb1, sheet = nombreHoja, x = titleSheet, startRow = 1, startCol = 1)
  writeData(wb = wb1, sheet = nombreHoja, x = resultDataSet, startRow = 2, startCol = 1, rowNames = TRUE)
  saveWorkbook(wb = wb1, file = "Resultados.xlsx", overwrite = TRUE)
}
```

- 1.6 Es necesario leer el catálogo en el cual se encuentran los valores o descripciones que pueden tomar todas las variables, nos sirve para realizar las comparaciones.

```
##### Leer Catálogo #####

catalogo <- list()
hoja <- excel_sheets(path = file.path("diee_catologo.xlsx"))
for(i in 1:length(hoja)){
  catalogo[[i]] <- read_excel("diee_catologo.xlsx", sheet = hoja[i], col_names = TRUE, na = "")}
}
```

2. Variables:

Las validaciones se realizan para las variables que conforman el Directorio de Empresas y Establecimientos, en total son 74 variables, que a continuación se detallan:

- ID EMPRESA
- ID RUC
- RAZON SOCIAL
- NOMBRE COMERCIAL
- FORMA INSTITUCIONAL
- TIPO DE UNIDAD LEGAL
- CLASE DE CONTRIBUYENTE
- OBLIGADO A LLEVAR CONTABILIDAD
- GRANDES SECTORES ECONOMICOS
- CODIGO SECCION
- CODIGO DIVISION
- CODIGO CLASE
- CODIGO ACTIVIDAD ECONOMICA
- ZONA DE PLANIFICACION
- PROVINCIA
- CANTON
- PARROQUIA
- FUENTE DE VENTAS
- FUENTE GEOGRAFIA

- FUENTE DIRECCION
- FUENTE FORMA INSTITUCIONAL
- FUENTE ZONIFICACION
- FUENTE COORDENADAS
- FUENTE ACTIVIDAD
- PLAZAS DE EMPLEO REGISTRADO PROMEDIO
- PLAZAS DE EMPLEO REGISTRADO PROMEDIO HOMBRES
- PLAZAS DE EMPLEO REGISTRADO PROMEDIO MUJERES
- PLAZAS DE EMPLEO REGISTRADO ULTIMO
- PLAZAS DE EMPLEO REGISTRADO ULTIMO HOMBRES
- PLAZAS DE EMPLEO REGISTRADO ULTIMO MUJERES
- EMPLEO REGISTRADO PROMEDIO
- EMPLEO REGISTRADO PROMEDIO HOMBRES
- EMPLEO REGISTRADO PROMEDIO MUJERES
- EMPLEO REGISTRADO ULTIMO
- EMPLEO REGISTRADO ULTIMO HOMBRES
- EMPLEO REGISTRADO ULTIMO MUJERES
- VENTAS TOTALES
- VENTAS NACIONALES
- VENTAS TARIFA 0%
- VENTAS TARIFA 12%
- INGRESOS CONSTRUCCION
- INGRESOS ARRENDAMIENTOS
- INGRESOS COMISIONES
- EXPORTACIONES NETAS
- SALARIOS
- SALARIOS HOMBRES
- SALARIOS MUJERES
- FECHA REGISTRO VENTAS
- FECHA FUENTE VENTAS
- FECHA FUENTE GEOGRAFIA
- FECHA FUENTE GEOGRAFIA
- FECHA REGISTRO DIRECCION
- FECHA FUENTE DIRECCION
- FECHA REGISTRO FORMA INSTITUCIONAL
- FECHA FUENTE FORMA INSTITUCIONAL
- FECHA REGISTRO ZONIFICACION
- FECHA FUENTE ZONIFICACION
- FECHA REGISTRO COORDENADAS
- FECHA FUENTE COORDENADAS
- FECHA REGISTRO ACTIVIDAD
- FECHA FUENTE ACTIVIDAD
- COORDENADAS
- ZONIFICACION
- ESTRATO PLAZAS PROMEDIO
- ESTRATO PLAZAS ULTIMO
- ESTRATO EMPLEO PROMEDIO
- ESTRATO EMPLEO ULTIMO
- ESTRATO VENTAS
- TAMANIO DE EMPRESA PLAZAS PROMEDIO
- TAMANIO DE EMPRESA PLAZAS ULTIMO
- TAMANIO DE EMPRESA EMPLEO PROMEDIO
- TAMANIO DE EMPRESA EMPLEO ULTIMO
- ORIGEN DE LA ESTRATIFICACION
- ESTRUCTURA DE LA ESTRATIFICACION

3. Formato:

El formato de las validaciones está unificado, lo que permite comprender su funcionamiento:

Está compuesto por el nombre de la variable, se da una breve explicación de lo que realiza la función, se realiza la selección y en el filtro se indica los parámetros que deben cumplir los datos, al final con el IF se van a guardar todos los casos que no cumplen con los estándares, resultando los errores que se van a guardar en el documento de Excel con el nombre de la hoja, indicando las novedades o error de estos casos.

