# Anexo 3 – Fase 3: Entendimiento de datos

|  |  |
| --- | --- |
| **Proceso 2** | **Diagrama de flujo** |
| **Nombre de Script**: ETL\_PPD\_Downloaded\_Data.ipynb  **Datos de entrada**: oc\_dl\_provincia\_canton.csv, oc\_dl\_familia\_sector\_perfil.csv, oec\_dl\_familia\_sector\_perfil.csv, oec\_dl\_provincia\_canton.csv, ci\_dl\_provincia\_canton.csv  **Datos de salida**: oc\_dl\_provincia\_canton.csv, oc\_dl\_familia\_sector\_perfil.csv, oec\_dl\_familia\_sector\_perfil.csv, oec\_dl\_provincia\_canton.csv, ci\_dl\_provincia\_canton.csv  **Descripción**: Se cargar los archivos .csv en formato dataframe de pandas donde son procesados cambiado su tipo de datos y se eliminan datos duplicados, posteriormente se guardan con el mismo nombre en la carpeta CLEANED\_DATA. | Text  Description automatically generated |