

```

# Archivo banda_anchafija.R

library(readr)
library(tidyverse)
library(readxl)

# Establece directorio de trabajo
setwd("/media/Box/Aprendizaje_Maquina/Proyecto")

# Crear las bases de datos para trabajar
source("creating_baf.R") # Accesos de banda ancha fija a junio/2019 BIT del IFT
source("creating_conapo.R") # Indice marginacion y porcentaje pob con menos de 2
salarios min, 2015 CONAPO
source("creating_inafed.R") # Superficie de municipios, INAFED
#source("creating_indiceinfraestructura.R") # Indice infraestructura TII - centro
de estudios IFT
source("creating_indicadores_serviciostelecom_viviendas_.R") # Indicadores de
disponiblidad de servicios de telecomunicaciones Encuesta intercensal 2015, INEGI
source("creating_hogares.R") # Hogares por municipios Encuesta intercensal 2015,
INEGI
source("creating_poblacion.R") # Poblacion por municipios Encuesta intercensal
2015, INEGI
source("creating_humandvelop_index.R") # Programa de las Naciones Unidas para el
Desarrollo (PNUD), datos del Índice de desarrollo humano 2015

# Elimina variables auxiliares
rm(index,
left_path,name_state,right_path,states_list,cleaing_hog_state,cleaing_pop_state)

# Consolidamos las bases
df <- left_join(hogares2015,poblacion2015, by = "K_ENTIDAD_MUNICIPIO")
df <- left_join(df,conapo, by = "K_ENTIDAD_MUNICIPIO")
df <- left_join(df,INAFED, by = "K_ENTIDAD_MUNICIPIO")
df <- left_join(df,indicadores_servicios2015, by = "K_ENTIDAD_MUNICIPIO")
df <- left_join(df,BAF_062019, by = "K_ENTIDAD_MUNICIPIO")

## El detalle de accesos a nivel municipal con NA se imputa con cero
df <- df %>% mutate_all(~replace(., is.na(.), 0))

df$SUPERFICIE<- as.numeric(df$SUPERFICIE)
df$P02SM<- as.numeric(df$P02SM)

# Se eliminan columnas sin interes para el analisis
df$K_ENTIDAD<-NULL
df$K_MUNICIPIO<-NULL
df$ANIO<-NULL
df$MES<-NULL

# df1 <- df
# Se cambian algunas variables de caracteres a numericas
# df1 <- df1 %>% select(-K_ENTIDAD_MUNICIPIO,-GM)
df$SPRIM <- as.numeric(df$SPRIM)
df$OVSD <- as.numeric(df$OVSD)
df$VHAC <- as.numeric(df$VHAC)
df$OVPT <- as.numeric(df$OVPT)
df$`PL<5000` <- as.numeric(df$`PL<5000`)

# Crea variable de densidad de hogares por kilometros cuadrados

```

```

df$DENS_HABS <- 0
for (i in 1:nrow(df)){
  df$DENS_HABS[i]<- df$HOGARES[i]/df$SUPERFICIE[i]*100
}

# Crea variable de penetracion de BAF por cada 100 hogares y la penetracion de
cable coaxial + fibra optica
#df <- df %>% mutate(PEN_BAF_HOGS = df$ALL_ACCESS/df$HOGARES*100)
df$PEN_BAF_HOGS_COAXFO <- 0
for (i in 1:nrow(df)){
  df$PEN_BAF_HOGS_COAXFO[i]<- df$COAX_F0[i]/df$HOGARES[i]*100
}

# Crea variable de penetracion de BAF por cada 100 habitantes y la penetracion de
cable coaxial + fibra optica
#df <- df %>% mutate(PEN_BAF_HABS = df$ALL_ACCESS/df$HOGARES*100)
df$PEN_BAF_HABS_COAXFO <- 0
for (i in 1:nrow(df)){
  df$PEN_BAF_HABS_COAXFO[i]<- df$COAX_F0[i]/df$POBLACION[i]*100
}

# Creamos una columna para clasificar los municipios segun su grado de penetracion.
#La clasificación de Penetracion de Fibra Óptica y Cable Coaxial:
# Sin cobertura=0
# Muy baja 0>25%
# Baja 25%>50%
# Media 50%>75%
# Alta 75%>100%
# Muy Alta 100%

df$CLASS_PEN_BAF_HOGS_COAXFO <- 0

for (index in 1:nrow(df)){
  df$CLASS_PEN_BAF_HOGS_COAXFO[index] <-
if_else(df$PEN_BAF_HOGS_COAXFO[index]==0,0,
      if_else(df$PEN_BAF_HOGS_COAXFO[index]<=25,1,
      if_else(df$PEN_BAF_HOGS_COAXFO[index]>25,2,
      if_else(df$PEN_BAF_HOGS_COAXFO[index]>50,3,
      if_else(df$PEN_BAF_HOGS_COAXFO[index]>75,4,
      if_else(df$PEN_BAF_HOGS_COAXFO[index]>100,5,0)
      )
      )
      )
      )
      )
}

# Creamos una columna para clasificar los municipios segun su grado de penetracion.
#La clasificación de Penetracion de Fibra Óptica y Cable Coaxial:
# Sin cobertura=0
# Muy baja 0>25%
# Baja 25%>50%
# Media 50%>75%
# Alta 75%>100%
# Muy Alta 100%

df$CLASS_PEN_BAF_HABS_COAXFO <- 0

for (index in 1:nrow(df)){
  df$CLASS_PEN_BAF_HABS_COAXFO[index] <-
if_else(df$PEN_BAF_HABS_COAXFO[index]==0,0,
      if_else(df$PEN_BAF_HABS_COAXFO[index]<=25,1,
      if_else(df$PEN_BAF_HABS_COAXFO[index]>25,2,
      if_else(df$PEN_BAF_HABS_COAXFO[index]>50,3,
      if_else(df$PEN_BAF_HABS_COAXFO[index]>75,4,
      if_else(df$PEN_BAF_HABS_COAXFO[index]>100,5,0)
      )
      )
      )
      )
      )
}

