

# ModeloEvaluacion

---

## Selección de las fuentes de información

Para los datos básicos y financieros de las empresas, tomamos los siguientes archivos de la página de la Supersociedades:

- datosBasicosComplete.xlsx
- Plenas - Individuales2017.xlsx
- Plenas - Individuales2018.xlsx
- Plenas - Individuales2019.xlsx

Primera iteración:

Código CIIU seleccionado: G4711

Macrosector: Comercio

Descripción: Comercio al por menor en establecimientos no especializados con surtido compuesto principalmente por alimentos, bebidas o tabaco.

Esta clase incluye:

- Los establecimientos no especializados de comercio al por menor de productos cuyo surtido está compuesto principalmente de alimentos (víveres en general), bebidas o tabaco. No obstante, expenden otras mercancías para consumo de los hogares tales como vestuario, electrodomésticos, muebles, artículos de ferretería, cosméticos, entre otros. Suelen realizar este tipo de actividad los denominados supermercados, cooperativas de consumidores, comisariatos y otros establecimientos similares. También se incluyen las tiendas, los graneros, entre otros, que se encuentran en los pueblos o en barrios tradicionales.

Esta clase excluye:

- El expendio de comidas preparadas en restaurantes, cafeterías y por autoservicio.

Al realizar los cargues iniciales de información, nos dimos cuenta de que cruzaban muy pocas empresas, el conjunto de datos seleccionado no era suficiente, por lo que decidimos utilizar otro CIIU.

Segunda iteración:

Código CIIU seleccionado: B0722

Descripción: Extracción de oro y otros metales preciosos

Esta clase incluye:

- La extracción de oro, plata y otros metales del grupo del platino (osmio, iridio, rodio, rutenio y paladio).
- Las actividades realizadas para extraer el oro existente en los lechos de los ríos sin importar el sistema de extracción empleado (barequeo, motobombas, draguetas, dragas, elevadores, monitores u otros).
- La extracción de los metales preciosos se realiza a través de dos métodos: de veta o filón, que consiste en la extracción manual, mecanizada o semimecanizada de oro y de plata presentes en las rocas formando venas, vetas o filones.

- Las actividades o procesos físicos necesarios para separar el oro de la roca que lo contiene, conocidos como procesos de beneficio del mineral, de los cuales los más comunes son la trituración y la molienda (pulverización).
- Otros procesos tales como lavado (mazamorreo) hasta separar el oro y la plata de otros elementos o impurezas, siempre y cuando se realicen por cuenta del explotador y en sitios cercanos a la mina.
- El segundo método consiste en la extracción de oro o platino de aluviones (concentración de mineral en el lecho de los ríos), el cual se realiza por diferentes sistemas de extracción, tales como: barequeo (mazamorreo); pequeña minería, representada por grupos de trabajadores que utilizan motobombas, elevadores y draguetas; mediana minería, utilizando maquinaria como retroexcavadoras y buldózers, y la gran minería que realiza la extracción de metales preciosos por medio de dragas de cucharas.

Esta clase excluye:

- Los servicios de apoyo para la extracción de oro y metales preciosos. Se incluyen en la clase 0990, «Actividades de apoyo para otras actividades de explotación de minas y canteras».

## Proceso de carga de los datos

```
library(tidyverse)
library("readxl")
library("dplyr")
```

1. Cargamos los datos básicos de las empresas

```
#Revisamos como son nuestros datos para saber si tenemos que realizar algún
#ajuste a la carga
#file.show("./data/datosBasicosComplete.xlsx")

#Como el archivo no tiene forma de tabla al principio, debemos realizar la
#carga, ignorando las primeras filas del archivo.

#Cargar los archivos a un dataframe
pd_datos_basicos <- read_excel("./data/datosBasicosComplete.xlsx",
                              sheet = "Reporte", skip=8, col_types = c("text",
"text", "text", "text", "text", "text", "text", "text", "text", "text",
"text", "text", "text", "text", "text", "date", "text", "date", "text", "date", "text",
"text"))

pd_datos_basicos %>%
  mutate(`Órgano Societario` = as.factor(`Órgano Societario`),
         `Etapa Situación` = as.factor(`Etapa Situación`)) -> pd_datos_basicos

head(pd_datos_basicos)

## # A tibble: 6 x 23
##   NIT   `Razón social` `Código CIU` `Tipo Societari` `Objeto Social`
##   <chr> <chr>          <chr>          <chr>          <chr>
## 1 1001~ NOREÑA MANRI~ 0          PERSONA NATURAL <NA>
## 2 1001~ PEÑA RAMIREZ ~ H5229      PERSONA NATURAL <NA>
## 3 1002~ GONZALEZ SANC~ G4731      PERSONA NATURAL <NA>
## 4 1002~ RODRIGO JAVIE~ L6810      PERSONA NATURAL <NA>
## 5 1002~ BUITRAGO GONZ~ H4923      PERSONA NATURAL <NA>
## 6 1005~ KAREN JULIETH~ M7500      PERSONA NATURAL <NA>
## # ... with 18 more variables: `Dirección Notificación Judicial` <chr>, `Ciudad
## #   Notificación Judicial` <chr>, `Departamento Notificación Judicial` <chr>,
```

```
## # `Teléfono Notificación Judicial` <chr>, `Dirección Domicilio` <chr>,
## # `Ciudad Domicilio` <chr>, `Departamento Domicilio` <chr>, `Apartado
## # Domicilio` <chr>, `E-Mail` <chr>, Web <chr>, Estado <chr>, `Fecha
## # Estado` <dtm>, Situación <chr>, `Fecha Situación` <dtm>, `Etap
## # Situación` <fct>, `Fecha Etapa` <dtm>, `Nombre Representante Legal` <chr>,
## # `Órgano Societario` <fct>
```

