



MEMORIA DE FINALIZACIÓN DE PRÁCTICAS

UNIVERSIDAD NEBRIJA GRADO EN
INGENIERÍA INFORMÁTICA MEMORIA
PRÁCTICAS EN EMPRESA

Guillermo Martínez de Hurtado Aricha

Enero de 2025



MEMORIA DE FINALIZACIÓN DE PRÁCTICAS

UNIVERSIDAD NEBRIJA GRADO EN
INGENIERÍA INFORMÁTICA MEMORIA
PRÁCTICAS EN EMPRESA

Guillermo Martínez de Hurtado Aricha

Tutor: Adrián Pradilla Portoles

Enero de 2025

Derechos

D. Guillermo Martínez de Hurtado Aricha autoriza a que el presente trabajo se guarde y custodie en los repositorios de la Universidad Nebrija y además NO autoriza a su disposición en abierto.

Contenido

Derechos	3
Índice de ilustraciones	5
Glosario de términos	6
1. RESUMEN	9
2. PRÁCTICAS	9
2.1. Antecedentes	9
2.2. Objetivos de las prácticas	14
2.3. Onboarding	15
2.4. Funciones y tareas en las prácticas. (evidencias)	18
2.5. Relaciones de problemas planteados y procedimientos para su resolución	26
2.6. Aprendizajes y desarrollo profesional (habilidades adquiridas)	29
2.7. Metodologías utilizadas	30
2.8. Herramientas utilizadas en las prácticas (evidencias)	30
2.9. Logros, resultados y discusión	34
3. DESARROLLO	35
3.1. Introducción	35
3.2. Marco técnico	38
3.3. Equipo de trabajo y metodología	38
3.4. Proyecto	39
3.5. Resultados y discusión	41
3.6. Conclusiones	41
3.7. Líneas futuras	42
4. BIBLIOGRAFÍA	43
5. ANEXOS	44
Anexo 1: Código de python	44
Anexo 2: PBI	49
Modelo de datos del PBI	49
Visualizaciones PBI	49

Índice de ilustraciones

Ilustración 1: Logotipo de Serveo. (Serveo, n.d.).....	10
Ilustración 2: Horario de onboarding en Serveo. (Elaboración propia).....	16
Ilustración 3: Jornada I, Eficiencia Operativa. (Elaboración propia).....	17
Ilustración 4: Jornada II, Eficiencia Operativa. (Elaboración propia).....	17
Ilustración 5: Aplicación OCR python en local, pruebas y errores. (Elaboración propia).....	19
Ilustración 6: Power Automate, Ejemplo de flujo. (Elaboración propia).....	20
Ilustración 7: Power Apps. (Elaboración propia).....	22
Ilustración 8: Listas Sharepoint. (Elaboración propia)	22
Ilustración 9: PDF Estándar Serveo. (Elaboración propia)	23
Ilustración 10: Interfaz Microsoft Fabric. (Elaboración propia)	24
Ilustración 11: Herramientas en Microsoft Fabric. (Elaboración propia).....	25
Ilustración 12: Lagos de Datos, Microsoft Fabric. (Elaboración propia)	25
Ilustración 13: Flujos de datos, Microsoft Fabric. (Elaboración propia).....	26
Ilustración 14: Librerías Python. (Elaboración propia)	32
Ilustración 15: XML. (Gobierno de España, 2025).....	36
Ilustración 16: Resultado terminal script. (Elaboración propia).	40
Ilustración 17: Power Query. (Elaboración propia)	40

Glosario de términos

API (Interfaz de Programación de Aplicaciones):

Conjunto de protocolos y herramientas que permiten la comunicación entre distintos sistemas. En este trabajo se utiliza para acceder a datos públicos, como los precios de carburantes ofrecidos por el gobierno.

CSV (Comma-Separated Values):

Formato de archivo que almacena datos en forma de texto plano, donde cada línea representa un registro y los campos están separados por comas (u otros delimitadores). Se emplea para almacenar y manipular datos extraídos de archivos XML.

Data Lake (Lago de Datos):

Repositorio centralizado que permite almacenar datos en su formato original, sean estructurados o no estructurados. En el proyecto con Microsoft Fabric se crean lagos de datos para integrar y transformar la información proveniente de diversas fuentes.

Delta (Formato Delta):

Formato de almacenamiento de datos optimizado para trabajar con lagos de datos, permitiendo actualizaciones y transacciones de manera eficiente.

Eficiencia Operativa:

Conjunto de prácticas y procesos enfocados en optimizar el rendimiento y la productividad dentro de una organización, a través de la automatización, mejora de procesos y uso de tecnología.

Excel:

Herramienta de Microsoft Office ampliamente utilizada para el manejo, análisis y visualización de datos en formato de hojas de cálculo.

IA (Inteligencia Artificial):

Campo de estudio de la informática que busca desarrollar sistemas capaces de aprender, razonar y tomar decisiones de manera autónoma. En el contexto del proyecto se emplea para entrenar modelos que permitan extraer información de documentos repetitivos.

Log (Registro de Eventos):

Archivo que almacena información y eventos ocurridos durante la ejecución de un programa. Se utiliza para el seguimiento y depuración de procesos en los scripts desarrollados.

OCR (Reconocimiento Óptico de Caracteres):

Tecnología que permite convertir diferentes tipos de documentos, como archivos PDF o imágenes, en datos editables y procesables. Se ha utilizado para extraer IBANs de documentos PDF.

Planner:

Herramienta de Microsoft 365 que facilita la gestión de tareas y proyectos, permitiendo asignar actividades, establecer fechas límite y coordinar el trabajo en equipo.

Power Automate:

Plataforma de Microsoft que permite la automatización de flujos de trabajo sin necesidad de escribir código (no-code). Se utiliza para tareas repetitivas, como la automatización del procesamiento de correos electrónicos.

Power BI:

Herramienta de visualización y análisis de datos de Microsoft que permite transformar datos en informes y dashboards interactivos, facilitando la interpretación de la información.

Power Query:

Componente de Microsoft que facilita la transformación, limpieza y preparación de datos, utilizando tanto una interfaz visual como el lenguaje de programación M para personalizaciones avanzadas.

PowerApps:

Plataforma de Microsoft que permite crear aplicaciones empresariales de forma rápida, con muy poco o sin necesidad de código. Se ha empleado para desarrollar aplicaciones internas, como el registro de EPIs.

Pipeline (Flujo de Datos):

Secuencia de procesos automatizados que permiten la ingesta, transformación y almacenamiento de datos. En Microsoft Fabric se utilizan pipelines para gestionar la actualización de datos.

Python:

Lenguaje de programación de alto nivel, muy utilizado en análisis y automatización de datos por su versatilidad y amplia disponibilidad de librerías.

PySpark:

API de Python para trabajar con Apache Spark, utilizada para el procesamiento y análisis de grandes volúmenes de datos de forma distribuida.

SQL (Structured Query Language):

Lenguaje de programación especializado en la gestión y manipulación de bases de datos relacionales. Se utiliza para realizar consultas y transformaciones en las bases de datos del proyecto.

SharePoint:

Plataforma de Microsoft para la gestión y almacenamiento de documentos y datos, que permite la colaboración y el acceso compartido a la información dentro de la empresa.

XML (Extensible Markup Language):

Lenguaje de marcado utilizado para almacenar y transportar datos de forma estructurada. En el proyecto se extraen datos en formato XML para luego transformarlos a CSV.

1. RESUMEN

Este trabajo documenta y analiza el recorrido y las experiencias acumuladas durante mis prácticas en el Departamento de Eficiencia Operativa y Análisis de Datos. En él se recoge el conjunto de proyectos que he llevado a cabo hasta la fecha, abarcando tanto la parte de eficiencia operativa como la de análisis de datos. Durante este periodo, he tenido la oportunidad de trabajar en la automatización de tareas, la transformación de datos y el desarrollo de aplicaciones que optimizan procesos internos. Entre los proyectos más destacados se encuentran la creación de un OCR para la extracción de IBANs desde PDFs, la automatización de la gestión de correos diarios para la gestora del contrato de Iberia mediante Power Automate, y la implementación de soluciones con PowerApps para el registro y seguimiento de EPIs. Asimismo, se ha avanzado en la consolidación de una base de datos en Microsoft Fabric, un proyecto ambicioso que promete centralizar los datos y facilitar el autoconsumo en los distintos departamentos de la empresa.

La segunda parte de este trabajo se dedicará en profundidad al "Conversor XML a CSV", una herramienta desarrollada con Python que automatiza la descarga de datos públicos en formato XML, su transformación a CSV y el posterior análisis mediante Power BI. Se explicará el proceso completo, desde la obtención de los datos a través de una API gubernamental, pasando por la transformación y almacenamiento en archivos CSV, hasta la creación de informes visuales que permiten interpretar de manera clara la evolución de los precios de carburantes. Este conversor no solo optimiza la consulta y el manejo de la información, sino que también contribuye a la generación de históricos y a la toma de decisiones basada en datos actualizados. En resumen, este trabajo recoge tanto las experiencias y aprendizajes adquiridos durante mis prácticas como la aplicación práctica de técnicas de automatización y análisis de datos en un entorno real, ofreciendo una visión integral del potencial de estas herramientas en el ámbito empresarial.

2. PRÁCTICAS

2.1. Antecedentes

2.1.1. ¿Qué entrevistas he hecho?

Empecé buscando entrevistas por LinkedIn, buscaba sobre todo un puesto que tuviera que ver con análisis de datos o ingeniería de datos, es decir, trabajar mucho con grandes cantidades de datos, pero no tanto solo enfocado a bases de datos, más bien la parte de trabajar el dato obtener información y así mejorar una empresa.

Aplicando a diversos sitios me llamarón de un par de ellos, el primero era Verisure, compañía de la alarmas y seguros de España y Portugal, este puesto no tenía nada que ver con nada de datos, consistía en ser el encargado del desarrollo de aplicaciones y todo tipo de herramientas necesarias para el equipo de Verisure, para este puesto hice únicamente la primera entrevista y no me dijeron hasta mitad de verano nada.

La segunda empresa a la que apliqué, que es la que estoy ahora, se llama Serveo y el puesto es Becario en análisis de datos y eficiencia operativa. Para entrar hice dos entrevistas, la primera con recursos humanos que me hizo la entrevista una chica que se llama Usune y la segunda con la jefa de mi equipo que se llama Inmaculada (más adelante hablo un poco más de ella).

2.1.2. Datos de la empresa



Ilustración 1: Logotipo de Serveo. (Serveo, n.d.)

Nombre y ubicación

La compañía en la que actualmente realizo mis prácticas se llama Serveo, anteriormente conocida como Ferrovial Servicios. Con sede en España y presencia en diversas localidades, Serveo forma parte del sector de servicios y gestión integral de infraestructuras.

Fechas de realización

Mis prácticas se han desarrollado desde septiembre y continuo en Serveo hasta junio, lo que me ha permitido involucrarme en distintos proyectos a lo largo del año.

Descripción de la actividad de la empresa

Serveo es una empresa del sector de servicios, especializada en el mantenimiento y la gestión integral de infraestructuras, facility management y servicios relacionados. La compañía presta servicios a clientes de

diversos sectores como Industria, Salud, Corporate, Gobierno e Infraestructuras. Entre sus contratos destacados se encuentra el mantenimiento y la jardinería de instalaciones para empresas reconocidas, como Airbus, en distintas localizaciones de España.

Descripción de la estructura organizativa y puesto ocupado en ella

Internamente, Serveo organiza su modelo de negocio en cinco áreas principales para facilitar la comprensión por parte de clientes y socios externos. Sin embargo, a nivel interno, la empresa se estructura en distintas delegaciones según la región o el sector: Delegación Norte, Delegación Sur, Delegación Centro, Delegación Industria y Delegación Servicios.

En mi caso, me he incorporado al equipo de Eficiencia Operativa dentro de la Delegación de Industria. Mi puesto es el de becario en eficiencia operativa y análisis de datos. Inicialmente, junto con otro becario, fuimos integrados en el mismo puesto, aunque debido a mi formación en ingeniería informática, se me asignaron proyectos orientados al análisis de datos, mientras que al otro becario se le orientó más hacia tareas de eficiencia operativa. La estructura jerárquica se caracteriza por una clara distribución de responsabilidades, donde la alta dirección establece las directrices estratégicas que se transmiten a través de las delegaciones y departamentos.

Tecnologías y metodologías utilizadas

Serveo apuesta por la innovación y la optimización de procesos a través del uso de diversas tecnologías y metodologías. Entre las herramientas empleadas se encuentran:

Power BI, Excel y Power Query para la visualización y análisis de datos.

Power Automate y PowerApps para la automatización y creación de aplicaciones sin necesidad de código.

Microsoft Fabric para la integración y transformación de datos en entornos modernos.

Además, se utilizan herramientas colaborativas como Planner y SharePoint, y se aplican metodologías ágiles que permiten gestionar proyectos de forma dinámica y eficiente.

Descripción del puesto

Mi puesto como becario en eficiencia operativa y análisis de datos me ha permitido trabajar en proyectos que van desde el desarrollo de scripts en Python para automatización y extracción de datos, hasta la creación de informes complejos en Power BI. Gracias a mi formación en ingeniería informática, he contribuido a optimizar procesos internos, desarrollando aplicaciones que mejoran la eficiencia operativa y facilitando la toma de decisiones basada en datos precisos y actualizados.

Descripción del Sector. Competidores, evolución, tendencias y oportunidades

El sector de servicios en el que opera Serveo es altamente competitivo y se caracteriza por una constante evolución tecnológica. La digitalización y automatización de procesos son tendencias clave, lo que obliga a las empresas del sector a adaptarse rápidamente. Entre los competidores se encuentran otras grandes compañías de facility management y servicios integrales, que compiten por ofrecer soluciones cada vez más eficientes e innovadoras. Esta evolución crea oportunidades para la implementación de tecnologías avanzadas en análisis y gestión de datos, permitiendo a las empresas mejorar su operatividad y respuesta a las necesidades del mercado.

