



Preparación de datos

Eleonora Rodríguez

- Realizar el proceso de validar, limpiar y consolidar los datos sin procesar antes de utilizarlos para realizar un análisis de negocio.



Objetivo



Preparación de datos.

- La preparación de datos, también conocida como “preprocesamiento”, es el acto de limpiar y consolidar los datos sin procesar antes de utilizarlos para realizar un análisis de negocio. Puede que no sea la tarea más valorada, pero efectuar una preparación de datos minuciosa es un componente clave para un correcto análisis de datos.
- Realizar el proceso de validar, limpiar y aumentar correctamente los datos sin procesar es fundamental para obtener insights precisos y significativos a partir de ellos. La validez y el poder de cualquier análisis de negocio dependen de la eficacia de la preparación de datos realizada en las etapas iniciales.

Importancia de la preparación de datos

- Las decisiones que toman los líderes dependen de los datos que las respaldan. Una preparación de datos cuidadosa y exhaustiva garantiza que los analistas se sientan seguros, tengan una mayor comprensión y hagan mejores preguntas sobre sus datos, lo que hace que sus análisis sean más precisos y significativos. A partir de un análisis de datos más significativo, se obtienen mejores insights y, por supuesto, mejores resultados.
- Para impulsar el nivel más profundo de análisis e insights, los equipos y las organizaciones exitosas deben implementar una estrategia de preparación de datos que priorice lo siguiente:

- **Accesibilidad:** cualquier persona, independientemente de sus habilidades, debe ser capaz de acceder a los datos de manera segura desde una única fuente confiable.
- **Transparencia:** cualquier persona debe ser capaz de ver, auditar y perfeccionar cualquier paso del proceso de preparación de datos integral que se llevó a cabo.
- **Capacidad de repetición:** la preparación de datos es conocida por ser lenta y repetitiva, y es por eso que las estrategias de preparación de datos exitosas invierten en soluciones diseñadas para ofrecer capacidad de repetición.

¿Qué pasos se deben seguir en los procesos de preparación de datos?

Data Preparation



Ref. Alteryx

- El proceso de preparación de datos puede variar según la industria o necesidad, pero normalmente consta de los siguientes pasos:

Adquisición de datos: determinación de los datos que se necesitan, recopilación de estos y establecimiento de un acceso consistente para crear análisis potentes y confiables.

- El primer paso en cualquier proceso de preparación de datos es adquirir los datos que un analista utilizará para llevar adelante su análisis. Normalmente provienen de un sistema de software de negocios o de un sistema de administración de datos. Por lo general, TI entrega estos datos en un formato accesible, como un documento de Excel o CSV.
- Un software analítico moderno puede eliminar la necesidad de depender de un intermediario para la búsqueda de datos a fin de acceder directamente a fuentes confiables como SQL, Oracle, SPSS, AWS, Snowflake, Salesforce y Marketo.

Exploración de datos: determinación de la calidad de los datos, revisión de su distribución y análisis de la relación entre cada variable para comprender mejor cómo elaborar un análisis.

Limpieza de datos: mejora de la calidad de los datos y de la productividad general para crear insights a prueba de errores.

- Durante la fase de exploración, es posible que los analistas verifiquen que los datos están mal estructurados y que se deben ordenar para mejorar su calidad. En este punto interviene la limpieza de datos. La limpieza de datos incluye los siguientes elementos:

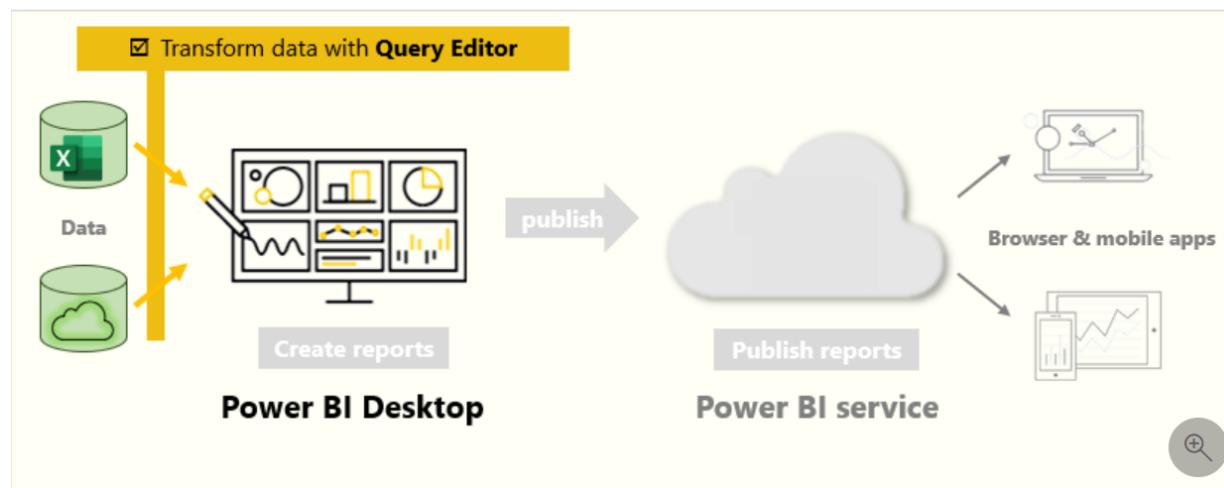
- ✓ Corrección de errores de entrada
- ✓ Eliminación de duplicados o valores atípicos
- ✓ Eliminación de datos faltantes
- ✓ Ocultación de información confidencial o sensible como nombres o direcciones

Transformación de datos: otorgar formato, orientación, adición y enriquecimiento de los conjuntos de datos utilizados en un análisis para producir insights más significativos.

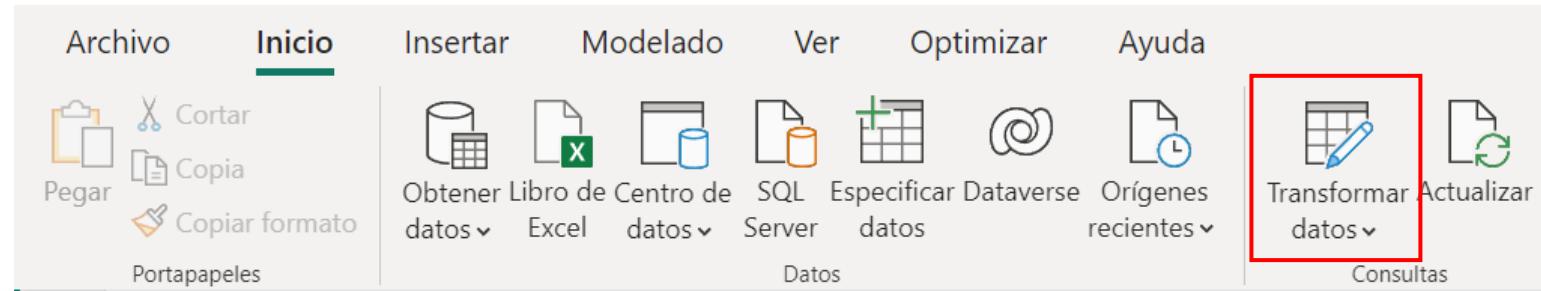
- Los datos están disponibles en muchas formas, tamaños y estructuras. Algunos están listos para el análisis, mientras que otros conjuntos de datos pueden verse como si estuvieran en un idioma extranjero.
- Un par de ejemplos comunes de transformaciones de datos son los siguientes:
 - ✓ Dinamizar o cambiar la orientación de los datos
 - ✓ Convertir los formatos de fecha
 - ✓ Agregar datos de ventas y rendimiento a través del tiempo

El Editor de Power Query

- El editor de Power Query de Power BI Desktop permite dar forma (transformar) los datos importados. Se pueden realizar acciones como, por ejemplo, cambiar el nombre de columnas o tablas, convertir texto en números, quitar filas, configurar la primera fila como encabezado entre otros. Es importante dar forma a los datos para asegurarse de que satisfacen las necesidades y son adecuados para su uso en los informes.



- Para empezar a dar forma a los datos, abra el Editor de Power Query. Para ello, seleccione la opción Transformar datos en la pestaña Inicio de Power BI Desktop.



- En el Editor de Power Query, los datos de la consulta seleccionada se muestran en el centro de la pantalla y, en el lado izquierdo, el panel de Consultas muestra las consultas disponibles (tablas).

