# **Descripción tratamiento ETL “Politics”.**

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

Cabe destacar que la explicación de los datos procesados a continuación no requirió de una manipulación compleja de manera previa hasta este punto, dado que estos se obtuvieron de datachile y solamente fue necesaria hacer una corrección de caracteres especiales no admitidos por el formato (tildes o virgulillas), la cual se realizó de forma manual mediante 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 las tablas que serán utilizadas para el análisis, correspondiente a los siguientes archivos y en el orden descrito a continuación:

* Presidenciales1-2013.csv
* Presidenciales2-2013.csv
* Presidenciales1-2017.csv
* Presidenciales2-2017.csv

Cada uno de los archivos mencionados anteriormente, se exporta al entorno con la finalidad de trabajar sobre los mismos.

**Join Table:**

En esta sección se procede a juntar todas las tablas exportadas en el paso anterior, y que tiene como finalidad principal consolidar todos los datos en una sola tabla, es decir, representar los votos de cada candidato por comuna para todos los procesos eleccionarios (considerando primera y segunda vuelta si corresponde). En este caso, todos los archivos se consolidaron bajo una variable común que las agrupa, obteniendo un archivo consolidado con un total de 10.020 filas.

**Delete *partido\_id*:**

Como se menciona en el informe, el partido al cual pertenece el candidato es una variable despreciable para nuestro análisis dado que es de carácter cualitativo, por ende, se procede a eliminar la columna del archivo generado anteriormente que cuenta con los resultados de ambos procesos de elección.

**Add Candidatos\*:**

Se exporta la tabla completa correspondiente a los datos de los candidatos, para luego añadir esta al archivo consolidado correspondiente a las elecciones. El acoplamiento de ambas tablas se logra por medio del *candidato\_id,* el cual que ahora el candidato pueda ser identificado con su nombre.

**Add *id\_region y nombre\_region*:**

Se exporta la tabla completa correspondiente a los datos de regiones, siguiendo la misma lógica que la sección anterior, esta se agrega al archivo consolidado de las elecciones. El acoplamiento de ambas tablas se logra por medio del atributo *comuna\_datachile\_id,* el cual que ahora el candidato pueda ser identificado con su nombre.

**Clean Table:**

En esta sección se procede a eliminar aquellos atributos que no son necesarios para el análisis a realizar, que para este caso corresponden a las columnas correspondientes a *comuna\_customs\_id* dado que la comuna ya cuenta con un identificador único (*comuna\_datachile\_id*)y *comuna\_tax\_office\_id*.

**Revisión CSV:**

En esta sección se verifica que el problema identificado al añadir la tabla de candidatos, este generó que la tabla final generada no tuviera los 10.020, si no que solo contara con 8.640 filas correspondiente a los resultados de los votos por candidato en cada ciudad/comuna.

Para verificar que este problema fue solucionado, se realiza la verificación de que todos los candidatos tengan votos asociados bajo su nombre, tanto para elecciones del año 2014 como 2017.

**Export CSV:**

Se genera el archivo de salida, el cual se encuentra debidamente especificado en el diseño del *dataset.*

**\*El proceso de adición de candidatos en la tabla correspondiente se encuentra debidamente explicado en el entregable.**