

Transformación de Datos y Preparación para Análisis en Power BI y Python - Isabel Mañero

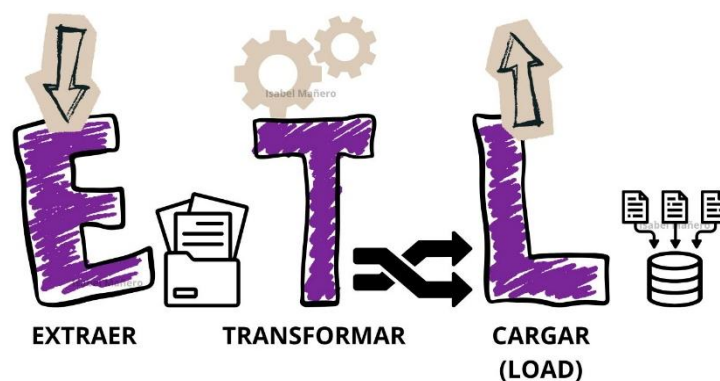
Proceso de transformación con PowerBI

El objetivo de este informe es mostrar cómo, mediante un **flujo de trabajo automatizado** en Power BI, es posible convertir datos en bruto en una estructura ordenada y preparada para el análisis, facilitando la creación de visualizaciones, dashboards y métricas avanzadas que apoyen una toma de decisiones más eficiente y fundamentada.

Partimos de un **archivo Excel (.xlsx)** con información ficticia sobre deportistas y eventos deportivos por año (cada hoja es un año diferente), que requiere transformaciones para ser analizable. Mediante **Power Query y/o Python**, se limpian y reestructuran los datos, unificando la información histórica y automatizando procesos para incorporar futuros archivos con el mismo formato. Gracias a esto, se obtiene una tabla organizada y lista para generar análisis temporales, visualizaciones y medidas DAX, optimizando el tiempo y mejorando la calidad de los resultados obtenidos.

TIPO DE DEPORTE													
DEPORTES		Fútbol		Tenis		Baloncesto		Natación		Hockey		Submarinismo	
Por grandes zonas	Por pequeñas zonas	Nº de Deportistas	Eventos	Nº de Deportistas	Eventos	Nº de Deportistas	Eventos	Nº de Deportistas	Eventos	Nº de Deportistas	Eventos	Nº de Deportistas	Eventos
ESPAÑA	ANDALUCÍA	240	7	120	61	205	3	200	21	50	2	70	4
	MADRID	400	12	260	34	400	56	150	14	500	12	15	21
	EXTREMADURA	205	14	56	68	167	22	32	15			6	4
	ASTURIAS	78	11	30	64			12	5			80	7
	CANTABRIA	190	2	30	12	23	204	7	25			50	2
	Total	923	44	466	227	772	81	394	55	550	14	171	36
INTERNACIONAL	PORTUGAL	200	12	306	4	306	23	306	5	306	4	223	85
	FRANCIA	45	2	277	19	277	14	277	4			38	11
	ITALIA	260	8	101	11	200	7	101	5	101	5	113	21
	REINO UNIDO	157	9	13	23	13	9	13	9	13	7		
	Total	662	31	697	57	796	53	697	23	420	16	374	117
Total general		1.585	75	1.163	284	1.568	134	1.091	78	970	30	545	153

Este tipo de tablas presentan varias limitaciones para el análisis directo, ya que los datos no están normalizados ni estructurados de forma óptima para obtener **insights significativos**. Para poder analizarlos de manera eficiente, es necesario aplicar transformaciones que permitan organizar la información, un proceso central en las **ETL (Extract, Transform, Load)**.



Vista general PowerBI - Editor/Transformaciones

Hojas cargas del archivo .xlsx

Estructura inicial del dataset

Transformaciones que se aplican al dataset

En este contexto, Power BI ofrece herramientas muy poderosas:

- **Power Query (lenguaje M):** permite realizar transformaciones en los datos antes de cargarlos en el modelo. Con Power Query podemos:
 - Limpiar datos (eliminar duplicados, corregir errores, homogeneizar formatos, elegir el tipo de dato (str, fecha, int...)).
 - Reestructurar tablas (unir hojas de distintos años, pivotar o despivotar columnas).
 - **Automatizar** procesos mediante scripts en el **Editor Avanzado**, de manera que cualquier nuevo archivo que cumpla con el mismo formato pueda incorporarse automáticamente al análisis.

