Preprocesamiento de datos sobre gestión de mantenimiento en una industria azucarera

Mauricio Ribas-García* Instituto Cubano de Investigaciones de los Derivados de la Caña de Azúcar (ICIDCA) Vía Blanca No. 804 y Carretera Central, San Miguel del Padrón. La Habana, Cuba. * mauricio.ribas@icidca.azcuba.cu

RESUMEN

En los últimos años, el crecimiento masivo en la escala de los datos es un factor clave en el actual escenario de procesamiento de datos. La eficacia de los algoritmos para la extracción de conocimiento depende, en gran medida, de la calidad de los datos, esto puede ser garantizado con el empleo de su preprocesamiento. En este trabajo se exponen los resultados de la aplicación de un preprocesamiento a datos originales, almacenados en un libro Excel, que contiene información diaria de las salidas de materiales del almacén para tareas de mantenimiento, en un central azucarero. Se identificaron los errores en la estructura de almacenamiento de datos, desarrollada por el cliente, que dificultan la extracción del conocimiento. Se evidenciaron las ventajas que ofrece la herramienta *Power Query* de *Excel,* para realizar el preprocesamiento de datos y se construyó un reporte interactivo y automatizado que permite visualizar la cantidad de materiales de cada tipo extraído por cada área, durante el período analizado.

Palabras clave: calidad de datos, preprocesamiento de datos, transformación de datos, Power Query.

ABSTRACT

In recent years, the massive growth in data scale is becoming a key factor in the current data processing scenario. The effectiveness of the algorithms for the extraction of knowledge depend on data quality, this can be guaranteed with the use of data pre-processing. In this work, the results of the processing application to original data, stored in an Excel book, which contains daily information on the material output from the store for maintenance tasks in a sugar mill, are exposed. Errors in the data storage structure developed by the user that make it difficult to extract knowledge were identified. The advantages offered by the *Power Query* tool to perform data preprocessing were shown and an interactive and automated report was built that allows visualizing the amount of materials of each type extracted by maintenance area during the analyzed period. **Key words:** data quality, data preprocessing, data transformation, *Power Query*.

INTRODUCCIÓN

El preprocesamiento de datos es una etapa fundamental en el proceso de extracción de conocimiento, cuyo objetivo principal consiste en obtener un conjunto de datos de calidad para la extracción del conocimiento (1).

La calidad del conocimiento extraído depende, en gran medida, de la calidad de los datos. Los datos, en muchas ocasiones, están afectados por factores negativos como: ruido, valores perdidos, inconsistencias, valores superfluos o un tamaño demasiado grande en cualquier dimensión (número de atributos e instancias). Está demostrado que una baja calidad de los datos conduce, en la mayoría de los casos, a una baja calidad del conocimiento extraído (1).

En este artículo se exponen los resultados del preprocesamiento realizado a los datos originales, almacenados en un libro Excel, que contiene información diaria, durante el mes de abril de 2022, de las salidas de materiales, para las tareas de mantenimiento en un central azucarero.

El objetivo de este trabajo es presentar la importancia del preprocesamiento de datos y el uso de la herramienta *Power Query* de Excel para obtener, de forma automática, una base de datos de calidad que permita extraer conocimiento de los datos.

Para ello, el trabajo se organiza como sigue. En la sección 2 se realiza una breve descripción de la herramienta *Power Query* de Excel. En la sección 3 se exponen las principales inconsistencias encontradas en el fichero original de datos Excel. En la sección 4 se describen los diferentes pasos del preprocesamiento de datos aplicado. En la sección 5 se realiza la extracción del conocimiento a partir de los datos preparados. Finalmente, en la sección 6, se presentan las conclusiones del trabajo.

DESARROLLO

Breve descripción de la herramienta power query de excel

Una de las principales características de *Excel* es su flexibilidad para generar un vínculo entre datos, tablas, gráficos y demás funcionalidades, que lo hacen la hoja de cálculo más importante en los últimos tiempos.

En muchas ocasiones, al analista de datos le solicitan crear un reporte de tabla de datos en Excel y se encuentra con una gran cantidad de anomalías e inconsistencias (respecto al formato tabular) como: filas en blanco, totales y subtotales, columnas combinadas, celdas vacías, estructura no adecuada de almacenamiento de datos. Esto implica horas frente a la PC, tratando con todo tipo de fórmulas, también las matriciales, técnicas de limpieza de datos, trucos de preparación de datos y hasta la programación de macros para, finalmente, obtener un conjunto de datos de calidad que permita generar el reporte solicitado con el uso de las tablas dinámicas.