```
#### ID EMPRESA ####

# Validar identificador de empresa con 11 caracteres #
# Verificar que id_empresa contenga 11 dígitos y no estén en blanco

error_id_empresa<-empresas %>%
  select( ruc_principal, anio, id_empresa) %>%
  filter(nchar(id_empresa) != 11 | is.na(id_empresa))

if(nrow(error_id_empresa)> 0){
  guardarExcel('error_id_empresa', 'id_empresa con cantidad de dígitos diferente de 11 o nulo', error_id_empresa)
}
rm(error_id_empresa)

# Validar id_empresa duplicado #
# Valida los Id empresa duplicados por todos los años

id_empr_dupl <- data.frame(empresas %>% select(id_empresa, anio))
id_dupl <- data.frame(table(id_empr_dupl$id_empresa, id_empr_dupl$anio))
id_dupl_resul <- id_dupl[id_dupl$Freq > 1, ]
names(id_dupl_resul)[1] <- 'Id_empresa'
names(id_dupl_resul)[2] <- 'Año'
names(id_dupl_resul)[3] <- 'Frecuencia'

if (nrow(id_dupl_resul) > 0) {
  guardarExcel('id_empr_dupl', 'Empresas que tienen id_empresa duplicado', id_dupl_resul)
}
rm(id_empr_dupl, id_dupl, id_dupl_resul)
```

4. Validaciones:

Cada variable puede tener dos o tres validaciones debido a que se debe revisar verificar varias novedades que se pueden presentar, pero siguiendo el formato en cada una:

4.1 En las primeras validaciones, como id empresa, id ruc, razón social y nombre comercial, se verifica que tengan los caracteres correspondientes, que no tengan caracteres especiales y casos duplicados.

Objeto validación	Función
ID EMPRESA	verificar_caracteres_id_empresa
	verificar_vacios_id_empresa
	verificar_duplicados_id_empresa
ID RUC	verificar_caracteres_ruc_principal
	verificar_duplicados_ruc_principal
RAZON SOCIAL	verificar_caracteres_especiales_razonsocial
NOMBRE COMERCIAL	verificar_caracteres_especiales_nombrecomercial

4.2 Las validaciones de las siguientes variables, se verifica que no tengan valores nulos y que se encuentren en el catálogo, es decir que no tengan valores diferentes a los parámetros permitidos, para comparar con el catálogo es necesario conocer en la hoja en la que se encuentra cada valor o descripción que deben tomar las variables:

```
#### FORMA INSTITUCIONAL ####

# Validar Forma Institucional y catálogo #
# Validar que forma institucional no sea nulo y este en catalogo

f_inst_error <- empresas %>%
  select(id_empresa, anio, forma_institucional) %>%
  filter(forma_institucional == "" | !forma_institucional %in% catalogo[[6]]$CODIGO)