Vista general PowerBI - Editor avanzado

En el Editor avanzado, encontramos la parte interna de esas transformaciones. Podemos comentar el código para reutilizarlo y hacerlo más legible.

```

let
    // 1. Cargar archivo Excel
    Origen = Excel.Workbook(File.Contents("C:\Users\lsau9\OneDrive\Escritorio\Deportes.xlsx"),
    // 2. Seleccionar la hoja 2012
    #2012_Sheet = Origen[[Item="2012",Kind="Sheet"]][Data],

    // 3. Promover encabezados
    #Encabezados promovidos1 = Table.PromoteHeaders(#2012_Sheet, [PromoteAllScalars=true]),
    #Encabezados promovidos2 = Table.PromoteHeaders(#Encabezados promovidos1, [PromoteAllScalars=true]),
    #Encabezados promovidos3 = Table.PromoteHeaders(#Encabezados promovidos2, [PromoteAllScalars=true]),

    // 4. Cambiar tipos de columnas iniciales (ajustar según tu tabla real)
    #Tipo cambiado1 = Table.TransformColumnTypes(#Encabezados promovidos3, {
        {"Por grandes zonas", type text}, {"Por pequeñas zonas", type text},
        {"Nº de Deportistas", Int64.Type}, {"Eventos", Int64.Type},
        {"Nº de Deportistas_1", Int64.Type}, {"Eventos_1", Int64.Type},
        {"Nº de Deportistas_2", Int64.Type}, {"Eventos_2", Int64.Type},
        {"Nº de Deportistas_3", Int64.Type}, {"Eventos_3", Int64.Type},
        {"Nº de Deportistas_4", Int64.Type}, {"Eventos_4", Int64.Type},
        {"Nº de Deportistas_5", Int64.Type}, {"Eventos_5", Int64.Type},
        {"Nº de Deportistas_6", Int64.Type}, {"Eventos_6", Int64.Type},
        {"Nº de Deportistas_7", Int64.Type}, {"Eventos_7", Int64.Type},
        {"Nº de Deportistas_8", Int64.Type}, {"Eventos_8", Int64.Type},
        {"Nº de Deportistas_9", Int64.Type}, {"Eventos_9", Int64.Type}
    }),

    // 5. Renombrar columnas de deportes
    #Columnas con nombre cambiado = Table.RenameColumns(#Tipo cambiado1, {
    
```

Por ejemplo, en este script:

```
// 8. Despivotar columnas de deportes
#"Otras columnas con anulaci3n de dinamizaci3n" = Table.UnpivotOtherColumns(
    #"Filas filtradas", {"Por grandes zonas", "Por peque1as zonas"}, "Atributo", "Valor"),
```

Despivotar columnas se refiere, a que transforma una tabla organizada de forma horizontal en una tabla organizada de forma vertical. En este proceso, las columnas que contienen valores de datos se convierten en filas. Es algo as3 como “girar” la tabla. Lo que antes era una sola fila con mucha informaci3n repartida hacia la derecha, ahora se convierte en **varias filas** (una por cada combinaci3n de deporte/m3trica) organizadas hacia abajo.

Gracias a estas herramientas, es posible generar un flujo de trabajo completamente **automatizado** que:

1. Limpia y transforma los datos desde el Excel.
2. Unifica la informaci3n hist3rica por a1os.
3. Permite realizar **visualizaciones y an3lisis temporales** de manera eficiente.

Finalmente obtenemos una tabla de este formato (con los datos de 2013 anexados tambi3n). C3mo se puede comprobar nuestra tabla tiene un formato ordenado y “analizable”.

As3, si los datos suelen llegarnos en este formato, es posible automatizar su transformaci3n y hacerla m3s eficiente, porque simplemente le pasaremos el script creado para el primer a1o, y el resto ser3 aplicar y anexar.

