# **Descripción tratamiento Población votante 2017 “Politics”.**

El presente resumen tiene como objetivo complementar el proceso de tratamiento y obtención de datos del archivo complementario correspondiente al ETL llamado “*Poblacion\_Votantes\_2017.csv”*, a utilizar por el grupo “Politics”, además este complemento busca entregar una mejor comprensión sobre el código utilizado para este fin.

Cabe destacar que la explicación del proceso realizado previamente para la manipulación de estos datos se encuentra en el archivo *“Descripcion\_Politics\_Población\_Votante.docx”,* documento el cual especifica el proceso de obtención tanto de la población correspondiente al año 2013 como 2017. Este proceso que requirió solamente hacer uso de la herramienta **Excel.**

Este documento se presenta por medio de secciones, que se encuentran debidamente comentadas en el extremo superior del segmento de código correspondiente, como se presenta a continuación:

**Data Read:**

En esta sección se genera una lectura de la tabla que será utilizada, para este caso la lectura se realizada corresponde al archivo de “*Población\_H\_M\_T\_2017.xlsx”*. Siendo guardado en una variable llamada pob2017.

**Filtrado:**

En esta sección se procede a filtrar el elemento cargado anteriormente, eliminando los intervalos de edad correspondiente a personas que no pueden votar, siendo estos los intervalos siguientes:

* Total, comunal.
* 0 a 4 años.
* 5 a 9 años.
* 10 a 14 años.
* 15 a 19 años.

Este último intervalo requiere un tratamiento especial dado que existen un cierto sub-intervalo contenido en el mismo que incluye edades que si pueden votar. El resultado de este filtro se guarda en una nueva variable la cual será utilizada más adelante (pob2017vot).

**Ajuste Población 15-19:**

Se genera nuevamente una variable a partir de la variable iniciar pob2017 (pob2013Ajust). Esta nueva variable solo incluye el intervalo correspondiente de 15 a 19 años para todas las comunas.

El ajuste realizado en esta sección asignó pesos a cada una de las edades del intervalo (20% para cada uno de los años), por ende, como las personas que tienen la potestad de votar son las de edad igual o mayor a 18 años, se consideró un 40% del intervalo tanto para hombres como mujeres. A partir de la suma del 40% correspondiente de hombres y mujeres se obtuvo el total de la población.

**Juntar tablas**

Al mismo archivo generado en la sección anterior, se unieron los valores ajustados (40%), de *hombres, mujeres,* y el *total*. Estas nuevas columnas fueron añadidas bajo el nombre de *hombres\_a, mujeres\_a y total\_a*.

**Eliminar columnas:**

Del archivo anterior generado, se eliminaron las columnas originales correspondiente a cantidades (por sexo) y totales, es decir, fueron eliminadas: *hombres, mujeres y total.*

**Renombrar columnas:**

En esta sección, se renombró el intervalo definido anteriormente de “15 a 19” como “18 a 19”. Luego las columnas ajustadas añadidas en la sección anterior bajo los nombres: *hombres\_a, mujeres\_a y total\_a*. Fueron renombradas como *hombres, mujeres y total.*

**Juntar Tablas:**

En esta sección se unieron las tablas generadas anteriormente. Para ello se utilizó la tabla filtrada en la primera sección sin los intervalos desde 0 a 19 años, llamada *pob2017vot* con la tabla final ajustada para las edades “18 a 19”, llamada *pob2017Ajust*.

**Export CSV:**

Se genera el archivo de salida, llamado *Población\_Votantes\_2017.csv2,* el cual se encuentra debidamente especificado en el diseño del *dataset.*