2. Filtramos los datos del CIU seleccionado

```
library(dplyr)

pd_datos_basicos_flt <- pd_datos_basicos[,c("NIT", "Razón social", "Código CIU",
      "Ciudad Domicilio", "Departamento Domicilio", "Estado", "Situación",
      "Órgano Societario", "Etapa Situación")]

names(pd_datos_basicos_flt) = c("NIT", "razon_social", "CIU", "ciudad",
      "departamento", "estado", "situacion",
      "organo_societario", "etapa_situacion")

pd_datos_basicos_flt <- filter(pd_datos_basicos_flt, CIU == "B0722" &
      situacion == "ACTIVA")

head(pd_datos_basicos_flt)

## # A tibble: 6 x 9
##   NIT   razon_social CIU   ciudad departamento estado situacion organo_societar~
##   <chr> <chr>         <chr> <chr>   <chr>         <chr> <chr>      <fct>
## 1 8002~ GRUPO DE BU~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 2 8110~ MINERA CROE~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 3 8110~ NUEVA CALIF~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 4 8110~ COLOMBIA GO~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 5 8110~ NEGOCIOS MI~ B0722 MEDEL~ ANTIOQUIA   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## 6 8300~ ECO ORO MIN~ B0722 BUCAR~ SANTANDER   INSPE~ ACTIVA   ACTIVIDAD ECONO~
## # ... with 1 more variable: etapa_situacion <fct>
```

3. Cargamos los datos financieros

```
pd_datos_fin_2017 <- read_excel("./data/Plenas - Individuales2017.xlsx",
      sheet = "Estado de Resultado Integral")

pd_datos_fin_2017 <- pd_datos_fin_2017[,c("Nit", "Periodo", "Costo de ventas",
      "Costos de distribución", "Gastos de administración", "Otros gastos, por función",
      "Costos financieros", "Gasto (ingreso) por impuestos, operaciones continuadas",
      "Ingresos de actividades ordinarias", "Otros ingresos", "Ingresos financieros")]

names(pd_datos_fin_2017) = c("NIT", "Periodo", "costo_ventas",
      "costo_distribucion", "gastos_administracion", "otros_gastos",
      "costos_financieros", "gasto_impuestos_operaciones",
      "ingresos_actividades_ordinarias", "otros_ingresos", "ingresos_financieros")

datos_completos_fin <- merge(pd_datos_basicos_flt, pd_datos_fin_2017,
      by.x="NIT", by.y="NIT")
```

Para efectos del ejercicio, no tomaremos el archivo de 2017, ya que el archivo 2018 tiene los datos de 2017

con la nueva norma.

```
pd_datos_fin_2018 <- read_excel("./data/Plenas - Individuales2018.xlsx",
                                sheet = "ERI" )

pd_datos_fin_2018 <- pd_datos_fin_2018[,c("Nit", "Periodo", "Costo de ventas",
    "Gastos de ventas", "Gastos de administración", "Otros gastos",
    "Costos financieros", "Ingreso (gasto) por impuestos",
    "Ingresos de actividades ordinarias", "Otros ingresos", "Ingresos financieros")]

names (pd_datos_fin_2018) = c("NIT", "Periodo", "costo_ventas", "gastos_ventas",
    "gastos_administracion", "otros_gastos", "costos_financieros", "gasto_impuestos",
    "ingresos_actividades_ordinarias", "otros_ingresos", "ingresos_financieros" )

datos_completos_2018 <- merge (pd_datos_basicos_flt, pd_datos_fin_2018,
                                by.x="NIT", by.y="NIT")

#Le damos formato a los periodos

datos_completos_2018$Periodo[datos_completos_2018$Periodo == "Periodo Anterior"] <-
    "2017"
datos_completos_2018$Periodo[datos_completos_2018$Periodo == "Periodo Actual"] <-
    "2018"
```

```
pd_datos_fin_2019 <- read_excel("./data/Plenas - Individuales2019.xlsx",
                                sheet = "ERI" )

#Revisar Costos de distribución
pd_datos_fin_2019 <- pd_datos_fin_2019[,c("Nit", "Periodo", "Costo de ventas",
    "Gastos de administración", "Otros gastos", "Costos financieros",
    "Ingreso (gasto) por impuestos", "Ingresos de actividades ordinarias",
    "Otros ingresos", "Ingresos financieros")]

names (pd_datos_fin_2019) = c("NIT", "Periodo", "costo_ventas",
    "gastos_administracion", "otros_gastos", "costos_financieros",
    "gasto_impuestos", "ingresos_actividades_ordinarias", "otros_ingresos",
    "ingresos_financieros" )

datos_completos <- merge (pd_datos_basicos_flt, pd_datos_fin_2019,
                                by.x="NIT", by.y="NIT")

datos_completos$Periodo[datos_completos$Periodo == "Periodo Actual"] <- "2019"

datos_completos <- filter(datos_completos, Periodo == "2019")

#Eliminamos variable diferente a 2019
datos_completos_2018 <- select(datos_completos_2018, -gastos_ventas)

#Unimos los 2 dataframes
datos_completos = rbind(datos_completos, datos_completos_2018)
datos_completos %>% mutate( razon_social = as.factor(razon_social),
    CIIU = as.factor(CIIU), ciudad = as.factor(ciudad),
    departamento = as.factor(departamento),
    estado =as.factor(estado), Periodo = as.factor(Periodo),
    situacion = as.factor(situacion)) -> datos_completos
```