**Documentación**

1. **Limpieza de datos:**

Para esta sección se tuvo que realizar diferentes cambios a las bases para poder utilizarlas en el proceso:

* Principalmente se cambio el formato de guardado con el que se llegó los archivos, pasándola a XLSX en caso de ser necesario, dado que un archivo tenia conflictos para poderlo leer desde R.
* Se detecto que las llaves principales que relacionan las tablas (nombre de municipio y departamento) tenían problemas ya sea por caracteres especiales, acentuación, dobles espacios y nombres cortados, por lo cual se procedió a solucionar cada problema como se ve en el código de R para poder realizar uniones entre los archivos.

1. **Utilización de SQL:**

Una vez se tenía limpia la información y las llaves funcionaban, se procedió a utilizar SQLite para poder generar el informe. Se utiliza el paquete RSQLite, que permite usar un motor de SQL para poder generar las consultas necesarias. Para poderlo usar se hizo lo siguiente:

* Se cargo las bases a la conexión de SQLite, las bases que se cargaron son las de municipios y prestadores.
* Se creo una tabla puente que permitía tener la relación de los prestadores, la superficie y población del territorio donde estaban ubicados. Esta tabla agrupa por cada territorio la cantidad de prestadores por cada tipo, esto se hizo para facilitar los cálculos.
* Se realizaron 5 consultas para poder sacar la información relevante para presentar el informe. Cada consulta se materializo en una tabla para poder generar gráficas, tablas o consultas puntuales.

1. **Procesamiento de la información:**

En esta sección se va a describir cada tabla que contiene:

* Tabla\_1\_1: Se realizo un conteo por cada tipo de prestador que este activo en Colombia, luego esta información se utilizo paga generar un diagrama de pastel.
* Tabla\_1: Se realizo un conteo por tipo y región a la que pertenece el prestador, luego se genero por cada tipo de prestador un diagrama de barras.
* Tabla\_2: Se genero un conteo por cada tipo de prestador y su razón social, luego se agregaron los totales por filas y columnas en esta tabla.
* Tabla\_3: Se cambio el formato de fecha que estaba en texto para poder hacer una resta contra el día actual para poder tener el tiempo en meses, días y años que ha estado vigente un prestador. Con esta información se utilizó la columna que calculo el tiempo en años de vigencia para generar un histograma y ver como se están distribuyendo los tiempos.
* Tabla\_4: Se generó unos cálculos para ver cuales son las densidades poblaciones, de presencial de prestadores por habitantes, por kilómetros cuadrados y por densidad poblacional. Todo esto teniendo como base el departamento. Esta tabla luego permitió generar consultas básicas para ver cuales son los top y los que se ubican debajo en estos cálculos.