Una vía más eficiente y rápida para realizar la preparación de datos consiste en el empleo del *Power Query* (Obtener y Transformar), que es una funcionalidad nativa y totalmente embebida de Excel a partir de 2016. Esta tecnología permite preparar, adecuar, limpiar, formatear, integrar, enriquecer, crear, refinar y, hasta resumir datos de forma veloz, intuitiva y coherente (2).

Power Query está estrechamente relacionado con un proceso ETL (*Extract-Transform-Load*). Permite extraer datos, a partir de múltiples fuentes y organizarlos en una lista. Posteriormente, se crean transformaciones de los datos (limpiarlos, formatearlos, integrarlos y enriquecerlos) para luego cargarlos desde *Power Query* a una hoja Excel listos para el análisis. En resumen, *Power Query* convierte datos en tablas que pueden ser analizadas y además permite realizar consultas sobre ellos (2).

Ventajas de Power Query

- Permite integrar tablas de una gama de orígenes extensa y constantemente creciente de modo intuitivo y eficaz.
- Brinda bondades para combinar, anexar e incrementar tablas con pasos sencillos y acertados.
- Proporciona una interfaz intuitiva para preparar tablas, sin necesidad de pasos complejo.
- Está incorporado en las tecnologías de Microsoft que manipulan datos (Microsoft Excel, Power BI de Escritorio (Desktop), Power BI en la nube (Power BI Services), Microsoft SQL Server Data Tools, Microsoft SQL Server Analysis Services, AzureAnalysis Services, Common Data Services for Apps, Microsoft flow).

- Permite crear modificaciones y automatizaciones en el proceso ETL, sin necesidad de conocimiento avanzado.
- Power Query está en constante actualización lo que expande el potencial de la herramienta.

La figura 1 muestra una imagen del panel de opciones de *Power Query (Obtener y transformar)* incluido dentro de la pestaña Datos en *Excel*. Estas opciones permiten obtener datos de diferentes fuentes. En este caso particular, los datos se obtienen desde un libro *Excel*. También es posible obtener datos desde una tabla, un rango de celdas, permite mostrar las consultas, actualizar y ver las conexiones a las fuentes de datos.

- 🖶 🐤 🗠	c† - 🔟	🖽 🕜 🔯 🔹					Libro	1 - Excel					8	- 0	• ×
Archivo In	icio I	nsertar Diseño de pl	igina Fó	rmulas Datos	Revisar	Vista	Desarrollador	Load Test	Four PDF	Power Pivot	Team	Q ¿Qué desea hacer?	Inicia	r sesión 🞗 Co	ompartir
Obtener datos externos *	Nueva consulta	Mostrar consultas Desde una tabla G Fuentes recientes	Actualizar todo -	Conexiones	21 A 2 A Ordenar	Filtro	📡 Borrar 🖅 Volver a aplicar 🎔 Avanzadas	Texto en columnas		Análisis de Previsi hipótesis -	ión Esquema	Análisis de datos	min z x s y x = 2 Model Solve	now/Hide Model uick Solve penSolver *	
	1	Ogsde un archivo		Desde un li	bro	har y	filtrar	Herramienta	is de datos	Previsión		Anālius	Opens	olver	-
G3)esde una <u>b</u> ase de datos	. •	Desde un a	rchivo <u>C</u> SV										^
A IN		Desde Azure	•	Desde un a	rchivo XML		H.	1	J	к	L	м	N O	р	(n
1 2 3		Desde otras fuentes	•	Desde un a	rchivo de <u>t</u> ext	to									
4 5	5	ombinar consultas	,	Desde una	carpeta		1							-	
6 7 8 9	Car Bús	queda en <u>c</u> atálogo de da consultas de catálogo d figuración del origen de ciones de <u>c</u> onsulta	tos e datos datos												
11 12 13															
14 15													_		
17 18	-												-		
19 20													_	-	
22 23															
Listo 17	Hoja	<u>1</u> ⊕								1 [4]		I	· · · ·	1	+ 100%

Figura 1. Imagen del panel de opciones de Power Query en Excel.

La figura 2 es una representación del editor de *Power Query* en *Excel*, con una consulta en blanco. En esta se aprecia un panel con 5 opciones.

