

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:

- Para que una factura sea válida y se pueda procesar, deberá incluir:
 - Número de factura **único**.
 - Nombre y razón social del Emisor
 - Nombre y razón social del Receptor
 - Contrato, asociado a un origen y un servicio facturado.
 - Fecha de facturación
 - Importe total y moneda correspondiente
- 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 del análisis del fichero de facturas
- **funciones_caso_practico.py**: Script de python con las funciones desarrolladas para la automatización de agregación de datos de facturas.

Adicionalmente y mediante observación de los datos, se ha intentado establecer una conexión entre los movimientos de trading y las facturas. En este caso, por falta de conocimiento de algunos datos de los archivos de la carpeta *Deals_Marzo_2024.xlsx*, no ha sido posible una identificación inequívoca de los mismo, pero se comparten las observaciones correspondientes más adelante.

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, debido a no contar con un diccionario de significado de cada uno de sus campos.

Se partía de la asunción de que los registros de los 4 ficheros de esta carpeta *Deals_Marzo_2024.zip*, pueden agregarse para resultar en las cantidades facturadas en cada uno de los registros del documento *Facturas_AGIkey.xlsx*, estableciendo una relación Factura:Deals de One:Many, es decir:

- Una factura, se relacionará con uno o más deals de trading, mientras que un deal, solo podrá estar relacionado con una factura.

En este caso, debido a los distintos formatos de los 4 subconjuntos de datos, la automatización del análisis de los mismo no ha sido posible por falta de tiempo y conocimiento del significado de los campos.

Se ha realizado un análisis más manual sobre los 4 archivos, observándose para cada uno de ellos lo explicado a continuación:

- **JVLNG_CTV_STOK_MARZO_2024_05_23_14_12.xlsx:**
Archivo más voluminoso de los 4, se compone de 42 registros correspondientes a transacciones y 180 columnas con información relativa a las mismas.
- **MEDG2_IAC_GETRA_MARZO_2024_05_23_14_12.xlsx:**
Archivo compuesto por únicamente 2 registros correspondientes a transacciones y 82 columnas con información relativa a las mismas.
- **SUGS2_AOC_STOK_MARZO_2024_05_23_14_12.xlsx:**
Archivo compuesto por 13 registros correspondientes a transacciones y 97 columnas con información relativa a las mismas.
- **SUGST_AOC_STOK_MARZO_2024_05_23_14_12.xlsx**
Archivo compuesto por 6 registros correspondientes a transacciones y 96 columnas con información relativa a las mismas.

De manera general a los 4 subconjunto, se identifican campos comunes pudiendo con más tiempo relacionar los deals con las facturas, debido a los importes individuales de cada uno de ellos, la identificación de la naturaleza del deal ,venta o compra, y su aportación a la factura y el Origen y el Servicio Facturado, obtenido a través del fichero *Mapping.xlsx*