**Resumen de la arquitectura de herramientas, librerías y tecnologías utilizadas en el estudio.**

Los datos se han obtenido, de las diversas fuentes, en formatos csv y Excel. Tanto estos ficheros como los utilizados durante el análisis, han sido almacenados en un proyecto público de github, consultable en el siguiente link:

<https://github.com/78ToniGil/Aire-Dexs>

El procesamiento y limpieza se ha realizado utilizando el lenguaje Python, en los frameworks Jupyterlab y Visual Code Studio. Principalmente, se han utilizado las librerías numpy y pandas para el análisis exploratorio de datos.

Finalmente, se muestran las conclusiones con gráficos realizados con Matplotlib y Seaborn.

De forma adicional, se ha incluido una pequeña muestra de cómo sería la funcionalidad montada sobre arquitectura cloud en Amazon Web Services. Para esto, los dataframes limpios y procesados, se han convertido en ficheros parquet, que se han subido a un bucket de S3 llamado “Dexs”. Una vez allí, el evento PUT dispara una función Lambda que:

* Lee el contenido de los ficheros
* Procesa los ficheros, insertando cada línea como un documento de una base de datos DocumentDB, que es la base de datos de documentos de AWS, compatible con MongoDB.
* Una vez procesados e insertados en la base de datos de documentos, se mueven a un segundo bucket llamado “Dexs-procesados”.

El resultado de este proceso puede verificarse en la base de datos:

Cuando esté terminado, añadimos aquí los datos de conexión a la base de datos