Posicionamiento de la empresa en el sector. Puntos fuertes y débiles. Aspectos mejorables

Serveo se posiciona como una de las principales empresas del sector de servicios en España, gracias a su estrategia de crecimiento sostenido y a su diversificación en áreas clave como Industria, Salud y Gobierno. Entre sus puntos fuertes destacan la solidez de su gestión de infraestructuras, la capacidad de innovación y la rápida adaptación a las nuevas tecnologías. No obstante, existen áreas de mejora, como la integración completa de sistemas de automatización y una mayor optimización de la comunicación interna entre delegaciones. Estos aspectos mejorables representan oportunidades para fortalecer aún más la competitividad de la empresa en un mercado en constante cambio.

2.1.3. Información previa

Dado que todo el entorno tecnológico de Serveo está basado en software de Microsoft, fue fundamental para mí contar con una base sólida en herramientas como Excel, Power Query y Power BI, que son esenciales para cualquier puesto de análisis de datos dentro de la empresa. Antes de comenzar las prácticas, realicé un curso en LinkedIn Learning en colaboración con Microsoft, titulado "Introducción al análisis de datos", que me proporcionó una formación inicial en estas herramientas. En este curso, se abordaron en profundidad las funciones avanzadas de Excel, algo crucial para la manipulación de grandes conjuntos de datos y, además, se enseñaron los fundamentos de Power BI, que es una de las herramientas más utilizadas para la visualización y análisis de datos empresariales.

Este curso fue muy útil para familiarizarme con el ecosistema de Microsoft y me permitió llegar con una buena base de conocimientos. Sin embargo, al iniciar mis prácticas en Serveo, la primera semana estuvo dedicada completamente a formaciones específicas dentro de la empresa, lo que me permitió perfeccionar mis habilidades en Power BI y Power Query. Estos programas son claves en la gestión de datos y la eficiencia operativa dentro de la compañía, ya que permiten automatizar procesos, generar informes visuales y extraer insights valiosos a partir de los datos disponibles. A lo largo de estas sesiones formativas, aprendí a utilizar herramientas avanzadas de Power Query

para la transformación de datos y el diseño de informes dinámicos en Power BI, los cuales se han convertido en herramientas habituales en mi día a día.

En cuanto a mis estudios universitarios, varias asignaturas me han proporcionado una base técnica sólida que me ha sido de gran utilidad durante mis prácticas. Una de las más relevantes ha sido la asignatura de Inteligencia Artificial, donde se puso un gran énfasis en la limpieza y el preprocesamiento de datos. Durante el curso, trabajamos intensamente en proyectos que requerían la preparación de grandes volúmenes de datos, algo que resulta fundamental en mi rol actual. Además, fue una de las pocas asignaturas en las que trabajamos con Python, lo que me ha permitido aplicar algunas de las técnicas de programación aprendidas en proyectos más complejos en el entorno laboral. La limpieza de datos es un paso crucial en el análisis de datos y haber practicado tanto esta habilidad en la universidad me ha facilitado muchísimo adaptarme a los procesos de análisis de datos en Serveo.

Otra asignatura que ha sido muy valiosa en mi formación es Programación Avanzada, donde aprendí a estructurar de manera correcta el código y a utilizar buenas prácticas en el desarrollo de software. Esta asignatura me enseñó la importancia de escribir código claro y bien documentado, algo esencial no solo para crear programas funcionales, sino también para trabajar de manera colaborativa en entornos empresariales. Durante esta asignatura, también aprendimos a diseñar diagramas de flujo de código y arquitectura de software, herramientas que, aunque en ese momento parecían algo teóricas, he descubierto que tienen una aplicación muy práctica en mi trabajo diario. Gracias a esos conocimientos, he podido familiarizarme rápidamente con herramientas como Power Automate y Power Apps, dos aplicaciones de Microsoft que se utilizan en Serveo para automatizar procesos y desarrollar aplicaciones internas de una manera eficiente.

Por último, la experiencia práctica adquirida en la universidad, tanto en las asignaturas como en los trabajos de clase, ha sido clave para afrontar los retos técnicos que se presentan en el día a día de mis prácticas. Haber trabajado en proyectos que simulan entornos reales me ha permitido desarrollar una capacidad de adaptación rápida a las herramientas y procesos dentro de la empresa. Además, el conocimiento adquirido en los cursos adicionales que he realizado de manera autónoma me ha dado una ventaja a la hora de integrar diferentes tecnologías en mi flujo de trabajo.

2.1.4. Motivos para elegir estas prácticas

Desde que entré en informática siempre he tenido mucho interés por todo lo que es la ciencia del dato y el manejo de grandes cantidades de datos, entonces cuando estaba buscando prácticas pensé que mejor forma de empezar en el mundo de la informática que en unas prácticas de análisis de datos y se presentó esta oportunidad, así que no dude ni un solo momento en decir que sí.

Otro motivo muy importante es que tal y como avanza la inteligencia artificial todo el mundo de los datos va a estar más cotizado a la vez que por estos mismos avances de la IA trabajos como los analistas de datos y otros van a ir desapareciendo o siendo cada vez más difícil entrar en este mundo, por eso aprovechando ahora las prácticas creo un trasfondo en mi curriculum y con suerte puedo haber entrado en este mundillo de los datos para quedarme mucho tiempo, ya que de momento me está gustando bastante todo lo que hago y me veo en un futuro siguiendo en este mundillo.

2.2. Objetivos de las prácticas

PROYECTO FORMATIVO, actividades a desarrollar por el alumno:

1. Ser participe en los proyectos de optimización y mejora de procesos en los que trabaja el departamento de Eficiencia Operativa, desarrollando e implantando mejoras operativas con los diferentes interlocutores (interno y externo).
2. Realizar cuadros de mando y análisis de datos, con la finalidad de obtener conclusiones y establecer recomendaciones que optimicen las operaciones.
3. Elaborar informes y reportes sobre el resultado de los análisis realizados con le objetivo de facilitar la toma de las decisiones a la dirección.
4. Dar soporte en la identificación e implantación de acciones de mejora en los diferentes servicios de la compañía.
5. Colaborar en la implantación y desarrollo de proyectos de implantación de metodología Lean /Six Sigma.
6. Trabajar realizando consultas y análisis en bases de datos en entornos SQL y Google Cloud Platform.

2.3. Onboarding

Voy a empezar con las reuniones de menor a mayor nivel y, poco a poco, iré aumentando el detalle. Para comenzar, mi departamento se divide en dos partes principales: la parte de eficiencia operativa y la parte de datos. Yo trabajo más con la parte de datos, y somos tres personas: Daniel, que es el líder del grupo (el Head) y la persona que nos guía; y Roberto, con quien trabajo directamente la mayor parte del tiempo. Entre los tres gestionamos una gran parte de los datos de la empresa. Esto lo explicaré mejor en el apartado de funciones y tareas en las prácticas en el siguiente punto. En cuanto al onboarding, quería mencionar las reuniones que tenemos todos los días, llamadas Daily. Todos los días, o prácticamente todos, nos reunimos a las 10:00 de la mañana, ya sea de manera presencial o conectándonos si alguno está teletrabajando. Estas reuniones duran unos 10 o 15 minutos, y en ellas hablamos de los proyectos en los que vamos a trabajar durante el día, los problemas que tuvimos el día anterior y las soluciones que les hemos dado, ya que estas soluciones pueden ser útiles para otras personas en sus propios proyectos.

En estas reuniones también se discuten los nuevos proyectos que llegan desde otros departamentos. La mayoría de estos proyectos llegan a Daniel, quien nos los comenta a Roberto y a mí, y luego los asignamos mediante un planner, distribuyendo las tareas entre nosotros según el tipo de proyecto. Por ejemplo, si se trata de un proyecto grande de datos, suele asignarse a Daniel. Con "proyecto grande de datos" me refiero a aquellos que se reportan directamente al CEO o dependen de datos financieros, por lo que requieren informes muy precisos y son más adecuados para alguien con mayor experiencia. Si el proyecto implica trabajo más rutinario, como informes en Power BI o Excel, normalmente se lo queda Roberto. A mí se me asignan proyectos variados, como el trabajo con código (especialmente Python, que es lo que más se utiliza), algunos proyectos en Power BI, PowerApps o flujos en Power Automate. Estas dos últimas herramientas, PowerApps y Power Automate, no las había utilizado antes, pero me parecen bastante interesantes. A pesar de que al principio tuve problemas, los fui comentando en las reuniones diarias y recibí ayuda de mis compañeros. Poco a poco, he ido adquiriendo más conocimientos sobre ellas hasta el punto de manejarlas mejor que ellos.

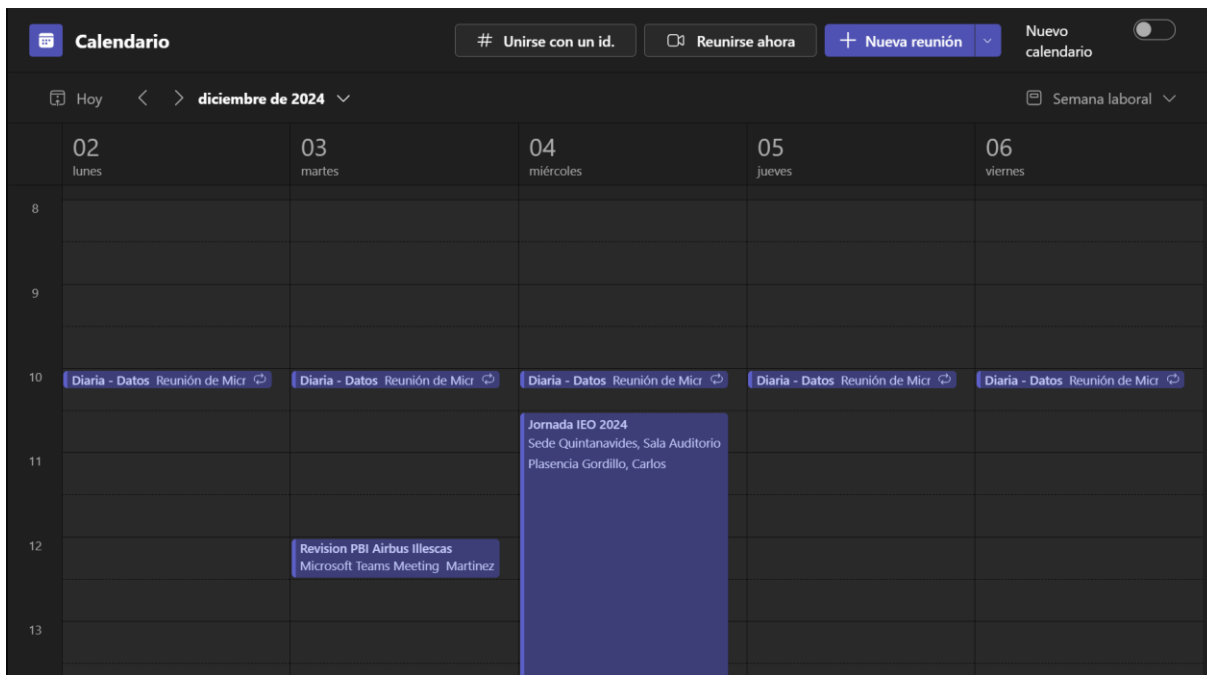


Ilustración 2: Horario de onboarding en Serveo. (Elaboración propia)

Cuando tengo un proyecto para otro departamento, organizo una reunión con la persona responsable de ese proyecto para acordar las pautas que debemos seguir y definir los requisitos. Como ejemplo, se puede ver en el calendario que adjunto.

Pasando a eventos más grandes, mencionaré las reuniones que tenemos con todo el departamento, es decir, tanto con la parte de eficiencia operativa como con la parte de datos en conjunto. Al principio de mis prácticas, en septiembre, los jefes decidieron crear una especie de jornada de team building para el departamento. Esta jornada consiste en que, cada mes, una persona del departamento debe preparar una comida, que puede ser un desayuno, una merienda o una comida. Normalmente, solemos quedar un lunes de cada mes y reservamos ese día para el departamento. Hasta ahora, tres personas han preparado las comidas, y a mí me toca en mayo (ya veremos qué tal me va).



Ilustración 3: Jornada I, Eficiencia Operativa. (Elaboración propia)



Ilustración 4: Jornada II, Eficiencia Operativa. (Elaboración propia)

En cuanto a eventos más amplios, tuvimos una jornada de todo el equipo de eficiencia operativa. Nos reunimos con el Departamento de Energía y el Departamento de Eficiencia Operativa, que es donde estoy yo. Ese día hicimos una comida, asistimos a un espectáculo de El Mago Pop, y después algunos de nosotros nos quedamos a tomar algo y salir de fiesta con el resto del departamento.

Finalmente, otro evento destacado fue la cena de Navidad, donde se reunió toda la empresa. Consistió en un cóctel con música en la planta baja.

2.4. Funciones y tareas en las prácticas. (evidencias)

Mi puesto en las prácticas es el de becario en eficiencia operativa y análisis de datos. Al comenzar, entramos dos becarios para el mismo puesto, o al menos con el mismo nombre. Ambos nos incorporamos al Departamento de Eficiencia Operativa como becarios en eficiencia operativa y analistas de datos. Dado que el otro becario tenía un perfil más enfocado a la ingeniería y yo venía de ingeniería informática, me asignaron principalmente a la parte de análisis de datos, mientras que a él lo orientaron más hacia la eficiencia operativa. Sin embargo, dado que vengo de ingeniería informática y mejorar el rendimiento en eficiencia operativa a menudo implica desarrollar aplicaciones, código o flujos, también me encargaron proyectos relacionados tanto con eficiencia operativa como con análisis de datos.