if (nrow(f_inst_error) > 0) {
  guardarExcel('f_inst_error', 'Forma Institucional con valor nulo o codigo no esta en catálogo', f_inst_error)
}
rm(f_inst_error)
```

Objeto validación	Función
FORMA INSTITUCIONAL	verificar_forma_inst
	verificar_catalogo_forma_inst
TIPO DE UNIDAD LEGAL	verificar_unidad_legal
	verificar_catalogo_unidad_legal
CLASE DE CONTRIBUYENTE	verificar_clase_contrib
	verificar_catalogo_clase_contrib
OBLIGADO A LLEVAR CONTABILIDAD	verificar_Obligado_llevar_cont
	verificar_catalogo_obligado_llevar_cont
GRANDES SECTORES ECONOMICOS	verificar_gsectores
	verificar_catalogo_gsectores
CODIGO SECCION	verificar_seccion
	verificar_catalogo_seccion
CODIGO DIVISION	verificar_division
	verificar_catalogo_division
CODIGO CLASE	verificar_clase
	verificar_catalogo_clase
CODIGO ACTIVIDAD ECONOMICA	verificar_actividad_econ
	verificar_catalogo_actividad_econ
ZONA DE PLANIFICACION	verificar_zona_planif
	verificar_catalogo_zona_planif
PROVINCIA	verificar_provincia
	verificar_catalogo_provincia
CANTON	verificar_canton
	verificar_catalogo_canton
PARROQUIA	verificar_parroquia
	verificar_catalogo_parroquia

4.3 Las variables de control que son las fuentes nos permiten conocer de dónde se obtuvo o actualizo cada dato por lo tanto se debe validar estas variables, lo que se verifica es que todas las empresas que tengan valores deben tener fuente y viceversa, además se verifica que los valores que tomen estas variables de control se encuentren en el catálogo:

```
#####VARIABLES DE CONTROL#####

##### FUENTE DE VENTAS #####
#Comprobar que todas las empresas que registran ventas totales tengan datos en fuente de ventas

validar_vent_fte <- empresas %>%
  select(id_empresa, ruc_principal, anio, ventas_totales, fuente_ventas)%>%
  filter(!is.na(ventas_totales) & (fuente_ventas == ""))

if (nrow(validar_vent_fte) > 0) {
  guardarExcel('cod_vent_fte', 'Empresas con ventas totales y sin fuentes de venta', validar_vent_fte)
}
rm(validar_vent_fte)

#Comprobar que todas las empresas que registran fuentes de ventas tengan datos en ventas totales#

validar_fte_ven <- empresas %>%
  select(id_empresa, ruc_principal, anio, ventas_totales, fuente_ventas)%>%
  filter((is.na(ventas_totales) & (fuente_ventas != "")))

if (nrow(validar_fte_ven) > 0) {
  guardarExcel('cod_fte_ven', 'Empresas con ventas totales vacias y con fuente de ventas', validar_fte_ven)
}
rm(validar_fte_ven)
#Comprobar que los valores que toma la variable se encuentre dentro del catalogo#

validar_fte_vent <- empresas %>%
  select(id_empresa, ruc_principal, anio, fuente_ventas)%>%
  filter(!(fuente_ventas %in% catalogo[[19]]$CODIGO) & !(fuente_ventas == ""))

if (nrow(validar_fte_vent) > 0) {
  guardarExcel('cod_fte_vent', 'Empresas con fuente ventas que se encuentran en el catalogo', validar_fte_vent)
}
rm(validar_fte_vent)
```

Objeto validación	Función
FUENTE DE VENTAS	verificar_fte_ventas
	verificar_fte_ventas_a
	verificar_fte_ventas_b
	verificar_catalogo_fte_ventas
FUENTE GEOGRAFIA	verificar_fte_geo
	verificar_fte_geo_a
	verificar_fte_geo_b
	verificar_catalogo_fte_geo
FUENTE DIRECCION	verificar_fte_direccion
	verificar_fte_direccion_a
	verificar_fte_direccion_b
	verificar_catalogo_fte_direccion
FUENTE FORMA INSTITUCIONAL	verificar_fte_forma_inst
	verificar_fte_forma_inst_a
	verificar_fte_forma_inst_b
	verificar_catalogo_fte_forma_inst
FUENTE ZONIFICACION	verificar_fte_zonific
	verificar_fte_zonific_a
	verificar_fte_zonific_b
	verificar_catalogo_fte_zonific
FUENTE COORDENADAS	verificar_fte_coordenadas
	verificar_fte_coordenadas_a
	verificar_fte_coordenadas_b
	verificar_catalogo_fte_coordenadas
FUENTE ACTIVIDAD	verificar_fte_actividad
	verificar_fte_actividad_a
	verificar_fte_actividad_b
	verificar_catalogo_fte_actividad

4.4 Se realizan validaciones para plazas de empleo y empleo registrado totales, para hombres y mujeres, se comprueban los totales de estas variables, se comprueban que los valores se encuentren en los límites permitidos es decir que sean correctos:

