
El presente documento es parte de la resolución del caso práctico propuesto.
Para su resolución se ha partido de los archivos de datos proporcionados, que han sido explorados con Python a través del código adjunto en formato Jupyter Notebook.

1.1 Introducción

1.1.1 Objetivo.

El objetivo del presente ejercicio se entiende como la búsqueda de la relación existente entre los registros del archivo *Facturas_AGIkey.xlsx* y la información contenida en los archivos de la carpeta *Deals_Marzo_2024.zip*, utilizando el fichero *Mapping.xlsx* como intermediario.

Se partirá de las siguientes premisas, obtenidas de los datos:

- Tras un primer análisis de la información contenida en los ficheros, se han identificado las siguientes relaciones entre los ficheros:
 - Los ficheros *Mapping.xlsx* y *Facturas_AGIkey.xlsx* se relacionan mediante los campos Origen y Servicio Facturado.
 - *Mapping.xlsx* y los archivos de la carpeta *Deals_Marzo_2024.zip* se relacionan mediante los campos de Portfolio, Comodity y Dealtype.

1.1.2 Alcance.

Como resolución de esta prueba, se ha desarrollado código Python para automatizar determinados procesos, para agilizar las tareas de BackOffice, de cara al tratamiento de las facturas. Adjuntos pueden encontrarse 2 archivos:

- **caso_practico.ipynb**: Jupyter Notebook con los resultados de la exploración de los dataset. ENTREGADO EN LA VERSIÓN 1.
- **caso_practico_2a.ipynb**: Jupyter Notebook con el código necesario para la reproducción de los resultados del apartado 2a: Conciliación de los datos de los ficheros *Facturas_AGIkey.xlsx* y los *Deals_Marzo_2024.zip*, correspondientes a servicios de las facturas de AGIKEY (C.I. Almería, Entrada PVB) con procedencia de los deals de TRS: **MEDG2_IAC_GETRA**. NOVEDAD DE LA VERSIÓN 2.
- **caso_practico_2b.ipynb**: Jupyter Notebook con el código necesario para la resolución del apartado 2b: Conciliación de los datos de los ficheros *Facturas_AGIkey.xlsx* y los *Deals_Marzo_2024.zip*, correspondientes a Servicios de (**TVB,Almacenamiento TVB**) con procedencia de TRS: **JVLNG_CTV_STOK**. NOVEDAD DE LA VERSIÓN 2.
- **funciones_caso_practico.py**: Script de python con las funciones desarrolladas para la automatización de agregación, análisis y conciliación de datos. INCLUYE FUNCIONES NUEVAS EN LA VERSIÓN 2.

1.1.3 Discrepancia identificada.

Tras analizar en profundidad la información adicional proporcionada por ENGIE el **Martes 4 de junio 2024**, vía mail, en el archivo **resultado prueba.txt** se ha identificado una discordancia en los datos:

Cuando se visualizan los datos correspondientes al Deal 37591650, según el archivo de resultado proporcionado, los subflows con Ids 93540227 y 93540226, tienen una fecha asociada de 2024-02-29 00:00:00.

Sin embargo, cuando se visualizan los mismos datos en el archivo excel, se puede ver como para el mismo Deal (37591650) hay diferencias en los datos.

- (a) Si de los 24 subflows coincidentes, basamos la selección de los mismos en los Ids en cuestión (93540227 y 93540226), la fecha asociada a esos registros es 2024-03-20 00:00:00+1 en ambos casos, lo que no concuerda con lo anterior. Si concordarían sin embargo, los importes de las subcantidades. Marcado en amarillo.
- (b) Si de los 24 subflows coincidentes, basamos la selección en la fecha (2024-02), lo cual podría ser lógico al tratarse de operaciones facturadas a mes vencido, no concuerdan los campos de subflow id ni los importes de las subcantidades. Marcado en verde.

Datos en resultado prueba.txt:

```
----> DEALS TRS <----  
Deal: 37591650, Cantidad: 503763.467648, StartDeliveryDate: 2024-01-01 00:00:00, EndDeliveryDate: 2024-12-31 00:00:00, Procedencia TRS: MEDG2_IAC_GETRA  
Subflow: 93540227, Date: 2024-02-29 00:00:00, Subcantidad: 120318.957648  
Subflow: 93540226, Date: 2024-02-29 00:00:00, Subcantidad: 383444.51
```

Datos en MEDG2_IAC_GETRA_MARZO_2024_05_23_14_12.xlsx:

Id	Date	Amount
93540225	2024-02-20T00:00:00+01:00	105279.586144
93540224	2024-02-20T00:00:00+01:00	409582.490000
93540227	2024-03-20T00:00:00+01:00	120318.957648
93540226	2024-03-20T00:00:00+01:00	383444.510000

Para la resolución del apartado 2.b, se ha tomado como premisa correcta la opción b, selección de subflows del mes anterior al mes en curso. Por lo que:

- Mes de las facturas Marzo, expresado como 2024-03.
- Mes de los subflows Febrero, expresado como 2024-02