```

```

for (index in 1:nrow(df)){
  df$IS_PEN_BAF_HABS_COAXFO[index] <- if_else(df$CLASS_PEN_BAF_HABS_COAXFO[index]!
=0,1,0)
}

#----- Agregamos una columna que nos diga la region socioeconomica a la que
pertenece el municipio

RegionesSocioEcono <- read_csv("RegionesSocioEcono.csv",
                                col_types = cols(NOM_ABRE_ENTIDAD = col_skip(),
                                                    NOM_ENTIDAD = col_skip(), NUM =
col_skip()))
df <- df %>% mutate(K_ENTIDAD = substr(K_ENTIDAD_MUNICIPIO,1,2))
df <- left_join(df,RegionesSocioEcono, by = "K_ENTIDAD")

df$K_ENTIDAD<-NULL

for (index in 1:nrow(df)){
  df$REG_SOCIOECONOM[index] <- if_else(df$REG_SOCIOECONOM[index]=="Centronorte",1,
                                     if_else(df$REG_SOCIOECONOM[index]=="Centrosur",2,
                                     if_else(df$REG_SOCIOECONOM[index]=="Norest
3,
                                     if_else(df$REG_SOCIOECONOM[index]=="
4,
                                     if_else(df$REG_SOCIOECONOM
5,
                                     if_else(df$REG_SOC
6,
                                     if_else(df$
7,8))))))
}

# Adjuntamos los datos de indices de derechos humanos
df <- left_join(df, hd_index2015, by = "K_ENTIDAD_MUNICIPIO")

# ---- Seleccionamos columnas para el analisis y escribimos la base para Python
df1<- df %>% select(K_ENTIDAD_MUNICIPIO, HOGARES, POBLACION, SUPERFICIE, DENS_HOGS,
DENS_HABS,
                ANALF, SPRIM, ANOS_PROMEDIO_DE_ESCOLARIDAD,OVSAE, OVSEE,
'PL<5000', P02SM, INGRESOPC_ANUAL,
                DISP_INTERNET, DISP_TV_PAGA, DISP_TEL_CELULAR, DISP_TEL_FIJO,
                NUM_OPS, CLASS_PEN_BAF_HABS_COAXFO, IS_PEN_BAF_HABS_COAXFO)

# Base para el problema de detecion de penetracion sin importar el nivel
write_csv(df1, "BAF_06209_P1.csv")

# Base para el problema de detecion del nivel de penetracion
write_csv(df1, "BAF_06209_P2.csv")

# Base completa para EDA
write_csv(df, "BAF_06209_EDA.csv")

```

```
# df1 <- df1 %>% select(ALL_ACCESS,COAX_F0,HOGARES,POBLACION, ANALF, SPRIM,  
OVSDE,OVSEE, OVSAE, VHAC, OVPT,`PL<5000`,P02SM,IM)
```

```
#df1 <- df1 %>%select(-PEN_CLASS,PEN_CLASS)
```

```
#library(corrplot)
```

```
#M<-cor(df1)
```

```
#corrplot(M, method="circle")#
```

```
#corrplot(M, method="number")
```