- **Archivo.** Es donde se puede cerrar el editor de *Power Query* y cargar los resultados de la consulta, tanto en el destino predeterminado (nueva hoja *Excel*) o en el destino que se especifique. También tiene opciones para descartar los resultados de la consulta realizada y cerrar el editor de *Power Query*, así como opciones de configuración para las consultas y orígenes de datos.
- Inicio. Permite realizar operaciones con los datos, como separar los elementos de una columna en varias, agrupa las filas de la tabla según los valores de la columna seleccionada, cambia el tipo de dato de la columna seleccionada, promueve la primera fila de la tabla en los encabezados de las columnas, reemplaza valores existentes en las columna seleccionadas por un nuevo valor especificado, combina los resultados de diferentes consultas para obtener una nueva, anexa los resultados de una consulta debajo de los de otra.
- **Transformar.** Se pueden realizar operaciones, como eliminar duplicados, eliminar filas vacías, transponer, invertir el orden de las filas en la tabla, concatenar columnas seleccionadas en una columna, contar el número de filas de la tabla y otras.
- Agregar columna. Permite crear una nueva columna, basada en una fórmula personalizada, crear una nueva columna que agregue condicionalmente los valores en la columna seleccio-

nada actualmente, crea una nueva columna que sea el duplicado de una seleccionada, crea nueva columna basada en la concatenación de columnas seleccionadas, permite realizar operaciones estadísticas, matemáticas básicas, científicas, trigonométricas sobre las columnas.

 Vista. Se puede configurar la ventana de consulta que aparece a la izquierda en el editor de Power Query (figura 2). En esta ventana Power Query graba todos los pasos aplicados al origen de datos y se puede mover a través de ellos, para ir atrás o hacía adelante en el preprocesamiento de los datos.



Figura 2. Imagen del editor de Power Query en Excel.

Exploración del libro de datos e identificación de inconsistencias

Se tiene información en un libro Excel sobre la cantidad de materiales que salieron del almacén, en un período de tiempo, para ser utilizados en cada área de mantenimiento industrial. Se desea construir un reporte que permita visualizar esta información.

El libro Excel entregado por el cliente contiene 12 hojas, una para cada área de mantenimiento del central azucarero: Laboratorio, Centros de acopio, Taller de maquinado, Instrumentación, Planta eléctrica, Generación de vapor, Centrífugas, Miscelánea, Basculador, Molinos y el área de Inversión de almacén.

En el rango de celdas seleccionado, de cada hoja, se ofrece la siguiente información:

- Área de mantenimiento industrial.
- Fecha de extracción del material del almacén.
- Número de norma o trabajo en el que se utilizó el material.
- Número de vale de extracción de materiales.
- Orden de trabajo asociada.
- Actividad a ejecutar.
- El encabezado de las otras columnas se corresponde con un tipo de material y la unidad de medida. En las celdas de estas columnas se escribe la cantidad de material utilizado de cada tipo.

La figura 3 muestra un ejemplo de los datos pertenecientes al área de *Centros de acopio*. Se pueden apreciar varias inconsistencias en la información que imposibilita tener datos de calidad para, posteriormente, ser analizados.

- La información está escrita en un rango de celdas y no en formato de tablas. En el formato de tablas cada columna representa una variable y, además, tiene la ventaja de que, si se adiciona una nueva fila o columna a la tabla, inmediatamente se actualiza el reporte solicitado de tabla dinámica.
- La información del Área de mantenimiento industrial está vacía y se utilizan varias columnas combinadas. Lo adecuado es emplear una sola columna para, posteriormente, hacer el reporte de tabla dinámica con la información solicitada.
- Los encabezados de las columnas están escritos en vertical, que dificulta su lectura y visualización.
- Los nombres de los materiales no están normalizados y validados. Un mismo material se puede encontrar escrito de diferentes formas.
- Los tipos de materiales empleados en cada área son diferentes por lo que el rango de celdas varía entre una hoja y otra. En total se emplearon 151 tipos distintos de materiales, por lo que se necesitaría, en cada hoja, una columna por cada material y no todos los materiales fueron utilizados en cada área de mantenimiento. Esto no permite consolidar la información de una manera eficiente.
- El formato de fecha seleccionado es *dd/mm/aaaa*y; sin embargo, los valores de fecha aparecen escritos como dd.mm.aa.
- En muchas hojas aparecen materiales que no fueron utilizados.

Estas mismas inconsistencias se repiten para todas las hojas del libro, lo que hace necesario un preprocesamiento de los datos.