```
##### PLAZAS DE EMPLEO REGISTRADO PROMEDIO#####

## Error en sumatoria plazas promedio ##
# Si plazas promedio total != plazas_hombres + plazas_mujeres.
# Valida que los valores que toma la variable plazas promedio total sea igual a la
# suma de plazas promedio hombres mas plazas promedio mujeres,
# si no es igual, muestra los casos que presentan el error.
# Se validara todos los años.

error_sum_plazas_prom <- empresas %>%
  select(id_empresa, ruc_principal, anio, plazas, plazas_hombres, plazas_mujeres) %>%
  filter(ifelse(is.na(plazas), 0, plazas) != ifelse(is.na(plazas_hombres), 0, plazas_hombres) +
    ifelse(is.na(plazas_mujeres), 0, plazas_mujeres))

if(nrow(error_sum_plazas_prom) > 0){
  guardarExcel('error_plazas_prom', 'Error en sumatoria plazas promedio', error_sum_plazas_prom)
}

rm(error_sum_plazas_prom)
```

Objeto validación	Función
PLAZAS DE EMPLEO REGISTRADO PROMEDIO	comprobacion_sum_plazas
	comprobacion_valores_plazas
PLAZAS DE EMPLEO REGISTRADO PROMEDIO HOMBRES	comprobacion_valores_plazas_hombres
PLAZAS DE EMPLEO REGISTRADO PROMEDIO MUJERES	comprobacion_valores_plazas_mujeres
PLAZAS DE EMPLEO REGISTRADO ULTIMO	comprobacion_sum_plazas_ult
	comprobacion_valores_plazas_ult
PLAZAS DE EMPLEO REGISTRADO ULTIMO HOMBRES	comprobacion_valores_plazas_ult_hombres
PLAZAS DE EMPLEO REGISTRADO ULTIMO MUJERES	comprobacion_valores_plazas_ult_mujeres
EMPLEO REGISTRADO PROMEDIO	comprobacion_sum_empleo
	comprobacion_valores_empleo
EMPLEO REGISTRADO PROMEDIO HOMBRES	comprobacion_valores_emp_hombres
EMPLEO REGISTRADO PROMEDIO MUJERES	comprobacion_valores_emp_mujeres
EMPLEO REGISTRADO ULTIMO	comprobacion_sum_emp_ult
	comprobacion_valores_emp_ult
EMPLEO REGISTRADO ULTIMO HOMBRES	comprobacion_valores_emp_ult_hombres
EMPLEO REGISTRADO ULTIMO MUJERES	comprobacion_valores_emp_ult_mujeres

4.5 Con la variable ventas y remuneraciones, se verifica que los totales deben estar acorde a los límites permitidos, se calculan las variables para ser comparadas con los valores de la base de datos:

```
#####VENTAS TOTALES#####

## Error en sumatoria ventas totales ##
# Si ventas totales != ventas nacionales + exportaciones netas.
# Valida que los valores que toma la variable ventas totales sea igual a la
# suma de ventas nacionales mas exportaciones netas,
# si no es igual, muestra los casos que presentan el error.
# Se validara todos los años.

error_sum_ventas_totales <- empresas %>%
  select(id_empresa, ruc_principal, anio, ventas_totales, ventas_nacionales, exportaciones_netas) %>%
  filter(ifelse(is.na(ventas_totales), 0, ventas_totales) != ifelse(is.na(ventas_nacionales), 0, ventas_nacionales) +
    ifelse(is.na(exportaciones_netas), 0, exportaciones_netas))

if(nrow(error_sum_ventas_totales) > 0){
  guardarExcel('error_ventas_totales', 'Error en sumatoria ventas totales', error_sum_ventas_totales)
}

rm(error_sum_ventas_totales)
```

Objeto validación	Función
VENTAS TOTALES	comprobacion_valores_ventas_totales
VENTAS NACIONALES	comprobacion_sum_ventas_nacionales
VENTAS TARIFA 0%	comprobacion_valores_ventas_tarifa0
VENTAS TARIFA 12%	comprobacion_valores_ventas_tarifa12
INGRESOS CONSTRUCCION	comprobacion_valores_ing_construccion
INGRESOS ARRENDAMIENTOS	comprobacion_valores_ing_arrendamientos
INGRESOS COMISIONES	comprobacion_valores_ing_comision
EXPORTACIONES NETAS	comprobacion_valores_exportaciones
SALARIOS	comprobacion_sum_salarios
	comprobacion_valores_salarios
SALARIOS HOMBRES	comprobacion_valores_salario_hombres
SALARIOS MUJERES	comprobacion_valores_salario_mujeres

4.6 Se valida la variable fecha para las variables que lo registran, se comprueba que la fecha tenga los dígitos necesarios además que se encuentren en un rango de años, se verifica que si tienen fuente deben tener fecha o viceversa:

```
#### VALIDAR FECHAS####
```

```
# Fecha registro ventas
fecha_reg_ventas <- empresas %>%
  select(id_empresa, ruc_principal, anio, fecha_registro_ventas)%>%
  filter( substr(fecha_registro_ventas, 1, 4) < 2005 |
         substr(fecha_registro_ventas, 1, 4) > as.numeric(format(Sys.Date(), "%Y")) |
         nchar(fecha_registro_ventas) != 8)

if (nrow(fecha_reg_ventas) > 0) {
  guardarExcel('fecha_r_ventas', 'Fecha registro ventas incorrecto', fecha_reg_ventas)
}
rm(fecha_reg_ventas)
```

Objeto validación	Función
FECHA REGISTRO VENTAS	verificar_rangos_fecha_reg_ventas
	verificar_caracteres_fecha_reg_ventas
FECHA FUENTE VENTAS	verificar_rangos_fecha_fte_ventas
	verificar_caracteres_fecha_fte_ventas
FECHA FUENTE GEOGRAFIA	verificar_rangos_fecha_reg_geo
	verificar_caracteres_fecha_reg_geo
FECHA FUENTE GEOGRAFIA	verificar_rangos_fecha_fte_geo
	verificar_caracteres_fecha_fte_geo
FECHA REGISTRO DIRECCION	verificar_rangos_fecha_reg_direc
	verificar_caracteres_fecha_reg_direc
FECHA FUENTE DIRECCION	verificar_rangos_fecha_fte_direc
	verificar_caracteres_fecha_fte_direc
FECHA REGISTRO FORMA INSTITUCIONAL	verificar_rangos_fecha_reg_forma
	verificar_caracteres_fecha_reg_forma
FECHA FUENTE FORMA INSTITUCIONAL	verificar_rangos_fecha_fte_forma
	verificar_caracteres_fecha_fte_forma
FECHA REGISTRO ZONIFICACION	verificar_rangos_fecha_reg_zonif
	verificar_caracteres_fecha_reg_zonif
FECHA FUENTE ZONIFICACION	verificar_rangos_fecha_fte_zonificacion
	verificar_caracteres_fecha_fte_zonificacion
FECHA REGISTRO COORDENADAS	verificar_rangos_fecha_reg_coordenadas
	verificar_caracteres_fecha_reg_coordenadas
FECHA FUENTE COORDENADAS	verificar_rangos_fecha_fte_coordenadas
	verificar_caracteres_fecha_fte_coordenadas
FECHA REGISTRO ACTIVIDAD	verificar_rangos_fecha_reg_actividad
	verificar_caracteres_fecha_reg_actividad
FECHA FUENTE ACTIVIDAD	verificar_rangos_fecha_fte_actividad
	verificar_caracteres_fecha_fte_actividad

4.7 Las variables coordenadas y zonificación se verifica que se encuentren dentro de los límites y en el catálogo:

```
#####COORDENADAS#####

# Validar latitud y longitud #
# Verificar que lat y long esten dentro de un rango de Ecuador (aproximado)
validar_coord <- empresas %>%
  select(id_empresa, ruc_principal, anio, punto_x, punto_y)%>%
  filter(!(between(as.numeric(punto_x), x0, x1) | between(as.numeric(punto_y), y0, y1)))

if (nrow(validar_coord) > 0) {
  guardarExcel('coord', 'Empresas con coordenadas erroneas', validar_coord)
}
rm(validar_coord)

