

Universidad del Valle de Guatemala
Data Science
Sección 10
Lynette Garcia



Excelencia que trasciende

DEL VALLE
GRUPO EDUCATIVO

CODEBOOK

Alfredo Quezada 191002
Estuardo Hernández 19202
Marco Ramirez 19588

VARIABLES:

Dentro de los datos se encuentran un listado detallado de los establecimientos educativos del país, cabe destacar que cada uno de estos conjuntos de datos se encuentran distribuidos por departamento, es decir que se encontraban en un conjunto los datos correspondientes al departamento de Guatemala, en otro conjunto de datos los correspondientes a Izabal, por hacer mención y así seguidos con cada uno de los 22 departamentos del país.

Cada conjunto de datos contenían las siguientes variables:

1. CODIGO: Esta es una variable de forma "chr", en la cual se encuentran listados los códigos de registro de cada uno de los centros educativos.
2. DISTRITO: Esta es una variable de forma "chr", en la cual podemos ver cómo se encuentran listados los centros educativos en base al distrito en el que se encuentran.
3. DEPARTAMENTO: Como su nombre lo indica, esta variable muestra el nombre del departamento en el que se encuentra el establecimiento educativo.
4. MUNICIPIO: Como su nombre lo indica, se listan de manera ordenada el nombre del municipio en el que se encuentran ubicados los establecimientos.
5. ESTABLECIMIENTO: En esta variable se lista de manera ordenada el nombre del centro educativo, el nombre bajo el cual está registrado.
6. DIRECCION: En esta variable se encuentra la ubicación física del centro estudiantil, la ubicación en la cual se encuentra establecido.
7. TELEFONO: Como su nombre lo indica, se encuentra listado el número de teléfono que tiene el centro educativo, en algunos datos no se fue proporcionado, por eso se encuentra bajo N/A.
8. SUPERVISOR: En esta variable se encuentra el nombre de la persona que designada por el MINEDUC para supervisar al centro estudiantil.
9. DIRECTOR: En esta variable se encuentra el nombre de la persona que funge como director del centro educativo, en algunos datos no se encuentra dicha persona, por eso se encuentra bajo N/A.
10. NIVEL: En esta variable se encuentra el nivel educativo que ofrece el establecimiento, puede ser: "PRIMARIO", "BASICO" o "DIVERSIFICADO"
11. SECTOR: En esta variable se encuentra establecido el sector al que pertenece el centro educativo, el cual puede ser ya sea "PRIVADO", "OFICIAL", "MUNICIPAL" o "COOPERATIVA"
12. AREA: En esta variable se encuentra listado a qué área pertenece el centro educativo, el cual puede ser: "RURAL" o "URBANO"
13. STATUS: En esta variable se encuentra en el estado en el que se encuentra el centro educativo, el cual puede ser "CERRADO TEMPORALMENTE", "ABIERTO" o "TEMPORALMENTE TITULOS"
14. MODALIDAD: En esta variable se encuentra establecido qué tipo de modalidad maneja el centro educativo, puede ser ya sea "MONOLINGUE" o "BILINGUE".
15. JORNADA: En la variable se encuentra listado qué jornadas maneja el centro educativo, dentro de estas variables se encuentra: "DOBLE", "VESPERTINA" o "MATUTINA".

16. PLAN: Como su nombre lo indica, en esta variable se encuentra establecido que plan maneja el centro, el cual puede ser "DIARIO(REGULAR)" o "FIN DE SEMANA".
17. DEPARTAMENTAL: En esta variable se encuentra establecido el departamental que se encarga de cada uno de los establecimientos (la departamental es la entidad que se encarga de supervisar al centro educativo, de allí sale la variable de supervisor.)

CODIGOS:

```
db$TELEFONO<-gsub("-", "", as.character(db$TELEFONO))

db$TELEFONO<-ifelse(nchar(db$TELEFONO)<8, NA, substr(db$TELEFONO, 1, 8))
```

Este código se utilizó para la limpieza de la variable teléfono, como se puede observar, la finalidad es que cualquiera de los valores dentro de la columna que cumpla con estas especificaciones, puede quedarse, de lo contrario se marcará como NA

```
#Reparando el director

# db[db$DIRECTOR == "-" & !is.na(db$DIRECTOR), ]$DIRECTOR <- NA
# db[db$DIRECTOR == "--" & !is.na(db$DIRECTOR), ]$DIRECTOR <- NA
```

De igual manera, para los datos del director, como se puede ver se está asegurando que dentro de la variable se encuentren todos los datos sin algún carácter no requerido, en caso existirá alguno, se marcará como NA