Para aclarar mejor esta parte, explicaré las tareas que he realizado durante las prácticas, primero enfocándome en la eficiencia operativa y luego en la parte de datos:

Eficiencia Operativa: Proyectos

Nada más comenzar las prácticas, no había muchos proyectos relacionados con datos, por lo que me asignaron algunos proyectos de eficiencia operativa. El más interesante fue la creación de un OCR para extraer los IBAN de cuentas bancarias desde distintos PDFs. No puedo mostrar el resultado de este proyecto debido a que incluye los IBAN de los proveedores que utilizamos, pero puedo explicar el código y el procedimiento seguido. Este proyecto surgió porque, en agosto, justo antes de que yo comenzara en septiembre, se hizo una migración del SAP de R1 a S3 (creo que es la denominación actual). Entonces, el departamento financiero, encargado de gestionar a los proveedores, acudió al departamento de eficiencia operativa con el problema de que debían transferir manualmente 500 IBAN a un Excel. Nuestra propuesta fue crear un script con OCR que permitiera cargar todos los PDFs a la vez, leerlos uno por uno, y crear una línea en el Excel para cada PDF procesado. En esta línea se almacenaría el banco del proveedor, su CIF o NIF, su cuenta bancaria y el nombre de la empresa. Como se puede ver en la captura lo que hago es crear una

aplicación que se sube a sharepoint y desde ahí la pueden ejecutar una vez sincronizan la carpeta en local.

» Codigo » AutomatizacionProveedores

Nombre	Fecha de modificación	Tipo	Tamaño
.venv	13/09/2024 13:19	Carpeta de archivos	
build	24/10/2024 8:56	Carpeta de archivos	
dist	24/10/2024 9:10	Carpeta de archivos	
DocumentosBanco	24/10/2024 9:50	Carpeta de archivos	
DocumentosBancoProcesados	23/10/2024 17:41	Carpeta de archivos	
DocumentosDuplicados	24/10/2024 9:55	Carpeta de archivos	
prueba	03/10/2024 10:46	Carpeta de archivos	
rutas	27/09/2024 14:12	Carpeta de archivos	
Procesador_PDF_IBAN.exe	24/10/2024 9:11	Aplicación	287.017 KB
crearExcelDesdeFacturas.py	04/11/2024 10:16	Archivo de origen Py...	22 KB
JuntarExcels.py	24/10/2024 10:30	Archivo de origen Py...	2 KB
Prueba.py	23/10/2024 18:50	Archivo de origen Py...	10 KB
READ.ME	24/09/2024 9:10	Archivo ME	1 KB
Procesador_PDF_IBAN.spec	24/10/2024 8:56	Archivo SPEC	1 KB
Prueba.spec	19/09/2024 16:29	Archivo SPEC	1 KB
PruebaCSV.spec	27/09/2024 14:04	Archivo SPEC	1 KB
EjecutableDptoFinanciero_Prueba.zip	19/09/2024 17:07	Carpeta comprimida ...	331.951 KB
Plantilla_Proveedores_10_filas.xlsx	01/10/2024 10:29	Hoja de cálculo de M...	11 KB
Tesseract_FINAL_log.xlsx	24/10/2024 9:50	Hoja de cálculo de M...	17 KB

Ilustración 5: Aplicación OCR python en local, pruebas y errores. (Elaboración propia)

Otro proyecto en el que participé al principio, a pesar de estar más orientado hacia datos, me lo asignó Roberto, un compañero del equipo de datos. Consistía en automatizar los correos diarios que recibía la gestora del contrato de Iberia. El problema era que diariamente recibía un par de correos con los registros de las órdenes de trabajo, divididos en operarios e instalaciones. Además, recibía un resumen semanal cada lunes y un histórico mensual. La gestora debía descargar manualmente estos archivos, comprobar que estuvieran correctos y luego subirlos a una carpeta de SharePoint para que el informe de Power BI, que dependía de esos archivos, funcionara correctamente. Nuestra solución fue crear un flujo de Power Automate que, como se puede ver en la captura, lee el correo, detecta la llegada de correos de una dirección específica con asuntos como "Oper. Diario" o "Mant. Diario", verifica que los datos sean correctos y, automáticamente, sube los archivos a la carpeta de SharePoint, evitando que la gestora tuviera que hacerlo manualmente.

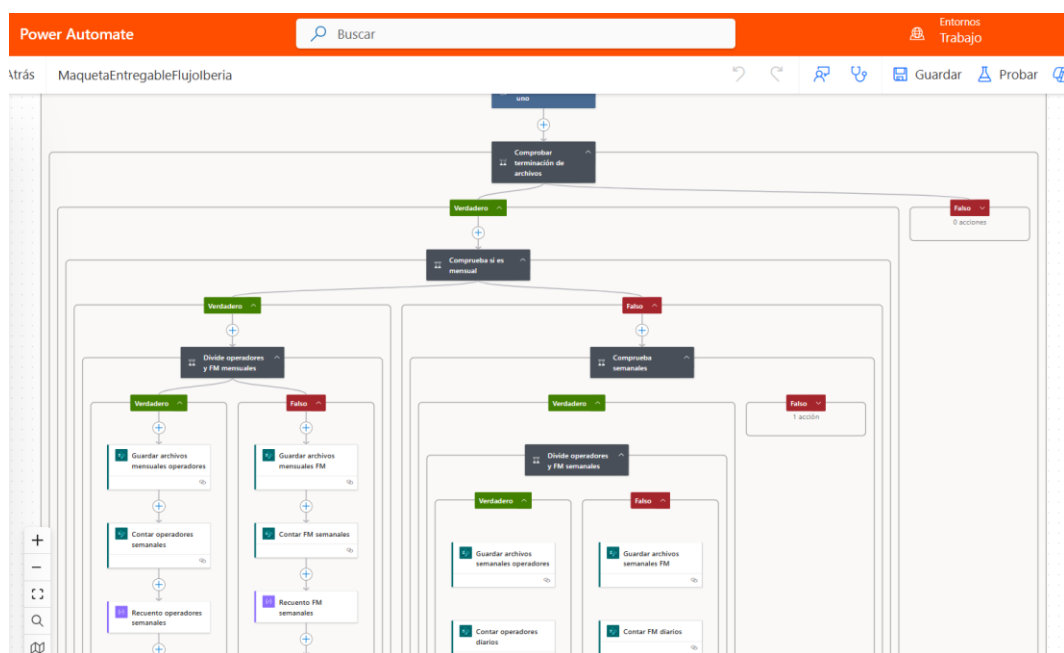


Ilustración 6: Power Automate, Ejemplo de flujo. (Elaboración propia)

En la parte de eficiencia operativa, he trabajado principalmente con Power Automate y con python. Aunque solo mencionaré el caso de la automatización de correos de Iberia y el OCR, he realizado muchos otros procesos similares de automatización, sobre todo relacionados con correos y otros proyectos que movían archivos de una carpeta de SharePoint a otra automáticamente cuando se creaban ciertos archivos o más extracciones de texto con código.

Con estos dos proyectos concluye un poco la parte de eficiencia operativa. Es verdad que no son todos los que he hecho de eficiencia operativa, pero sí que es lo más principal y más importantes que he realizado en esa parte del departamento. A partir de octubre empecé a trabajar con Dani, que es el jefe de datos de la compañía, por así decirlo, aunque no hay un departamento de datos como tal, por eso meto todos los demás proyectos en la parte de datos, aunque hay bastantes que son de eficiencia operativa, y optimización de las horas de la gente.

Análisis de Datos: Proyectos

El primer proyecto del que quiero hablar de la parte de datos son varios proyectos en sí porque voy a hablar un poco de Power BI en general y todos los proyectos de Power BI que he realizado en la empresa. De la mayoría de estos proyectos no puedo enseñar los informes como tal, ya que son en

sí datos confidenciales de la compañía. Pero voy a intentar poner alguna captura de los que sí que puedo mostrar. El principal proyecto del que quiero hablar, y el más importante me ha parecido que he realizado, es el de la compañía Airbus Illescas. Airbus es una empresa bastante conocida de fabricación de aeronaves, tanto civiles como militares, y que restringe bastante los datos que se pueden mostrar. Serveo se encarga en sí del mantenimiento y la jardinería de las instalaciones de Airbus, desde Airbus Illescas, Airbus Getafe, incluso Airbus Sevilla y otras localizaciones de España. Este proyecto era el de seguimiento de órdenes de trabajo en Maximo. Maximo es la plataforma de gestión de órdenes de trabajo que maneja Airbus. Básicamente, querían controlar las órdenes de trabajo que se están realizando, las que no han llegado a tiempo a realizar y, así, poder mejorar en esas áreas de trabajo. Entonces, el Power BI que se creó servía un poco para tener ese control. Este Power BI es bastante grande, ya que es un informe de todas las órdenes de trabajo que se realizan en la planta de Airbus Illescas y, en principio, es un proyecto que se quiere agrandar para utilizar en todas las plantas de Airbus de España que maneja Serveo.

-

Ahora voy a hablar de uno de los proyectos más importantes para mí que he realizado en la empresa. Es uno de los proyectos con los que más he aprendido, ya que utilizaba una herramienta que es PowerApps, y no la había utilizado en mi vida, no tenía ni idea de lo que era. Pero me ha servido para aprender un poco más sobre cómo crear aplicaciones, mantener un orden limpio de las cosas y organizar mejor mis tareas. PowerApps es una herramienta que explicaré más adelante, pero te pide que sigas un orden estricto de las cosas. Este proyecto lo retomé yo de un proyecto que estaba haciendo Roberto, que consistía en una aplicación para el registro de EPI, que son los equipos de protección individual, de forma más sencilla para los gestores de contrato. Básicamente, lo que se hacía hasta antes de tener la aplicación era rellenar a mano unas hojas con tablas y, usando un formato estándar de Serveo, se registraban cada uno de los EPIs. Con esta aplicación, lo que se hace, por un lado, es conectarse a una de las listas de SharePoint. De esta lista saco los valores de cada EPI, de cada gestor de contrato, de cada gestor de personal, de los distintos EPIs y de las distintas tallas para cada EPI. Entonces, por un lado, la aplicación se conecta a esta lista de la que extrae los datos. En esta aplicación también se van a guardar los distintos registros que se hagan. Por otro lado, en una carpeta de SharePoint, va a guardar un PDF con el formato estándar de Serveo. Lo que hago es utilizar un código HTML para replicar ese PDF estándar de Serveo que utilizaban antes de la aplicación. Este proyecto, la verdad, es bastante interesante y me gustó un montón. Pronto, una vez lo presentemos a los jefazos, queremos meterle una actualización para las fechas de caducidad de cada EPI, y que dé un aviso cuando quede poco tiempo para que caduque el EPI de alguno del personal.

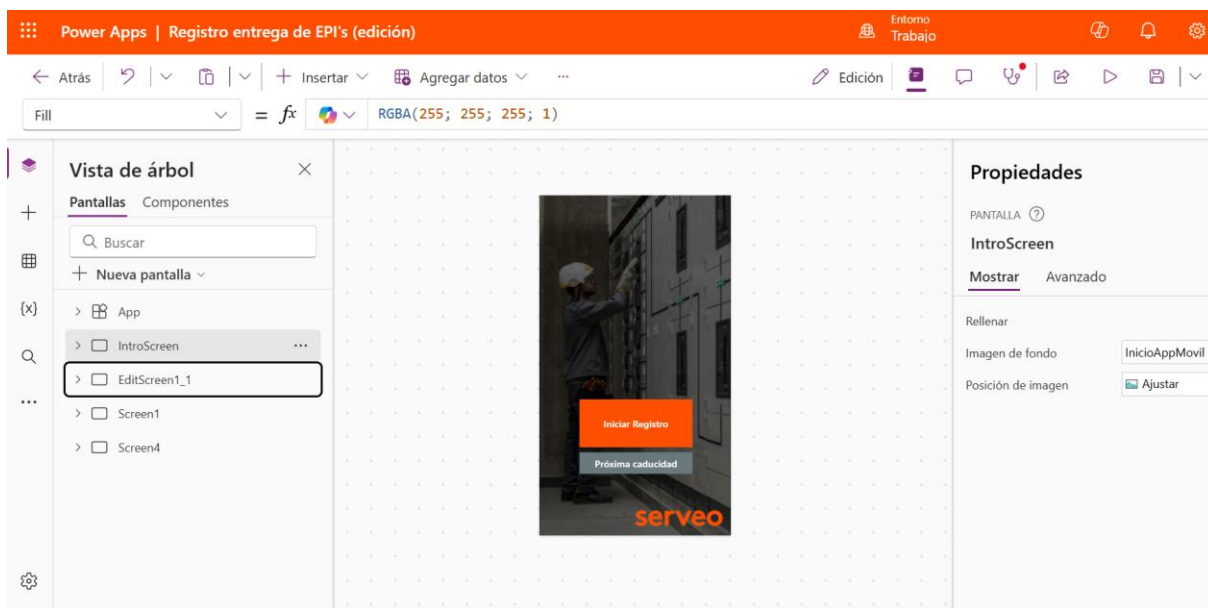


Ilustración 7: Power Apps. (Elaboración propia)

La lista de Sharepoint de donde se extraen y se guardan los registros de la app:

RESPONS...	FECHA...	RECEPTOR	DNI	CATEGORÍA	CALZADO	TALLA C...	TORSO	TALLA T...	PANTALÓN	TALLA PA...	OTROS	FIRMA TRABAJ...	FIRMA R...
12/12/2024 10:30				LIMPIADOR/A	ZUECOS LIMP.	36	CASACA LIMP.	S	PANTALÓN BLA...	S			
12/12/2024 16:49				PEÓN ESPECIAL...	BOTAS SEGURID...	43	POLO MANGA LAR...	M	PANTALÓN ALT...	L	PANTALLA		
											ARNÉS		
											CASCO		

Ilustración 8: Listas Sharepoint. (Elaboración propia)

El PDF que se guarda una vez se ha registrado un EPI:

serveo

ANEXO I - PRO-SSU-12
 Registro de entrega de
 equipos de protección individual
 Rev. 03

ANEXO I Registro de entrega de equipos de protección individual

Nombre y Apellidos	[Redacted]	DNI	[Redacted]
Fecha de entrega	27/12/2024		
Responsable	MAITE AZPURI		

Por la presente declara:


Haber recibido los siguientes equipos de protección individual, que a continuación se relacionan para cubrir los riesgos especificados en la evaluación de riesgos, comprometiéndome a consultar cualquier duda sobre la correcta utilización del equipo, conservar los mismos en buen estado y a comunicar cualquier deterioro o extravío.