1.2 Descripción de datasets de partida

1.2.1 Facturas Agiekey

Este dataset contiene información relativa facturas. Contiene un total de 19 columnas, estando el 15 de los mismos totalmente informados (78.9% completitud).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   NumeroFactura         14 non-null    object
1   Emisor                 14 non-null    object
2   RazonSocialEmisor     14 non-null    object
3   DepEmisor             13 non-null    object
4   Receptor              14 non-null    object
5   RazonSocialReceptor   14 non-null    object
6   DepReceptor           0 non-null     object
7   FechaFactura          14 non-null    object
8   Importe               14 non-null    object
9   Moneda                14 non-null    object
10  Estado                14 non-null    object
11  FechaEstado           14 non-null    object
12  FechaRegistro         14 non-null    object
13  Destino               14 non-null    object
14  Contrato              14 non-null    object
15  Origen                14 non-null    object
16  ServicioFacturado     14 non-null    object
17  ¿Verificada por BO?   0 non-null     object
18  Observaciones         0 non-null     object
dtypes: object(19)
memory usage: 2.2+ KB
```

Los 4 campos con valores vacíos, presentan distintos porcentajes de datos vacíos: DepEmisor [86.6% informado], DepReceptor [0% informado], ¿Verificada por BO? [0% informado] y Observaciones [0% informado] tienen valores vacíos.

Adicionalmente se extrae la siguiente información:

- Todas las facturas son válidas según los criterios establecidos anteriormente y presentan una facturación total:

NumeroFacturasValidas	NumeroClientes	FacturacionTotal
14	2	811357.98EUR

- La facturación por cliente se divide según:

RazonSocialEmisor	Emisor	NumeroOrigenes	NumServiciosFacturados	NumeroFacturas	FacturacionTotal
ENAGAS GTS, SAU	ESA86484292	2	2	13	307594.51EUR
ENAGAS TRANSPORTE, SAU	ESA86484334	1	1	1	503763.47EUR

- ENAGAS GTS, SAU, presenta facturas asociadas a Orígenes y Servicios Facturados distintos.
- ENAGAS TRANSPORTE, SAU presenta facturas asociadas a un único Origen y Servicio Facturado.
- Ambos clientes, asocian sus facturas a un único código de Contrato

1.2.2 Mapping


Este dataset **es la clave para conciliar** los datos de Facturación con los deals registrados en los archivos de la carpeta *Deals_Marzo_2024.zip*. El archivo contiene 5 columna y 47 registros, todos ellos altamente informados, por encima del 91%.


```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 47 entries, 0 to 46
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Origen                47 non-null    object
1   Servicio facturado    47 non-null    object
2   Portfolio             43 non-null    object
3   Commodity             43 non-null    object
4   Dealtype              43 non-null    object
dtypes: object(5)
memory usage: 2.0+ KB
```


Este conjunto de datos se relaciona con:


- **Facturas_AGIkey.xlsx:** Mediante los campos Origen y Servicio Facturado, presente en ambos, con la misma estructura y con valores coincidentes. Todos los servicios y orígenes descritos en el fichero de Facturas, concuerdan con alguno de los valores presentes en este conjunto de datos.
- **Archivos de Deals_Marzo_2024.zip:** Los archivos contenidos en esta carpeta, se nombran mediante la combinación de 3 campos de esta tabla según la siguiente estructura:

PORTFOLIO_COMODITY_DEALTYPE_MES_AÑO_XX_XX_XX_XX.xlsx

 JVLNG_CTV_STOK_MARZO_2024_05_23_14_12

 MEDG2_IAC_GETRA_MARZO_2024_05_23_14_12

 SUGS2_AOC_STOK_MARZO_2024_05_23_14_12

 SUGST_AOC_STOK_MARZO_2024_05_23_14_12

1.2.3 Deals_Marzo_2024.zip

Los archivos contenidos en esta carpeta presentan distintas estructuras de datos y su comprensión ha sido más complicada. Con la información adicional proporcionada en el mail del 4 de junio 2024, se ha podido comprender mejor estos archivos:

- Cada archivo, contiene información sobre los Deals de un concepto, entendiéndose este como las combinaciones de Origen y Servicio Facturado.
- Cada uno de los registros corresponde con un Deal, el cual tiene asociados diferentes Flows. La información de los Flows, se encuentra en la columna **Flow** en formato tipo json.

1.3 Resultados de la conciliación

Según lo indicado en el mail de 4 de junio de 2024, se adjunta dos archivos adicionales, son los resultados de la conciliación.

- **Respuesta_2a.txt:**
Reproduce la resolución facilitada en el mail, para los datos de los servicios de las facturas de AGIKEY (C.I. Almería, Entrada PVB) con procedencia de los deals de TRS: MEDG2_IAC_GETRA. No pasar por alto la discrepancia comentada en el punto 1.1.3.
Datos obtenidos con la ejecución del archivo caso_practico_2a.ipynb
- **Respuesta_2b.txt:**
Aplica la solución creada para el apartado anterior y lo aplica sobre los datos de los servicios de (TVB, Almacenamiento TVB) con procedencia de TRS: JVLNG_CTV_STOK.
Datos obtenidos con la ejecución del archivo caso_practico_2b.ipynb