3			2101710-0-01								
4	ÁREA	S	-	No	MA	TERIALES		~T	UM	-	UM -
5	Labo	ratorio		1	Ace	ite 15w4	0		It		It
6	Cent	ros de Acopi	D	2	Ace	ite cilind	ro		It		mª
7	Talle	r Maquinado		3	Ace		U				
8	Instru	umentación		4	Ace	ite EP140)		It		kg
9	Plant	a Eléctrica		5	Ace	ite guijo			It		par
10	Gene	ración de Va	por	6	Ace	ite hidrá	ulico 68		It		m
11	Cent	rífugas		7	Ace	ite solub	le		It		
12	Casa	Calderas		8	Ace	ite turbo	68		It		
13	Miscelánea Basculador			9	Ace	ite visco	pren		It		
14				10	Acetileno			mª			
15	Molie	enda		11	Acid	lo			It		
16	Inver	sión Almacé	n	12	Acid	lo clorhid	drico		It		
17				13	Ara	ndela 24	nm		U		
18				14	Aran	ndela de	apresion 12m	m	U		
19				15	Ara	ndela pla	na 12mm		U		
20				16	Ara	ndela pla	na 16mm		U		
21				17	Ara	ndela pla	na 24mm		U		
22				18	Ara	ndela pre	sion 20mm		U		
23				19	Ara	ndela pre	sion 10mm		U		

Figura 3. Hoja de cálculo correspondiente al área de Centros de acopio.

Pasos aplicados para el preprocesamiento de datos

Paso 1. Consistió en normalizar los nombres de las áreas de mantenimiento, los materiales y las unidades de medida. Para esto se creó una nueva hoja Excel llamada *00_Listas* que contiene 3 listas en formato de tablas (no rangos de celdas) con nombres *Lista Áreas, Lista Materiales y Lista UM*. Esto permite agregar nuevos elementos que se actualizan automáticamente en las validaciones de datos y reportes de tablas dinámicas. La figura 4 muestra una imagen de la hoja Excel.



Figura 4. Hoja de cálculo con el listado de las áreas de mantenimiento, materiales y unidades de medida.

Paso 2. Se eliminaron las columnas combinadas, en cada una de las 12 hojas que tiene el libro y quedó una columna para cada encabezado. Para esto se seleccionaron las columnas combinadas y se pinchó en la opción *Combinar y central > Separar celdas.* Se eliminaron las columnas sobrantes y se le asignó un encabezado por cada columna. El texto de este se orientó en la dirección horizontal. Se completaron los datos de la columna Área con el nombre del área en cada fila (figura 5).

Paso 3. Se seleccionó el rango de celdas que contiene la información y se le aplicó formato de tabla, mediante la opción *Insertar>Tabla* del menú de opciones de Excel. La figura 5 muestra la imagen de la tabla creada con los datos del área de *Centros de acopio*. Este mismo procedimiento se repitió para cada una de las 12 hojas del libro Excel.

8 5. 6. 0	a :					BORIS LUIS ABRIL 2023	atex - Excel :							.00	- Ø	\mathbf{x}
Archivo Inicio Ins	ertar Dise	no de págin	a Fórmul	as C	atos Revisar Vista D	esarrollador Powe	r Pivot 🛛 🖓							Iniciar sesiór	A Compa	rtiel
Pegar S Portapapetes 15	• 11 1 • 田 • Fuente	A A -	# =	#> - 111 3 	Ajustar texto	General C - 96 soo 58 &	Formate	Dar format al = como tabla Estilos	o Estilos de celda	insertar	Eliminar For	tinato •	Autosuma Rellenar * Borrar *	Ordenar y filtrar *	Buscar y eleccionar *	~
H2 *	× ×	f														î
A 1		c	D	F	F	6	H	i an ii		к	i ne i	м	N	0	p	
1 UEB CENTRAL AZUCA 3 MES: ABRIL 2022	ARERO BORI	S LUIS SAN	TA COLOMA		SALIDA DE MATERIALES	-										
Area:	Fecha	Norma	No. vale	OT	Actividad a Ejecutar	PETROLEO	guantes ref. Par	aceite 15w40 Its	aceite EP140 It.	aceite h.68 Ita	aceito guijo Ita	grasa lisant li kg	oxigeno	acetileno	electrodo 6013 4mm kg	el 70
5 Centros de Acopio	01.04.22		1168		T.3		7.0									
6 Centros de Acopio	04.04.22		88	31	relleno 1436 fusible	4		5.0								
7 Centros de Acopio	05.04.22		1211		winche perchero izaje cayaja	abo							7.5	6.6	5.0	
8 Centros de Acopio	07.04.22		92		1411.3			5.0								
9 Centros de Acopio	09.04.22		1253	85	gallego fusible		-						7.5	6.6		
10 Centros de Acopio	09.04.22		93		1411.32 fusible				5.0	2.0		2.0				
11 Centros de Acopio	12.04.22		1678										U			
12 Centros de Acopio	12.04.22				estera A				-							
13 Centros de Acopio	13.04.22		1688	107	estera A, fusible						()	()	7.5			
14 Centros de Acopio	14.04.22		368		banda transp caña, cayajabos	£			_		20.0		-			
15 Centros de Acopio	14.04.22		369		banda transp caña, fusible						20.0					
16 Centros de Acopio	18.04.22	1	103		1430.03 fusible		-	20.0								
17 Centros de Acopio	18.04.22		327	-	winche izaje		-			10.0				1		
18 Centros de Acopio	20.04.22		1790	99	fusible bascula							_				
Datos	Listas	Laborator	Centr	os de A	copio Taller Maquinado	Instrumentación	Plar	• • •		0						

Figura 5. Información en formato de tabla del área Centros de acopio.