Igualmente, declaro haber recibido las oportunas explicaciones sobre el correcto uso y mantenimiento (según instrucciones del fabricante) de los citados Equipos de Protección Individual, habiendo recibido las indicaciones oportunas acerca de la obligatoriedad de su uso en los procesos en que sean necesarios.

Relación de equipos de protección individual asignados al puesto


Calzado	Talla Calzado	Torso	Talla Torso	Pantalón	Talla Pantalón	Otros
	42	CASACA LIMP. POLO MANGA LARGA AV	M	PANTALÓN BLANCO LIMP.	M	

Firma Trabajador



[Redacted]

Firma Responsable



[Redacted]


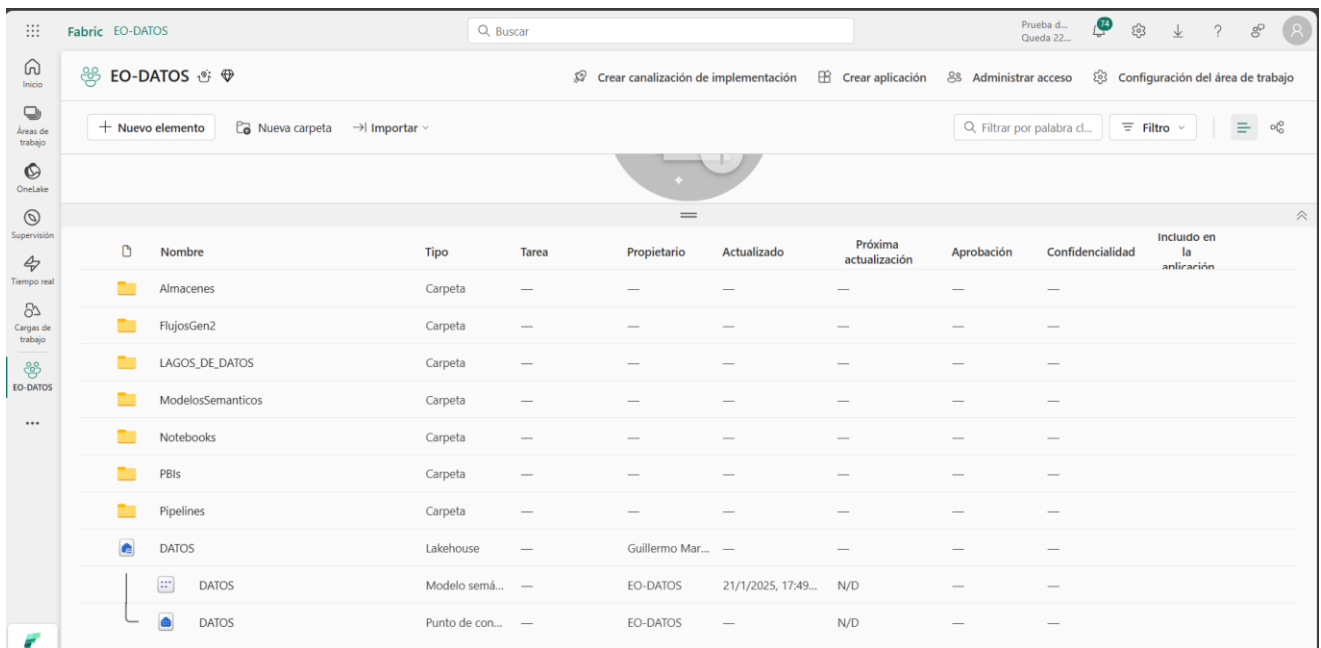


Ilustración 9: PDF Estándar Serveo. (Elaboración propia)

- - El último proyecto del que quiero hablar, que he realizado hasta la fecha, es el más importante que me han encargado. Consiste en crear una base de datos en Microsoft Fabric, que es una herramienta relativamente nueva. Este proyecto surge porque el equipo de datos, en sí, queremos tener un poco más de independencia, ya que en estos momentos dependemos de la parte de sistemas para todo lo relacionado con las descargas de datos de SAP y demás plataformas. Entonces, creando esta base de datos, lo que conseguiríamos sería tener estos datos centralizados, no depender de sistemas para esas descargas de SAP. Y, sobre todo, lo más importante es que queremos llegar a conseguir el autoconsumo de los departamentos, es decir, nosotros crearíamos un modelo de datos específico para cada uno de los proyectos

que nos pidan. Lo que se conseguiría sería que los departamentos, a raíz de esos modelos de datos, creen sus propios informes en Power BI, puedan hacer Excel con esos datos y tenerlos a mano. El alcance todavía no está bien definido, puesto que es una herramienta relativamente nueva, como ya he dicho. De momento, tenemos muchas esperanzas en lo que podemos conseguir con esta herramienta, pero seguimos un poco pendientes de ver qué se nos permite. Para la memoria del segundo cuatrimestre seguro que tengo nuevas actualizaciones sobre este proyecto, ya que solo llevo con él desde diciembre y es el proyecto más extenso de todas mis prácticas hasta el momento.

La interfaz de Microsoft Fabric es muy similar a la de Power BI online, y es porque está inspirado en esa interfaz, pero implementando más herramientas como análisis de datos en tiempo real, ingeniería de datos, ...



The screenshot shows the Microsoft Fabric 'EO-DATOS' workspace. The interface includes a left sidebar with navigation options like 'Inicio', 'Áreas de trabajo', 'OneLake', 'Supervisión', 'Tiempo real', and 'Cargas de trabajo'. The main area displays a table of data assets with columns for 'Nombre', 'Tipo', 'Tarea', 'Propietario', 'Actualizado', 'Próxima actualización', 'Aprobación', 'Confidencialidad', and 'Incluido en la aplicación'. The table lists various folders like 'Almacenes', 'FlujosGen2', 'LAGOS_DE_DATOS', 'ModelosSemanticos', 'Notebooks', 'PBIs', and 'Pipelines', as well as specific data assets like 'DATOS' (Lakehouse, Modelo semántico, and Punto de conexión).

Nombre	Tipo	Tarea	Propietario	Actualizado	Próxima actualización	Aprobación	Confidencialidad	Incluido en la aplicación
Almacenes	Carpeta	—	—	—	—	—	—	—
FlujosGen2	Carpeta	—	—	—	—	—	—	—
LAGOS_DE_DATOS	Carpeta	—	—	—	—	—	—	—
ModelosSemanticos	Carpeta	—	—	—	—	—	—	—
Notebooks	Carpeta	—	—	—	—	—	—	—
PBIs	Carpeta	—	—	—	—	—	—	—
Pipelines	Carpeta	—	—	—	—	—	—	—
DATOS	Lakehouse	—	Guillermo Mar...	—	—	—	—	—
DATOS	Modelo semántico	—	EO-DATOS	21/1/2025, 17:49...	N/D	—	—	—
DATOS	Punto de conexión	—	EO-DATOS	—	N/D	—	—	—

Ilustración 10: Interfaz Microsoft Fabric. (Elaboración propia)

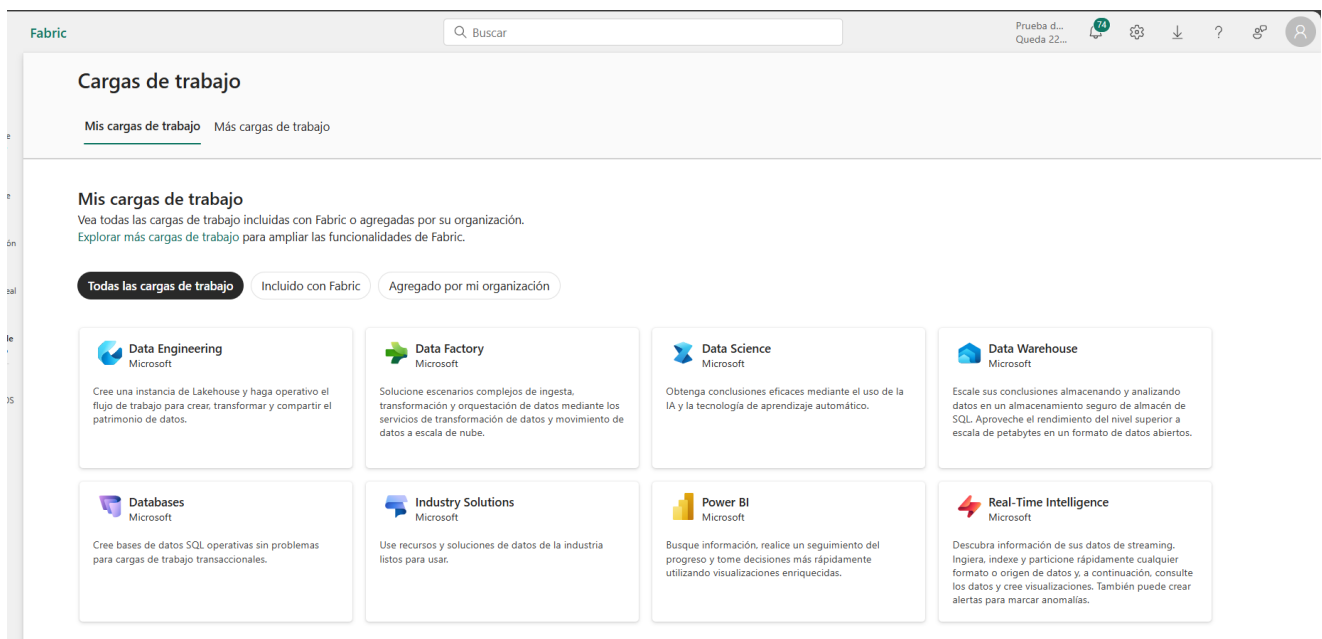


Ilustración 11: Herramientas en Microsoft Fabric. (Elaboración propia)

La forma en la que lo hemos montado es que de cada una de las distintas áreas de trabajo que hay en la compañía creamos un lago de datos y de esos lagos de datos vamos haciendo referencias a sus tablas en el lago de datos genérico que es el que aparece arriba como DATOS. La ingesta de datos en Fabric la hacemos como ya he dicho con pipelines o flujos de datos Gen 2 que son unos flujos que tienen incorporado Power Query por si hay que hacer alguna transformación antes de guardar las tablas en formato delta en los lagos de datos, además todos estos flujos y pipelines se pueden dejar programados para que se actualicen a distintas horas sin tener que estar pendiente de actualizarlos manualmente:

Nombre	Tipo	Tarea	Propietario	Actualizado	Próxima actualización	Aprobación	Confidencialidad	Incluido en la antelación
EO_GIM	Lakehouse	—	Guillermo Mar...	—	—	—	—	—
EO_GIM	Modelo semá...	—	EO-DATOS	20/1/2025, 15:23...	N/D	—	—	—
EO_GIM	Punto de con...	—	EO-DATOS	—	N/D	—	—	—
EO_JERARQUIA	Lakehouse	—	Guillermo Mar...	—	—	—	—	—
EO_JERARQUIA	Modelo semá...	—	EO-DATOS	15/1/2025, 10:16...	N/D	—	—	—
EO_JERARQUIA	Punto de con...	—	EO-DATOS	—	N/D	—	—	—

Ilustración 12: Lagos de Datos, Microsoft Fabric. (Elaboración propia)

The screenshot shows the Microsoft Fabric EO-DATOS interface. At the top, there's a search bar and navigation links like 'Crear canalización de implementación', 'Crear aplicación', 'Administrar acceso', and 'Configuración del área de trabajo'. Below this, there's a section for 'FlujosGen2' with a table of data flows.

Nombre	Tipo	Tarea	Propietario	Actualizado	Próxima actualización	Aprobación	Confidencialidad	Incluido en la aplicación
Datos_GIM	Flujo de datos...	—	Guillermo Mar...	28/1/2025, 15:07...	N/D	—	—	
Datos_JERARQUIA	Flujo de datos...	—	Guillermo Mar...	28/1/2025, 9:20:32	N/D	—	—	

Ilustración 13: Flujos de datos, Microsoft Fabric. (Elaboración propia)

2.5. Relaciones de problemas planteados y procedimientos para su resolución

El primer problema con el que me encontré al comenzar las prácticas fue que esta era mi primera experiencia en el entorno de oficina y, especialmente, en el sector tecnológico. Esto me generaba muchas dudas sobre lo que se esperaba de mí: no tenía claro qué tareas tendría que realizar, cuál sería mi carga de trabajo, o si me estarían limitando por ser becario. No sabía muy bien cómo abordar los proyectos ni cómo gestionar las relaciones con otras personas dentro de la empresa. Además, al tratarse de una compañía tan grande, con cerca de 47.000 empleados, me enfrentaba a una gran cantidad de regulaciones y normativas internas que desconocía y que fui descubriendo poco a poco conforme avanzaban los proyectos.