##### ZONIFICACION #####

# Validar Zonificacion #
# Verifica codigo de "Zona" que este en el catalogo
validar_zona <- empresas %>%
  select(id_empresa, ruc_principal, anio, zona_censal)%>%
  filter(!(as.numeric(zona_censal) %in% catalogo[[22]]$CODIGO) & !is.na(as.numeric(zona_censal)))

if (nrow(validar_zona) > 0) {
  guardarExcel('z_censal', 'Empresas con codigo z.censal q no esta en catalogo', validar_zona)
  message(validar_zona)
}
rm(validar_zona)
```

Objeto validación	Función
COORDENADAS	verificar_longitud
	verificar_latitud
ZONIFICACION	verificar_catalogo_zona
	verificar_catalogo_sector
	verificar_catalogo_manzana

4.8 Las variables de estratos, tamaños y estructura de la estratificación se vuelven a calcular para realizar una comparación con los valores de la base de datos, además se verifica que se encuentren en el catálogo:

```
##### ESTRATO PLAZAS PROMEDIO #####

## Error en calculo de estrato plazas promedio ##
# Valida que el estrato plazas promedio se encuentre correctamente calculado
# en función de los valores de plazas de empleo promedio, para ello calcula
# un nuevo estrato plazas promedio y compara con la variables predeterminada en la bdd.
# Se validara para todos los años.

estrato_plazas_prom <- empresas %>%
  select(id_empresa, ruc_principal, anio, plazas, estrato_plazasp)

estrato_plazas_prom <- estrato_plazas_prom %>%
  mutate(nuev_estrato_plazasp =
    case_when((is.null(plazas) | is.na(plazas) | plazas <= 0) ~ 0,
              between(plazas, 1, 9) ~ 1,
              between(plazas, 10,49) ~ 2,
              between(plazas, 50, 99) ~ 3,
              between(plazas, 100, 199) ~ 4,
              between(plazas, 200, 9999999999) ~ 5
    )
  )

error_estrato_plazas <- estrato_plazas_prom %>%
  filter(estrato_plazasp!= nuev_estrato_plazasp)

if(nrow(error_estrato_plazas) > 0){
  guardarExcel('Comprob_estrato_plazasp', 'Valida si estrato plazas promedio esta bien calculado',error_estrato_plazas)
}

rm(estrato_plazas_prom, error_estrato_plazas)
```

Objeto validación	Función
ESTRATO PLAZAS PROMEDIO	comprobacion_estrato_plazas_prom
	verificar_catalogo_estrato_plazas_prom
ESTRATO PLAZAS ULTIMO	comprobacion_estrato_plazas_ult
	verificar_catalogo_estrato_plazas_ult
ESTRATO EMPLEO PROMEDIO	comprobacion_estrato_empleo_prom
	verificar_catalogo_estrato_empleo_prom
ESTRATO EMPLEO ULTIMO	comprobacion_estrato_empleo_ult
	verificar_catalogo_estrato_empleo_ult
ESTRATO VENTAS	verificar_catalogo_estrato_ventas
TAMANIO DE EMPRESA PLAZAS PROMEDIO	comprobacion_tamano_plazas_prom
	verificar_catalogo_tamano_plazas_prom
TAMANIO DE EMPRESA PLAZAS ULTIMO	comprobacion_tamano_plazas_ult
	verificar_catalogo_tamanou_plazas
TAMANIO DE EMPRESA EMPLEO PROMEDIO	comprobacion_tamano_empleo_prom
	verificar_catalogo_tamano_empleo_prom
TAMANIO DE EMPRESA EMPLEO ULTIMO	comprobacion_tamano_empleo_ult
	verificar_catalogo_tamanou_empleo
ORIGEN DE LA ESTRATIFICACION	comprobacion_origen_estratif
	verificar_catalogo_origen
ESTRUCTURA DE LA ESTRATIFICACION	comprobacion_estructura_estratif
	verificar_catalogo_estructura

Nota Importante:

Es necesario tener en cuenta que las validaciones nos permiten verificar variables de ubicación y dirección, para facilidad de su uso se propone correr primero la base del año proceso con todas las variables necesarios y luego correr al base de todos los años debido a que el script detecta errores en todos los años de las empresas activas.

Elaborado por:	Jacqueline Imbaquingo	Firma: 
Revisado por:	Alejandro Balseca	Firma: 
Aprobado por:	Libertad Trujillo	Firma: 
Fecha de presentación del presente informe:		28/05//2020