Paso 4. Se creó una consulta en *Power Query* por cada tabla seleccionada (Hacer click sobre la opción *Datos > Desde una tabla*). La figura 6 muestra el resultado de la consulta para la tabla

TCentros acopio. Dentro del editor de *Power Query* se realizaron transformaciones al formato de las columnas. Observe como los valores de la columna *Fecha* ya aparecen con el formato adecuado, de fecha *dd/mm/aaaa*. En esta tabla todavía no está resuelto el problema de dedicar una columna para cada tipo de material empleado y los tipos de materiales que se utilizan en las áreas pueden ser diferentes. Con esta estructura no se puede crear, de forma automatizada, una matriz de datos que permita consolidar, en una tabla dinámica, la información de todas las áreas.

vis	Actualizar ita previa - Admini Consulta	dades avanzado istrar • colum Admi	gir Quitar nas • columnas	Conser filas s Red	var Quitar filas ucir filas	21 11 Dividi column	Agrupar - por - tor - tor	o • fila co valore	omo encabezado • es	Combinar consultas • Combinar archivos Combinar	Administrar parámetros Parámetros	Configuración de origen de datos Orígenes de datos	Nuevo origen •	es •
×	√ f _X = T	able.Transform	olumnTypes(#	"Columnas	quitadas",(("Norma", In	t64.Type}, {"No. vale", 1	Int64	l.Type}})					`
-	A ^B _C Area -	Fecha -	123 Norma	~ 123 N	vale 🔫	123 OT .	A ^B C Actividad a Ejecutar		123 guantes r	123 aceite 15w * 123 ac	eite EP1 💌	123 aceite h.68 lts	* 123 aceite gu	ijo)
1	Centros de Acopio	01/04/202	2	nuil	1168	nul	T.3		7	null	null		null	
2	Centros de Acopio	04/04/202	2	null	88	31	relieno 1436 fusible		nult	5	null		null	2
3	Centros de Acopio	05/04/202	2	null	1211	nul	winche perchero izaje cavajabo		null	null	null		null	
4	Centros de Acopio	07/04/202	2	null	92	nut	1411,3		null	5	mult		null	
5	Centros de Acopio	09/04/202	2	null	1253	85	gallego fusible		null	null	null		null	
6	Centros de Acopio	09/04/202	2	null	93	nul	1411.32 fusible		null	nuit	5		2	
7	Centros de Acopio	12/04/202	2	nult	1678	nul		null	null	null	nult		null	
8	Centros de Acopio	12/04/202	2	null	null	nul	estera A		null	nult	null		null	
9	Centros de Acopio	13/04/202	2	null	1688	107	estera A,fusible		nu0	null	mult		null	
10	Centros de Acopio	14/04/202	2	null	368	nul	banda transp caña, cayajabos		nuti	null	null		null	
11	Centros de Acopio	14/04/202	2	null	369	nul	banda transp caña,fusible		null	null	mult		null	
12	Centros de Acopio	18/04/202	2	null	103	nul	1430.03 fusible		null	20	null		null	
13	Centros de Acopio	18/04/202	2	null	327	nul	winche izaje		null	null	null		10	
14	Centros de Acopio	20/04/202	2	null	1790	99	fusible bascula		null	null	null		null	
15	Centros de Acopio	20/04/202	z	null	1798	nul		null	null	null	null		null	
16	Centros de Acopio	19/04/202	2	null	1769	93	ventilador 2		null	null	null		null	
17	Centros de Acopio	21/04/202	2	nult	1807	102	tub.lubricacion		null	null	null		null	
18	Centros de Acopio	21/04/202	2	null	1813	102	tub.lubricacion		null	nutt	nult		null	
19	Centros de Acopio	22/04/202	2	nuit	1837	109	winche		null	null	null		null	
20	Centros de Acopio	23/04/202	2	null	1853	91	ventilador 2 fusible		null	null	null		null	
21	Centros de Acopio	23/04/202	2	null	1854	76	winche cayajabos		null	null	null		null	1
22	Centros de Acopio	26/04/202	2	mult	1879	118	estera BC		null	null	null		null	
23	Centros de Acopio	25/04/202	2	null	108	nul	1411,32		nutt	4	mult		null	- 1

Figura 6. Información en formato de tabla del área Centros de acopio.

