

```

# Archivo creating_baf.R

library(readr)
library(tidyverse)
library(ggplot2)
#library("tidyverse")

#####---- Carga base de datos de accesos de banda ancha fija ----#####

BAF_raw <- read_csv("TODO_BAF/TD_ACC_BAF_ITE_VA.csv",
                    col_types = cols(ANIO = col_character(),
                                     K_ACCESO_INTERNET = col_character(),
                                     K_ENTIDAD = col_character(),
                                     K_MUNICIPIO = col_character(),
                                     MES = col_character()), locale =
                    locale(encoding = "ISO-8859-1"))

#####---- Modificamos los datos atipicos de de Mayapan (Yucatan) y Rayon (Estado
de Mexico) ----#####

# Los municipios circundantes a Mayapan no tienen penetracion de BAF en cable
coaxial ni fibra optica.
# Se considera que Mayapan tampoco (error en la base del BIT)

indice_mayapan = (BAF_raw$ANIO == "2019")*(BAF_raw$ENTIDAD ==
"Yucatan")*(BAF_raw$MUNICIPIO == "Mayapan")
for (i in which(indice_mayapan==1)){
  BAF_raw$A_TOTAL_E[i]<-0
}

rm(indice_mayapan)
# El nivel reportado de penetracion para Rayon es muy alto debido a inconsistencias
con los accesos
# reportados por Megacable en los meses 04, 05 y 06 de 2019. Se les imputa el valor
# historico mas reciente, que es congruente con los datos historicos previos.

indice_rayon = (BAF_raw$ANIO == "2019")*(BAF_raw$ENTIDAD ==
"Mexico.")*(BAF_raw$MUNICIPIO == "Rayon.")*(BAF_raw$EMPRESA ==
"MEGACABLE")*(BAF_raw$TECNO_ACCESO_INTERNET == "Cable Coaxial")

indice_rayon_03 = which((indice_rayon*(BAF_raw$MES == "03"))==1)
indice_rayon_04 = which((indice_rayon*(BAF_raw$MES == "04"))==1)
indice_rayon_05 = which((indice_rayon*(BAF_raw$MES == "05"))==1)
indice_rayon_06 = which((indice_rayon*(BAF_raw$MES == "06"))==1)

BAF_raw$A_TOTAL_E[indice_rayon_04] <- BAF_raw$A_TOTAL_E[indice_rayon_03]
BAF_raw$A_TOTAL_E[indice_rayon_05] <- BAF_raw$A_TOTAL_E[indice_rayon_03]
BAF_raw$A_TOTAL_E[indice_rayon_06] <- BAF_raw$A_TOTAL_E[indice_rayon_03]

#####---- Transformacion y limpieza de la base de accesos ----#####

## Crea una llave para cruzar con las otras bases
BAF_study <- BAF_raw
BAF_study$K_MUNICIPIO <- substr(BAF_study$K_MUNICIPIO,3,5)
BAF_study <- BAF_study %>% mutate(K_ENTIDAD_MUNICIPIO = paste(K_ENTIDAD,
K_MUNICIPIO, sep = ""))

## seleccion variable y alarga la bases de accesos por tecnologia

```

```

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

# Renombramos columnas de accesos segun su tecnologia
names(BAF_study)[6] <- "CABLE_COAXIAL"
names(BAF_study)[7] <- "DSL"
names(BAF_study)[8] <- "FIBRA_OPTICA"
names(BAF_study)[9] <- "SATELITAL"
names(BAF_study)[10] <- "TERRESTRE_FIJO_INALAMBRICO"
names(BAF_study)[11] <- "OTRAS_TECNOLOGIAS"
names(BAF_study)[12] <- "SIN_TECNOLOGIA_ESPECIFICADA"

# Agregamos columna de todos los accesos del municipio
BAF_study <- BAF_study %>% mutate(ALL_ACCESS =
CABLE_COAXIAL+DSL+FIBRA_OPTICA+SATELITAL+TERRESTRE_FIJO_INALAMBRICO+OTRAS_TECNOLOGI

# Agregamos columna de todos los accesos cable coaxial y fibra optica del municipio
BAF_study <- BAF_study %>% mutate(COAX_F0 = CABLE_COAXIAL+FIBRA_OPTICA)

# Se excluyen los datos de accesos que no tienen ubicacion de municipio
BAF_062019 <- subset(BAF_study, ANIO == "2019" & MES == "06" & K_ENTIDAD != "99"
& K_MUNICIPIO != "999" )

# ----- Contamos la cantidad de empresas presentes en el municipio

# Filtramos la base en crudo para junio de 2019
BAF_raw062019<- BAF_raw %>%filter(ANIO=="2019" & MES == "06" & K_ENTIDAD != "99" &
K_MUNICIPIO != "999" )

## Crea una llave para cruzar con las otras bases
BAF_raw062019$K_MUNICIPIO <- substr(BAF_raw062019$K_MUNICIPIO,3,5)
BAF_raw062019 <- BAF_raw062019 %>% mutate(K_ENTIDAD_MUNICIPIO = paste(K_ENTIDAD,
K_MUNICIPIO, sep = ""))

# Crea la base auxiliar que tiene por clave de municipio la cantidad de empresas de
BAF en esta
BAF_study_ops <- BAF_raw062019 %>% select(K_ENTIDAD_MUNICIPIO) %>% unique()
BAF_study_ops$NUM_OPS <-NA

for (i in 1:nrow(BAF_study_ops)){
  folio =BAF_study_ops$K_ENTIDAD_MUNICIPIO[i]
  n <- BAF_raw062019 %>% select(EMPRESA,K_ENTIDAD_MUNICIPIO) %>%
filter(K_ENTIDAD_MUNICIPIO == BAF_study_ops$K_ENTIDAD_MUNICIPIO[i]) %>% unique()
%>% nrow()
  BAF_study_ops$NUM_OPS[i]= n
}

# Agrega un columna de la cantidad total de empresas que cuentan con al menos un
acceso de BAF en cada municipio
BAF_062019 <- left_join(BAF_062019,BAF_study_ops, by = "K_ENTIDAD_MUNICIPIO")

#----- Escribe la base final
write_csv(BAF_062019, "BAF_06209.csv")

# Eliminamos objetos auxiliares
rm(BAF_raw,BAF_study,folio,i,n,BAF_study_ops,BAF_raw062019)

####---- Resúmenes