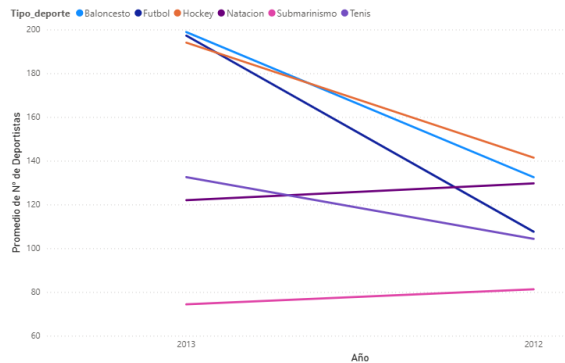
	Por grandes zonas	Por peque1as zonas	Tipo_deporte	Nº de Deportistas	Eventos	A1o
1	ESPA1A	ANDALUC3A	Baloncesto	85	3	2012
2	ESPA1A	ANDALUC3A	Futbol	120	7	2012
3	ESPA1A	ANDALUC3A	Hockey	12	2	2012
4	ESPA1A	ANDALUC3A	Natacion	115	21	2012
5	ESPA1A	ANDALUC3A	Submarinismo	130	4	2012
6	ESPA1A	ANDALUC3A	Tenis	54	61	2012
7	ESPA1A	ASTURIAS	Futbol	34	11	2012
8	ESPA1A	ASTURIAS	Natacion	52	5	2012
9	ESPA1A	ASTURIAS	Submarinismo	60	7	2012
10	ESPA1A	ASTURIAS	Tenis	12	64	2012
11	ESPA1A	CANTABRIA	Baloncesto	13	204	2012
12	ESPA1A	CANTABRIA	Futbol	78	2	2012
13	ESPA1A	CANTABRIA	Natacion	13	25	2012
14	ESPA1A	CANTABRIA	Submarinismo	45	2	2012
15	ESPA1A	CANTABRIA	Tenis	13	12	2012
16	ESPA1A	EXTREMADURA	Baloncesto	80	22	2012
17	ESPA1A	EXTREMADURA	Futbol	98	14	2012
18	ESPA1A	EXTREMADURA	Natacion	90	15	2012
19	ESPA1A	EXTREMADURA	Submarinismo	15	4	2012
20	ESPA1A	EXTREMADURA	Tenis	23	68	2012
21	ESPA1A	MADRID	Baloncesto	185	56	2012
22	ESPA1A	MADRID	Futbol	200	13	2012
23	ESPA1A	MADRID	Hockey	275	12	2012
24	ESPA1A	MADRID	Natacion	200	14	2012
25	ESPA1A	MADRID	Submarinismo	26	21	2012
26	ESPA1A	MADRID	Tenis	140	34	2012
27	INTERNACIONAL	FRANCIA	Baloncesto	277	14	2012
28	INTERNACIONAL	FRANCIA	Futbol	58	3	2012

Después de las transformaciones, podemos desde el panel de análisis de PowerBI, realizar gráficas y crear dashboards valiosos, y si nuestros datos nos lo permiten, crear medidas DAX para resultados más completos.

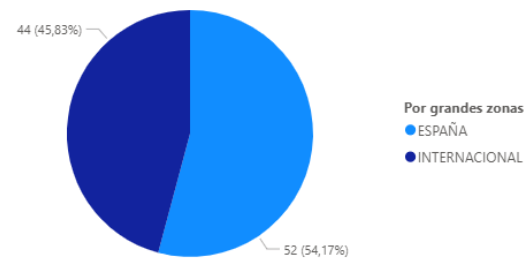
De este modo, Power BI no solo facilita la creación de dashboards, sino que asegura que los datos estén siempre listos para el análisis, optimizando el tiempo y mejorando la calidad de las decisiones basadas en información mediante la **automatización** de todo el proceso.

Ejemplos de gráficos rápidos en PowerBI:

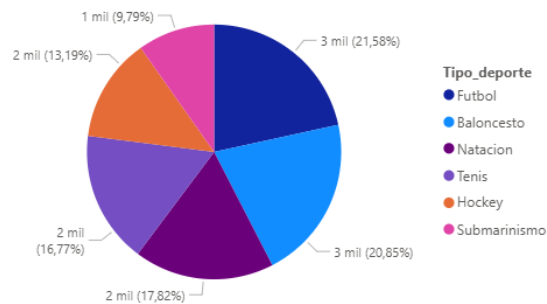
Promedio de N° de Deportistas por Año y Tipo_deporte



Recuento de Tipo_deporte por Por grandes zonas



Suma de N° de Deportistas por Tipo_deporte



Proceso de transformación con Python

Adicionalmente, este mismo proceso de transformación puede realizarse **mediante Python**, creando funciones reutilizables e integrándolas dentro de un flujo **ETL automatizado** previo a la carga en Power BI u otra herramienta de visualización.