.	A ^B C ÁREA	FECHA -	123 NORMA -	123 No. VALE -	123 OT -	A ^B _C ACTIVIDAD A EJECUTAR	A ^B C MATERIALES	1.2 VALOR -		
1	Centros de Acopio	01/04/2022	null	1168	null	T.3	guantes ref. Par	7		
2	Centros de Acopio	04/04/2022	null	88	31	relleno 1436 fusible	aceite 15w40 Its	5		
3	Centros de Acopio	05/04/2022	null	1211	null	winche perchero izaje cayajabo	oxígeno	7.5		
4	Centros de Acopio	05/04/2022	null	1211	null	winche perchero izaje cayajabo	acetileno	6.6		
5	Centros de Acopio	05/04/2022	null	1211	null	winche perchero izaje cayajabo	electrodo 6013 4mm kg	5		
6	Centros de Acopio	07/04/2022	null	92	null	1411,3	aceite 15w40 Its	5		
7	Centros de Acopio	09/04/2022	null	1253	85	gallego fusible	oxígeno	7.5		
8	Centros de Acopio	09/04/2022	null	1253	85	gallego fusible	acetileno	6.6		
9	Centros de Acopio	09/04/2022	null	93	null	1411.32 fusible	aceite EP140 It.	5		
10	Centros de Acopio	09/04/2022	null	93	null	1411.32 fusible	aceite h.68 Its	2		
11	Centros de Acopio	09/04/2022	null	93	null	1411.32 fusible	grasa lisant II kg	2		
12	Centros de Acopio	12/04/2022	null	null	null	estera A	electrodo 7018 4mm kg	5		
13	Centros de Acopio	13/04/2022	null	1688	107	estera A,fusible	oxígeno	7.5		
14	Centros de Acopio	14/04/2022	null	368	null	banda transp caña,cayajabos	aceite guijo Its	20		
15	Centros de Acopio	14/04/2022	null	369	null	banda transp caña, fusible	aceite guijo Its	20		
16	Centros de Acopio	18/04/2022	null	103	null	1430.03 fusible	aceite 15w40 Its	20		
17	Centros de Acopio	18/04/2022	null	327	null	winche izaje	aceite h.68 Its	10		
18	Centros de Acopio	20/04/2022	null	1790	99	fusible bascula	tornillo 24 x 100 U	4		
19	Centros de Acopio	20/04/2022	null	1790	99	fusible bascula	tuerca 24mm U	8		
20	Centros de Acopio	20/04/2022	null	1790	99	fusible bascula	arandela 24mm U	8		
21	Centros de Acopio	20/04/2022	null	1798	null	null	capas U	4		
22	Centros de Acopio	19/04/2022	null	1769	93	ventilador 2	oxígeno	7.5		
23	Centros de Acopio	19/04/2022	null	1769	93	ventilador 2	acetileno	6.6		
24	Centros de Acopio	19/04/2022	null	1769	93	ventilador 2	electrodo 7018 4mm kg	5		

Figura 7. Información en formato de tabla del área Centros de acopio.

Paso 5. Para resolver el problema de estructura de los datos se realizó otra transformación de estos dentro del editor de *Power Query*, que permitió resumir la información de todas las columnas asociadas a los materiales en dos nuevas columnas con nombres *Materiales* y Valor. En la primera aparecen los nombres de los materiales y, en la segunda, la cantidad solicitada por cada vale.

Esta operación se realiza al seleccionar todas las columnas con nombres de materiales y pulsar la tecla derecha del mouse, entonces, aparece un menú emergente con la opción *Anular dinamiza-ción de las columnas seleccionadas únicamente*. La figura 7 presenta el resultado de esta transformación. Se observa que la tabla de datos del área de *Centros de acopio* tiene solamente 8 columnas.

A continuación, se procede a seleccionar la opción *Cerrar y cargar en...* y aparece el siguiente menú (figura 8) en el que el usuario indica dónde va cargar los datos. En este caso se debe indicar *Crear solo conexión*.

Este proceso se repite para todas las tablas y, finalmente, se obtienen 12 nuevas, con 8 columnas cada una.