La solución a este problema fue bastante sencilla: preguntar. Me di cuenta de que hacer más preguntas me ayudaba a entender mejor lo que se esperaba de mí y a aclarar las dudas que me surgían sobre el funcionamiento interno de la empresa. Preguntaba lo necesario sin interrumpir demasiado, y siempre recibía respuestas rápidas que me permitían seguir avanzando en los proyectos sin quedarme atascado. Esto me ayudó a manejar mejor las tareas y a ganar confianza en mi trabajo.

Otro de los problemas más importantes con los que me encontré al comenzar fue que el resto del equipo tenía un nivel muy alto en el manejo de Power BI y, sobre todo, de Excel. Comparado con ellos, me sentía un poco atrás, ya que no alcanzaba el nivel de destreza que ellos tenían en estas

herramientas. Sin embargo, con el tiempo y la práctica diaria, fui mejorando y adquiriendo más confianza. Manejar Excel y Power BI todos los días me ha permitido mejorar en las transformaciones de datos y en los aspectos más técnicos que dominan mis compañeros. Es cierto que ellos tienen mucha más experiencia, pero poco a poco he ido cogiéndole el ritmo.

La solución a este problema ha sido precisamente la experiencia, el uso de estas herramientas día a día, familiarizándome cada día más con las herramientas que usamos en el departamento. Creo que, con el tiempo, estaré al nivel de los demás del departamento, ya que cada vez me siento más suelto y más libre sin preguntar dudas ni nada.

Ahora voy a centrarme un poco más en los problemas de los técnicos, más que los personales y de adaptación mía al departamento. Siguiendo sobre todo un poco el orden de los proyectos que he mencionado en el punto anterior y así podemos seguir los problemas que me surgieron en esos proyectos con las soluciones que fui abordando.

Comenzando por el proyecto del OCR y teniendo sobre todo en cuenta que era la primera vez que utilizaba un OCR, surgieron diversos problemas. Este proyecto estaba escrito en Python y consistía en leer los distintos IBAN de los proveedores de las cuentas de banco. Con este proyecto estuve más o menos una semana peleándome, puesto que quería perfeccionar ese OCR para que detectara todos los IBAN posibles. Como era mi primer proyecto en una compañía, yo no contaba con el factor humano, y resulta que bastantes de los PDFs que estaba tratando de leer no encontraban el IBAN simplemente porque no tenían IBAN. Entonces estaba perfeccionando una herramienta que no iba a encontrar un dato porque ese dato no existía. Esto me enseñó una gran lección y era que hay que revisar el origen de los datos antes de ponerte con un proyecto, puesto que si estás intentando obtener datos de donde no existen, no vas a lograr nada. Luego, es cierto que sí, que me sirvió todas las predicciones que hizo el código para detectar demás IBANs que sí que existían, pero no los detectaba al principio, y cosas así. Es decir, que la pelea que tuve con el código sí que me sirvió bastante para perfeccionar ese proyecto y yo creo que es lo que me puso con un buen pie en la empresa y, sobre todo, en el departamento, vieron que tenía capacidades más que suficientes para estar en ese departamento. Esto lo cuento como un problema, puesto que de ahí aprendí una lección que era aprender a mirar el origen de los datos.

Otro problema con el siguiente proyecto que tuve que hacer de la automatización de correos diarios de Iberia. Fue que era la primera vez que utilizaba la herramienta de power automate. Y como los

demás de mi departamento de datos estaban un poco liados, tuve que aprenderla yo solo. Lo cual no fue un gran problema, simplemente tardó un poco más el proyecto en salir adelante. Pero no hubo ningún problema grande por esto. La solución fue tiempo y paciencia. De hecho, la gestora del contrato de Iberia con este proyecto se quedó muy contenta con lo que hicimos. Con el tiempo que le ahorramos de los correos y demás. Y ahora la estamos un poquito utilizando de conejillo de Indias. Sobre todo para el proyecto de Microsoft Fabric, puesto que es una propuesta bastante grande para la empresa. Y necesitamos gente que nos apoye en este proyecto.

Llegamos a la parte más complicada y donde está todo lo más importante de mi trabajo, que en sí son los informes de Power BI. Digo que es la parte más complicada, puesto que hay algunos Power BI que salen al momento, y hay otros con los que puedes estar perdiéndote durante horas con los datos que te llegan. Pero, al final, la calidad de los datos influye demasiado en cómo sale este informe o reporte. Por ejemplo, con el que menciono en el punto anterior de Airbus Illescas, nos dieron bastante guerra, puesto que cada vez que les subíamos una versión del Power BI, cambiaban el tipo de archivo en el que querían hacerlo. Es decir, una semana me subían un archivo XLS, a la siguiente semana me subían una hoja de cálculo normal y, a la siguiente semana, un archivo CSV. Entonces, al final tuvimos que hablar con ellos en distintas reuniones y buscar un punto en común sobre qué era lo más sencillo para ellos de descargar de máximo, y así nosotros les haríamos las transformaciones sobre el archivo que les fuera más sencillo de subir a la carpeta de SharePoint para alimentar el Power BI. Luego, por ejemplo, en otros muchos Power BI que hemos hecho, como el de órdenes de trabajo de mantenimiento de GYM, que gestiona todas las órdenes de trabajo de la compañía en general, el dato viene más en crudo. Son descargas automáticas de SAP que no necesitan muchas transformaciones y ese no nos da problema nunca. Otros con los que no hemos tenido muchos problemas son los informes que les hacemos a Recursos Humanos, ya que, como son también descargas directas de las bases de datos de SAP y demás, no nos supone un gran problema, ya que sabemos que los archivos los van a mantener iguales más o menos siempre. Es verdad que, si hay una migración de sistema o cualquier cosa, tenemos que cambiar todos los informes que hayamos hecho para esos archivos, pero al final eso entra dentro de la parte de mantenimiento de informes.

Los siguientes problemas surgieron cuando me metieron en el proyecto de PowerApps, ya que sucedió un poco igual que con Power Automate. Era la primera vez que utilizaba PowerApps, así que necesitaba ese empujón inicial para aprender todo lo relacionado con la herramienta y cómo crear aplicaciones desde cero. Por suerte, para cuando me asignaron este proyecto, Roberto estaba

un poco más libre y me sirvió de guía para aprender PowerApps. Además, como este proyecto lo cogí de su carga de trabajo, simplemente él se sentaba a mi lado, me enseñaba un poco lo que había hecho hasta ese momento y yo iba aprendiendo cómo manejar PowerApps. A la vez, fui mejorando algunas de sus implementaciones, ya que vengo de ingeniería informática y Roberto había estudiado INEF con un máster posterior en ciencia de datos, por lo que toda esta lógica de implementación de aplicaciones yo la tengo más integrada gracias a mi formación. Este proyecto sigue activo y, de momento, estoy teniendo algunos problemas para entender cómo PowerApps gestiona los datos de las listas de SharePoint. Es por eso que, en la siguiente actualización de la aplicación, una vez la presentemos a los jefes y demás, vamos a querer incluir fechas de caducidad para cada EPI. Sin embargo, no he logrado implementar esto aún, ya que no consigo que funcione correctamente. Nadie en el departamento controla PowerApps a un nivel avanzado, así que soy ahora mismo la persona más capacitada para llevar este proyecto, pero no puedo completarlo al 100%. Es un proyecto en el que tendré que seguir aprendiendo mucho más y estoy muy motivado con él, ya que también es una oportunidad personal de superación con la que tengo muchas ganas de seguir creciendo.

Con Microsoft Fabric sucede algo similar a lo que ocurrió con PowerApps. Es un proyecto que inicié desde cero en diciembre y, desde entonces, estoy buscando constantemente cómo mejorarlo, cómo lograr que todo el proceso, desde el almacenamiento y la transformación de los datos hasta su análisis y la presentación al cliente interno o externo, sea lo más óptimo posible. Esta herramienta es bastante nueva y, de momento, soy el único en el departamento que la utiliza de esta forma; a Roberto y a Dani les estoy enseñando las mejores opciones que ofrece, y confían bastante en mí. Espero poder demostrarles que realmente pueden contar conmigo para estos proyectos.

2.6. Aprendizajes y desarrollo profesional (habilidades adquiridas)

Durante estas prácticas he aprendido a gestionar equipos y a organizar mejor mi tiempo, lo que me ha permitido enfrentar de manera más efectiva los distintos proyectos. Además, he profundizado en el uso de metodologías para la transformación de datos, lo que me ha ayudado a estructurar y procesar grandes volúmenes de información de forma ordenada. Sin embargo, lo que más me ilusiona es la parte del proyecto de Microsoft Fabric, donde combino bases de datos SQL con transformaciones de datos en SQL y PySpark, y también todo el trabajo relacionado con el análisis de datos en tiempo real. Esta área me ha permitido explorar nuevas técnicas y herramientas, lo que

no solo amplía mis conocimientos técnicos, sino que también refuerza mi pasión por el mundo del dato y la innovación.

2.7. Metodologías utilizadas

En cuanto a la organización de los proyectos, en el departamento usamos principalmente el Planner de Microsoft 365. Cada vez que llega un nuevo proyecto, se asigna directamente a la persona correspondiente del equipo a través del Planner. Esto nos permite mantener un seguimiento claro de todas las tareas, con fechas de entrega y prioridades bien definidas, y facilita la coordinación y el trabajo en equipo. Con esta metodología, todo el mundo sabe qué tiene que hacer y cuándo, lo que ayuda a que el flujo de trabajo sea más ordenado y eficiente.

2.8. Herramientas utilizadas en las prácticas (evidencias)

Para tratar este tema de herramientas utilizadas voy a hacer un orden de más conocidas a menos conocidas o como pienso yo, que se conoce más a las que menos se conocen.

La primera herramienta de la que voy a hablar, y que prácticamente todo el mundo en el entorno de oficina sabe utilizar, es Excel. En mi oficina, todos tienen conocimientos de Excel, ya sea a un nivel más básico o avanzado. Principalmente, Excel se utiliza para almacenar y organizar datos, facilitando la creación de tablas, gráficos y análisis sencillos. Además, es habitual que estos archivos se suban a carpetas en SharePoint para centralizar la información. Una vez allí, se pueden crear informes, reportes o incluso analizar el dato en bruto directamente desde los archivos, usando funciones avanzadas que permiten hacer cálculos complejos y automatizar tareas mediante macros.

Esto nos lleva a la segunda herramienta, Power BI, que es esencial para la creación de informes visuales y dinámicos. Aunque no es una herramienta que se use a diario en todos los puestos, un gran número de personas la conoce y reconocen su importancia para generar informes atractivos y comprensibles. Power BI permite transformar datos en dashboards interactivos, facilitando la toma de decisiones a partir de análisis en tiempo real. En mi oficina, utilizamos Power BI para reportar y analizar información de forma detallada, tanto para clientes internos como externos.

Habiendo hablado ya de Excel y Power BI, quiero explicar una herramienta que, aunque no es tan conocida, es esencial en mi puesto de trabajo: Power Query. Power Query es un entorno de trabajo

basado en el lenguaje de programación M de Microsoft, diseñado específicamente para la transformación, limpieza y preparación de datos. Gracias a Power Query, es posible automatizar procesos de extracción y transformación, lo que permite integrar información de múltiples fuentes y asegurar que los datos estén en el formato adecuado antes de su análisis. Esto no solo optimiza el tiempo invertido en la preparación de los datos, sino que también ayuda a hacer mejores Power BIs.

La siguiente herramienta que he utilizado bastante ha sido Python, principalmente en Visual Studio. Esta herramienta me ha sido de gran ayuda, sobre todo para crear programas y scripts que faciliten las tareas repetitivas del día a día, haciendo el trabajo mucho más eficiente. Por ejemplo, en el proyecto de OCR, donde había 500 archivos que necesitaban ser abiertos manualmente, creé una aplicación que, con solo un clic, leía los archivos y extraía los IBANs de manera automática. Otro caso fue con la herramienta iLovePDF, que se utilizaba mucho en la empresa para juntar varios archivos PDF. Vi que esto generaba dependencias externas, por lo que decidí crear una pequeña aplicación en Python que realizaba la misma tarea sin tener que depender de aplicaciones o páginas web externas. Esto no solo optimizó el proceso, sino que también permitió tener todo bajo control dentro de la infraestructura interna de la empresa. Para estos proyectos, he trabajado con diferentes librerías en Python, tanto para tareas básicas como para tareas más complejas. Por ejemplo, utilicé pandas y numpy para la manipulación de datos, tesseract para la parte de OCR (reconocimiento óptico de caracteres) y tensorflow para desarrollar modelos de aprendizaje automático. Actualmente, estamos explorando proyectos de reconocimiento de lenguaje natural (NLP), para los cuales se están utilizando nuevas librerías, pero todavía no he profundizado lo suficiente en estos proyectos. Sin embargo, ya incluiré más detalles sobre ellos en la memoria del segundo cuatrimestre, ya que de momento no he tenido mucha interacción con estas tecnologías. (Ordenador de la empresa para el proyecto de los bancos:)

```
Prueba.py 3 x  CN43N Untitled-1
C:\Users\A03427\Desktop\Codigo\AutomatizacionProveedores> Prueba.py > ...
1  import os
2  import pytesseract
3  import re
4  import cv2
5  import pandas as pd
6  import numpy as np
7  from PIL import Image
8  from pdf2image import convert_from_path
9
10 # Rutas necesarias para tesseract y poppler
11 tesseract_cmd = r".\rutas\Tesseract-OCR\tesseract.exe"
12 poppler_path = r".\rutas\poppler-24.07.0\Library\bin"
13 pytesseract.pytesseract.tesseract_cmd = tesseract_cmd
14
15 # Diccionario para sacar el banco del IBAN
16 bancos = {
17     '2080': 'ABANCA',
18     '0061': 'BANCA MARCH',
19     '0078': 'BANCA PUEYO',
20     '0081': 'BANCO SABADELL',
21     '0128': 'BANKINTER',
22     '0182': 'BBVA',
23     '3060': 'C.R.BURGOS',
24     '3159': 'CAIXA POPULAR',
25     '2100': 'CAIXABANK',
26     '3140': 'CAIXAGUISSONA',
27     '3005': 'CAJA RURAL CENTRAL',
28     '3191': 'CAJA RURAL DE ARAGÓN',
29     '3017': 'CAJA RURAL DE SORIA',
30     '3187': 'CAJA RURAL DEL SUR',
31     '3023': 'CAJA RURAL NAVARRA',
32     '3058': 'CAJAMAR',
33     '3076': 'CAJASIEETE',
34     '0019': 'DEUTSCHE BANK',
35     '3190': 'GLOBALCAJA',
36     '2085': 'IBERCAJA',
37     '1465': 'ING',
38     '2095': 'KUTXABANK',
39     '3035': 'LABORAL KUTXA',
40     '3059': 'RURALVÍA',
41     '0075': 'SANTAMER'
```

Ilustración 14: Librerías Python. (Elaboración propia)

Voy a hablar de varias de las herramientas de Microsoft que he utilizado a lo largo de las prácticas, excepto Power BI, Excel y Power Query, que ya mencioné antes. Empezaré de lo más sencillo a lo más complejo.