El uso de Python aporta ventajas adicionales como mayor escalabilidad, versionado del código, pruebas automatizadas y facilidad de integración con otros sistemas, complementando perfectamente las capacidades de Power Query. De este modo, Power BI puede centrarse en la capa de análisis y visualización, mientras que Python se encarga de la preparación avanzada de los datos dentro de un proceso ETL robusto y reproducible.

Utilizando la librería pandas, como se puede ver en el siguiente ejemplo:

Se ha creado esta función:

```
import pandas as pd

def transformar_deportes_excel(ruta_excel: str, hoja: str, anio: int) -> pd.DataFrame:
    # 1. Cargamos el Excel sin procesar nada todavía [cite: 104]
    df = pd.read_excel(ruta_excel, sheet_name=hoja, header=None)

    # 2. LIMPIEZA DE CABECERAS (El truco maestro)
    # En la fila 0 están los nombres de los deportes
    # Usamos ffill() para que el nombre del deporte se repita en la columna de 'Eventos'
    # Rellenar hacia abajo
    deportes = df.iloc[0].ffill()

    # En la fila 2 están las métricas: "N de Deportistas", "Eventos"
    metricas = df.iloc[2]

    # 3. UNIÓN
    # Solo unimos si es un deporte real. Si es una zona, la dejamos igual.
    nuevos_nombres = []
    for dep, met in zip(deportes, metricas):
        # Si la columna es de zonas, no le pegamos ningún deporte
        if "zona" in str(met).lower() or pd.isna(dep) or dep == "TIPO DE DEPORTE":
            nuevos_nombres.append(met)
        else:
            # Esto creará algo como "N de Deportistas_Futbol"
            nuevos_nombres.append(f"{met}_{dep}")
```

```

# 4. QUITAR LAS FILAS QUE YA NO SIRVEN
# Quitamos las primeras 3 filas que eran los títulos [cite: 107]
df = df.iloc[3:].reset_index(drop=True)

# 5. RELLENAR ZONAS (España, Internacional...) [cite: 119]
df['Por grandes zonas'] = df['Por grandes zonas'].ffill()

# 6. QUITAR FILAS DE TOTALES
df = df[~df['Por pequeñas zonas'].str.contains('Total', na=False)]

# 7. DESPIVOTAR (Pasar de horizontal a vertical) [cite: 123]
"""Escalabilidad: Al usar el melt sin listar todos los deportes a mano,
si el archivo pasa de tener 6 deportes a tener 20, tu función seguirá
funcionando sin tocar ni una coma."""

df_unpivot = df.melt(
    id_vars=['Por grandes zonas', 'Por pequeñas zonas'],
    var_name='Atributo',
    value_name='Valor'
)

# 8. Introducimos una nueva función:

```

```

# Si la columna es "N de Deportistas_Fitoor", divide por el "_"
def separar_deporte(texto):
    if "_" in str(texto):
        partes = str(texto).split("_")
        return pd.Series([partes[0], partes[1]]) # [Metrica, Deporte]
    return pd.Series([texto, None])

df_unpivot[['metrica', 'Tipo_deporte']] = df_unpivot['Atributo'].apply(separar_deporte)

# 9. PIVOTAR PARA TENER COLUMNAS LIMPIAS
df_final = df_unpivot.pivot_table(
    index=['Por grandes zonas', 'Por pequeñas zonas', 'Tipo_deporte'],
    columns='metrica',
    values='Valor',
    aggfunc='first'
).reset_index()

# 10. AÑADIR EL AÑO
df_final['Año'] = anio

return df_final.head()

```

Este proyecto demuestra la capacidad de transformar y preparar datos partiendo de archivos Excel con estructuras no óptimas, aplicando un proceso ETL mediante **Excel**, **Power Query en Power BI** y **Python**, con el objetivo de obtener un modelo de datos limpio, automatizado y listo para el análisis. El resultado permite la creación de visualizaciones y medidas DAX, evidenciando el manejo de distintas herramientas y tecnologías clave orientado a la calidad del dato y al análisis eficiente.

Isabel Mañero