Cargar en Selecciona cómo se quieren ver estos datos en el libro. Tabla Crear solo conexión Seleccione dónde se deben cargar los datos. Hoja de cálculo nueva Hoja de cálculo existente: SFS4	argar en	
Selecciona cómo se quieren ver estos datos en el libro.		
 Tabla Crear solo conexión Seleccione dónde se deben cargar los datos. Hoja de cálculo nueva Hoja de cálculo existente: \$F\$4 	lecciona cómo se quieren ver estos datos	en el libro.
 Crear solo conexión Seleccione dónde se deben cargar los datos. Hoja de cálculo nueva Hoja de cálculo existente: \$F\$4 	⊖ Tabla	
Seleccione dónde se deben cargar los datos. Hoja de cálculo nueva Hoja de cálculo existente: \$F\$4	Orear solo conexión Orear solo conexión	
SF\$4	leccione dónde se deben cargar los datos Hoja de cálculo nueva	i.
31.34	Hoja de calculo existente:	
	3134	
	Cargat	r Cancelar
Cargar Cancelar		

Figura 8. Diferentes opciones para cargar los datos ya transformados en *Power Query.*

Paso 6. Ya se está en condiciones de anexar las tablas, una debajo de la otra, para crear una matriz de datos de calidad, que permita extraer la información solicitada. Esta operación se realiza de manera automática con el editor de *Power Query*, mediante la opción *Inicio > Anexar consultas Anexar consultas para crear una nueva*. Esto permite crear una nueva consulta que contiene las tablas anexadas, de cada área, una debajo de la otra y con el mismo encabezamiento. Finalmente se pincha la opción *Cerrar y cargar* de *Power Query*, para cargar la matriz de datos en una nueva hoja Excel que se nombró *Datos* (figura 9).

MES	S CENTRAL AZUCARER S: ABRIL 2022	D BORIS LUIS S	ANTA COLOMA							SALIDA DE MATERIALES
1	AREA	FECHA	NORMA No. VAL	ORDEN RABAJ	ACTIVIDAD A EJECUTAR	1.	MATERIAL	, VALOP,	UN	MATERIAL (UM)
1	Laboratorio	14/04/2022	1714	1	sin actividad	Pa	pel filtro	10.0	U	Papel filtro (U)
2	Laboratorio	14/04/2022	1716		batas sanitarias	Ba	tas sanitarias	20.0	U	Batas sanitarias (U)
3	Laboratorio	16/04/2022	1740		sonda	GL	iantes latex	6.0	par	Guantes latex (par)
4	Laboratorio	20/04/2022	1803	29	sin actividad	Va	rillas niquel	3.0	U	Varillas niquel (U)
5	Laboratorio	22/04/2022	1828		sin actividad	Sa	cos	100.0	U	Sacos (U)
6	Laboratorio	25/04/2022	402		prensa	Ac	eite hidráulico 68	4.0	lt	Aceite hidráulico 68 (lt)
7	Centros de Acopio	01/04/2022	1168		T.3	GL	ante reforzado	7.0	par	Guante reforzado (par)
8	Centros de Acopio	04/04/2022	88	31	relleno 1436 fusible	Ac	eite 15w40	5.0	lt	Aceite 15w40 (lt)
9	Centros de Acopio	05/04/2022	1211		winche perchero izaje cayajabo	0)	igeno	7.5	m³	Oxígeno (m ³)
10	Centros de Acopio	05/04/2022	1211		winche perchero izaje cayajabo	Ac	etileno	6.6	m³	Acetileno (m3)
11	Centros de Acopio	05/04/2022	1211		winche perchero izaje cayajabo	Ele	ectrodo 6013 4mm	5.0	kg	Electrodo 6013 4mm (kg)
12	Centros de Acopio	07/04/2022	92		1411,3	Ac	eite 15w40	5.0	lt	Aceite 15w40 (lt)
13	Centros de Acopio	09/04/2022	1253	85	gallego fusible	0)	igeno	7.5	m³	Oxígeno (m ³)
14	Centros de Acopio	09/04/2022	1253	85	gallego fusible	Ac	etileno	6.6	m³	Acetileno (m³)
15	Centros de Acopio	09/04/2022	93		1411.32 fusible	Ac	eite EP140	5.0	lt	Aceite EP140 (It)
16	Contror do Aconio	00/04/2022	02		1411 22 fucible		oito hidróulico 69	2.0	1+	Acoito hidróulico 69 (lt)

Figura 9. Matriz final de datos en Excel.

La matriz de datos de la figura 9 tiene 389 filas y 11 columnas. Si se realiza un cambio en los datos de algunas de las tablas del libro, estos se pueden actualizar en ella al pulsar la opción Actualizar, con la tecla derecha del mouse, sin necesidad de repetir otra vez todo el proceso.

Paso 7. Se aplicó una validación de datos en la columna *Material* para hacer coincidir el nombre de los materiales con los que aparecen en la tabla *Lista Materiales.*

Paso 8. Se empleó la fórmula *BUSCARV* para completar la columna *UM*, que contiene la unidad de medida del material (Ec. 1) y concatenó la columna Material con la columna UM para obtener la columna *MATERIAL (UM)* (Ec. 2).