Primero, las listas de SharePoint. Estas listas son muy sencillas de usar: simplemente vas creando columnas y eliges el tipo de dato que quieres guardar, ya sea un número entero, un string, un objeto, una imagen, etc. Es bastante práctico para almacenar datos de forma organizada y fácil de integrar con otras aplicaciones.

La siguiente herramienta de la que voy a hablar es Visio. Visio sirve para crear flujos y diagramas de funcionamiento, lo que facilita ver de forma clara cómo funciona un código o un proceso. Es muy útil para representar procesos complejos de almacenamiento o transformación de datos.

También usamos la herramienta Planner de Microsoft 365. Con Planner, básicamente asignamos las tareas que le corresponden a cada uno del equipo, ponemos fechas de entrega, establecemos prioridades y todo lo que hace falta para organizar el trabajo. Es muy práctico para tener un seguimiento de lo que cada persona está haciendo.

Me voy a centrar en las dos herramientas principales que más he utilizado: Power Automate y PowerApps.

Power Automate es una versión no-code para crear flujos de datos y automatizar procesos repetitivos del día a día. La he utilizado, como ya he comentado, para la automatización de los correos diarios de la gestora de contrato de Iberia. Y también, lo bueno que tiene es que te incluye una especie de modelo de IA en el que, por ejemplo, si le pasas distintos PDF que tienen el mismo formato, lo puedes entrenar con 5 documentos más o menos y te puedes traer los datos si son todos bastante parecidos.

Por último, PowerApps sirve para crear aplicaciones de forma rápida, con muy poco código. En mi caso, el único código que suelo usar es HTML, que utilizo para crear documentos de salida, los cuales luego se guardan como PDF o en otros formatos.

Estas herramientas de Microsoft me han permitido agilizar muchos procesos en el departamento y facilitar la integración de datos en nuestras tareas diarias.

Ya por último, la última de las herramientas que he utilizado hasta el momento es Microsoft Fabric, que sigo utilizando en este momento. Es el proyecto más grande en el que estoy metido; cada día descubro una novedad sobre esta aplicación. Es muy interesante, ya que te permite desde almacenar el dato en bruto hasta mostrarlo en un informe, en un Excel o reportarlo directamente desde Fabric. El funcionamiento es básicamente: ingestas de datos, ya sea mediante un pipeline, un flujo de datos o una consulta SQL; esos datos, una vez almacenados en las tablas de Fabric, se pueden transformar, ya sea con query, con notebooks de PySpark, SQL, R u otros lenguajes de

programación. Una vez has hecho todas las transformaciones, de momento estamos probando a generar únicamente informes, o incluso a exportarlos como Excel, por si algún departamento quiere utilizar el dato tal y como sale de Fabric. Dentro de Fabric se incluyen un montón de herramientas que todavía no he llegado a probar; por ejemplo, hay un área de análisis en tiempo real, un área de análisis de datos y un área de ingeniería de datos que, de momento, no he tocado. Yo, sobre todo, estoy con la parte de ingeniería de datos y Power BI, que es la visualización de datos, pero estoy deseando poder llegar a todo lo demás, especialmente al análisis de datos en tiempo real, que es lo que más me llama la atención.

2.9. Logros, resultados y discusión.

Uno de los mayores logros para mí ha sido confirmar mi pasión por el mundo del dato, algo que siempre me ha fascinado desde mis que empecé en informática. Poder estar en esta beca como analista de datos ya es, en sí mismo, un gran logro personal, y cada día me siento más convencido de que estoy donde quiero estar.

Dentro del departamento, he notado que mi trabajo se valora y eso se refleja en la confianza que han depositado en mí, por ejemplo, dejándome a mí solo montar un BBDD que se pretende que use toda la compañía. Cumpló con las fechas de entrega e incluso a veces se sorprenden con la velocidad que termino alguna tarea. Me encanta ver cómo, desde la automatización de procesos con Power Automate hasta la creación de informes complejos, puedo aportar mi granito de arena para mejorar la eficiencia de la empresa.

Algo que me ha sorprendido mucho es observar la diversidad en la forma de trabajar con los datos. Aunque todos estamos en el mismo equipo, se nota que cada persona tiene su propio estilo y sus propios métodos para trabajar y transformar los datos o los flujos de trabajo. Algunas personas son súper meticulosas y otras más espontáneas, lo que me ha enseñado a valorar la importancia de convertir esos datos en algo fiable y legible para todos. Este contraste me inspira a seguir perfeccionando mis técnicas, porque sé que de ahí nace el verdadero valor de nuestro trabajo.

Mirando hacia el futuro, me gustaría la idea de profundizar aún más en áreas innovadoras como el análisis en tiempo real. Buscando implementar nuevas tecnologías para mejorar el rendimiento del equipo, como por ejemplo he notado mucho la falta de IA en la empresa y ahorraría tiempo de tareas redundantes que se hacen. Personalmente, estoy ansioso por aprender y experimentar, y creo firmemente que, a medida que sigamos explorando estas herramientas, podremos transformar por completo la manera en la que trabajamos con la información.

Creo que con la gente que me rodea en este equipo voy a mejorar mucho, ya que son gente que apoya todas las ideas, mientras no sean ideas locas, y dejan crecer a la gente mucho estando con ellos, aparte de grandes profesionales son buena gente y eso hace que haya algo más de motivación por querer hacer cosas, porque nadie quiere ver a gente buena haciendo de más simplemente porque haya gente que no quiere trabajar.

3. DESARROLLO

3.1. Introducción

Mi proyecto se estructura en tres partes que, en conjunto, permiten obtener, transformar y visualizar datos de forma integral. La primera parte consiste en el desarrollo de un script en Python. Con este script, extraigo datos en formato XML desde una fuente pública, en este caso, la página de carburantes del gobierno español (<https://datos.gob.es/en/catalogo/e05068001-precio-de-carburantes-en-las-gasolineras-espanolas>). La motivación de utilizar Python radica en su versatilidad para automatizar tareas repetitivas y en su capacidad para manejar diferentes formatos de datos.

La segunda parte del proyecto se centra en la transformación de esos datos obtenidos. Con el script, convierto el XML en un archivo CSV personalizado. Este proceso de transformación no solo optimiza la consulta reduciendo, por ejemplo, el peso de la query en comparación con la extracción directa del XML, sino que también permite generar un histórico de datos que se adapta a las necesidades específicas del análisis. De este modo, puedo disponer de una base de datos ordenada y actualizada que recoge la evolución de los precios de los carburantes a lo largo del tiempo.

La tercera parte consiste en la creación de un informe en Power BI, donde se representan visualmente los datos procesados. Mediante diversas visualizaciones, el informe permite interpretar de manera clara y dinámica la información contenida en el CSV, facilitando la identificación de tendencias y la toma de decisiones. Este enfoque integral –desde la extracción y transformación hasta la visualización– proporciona una solución completa para el análisis de datos en un entorno real.

Esto es un xml:

02/02/2025 2:04:4102250AVENIDA CASTILLA LA MANCHA, 26L-D: 07-00-22:0039,211417ABENGIBRE-1,539167DABengibre1,4590,9991,499ALBACETEIdmN° 10.935P<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0437552020702152CR CM-332, 46.4L-D: 7:00-23:0039,100389ALATOZ-1,3460831Alatoz1,5891,7091,699ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0512253020702001AVENIDA 1º DE MAYO, S/NL-S: 08:00-22:00, D: 09:00-21:0038,98567ALBACETE-1,885300NABacete1,5391,5791,6991,799ALBACETEIdmCARFORURP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0107855402070200CALLE PRINCIPE DE ASTURIAS (POLIGONO DE ROMICA), 5L-D: 06:00-23:0039,054609ALBACETE-1,832000ABacete1,5691,6391,6991,751ALBACETEIdmRBP ROMICAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0120545402070200CALLE FEDERICO GARCIA LORCA, 5L-S: 05:00-23:0038,99772ALBACETE-1,8463611ABacete1,589ALBACETEIdmCEPSAR<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0443854020702001CALLE FEDERICO GARCIA LORCA, 1L-D: 24H39,000861ALBACETE-1,849813DABacete1,5791,427ALBACETEIdmPLENOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0139354020702000CALLE CONSTANTINO ROMERO, S/NL-D: 24H39,005528ALBACETE-1,884440DABacete1,5791,427ALBACETEIdmPLENOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01601154020702000C/ PASO DE LA CUBA, 15L-D: 06:00-22:0038,99972ALBACETE-1,854556DABacete0,5391,5291,5891,6191,729ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0519554020702000AVENIDA ESCRITOR RODRIGO RUBIO, 3L-S: 09:00-21:3039,006889ALBACETE-1,885361DABacete0,8181,3791,4391,427ALBACETEIdmARAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.015000540207020005AVENIDA MENENDEZ PIDAL, 58L-D: 24H39,003335ALBACETE-1,864917NABacete1,5491,6291,5991,729ALBACETEIdmTAMOSP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0436954020702000PASO DE CUBA (LA), 36L-D: 06:00-22:0039,003081ALBACETE-1,8359917NABacete0,5391,5491,6491,5991,751ALBACETEIdmCEPSAR<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0141254020702000CALLE HERMANOS FALCÓ, 2L-D: 06:30-22:3038,989250ALBACETE-1,849028DABacete1,5691,6491,6991,759ALBACETEIdmCEPSAR<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0531354020702006PG CAMPOLLANO P-2-CT, S/NL-V: 06:00-22:00, S: 07:30-15:30, D: 09:00-15:0039,012583ALBACETE-1,874972DABacete1,5291,6091,6191,739ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0519154020702006CR N-301, 244L-D: 24H39,007778ALBACETE-1,892978ABacete1,5291,6091,6991,729ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0520854020702006CARRETERA N-322 KM. 359L-V: 07:00-21:00, S-D: 09:00-14:0039,033139ALBACETE-1,845083DABacete1,5391,5991,6191,739ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0528954020702006AVENIDA 1º DE MAYO, S/NL-D: 07:00-23:0038,985194ALBACETE-1,866806NABacete1,4891,559ALBACETEIdmPETROCAMP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0531854020702006AVDA. DE ESPAÑA (PROX. A C/ SAN JUAN), S/NL-D: 07:00-23:0038,982167ALBACETE-1,853366NABacete1,4691,529ALBACETEIdmINPEALSAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0441654020702006CARRETERA JAEN KM. 2L-D: 24H38,99467ALBACETE-1,871772DABacete1,5691,6391,6991,799ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.052454020702006AVDA. DE LOS TOREROS, S/NL-D: 07:00-23:0038,998222ALBACETE-1,869898NABacete1,4691,5191,529ALBACETEIdmINPEALSAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0441554020702006CALLE CTRA. DE JAEN, 79L-D: 06:00-22:0038,988111ALBACETE-1,883694IABacete1,5691,6391,6991,799ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0522254020702006CR N-322, 337/8L-S: 06:30-21:3038,923833ALBACETE-2,031861IABacete1,5691,6391,6991,799ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0511654020702006CALLE ALCALDE CONANGLA (C.C. EROSKI), S/NL-S: 07:00-23:00, D: 08:00-23:0038,988972ALBACETE-1,847361NABacete1,3791,4391,4291,559ALBACETEIdmFAMILY ENERGY<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0442854020702006CARRETERA NACIONAL 322 PTO KM 347 KM. 347L-S: 07:00-22:0038,965639ALBACETE-1,935500DABacete1,459ALBACETEIdmP47P<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01437754020702099CARRETERA NACIONAL 301 KM. 256L-D: 24H38,929389ALBACETE-1,831639DABacete1,4591,529ALBACETEIdmINPEALSAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0512954020702004CARRETERA N-322 KM. 349L-D: 24H38,970167ALBACETE-1,919778DABacete1,3891,6491,6991,799ALBACETEIdmCEPSAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0110485402070238AVENIDA PRINCIPE, 2L-D: 08:00-16:0038,900944SANTA ANA, 1,994028DABacete1,5091,2491,5591,5691,679ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0527254020702099CTRA AGUASNEUVAS KM 0.25L-D: 24H38,870278SALOBRAAL (EL), 1,922722DABacete1,030ALBACETEIdmCORYMARR<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0891554020702007POLIGONO CAMPOLLANO, 55L-V: 05:30-23:00, S: 06:30-22:30, D: 07:30-21:0039,015444ALBACETE-1,871361NABacete1,5391,5991,689ALBACETEIdmCEPSAR<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0136254020702007CARRETERA C-M 332 KM. 2,6L-D: 06:00-23:3038,998111ALBACETE-1,831611DABacete1,5791,6391,6991,789ALBACETEIdmINLNCOR S.L. "CEPSA" P<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0479054020702007POLIGONO AVDA. 3º, ESQUINA C/AUTOVIA, 69L-D: 24H39,01800ALBACETE-1,874333DABacete1,4791,5791,591,699ALBACETEIdmENERCOM CARBURANTES<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01280154020702007POLIGONO CAMPOLLANO AVENIDA 1, 0L-V: 06:00-22:00, S: 06:00-14:0039,016399ALBACETE-1,893948NABacete1,4391,5091,529ALBACETEIdmINPEALSAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0441754020702007SECTOR PARCELA T-3 LOCAL-52, 14L-S: 06:00-23:00, D: 08:00-23:0039,009639ALBACETE-1,878111NABacete1,3791,4391,4381,559ALBACETEIdmALCAMPOP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0960154020702007MADRID-CARTAGENA km 248,5L-D: 06:00-22:0038,971722ALBACETE-1,845278DABacete1,5391,6091,6991,6991,729ALBACETEIdmREPSPOL. LOS ANGELES DE LA MANCHA, S.L.P<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0954654020702007MADRID-CARTAGENA km 248,5L-V: 06:00-22:0038,9717717ALBACETE-1,843306IABacete1,5391,6091,6991,729ALBACETEIdmREPSPOL N° ESTACION 971179P<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01129754020702007AUTOVIA A-31 KM. 69L-D: 06:00-22:0039,015283ALBACETE-1,891361IABacete0,9491,5291,5991,6191,740ALBACETEIdmREPSPOLP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01108754020702007POLIGONO CAMPOLLANO C/F, 13L-D: 24H39,012472ALBACETE-1,889417DABacete1,389ALBACETEIdmAVANZA OILP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01012954020702007AVENIDA PRIMERA, S/NL-D: 24H39,002139ALBACETE-1,882056DABacete1,3691,419ALBACETEIdmGMOLIP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01336954020702007CALLE AUTOVIA, SNL-D: 24H39,029167ALBACETE-1,888333NABacete1,4691,539ALBACETEIdmLA GAYOTIA GASOLINERA 24 HORAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01534754020702007CALLE C/ AUTOVIA, C/ AVDA. SEXTA, S/NL-D: 24H39,029583ALBACETE-1,888083DABacete1,4691,529ALBACETEIdmNATUROY ES GNV ALBACETP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01461354020702007AVENIDA 3ª ESQUINA C/ F Nº9, 3L-D: 24H39,009417ALBACETE-1,883694NABacete1,4191,499ALBACETEIdmFAST & GASPC<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01276354020702007AVENIDA POLIGONO CAMPOLLANO AVENIDA 0, 67L-D: 24H39,021583ALBACETE-1,886556NABacete1,4891,549ALBACETEIdmVALERO 24HP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0139354020702007POLIGONO CAMPOLLANO AVENIDA -0-, 69L-V: 07:00-21:00, S-D: 09:00-14:0039,0121972ALBACETE-1,879639DABacete1,4791,539ALBACETEIdmABOILP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0441354020702008CARRETERA DE LAS PEÑAS KM 0-238 KM. 0,238L-D: 07:00-22:0038,970778ALBACETE-1,873194IABacete1,4971,559ALBACETEIdmAREA LOS LLANOS<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0144654020702008POLIGONO Nº 71, PARCELA 678, PARAJE HONDO DE LA MORENA S/N, S/NL-D: 06:30-22:3038,969639ALBACETE-1,823639DABacete1,5691,6491,6991,729ALBACETEIdmCEPSAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.01428754020702008AUTOVIA A-30 KM. 239,4L-D: 07:00-23:0038,99806ALBACETE-1,933230DABacete1,5291,6091,5891,721ALBACETEIdmCEPSAP<_x0025_x0020_BioEtanol=0.0<_x0025_x0020_Ester_x0020_metilico=0.0517354020702008AUTOVIA A-30 KM. 239,4L-D:

Ilustración 15: XML. (Gobierno de España, 2025)

3.1.1. Motivación

La motivación principal detrás de este proyecto es la necesidad de optimizar el manejo de grandes volúmenes de datos que, en su formato original, no resultan fácilmente aprovechables. La posibilidad de automatizar la conversión de un XML a un CSV, junto con la generación de informes visuales en Power BI, abre la puerta a un análisis más eficiente y profundo de la información disponible. Esto no solo aporta valor a nivel técnico, sino que también contribuye a la toma de decisiones estratégicas al ofrecer datos procesados y fácilmente interpretables.

3.1.2. Análisis de mercado y necesidades

En la actualidad, la demanda de soluciones que permitan procesar y visualizar datos de manera rápida y eficaz está en aumento. Las organizaciones necesitan herramientas que conviertan datos brutos en información útil para la toma de decisiones. En este sentido, el proyecto responde a una necesidad real: transformar datos de una fuente pública en información histórica y visualmente accesible. Este enfoque permite a las empresas optimizar el consumo de recursos, reducir tiempos de respuesta y mejorar la calidad de sus análisis.

3.1.3. Objetivos

El objetivo principal es desarrollar una aplicación integral que combine tres componentes esenciales:

- **Extracción de datos:** Mediante un script en Python se extrae información en formato XML desde la web.
- **Transformación de datos:** Se convierte el XML en un CSV personalizado que facilite la creación de históricos y la manipulación de la información.
- **Visualización de datos:** Se construye un informe interactivo en Power BI que permita analizar y presentar la información de manera clara y dinámica.

3.1.4. Requisitos técnicos

Para llevar a cabo el proyecto se utilizan diversas herramientas y tecnologías:

- **Python:** Para el desarrollo del script de extracción y transformación de datos.
- **XML y CSV:** Como formatos de entrada y salida, respectivamente, facilitando la manipulación y el almacenamiento de la información.
- **Power BI:** Para la creación de informes visuales e interactivos.

- Conexión a la fuente de datos pública de la página de carburantes del gobierno, asegurando la actualización y veracidad de la información.

3.1.5. Análisis de mercado

El análisis de mercado muestra una creciente necesidad de soluciones automatizadas para el procesamiento de datos. La transformación de información bruta en datos estructurados y visualmente interpretables es fundamental para la toma de decisiones en entornos empresariales. Este proyecto se alinea con la tendencia actual, demostrando que mediante la integración de Python y Power BI es posible crear soluciones personalizadas que optimicen la extracción, transformación y visualización de datos, aportando un valor añadido a la organización.

3.2. Marco técnico

El marco técnico del proyecto se apoya en tres pilares clave: el lenguaje de programación Python, la API proporcionada por el gobierno español y las herramientas de visualización de datos, como Power BI.

- **Python:** Utilizado para el desarrollo de un script que automatiza la extracción, transformación y almacenamiento de datos. Python es una herramienta versátil y potente para el procesamiento de grandes volúmenes de datos, y en este proyecto aprovecho librerías como requests para realizar solicitudes HTTP, csv para la escritura de archivos y logging para la captura de eventos importantes durante la ejecución del script.
- **API gubernamental:** Los datos de los precios de carburantes provienen de la API pública del gobierno español (Gobierno de España, 2025). Esta API ofrece datos en formato JSON que contienen información de las estaciones de servicio y sus precios actualizados.
- **CSV (Comma-Separated Values):** Los datos extraídos de la API en formato JSON se transforman en archivos CSV, un formato accesible y ampliamente compatible para su posterior análisis.
- **Power BI:** Una vez que los datos están almacenados en CSV, se utilizan herramientas de visualización como Power BI para crear informes visuales, interactivos y dinámicos, que permiten interpretar fácilmente los datos transformados.

La arquitectura técnica sigue una lógica de pipeline, donde los datos se extraen mediante el script en Python, se procesan y almacenan en CSV, y finalmente se presentan en Power BI.

3.3. Equipo de trabajo y metodología

El proyecto fue desarrollado principalmente de forma individual, aunque siguiendo un poco las recomendaciones de Adrián de donde mejorar el trabajo y que se puede optimizar más. También he recibido un poco de ayuda de mi equipo de prácticas en algún ajuste del Power BI sobre todo para la parte de los clusters.

En cuanto a la metodología, he utilizado un enfoque **ágil**, dividiendo el proyecto en pequeñas tareas que he ido completando de forma incremental. Cada fase del desarrollo (extracción de datos, transformación, almacenamiento, visualización) la he hecho meticulosamente y revisando todos los posibles casos de error que se me iban ocurriendo, asegurando así el mejor resultado final de proyecto.

3.4. Proyecto

3.4.1. Resumen de contribuciones y productos desarrollados

Este proyecto ha resultado en el desarrollo de una solución automatizada para la extracción, transformación y análisis de datos públicos sobre los precios de carburantes en España. Las principales contribuciones incluyen:

- **Script en Python:** El script automatiza la descarga de datos en formato JSON, su transformación a CSV y la generación de un registro histórico.
- **Almacenamiento de datos:** Los datos son almacenados en dos archivos CSV, uno con un registro histórico de todas las extracciones realizadas (con una columna adicional de FechaRegistro) y otro con los datos más actuales.
- **Transformaciones Power Query:** Para transformar los datos y hacerlos más fácil de usar en el informe interactivo.
- **Informe en Power BI:** Creación de un informe interactivo en Power BI que muestra de manera clara y visual la evolución de los precios de carburantes.

3.4.2. Planificación temporal

La planificación del proyecto ha seguido el siguiente esquema:

1. **Primeros pasos:** Analizar los archivos XML y ver que es lo más óptimo, si transformarlos con Power Query o transformar el archivo a uno más sencillo
2. **Segundos pasos:** Desarrollo del script de extracción de datos y conversión a CSV.

- ```
Clean sys.stdout
PS D:\Trabajos Practicas> python .\XMLconverter.py
2025-02-02 19:53:45,480 - INFO - Iniciando el procesamiento desde la URL: https://sedesaplicaciones.minetur.gob.es/ServiciosRESTCarburantes/PreciosCarburantes/EstacionesTerrestres/
2025-02-02 19:53:47,457 - INFO - Tipo de contenido recibido: application/json; charset=utf-8
2025-02-02 19:53:47,468 - INFO - Contenido descargado: [{"C.P.":"02250","Dirección":"AVENIDA CASTILLA LA MANCHA, 26","Horario":"L-D: 07:00-22:00","Latitud":39,211417,"Localidad":"ABENGIBRE","Longitud (WGS84)": "-1,539167","Margen":"D","Municipio":"Abengibre","Precio Biodiesel":"","Precio Bioetanol":"","Precio Gas Natural Comprimido":"","Precio Gas Natural Licuado":"","Precio Gasos Licuados del petroleo":"","Precio Gasoleo A":"1,459","Precio Gasoleo B":"0,999","Precio Gasoleo Premium":"","Precio Gas
2025-02-02 19:53:47,570 - INFO - Procesando JSON desde URL
2025-02-02 19:53:47,837 - INFO - JSON procesado con éxito.
2025-02-02 19:53:47,859 - INFO - Finalizando el script de procesamiento JSON desde URL a CSV.
```



Ilustración 16: Resultado terminal script. (Elaboración propia).

3. **Terceros pasos:** Optimización del script y creación de logs para el seguimiento del proceso.
  - (Adjunto el código en la parte de abajo en el anexo)
4. **Cuartos pasos:** Hacer las transformaciones necesarias en Power Query demostrando la comparativa de porque desde web es menos optimo que transformando el archivo.