```

```

  group_by(TECNO_ACCESO_INTERNET) %>% summarize(n=sum(A_TOTAL_E,na.rm = TRUE)/
1000000) %>%
  ungroup() %>% mutate(distrib_n = n/18.85439)

# Numero de accesos de BAF basados en DSL en 06/2019 (11.34751 millones)
BAF_raw %>% filter(ANIO=='2019' & MES=='06' & TECNO_ACCESO_INTERNET=='DSL' ) %>%
select(A_TOTAL_E) %>% sum()/1000000

# Numero de accesos de BAF basados en fibra óptica o cable coaxial en 06/2019
(11.34751 millones)
BAF_raw %>% filter(ANIO=='2019' & MES=='06' & (TECNO_ACCESO_INTERNET=='Cable
Coaxial' | TECNO_ACCESO_INTERNET=='Fibra Óptica' ) ) %>% select(A_TOTAL_E) %>%
sum()/1000000

# Empresas y concesionarios de toda la DB (una empresas esta formada de varios
concesionarios)
unique(BAF_raw$EMPRESA)
unique(BAF_raw$CONCESIONARIO)

# Empresas de Grupo Televisa
BAF_raw %>% select(GRUPO, EMPRESA) %>% unique() %>% filter(GRUPO=="GRUPO TELEVISA")

# Empresas de grupo America Movil
BAF_raw %>% select(GRUPO, EMPRESA) %>% unique() %>% filter(GRUPO=="AMÉRICA MÓVIL")

# Empresas de grupo MEGACABLE-MCM
BAF_raw %>% select(GRUPO, EMPRESA) %>% unique() %>% filter(GRUPO=="MEGACABLE-MCM")

# Empresas de grupo TOTALPLAY
BAF_raw %>% select(GRUPO, EMPRESA) %>% unique() %>% filter(GRUPO=="TOTALPLAY")

# Distribución de acceso por grupos economicos
BAF_raw%>% filter(ANIO=='2019' & MES=='06') %>% select(GRUPO,A_TOTAL_E) %>%
  group_by(GRUPO) %>% summarize(n=sum(A_TOTAL_E,na.rm = TRUE)/1000000) %>%
  ungroup() %>% mutate(distrib_n = n/18.85439) %>% arrange(desc(distrib_n))

# Distribucion de acceso por tecnologia y por grupos de empresas
BAF_raw%>% filter(ANIO=='2019' & MES=='06') %>%
select(GRUPO,TECNO_ACCESO_INTERNET,A_TOTAL_E) %>%
  group_by(GRUPO,TECNO_ACCESO_INTERNET) %>% summarize(n=sum(A_TOTAL_E,na.rm =
TRUE)/1000000) %>%
  ungroup() %>% spread(TECNO_ACCESO_INTERNET, n)%>% mutate_all(~replace(.,
is.na(.), 0)) %>%
  mutate(Sin_tecnologia_especificada = `Sin tecnología especificada`+`Sin Tecnología
especificada`) %>%
  select(-c(`Sin tecnología especificada`,`Sin Tecnología especificada`))

# Distribucion de acceso por tecnologia y por por grupos de empresas mas
importantes
BAF_raw%>% filter(ANIO=='2019' & MES=='06') %>%
select(GRUPO,TECNO_ACCESO_INTERNET,A_TOTAL_E) %>%
  group_by(GRUPO,TECNO_ACCESO_INTERNET) %>% summarize(n=sum(A_TOTAL_E,na.rm =
TRUE)/1000000) %>%
  ungroup() %>% spread(TECNO_ACCESO_INTERNET, n)%>% mutate_all(~replace(.,
is.na(.), 0)) %>%
  mutate(Sin_tecnologia_especificada = `Sin tecnología especificada`+`Sin Tecnología
especificada`) %>%
  select(-c(`Sin tecnología especificada`,`Sin Tecnología especificada`)) %>%
  filter(GRUPO == "AMÉRICA MÓVIL"|GRUPO == "GRUPO TELEVISA"| GRUPO == "AMÉRICA

```