=BUSCARV([@MATERIAL];ListaMateriales[[MATERIALES]:[UM]];2;FALSO) Ec. 1 =[@MATERIAL]&" "&"("&[@UM]&")" Ec. 2

Extracción del conocimiento, a partir de la matriz de datos generada

A partir de la matriz de datos anterior se pulsó la opción Insertar > Tabla dinámica para generar un reporte, como el que muestra la figura 10. En el mes de abril de 2022 se consumieron 70 L de Aceite circulación 220 y fue el área de Molienda la de mayor demanda de este tipo de aceite, con 67 L en el mes. Se consumieron en el mes 113 L de Aceite hidráulico 68, de ellos, 72 L fueron entregados al área de Centrífugas. Esta tabla tiene incluida a la derecha, una barra de desplazamiento vertical, para moverse entre los diferentes materiales. Se programó una macro para actualizar automáticamente la tabla dinámica, cada vez que ocurra un cambio en la matriz de datos.

No	ÁREAS/MATERIAL	Laboratorio	Centros de Acopio	Taller Maquinado	Instrumen tación	Planta Eléctrica	Generación de Vapor	Centrífugas	Casa Calderas	Miscelánea	Basculador	Molienda	Inversión Almacén	TOTAL	^
2	Aceite circulacion 220 (lt)	1.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	67.0	0.0	70.0	
3	Aceite EP140 (It)	0.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	
4	Aceite guijo (It)	0.0	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	38.0	0.0	78.0	
5	Aceite hidráulico 68 (lt)	4.0	12.0	0.0	0.0	0.0	0.0	72.0	0.0	0.0	25.0	0.0	0.0	113.0	
6	Aceite soluble (It)	0.0	0.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	10.0	0.0	0.0	15.0	
7	Aceite turbo 68 (lt)	0.0	0.0	0.0	0.0	208.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	208.0	
8	Aceite viscopren (It)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2704.0	0.0	2704.0	
9	Acetileno (m ³)	0.0	26.4	0.0	6.6	0.0	0.0	39.6	66.9	0.0	6.6	0.0	6.6	152.7	
10	Acido (lt)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1000.0	0.0	0.0	0.0	0.0	1000.0	
11	Acido clorhidrico (It)	0.0	0.0	0.0	0.0	220.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	220.0	
12	Arandela 24mm (U)	0.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.0	
13	Arandela de apresion 12mm (U)	0.0	0.0	0.0	0.0	0.0	351.0	0.0	0.0	0.0	0.0	0.0	0.0	351.0	
14	Arandela plana 12mm (U)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0	0.0	15.0	
15	Arandela plana 16mm (U)	0.0	0.0	0.0	0.0	0.0	20.0	10.0	0.0	0.0	0.0	0.0	0.0	30.0	
16	Arandela plana 24mm (U)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	2.0	۷

Figura 10. Reporte de tabla dinámica con la cantidad de materiales solicitados por área.



Figura 11. Reporte interactivo con la cantidad de materiales solicitados por fecha.

Otro reporte útil consiste en conocer la cantidad de materiales solicitados para trabajos de mantenimiento, por fechas. En este caso, se generó la tabla dinámica de la figura 11, que tiene asociada una segmentación de datos para seleccionar el tipo de material y tener la información por fecha. Paralelamente se construyó un gráfico dinámico de barras, para visualizar de manera interactiva la información solicitada.

CONCLUSIONES

- Se demostró la necesidad de realizar un preprocesamiento a los datos obtenidos directamente del problema.
- El análisis exploratorio de los datos permitió identificar varias inconsistencias en el libro Excel las que dificultan obtener los reportes solicitados.
- Con las transformaciones realizadas en *Power Query* se logró obtener una matriz de datos de calidad, a partir de la cual se obtienen los reportes de resultados.
- Se construyó un reporte de tabla dinámica que brinda información sobre la cantidad de materiales de cada tipo, solicitado por área, en el período analizado, así como el total general.
- El reporte interactivo permite consolidar la información por fecha de solicitud de materiales.

REFERENCIAS BIBLIOGRÁFICAS

- García, S., Ramírez, S., Luengo, J., Herrera, F. *Big Data:* Preprocesamiento y calidad de datos. Big Data monografía., julio-octubre 2016, España, ATI, Novática No. 237, pp.17-23. ISSN: 2444-6629.
- Caballero, M., Torres, F.EL ADN de Power Query. Power Bi Aplicado a Limpieza de Datos Inteligente. Primera Edición COLECCIÓN ADN, Colombia 2019. ISBN: 978-958-46-8524-7, pp. 638.