|    | Fecha               | Código postal | Dirección                                          | Horario                                          | Localidad               |
|----|---------------------|---------------|----------------------------------------------------|--------------------------------------------------|-------------------------|
| 1  | 28/01/2025 22:04:07 | 02250         | AVENIDA CASTILLA LA MANCHA, 25                     | L-D: 07:00-22:00                                 | 39,211417 ALBACETE      |
| 2  | 28/01/2025 22:04:07 | 02152         | CR CM-332, 46,4                                    | L-D: 7:00-23:00                                  | 39,100389 ALATÓZ        |
| 3  | 28/01/2025 22:04:07 | 02001         | AVENIDA 1ª DE MAYO, S/N                            | L-S: 08:00-22:00; D: 09:00-21:00                 | 38,985667 ALBACETE      |
| 4  | 28/01/2025 22:04:07 | 02001         | CALLE PRINCEPE DE ASTURIAS (POLÍGONO DE ROMICA), 5 | L-D: 06:00-22:00                                 | 39,054694 ALBACETE      |
| 5  | 28/01/2025 22:04:07 | 02001         | CALLE FEDERICO GARCIA LORCA, 5                     | L-S: 05:00-23:00                                 | 38,997722 ALBACETE      |
| 6  | 28/01/2025 22:04:07 | 02001         | CALLE FEDERICO GARCIA LORCA, 1                     | L-D: 24H                                         | 39,000861 ALBACETE      |
| 7  | 28/01/2025 22:04:07 | 02005         | CL PASEO DE LA CUBA, 15                            | L-D: 06:00-22:00                                 | 38,999722 ALBACETE      |
| 8  | 28/01/2025 22:04:07 | 02005         | AVENIDA MENÉNDEZ PIDAL, 58                         | L-D: 24H                                         | 39,003333 ALBACETE      |
| 9  | 28/01/2025 22:04:07 | 02005         | AVENIDA ESCRITOR RODRIGO RUBIO, 3                  | L-S: 09:00-21:30                                 | 39,006889 ALBACETE      |
| 10 | 28/01/2025 22:04:07 | 02005         | PASEO CUBA (LAL), 38                               | L-D: 06:00-22:00                                 | 39,000583 ALBACETE      |
| 11 | 28/01/2025 22:04:07 | 02005         | CALLE CONSTANTINO ROMERO, S/N                      | L-D: 24H                                         | 39,005539 ALBACETE      |
| 12 | 28/01/2025 22:04:07 | 02002         | CALLE HERMANOS FALCÓ, 2                            | L-D: 06:30-22:30                                 | 38,98925 ALBACETE       |
| 13 | 28/01/2025 22:04:07 | 02006         | CARRETERA JAÉN KM. 2                               | L-D: 24H                                         | 38,994667 ALBACETE      |
| 14 | 28/01/2025 22:04:07 | 02006         | AVDA. DE LOS TOREROS, S/N                          | L-D: 07:00-23:00                                 | 38,988222 ALBACETE      |
| 15 | 28/01/2025 22:04:07 | 02006         | CALLE ALMAMSA, 2                                   | L-D: 06:00-23:00                                 | 38,984139 ALBACETE      |
| 16 | 28/01/2025 22:04:07 | 02006         | CR N-322, 337,8                                    | L-S: 06:30-21:30                                 | 38,925833 ALBACETE      |
| 17 | 28/01/2025 22:04:07 | 02006         | CALLE ALCALDE CONANGLA (C.C. EROSKI), S/N          | L-S: 07:00-23:00; D: 08:00-23:00                 | 38,988972 ALBACETE      |
| 18 | 28/01/2025 22:04:07 | 02006         | CALLE CTRA. DE JAÉN, 79                            | L-D: 06:00-22:00                                 | 38,988111 ALBACETE      |
| 19 | 28/01/2025 22:04:07 | 02006         | PG CAMPOLLANO-P. 2-CT, S.N.                        | L-V: 06:00-22:00; S: 07:30-15:30; D: 09:00-15:00 | 39,012583 ALBACETE      |
| 20 | 28/01/2025 22:04:07 | 02006         | CR N-301, 244                                      | L-D: 24H                                         | 39,027778 ALBACETE      |
| 21 | 28/01/2025 22:04:07 | 02006         | CARRETERA N-322 KM. 359                            | L-V: 07:00-21:00; S-D: 09:00-14:00               | 39,033139 ALBACETE      |
| 22 | 28/01/2025 22:04:07 | 02006         | AVDA. DE ESPAÑA (PRÓX. A C/ SAN JUAN), S/N         | L-D: 07:00-23:00                                 | 38,962167 ALBACETE      |
| 23 | 28/01/2025 22:04:07 | 02006         | AVENIDA 1ª DE MAYO, S/N                            | L-D: 07:00-23:00                                 | 38,965194 ALBACETE      |
| 24 | 28/01/2025 22:04:07 | 02006         | CARRETERA NACIONAL 322 PTO KM 347 KM. 347          | L-S: 07:00-22:00                                 | 38,965639 ALBACETE      |
| 25 | 28/01/2025 22:04:07 | 02009         | CARRETERA NACIONAL 301 KM. 256                     | L-D: 24H                                         | 38,929389 ALBACETE      |
| 26 | 28/01/2025 22:04:07 | 02004         | CARRETERA N-322 KM. 349                            | L-D: 24H                                         | 38,970167 ALBACETE      |
| 27 | 28/01/2025 22:04:07 | 02338         | AVENIDA PRINCEPE, 2                                | L-D: 08:00-16:00                                 | 38,900944 SANTA ANA     |
| 28 | 28/01/2025 22:04:07 | 02140         | CM. 3203 km 14,1                                   | L-S: 07:00-14:00; D: 09:00-14:00                 | 38,867556 SALOBRAL (EL) |
| 29 | 28/01/2025 22:04:07 | 02099         | CTRA. AGUASNEUVAS KM 0,25                          | L-D: 24H                                         | 38,870278 SALOBRAL (EL) |

Ilustración 17: Power Query. (Elaboración propia)

5. **Quintos pasos:** Integración de Power BI para la visualización de los datos almacenados.
  - (Adjunto capturas de todas las visualizaciones abajo en el anexo)

### 3.4.3. Recursos empleados

- **Python 3.12.6:** Para el desarrollo del script.
- **Librerías Python:** requests, csv, datetime, logging, tqdm (para el seguimiento del progreso).
- **Power BI:** Para la creación del informe visual.
- **API de precios de carburantes:** Fuente de datos en formato JSON.
- **Git:** Para el control de versiones del código.

### 3.4.4. Trabajo desarrollado



El trabajo principal ha sido el desarrollo del script en Python, el cual está diseñado para ejecutar automáticamente la extracción de datos desde la API del gobierno y generar archivos CSV que se pueden analizar posteriormente en Power BI.

El flujo de trabajo incluye las siguientes tareas:

1. **Descarga de datos:** El script hace una solicitud HTTP a la API, obtiene los datos en formato JSON y verifica que el contenido sea correcto.
2. **Transformación:** El JSON descargado es transformado en CSV. Se guardan dos versiones: una para el registro histórico, con la fecha de extracción, y otra para los datos más recientes.
3. **Almacenamiento:** Los datos se almacenan en archivos CSV y se generan logs para monitorear cualquier error o archivo fallido durante el proceso.
4. **Visualización:** Los CSV se importan a Power BI para la creación de informes visuales que permiten analizar los precios de carburantes en función de distintos parámetros.

### 3.5. Resultados y discusión

Hasta ahora, los resultados obtenidos han sido muy positivos. El script en Python funciona correctamente y logra extraer y almacenar los datos sin mayores complicaciones. La automatización de este proceso ha permitido ahorrar una cantidad significativa de tiempo, eliminando la necesidad de realizar estas tareas manualmente. Además, los informes en Power BI permiten analizar los precios de carburantes de manera visual, facilitando la interpretación de los datos para la toma de decisiones.

Una de las principales ventajas del proyecto es su escalabilidad: el script puede seguir ejecutándose periódicamente para actualizar los datos, creando un registro histórico que puede ser consultado en cualquier momento. Esto ofrece una visión en el tiempo de la evolución de los precios de carburantes, algo que no sería posible si solo se analizara la información actual.

Y como se esperaba alimentando el Power BI únicamente con el archivo histórico resultado de la ejecución del script se logra que tarde menos en actualizar los datos.

### 3.6. Conclusiones

Este proyecto ha demostrado que es posible integrar diferentes herramientas tecnológicas (Python, CSV, Power BI) para desarrollar soluciones completas de análisis de datos. La automatización del proceso de extracción y transformación de datos desde una API pública ha sido clave para ahorrar tiempo y mejorar la eficiencia en la gestión de grandes volúmenes de información. El uso de Power

BI como herramienta de visualización ha permitido dar valor agregado al proyecto, ofreciendo una forma clara y comprensible de interpretar los datos obtenidos.

### 3.7. Líneas futuras

De cara al futuro, una posible mejora del proyecto sería la incorporación de análisis en tiempo real, para que los datos de carburantes sean actualizados de manera más frecuente y las visualizaciones en Power BI se sincronicen automáticamente con los datos más recientes. Además, podría explorarse el uso de técnicas de machine learning para prever la evolución de los precios de carburantes en base a históricos, lo que agregaría un componente predictivo a la solución. Este proyecto también se puede escalar para predecir precios de otros países e intentar hacerlo un proyecto internacional.

## 4. BIBLIOGRAFÍA.

Precio de carburantes en las gasolineras españolas

Gobierno de España. (2025).

<https://datos.gob.es/en/catalogo/e05068001-precio-de-carburantes-en-las-gasolineras-espanolas>

Sobre Serveo

Serveo

<https://serveo.com/quienes-somos/sobre-serveo>

## 5. ANEXOS

### Anexo 1: Código de python

```
import requests
import os
import csv
import datetime
import logging
from tqdm import tqdm

logger = logging.getLogger(__name__)
logger.setLevel(logging.INFO)

formatter = logging.Formatter('%(asctime)s - %(levelname)s - %(message)s')

stream_handler = logging.StreamHandler()
stream_handler.setFormatter(formatter)

file_handler = logging.FileHandler('logfile.log')
file_handler.setFormatter(formatter)

logger.addHandler(stream_handler)
logger.addHandler(file_handler)

max_files_process = 60000

def download_json_from_url(url):
 """
 Descarga el archivo JSON desde una URL.
 """
 try:
 response = requests.get(url)
 response.raise_for_status()
```

```

content_type = response.headers.get('Content-Type', "")
logger.info(f"Tipo de contenido recibido: {content_type}")

logger.info(f"Contenido descargado: {response.text[:500]}")

if 'json' in content_type:
 return response.json()
else:
 logger.error("El contenido no es JSON.")
 return None
except requests.exceptions.RequestException as e:
 logger.error(f"Error al descargar el JSON desde {url}: {e}")
 return None

```

```

def json_to_csv_writer(json_data, csv_writer_historical, csv_writer_current, processed_files,
failed_files, file_count):

```

```

 """

```

Procesa el JSON y escribe los datos extraídos en dos archivos CSV: uno histórico (con FechaRegistro) y otro actual (sin FechaRegistro).

```

 """

```

```

unique_id = "from_url"
reg_date = datetime.datetime.now().strftime('%Y-%m-%dT%H:%M:%S')
file_count[0] += 1

```

```

 """

```

```

if unique_id in processed_files:
 logger.info(f'Archivo ya procesado anteriormente. Omitiendo...')
 return
"""

```

```

try:
 logger.info(f"Procesando JSON desde URL")

```

```

estaciones = json_data.get("ListaEESSPrecio", [])

headers = set()
for estacion in estaciones:
 headers.update(estacion.keys())

headers = sorted(headers)

if file_count[0] == 1:
 csv_writer_historical.writerow(['FechaRegistro'] + headers)
 csv_writer_current.writerow(headers)

for estacion in estaciones:
 row_with_date = [reg_date]
 row_without_date = []
 for header in headers:
 value = estacion.get(header, "")
 row_with_date.append(value)
 row_without_date.append(value)
 csv_writer_historical.writerow(row_with_date)
 csv_writer_current.writerow(row_without_date)

processed_files.add(unique_id)
logger.info(f"JSON procesado con éxito.")

except Exception as e:
 logging.error(f"Error inesperado en JSON: {e}")
 failed_files.add(f"JSON desde URL - {datetime.datetime.now().strftime('%d-%m-%Y %H:%M:%S')}")

def process_from_url(url, csv_file_historical_path, csv_file_current_path, processed_files_path,
failed_files_path, file_count):

```

```
"""
```

Procesa el JSON desde la URL y escribe los resultados en dos archivos CSV: uno histórico (con FechaRegistro) y otro actual (sin FechaRegistro).

```
"""
```

```
processed_files = set()
```

```
failed_files = set()
```

```
if os.path.exists(processed_files_path):
```

```
 with open(processed_files_path, 'r', encoding='utf-8') as file:
```

```
 processed_files = set(file.read().splitlines())
```

```
with open(csv_file_historical_path, 'a', newline="", encoding='utf-8') as csv_file_historical, \
```

```
 open(csv_file_current_path, 'w', newline="", encoding='utf-8') as csv_file_current:
```

```
 csv_writer_historical = csv.writer(csv_file_historical)
```

```
 csv_writer_current = csv.writer(csv_file_current)
```

```
 logger.info(f"Iniciando el procesamiento desde la URL: {url}")
```

```
 json_data = download_json_from_url(url)
```

```
 if json_data:
```

```
 json_to_csv_writer(json_data, csv_writer_historical, csv_writer_current, processed_files,
failed_files, file_count)
```

```
with open(processed_files_path, 'w', encoding='utf-8') as file:
```

```
 for file_path in processed_files:
```

```
 file.write(file_path + '\n')
```

```
with open(failed_files_path, 'w', encoding='utf-8') as file:
```

```
 for file_path in failed_files:
```

```
 file.write(file_path + '\n')
```

```
def main():
```

```

url =
"https://sedeaplicaciones.minetur.gob.es/ServiciosRESTCarburantes/PreciosCarburantes/EstacionesTerrestres/"
file_count = [0]

process_from_url(url, 'historico.csv', 'actual.csv', 'Archivos_Procesados.txt', 'Errores.txt',
file_count)

logger.info("Finalizando el script de procesamiento JSON desde URL a CSV.")

if __name__ == '__main__':
 main()

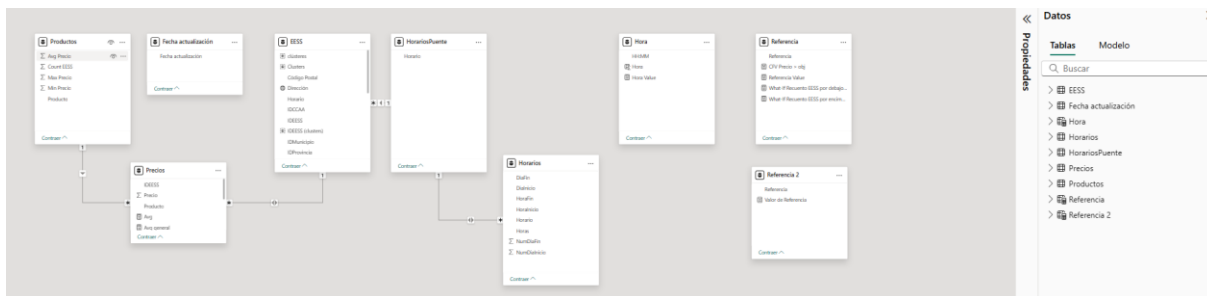
"""
Web del gobierno : https://datos.gob.es/en/catalogo/e05068001-precio-de-carburantes-en-las-gasolineras-espanolas
"""

```



## Anexo 2: PBI

## Modelo de datos del PBI



## Visualizaciones PBI

