

Ejercicios Tidyverse 2

Laboratorio Políticas Públicas

28/4/2020

Practiquemos lo que vimos hoy!:

Generemos una tabla única que aglutine la información de “data_clase_final”, “poblacion_edad” y “casos_muertos”!

Para esto vamos a trabajar con los dataset presentes en el siguiente link: https://github.com/labpoliticasuba/Clases_2020/tree/master/Clase%2003/Ejercicio_clase.

1. Llamemos a Tidyverse!

2.a. Carguemos nuestro dataset llamado “data_clase_final.csv” en una nueva variable llamada “data_clase_final” (acordate de modificar el encoding en “Latin1”)

2.b. Demos un pantallazo a nuestro dataset usando la función tail().

2.c. Conozcamos cuáles son las columnas que lo integran.

3.a. Traigamos nuestro dataset “poblacion_edad.csv” y llamemoslo “poblacion_edad”.

3.b. Examinemos las columnas con un names(). Cada columna tiene la siguiente información:

- PROVINCIA: contiene los datos del distrito
- X0.19: Cantidad de hab. de 0 a 19 años
- X20.39: Cantidad de hab. de 20 a 39 años
- X40.59: Cantidad de hab. de 40 a 59 años
- X60.79: Cantidad de hab. de 60 a 79 años
- X80.99: Cantidad de hab. de 80 a 99 años

3.c. Utilicemos la función gather() para generar una columna llamada “grupo_etario” que aglutine a las diferentes columnas que tienen la cantidad de hab x provincia.

3.d. Modifiquemos como estan redactadas la observaciones/filas en la columna “grupo_etario” con la función case_when(). Deben ser reemplazadas así:

- X0.19: “De 0 a 19 años”
- X20.39: “De 20 a 39 años”
- X40.59: “De 40 a 59 años”
- X60.79: “De 60 a 79 años”
- X80.99: “De 80 a 99 años”

3.e. ¿Cómo podemos unir el dataset de “data_clase_final” con el de “poblacion_edad” ? Vamos paso a paso.

3.e.i. ¿Cuál es nuestra primary key? Es decir, ¿cuál es nuestra columna que nos va a permitir unir ambas tablas ya que tienen los mismos valores?

3.e.ii. Fijemosnos el nombre de las columnas, ¿tiene el mismo nombre para ambos casos? Modifiquemoslas si es necesario.

3.e.iii. Revisemos que tengan los mismos valores en ambas columnas -de ambos datasets- para que la unión se dé bien entre todas las observaciones.

3.e.iv. Modificá los siguientes valores utilizando el siguiente fragmento:

```
# cambiamos el tipo de objeto a character:
poblacion_edad$DISTRITO <- as.character(as.factor(poblacion_edad$DISTRITO))

# le pedimos que cada vez que encuentre una observacion en la columna
# DISTRITO y que sea igual a "Ciudad A..." la reemplace por "CABA":
poblacion_edad$DISTRITO[poblacion_edad$DISTRITO == "Ciudad Autónoma de Buenos Aires"] <- 'CABA'

poblacion_edad$DISTRITO[poblacion_edad$DISTRITO == "Entre Rios"] <- 'Entre Ríos'
poblacion_edad$DISTRITO[poblacion_edad$DISTRITO == "Neuquen"] <- 'Neuquén'
```

4. Realicemos el JOIN entre “data_clase_final” y “poblacion_edad” en un nuevo dataset llamado “data_poblacion”.

5. Ahora realice un right_join() entre “data_poblacion” y “caso_muertos.csv” utilizando lo aprendido! Genere un nuevo dataset llamado “data_poblacion_casos” donde haya una columna para la cantidad de infectados y otra para la cantidad de fallecidos.

6.a. Ordenemos nuestras columnas de “data_poblacion_casos” y eliminemos la columna “X”

6.b. Reemplacemos las columnas para que queden todos en mayusculas (y cambiemos el nombre de la columna “value” por “CANT_HAB_X_GRUPO”).

7. A. Supongamos que en Argentina se va a contagiar con coronavirus el 10% de la población mayor de 60 años. En base al dataset generado anteriormente, ¿cuántas camas necesitaríamos para atenderlos si todos se enfermaran a la vez? Para responder la pregunta supongamos que el contagio respeta los patrones que se han observado en otras experiencias: el 5% de los contagiados son casos graves y el 15% de los contagiados son casos que necesitan aislamiento y cuidados paliativos. (atención: en la respuesta tiene que estar distinguida la cantidad de camas para casos graves y leves) B. ¿Cuántas camas faltan (o sobran) en cada distrito para atender a esta demanda?