Laboratorio 7 - Editor de Power Query

Archivo Inicio Transformar Agregar columna Vista Herramientas Ayuda

Cerrar y aplicar Nuevo Orígenes recientes Especificar datos Configuración de origen de datos Administrar parámetros Actualizar vista previa Propiedades Editor avanzado Consulta Administrar Elegir columnas Quitar columnas Conservar filas Quitar filas Ordenar Dividir columna Agrupar por Usar la primera fila como encabezado Reemplazar los valores Tipo de datos: Número entero Combinar consultas Anexar consultas Combinar archivos Combinar

Cerrar Nueva consulta Orígenes de datos Parámetros Administrar Consulta Administrar columnas Reducir filas Ordenar Transformar Combinar

Consultas [2]

= Table.ReplaceValue(#"Filas filtradas", "CHO", "CHA", Replacer.ReplaceValue, {"Abreviación Oponente"})

	Season	Top.px.(Location)	Left.px.(location)	Fecha	Abreviación Oponente	Oponente	Location
1	2015	296	329	15/11/14	CHA		null HOME
2	2015	256	74	15/11/14	CHA		null HOME
3	2015	136	82	15/11/14	CHA		null HOME
4	2015	133	74	15/11/14	CHA		null HOME
5	2015	192	104	15/11/14	CHA		null HOME
6	2015	255	57	15/11/14	CHA		null HOME
7	2015	157	240	15/11/14	CHA		null HOME
8	2015	54	230	15/11/14	CHA		null HOME
9	2015	179	16	15/11/14	CHA		null HOME
10	2015	280	74	15/11/14	CHA		null HOME
11	2015	179	375	15/11/14	CHA		null HOME
12	2015	190	381	15/11/14	CHA		null HOME
13	2015	67	253	15/11/14	CHA		null HOME
14	2015	81	30	15/11/14	CHA		null HOME
15	2015	316	208	15/11/14	CHA		null HOME
16	2015	215	446	28/11/14	CHA		null AWAY
17	2015	294	397	28/11/14	CHA		null AWAY
18	2015	78	241	28/11/14	CHA		null AWAY
19	2015	292	139	28/11/14	CHA		null AWAY
20	2015	185	348	28/11/14	CHA		null AWAY
21	2015	80	214	28/11/14	CHA		null AWAY
22	2015	239	414	28/11/14	CHA		null AWAY
23	2015	72	176	28/11/14	CHA		null AWAY
24	2015	269	113	28/11/14	CHA		null AWAY
25	2015	51	225	28/11/14	CHA		null AWAY
26	2015	272	386	28/11/14	CHA		null AWAY
27	2015	160	293	28/11/14	CHA		null AWAY
28	2015	138	240	28/11/14	CHA		null AWAY
29	2015	76	479	28/11/14	CHA		null AWAY
30	2015	253	387	28/11/14	CHA		null AWAY
31	2015	219	413	28/11/14	CHA		null AWAY

15 COLUMNAS, 116 FILAS Generación de perfiles de columnas basada en las 1000 primeras filas

- En el lado derecho, se registran todos los pasos que se realizan para dar forma a los datos. Después, cada vez que la consulta se conecta al origen de datos, aplica automáticamente los pasos, por lo que los datos siempre tendrán la forma especificada.
- Si se han realizado todos los cambios se selecciona Cerrar y aplicar, para cerrar el Editor de Power Query y aplicar los cambios al modelo de datos.

The screenshot shows the Microsoft Power Query Editor interface. On the left, there's a preview of the data with columns labeled: Season, Top.px. (Location), Left.px. (location), Fecha, Abreviación Oponente, Oponente, and Location. The data consists of 31 rows of NBA team statistics. On the right, the ribbon menu is visible with tabs like Agregar columna, Vista, Herramientas, and Ayuda. Below the ribbon, there are various tools: Clasificar, Configuración de origen de datos, Administrar parámetros, Actualizar vista previa, Propiedades, Editor avanzado, Administrar consultas, Elegir columnas, Quitar columnas, Conservar filas, Quitar filas, Reducir filas, Ordenar, Dividir columna, Agrupar por, Reemplazar los valores, Transformar, Combinar consultas, Anexar consultas, Combinar archivos, and Combinar. A status bar at the bottom indicates "VISTA PREVIA DESCARGADA A LAS 23:33".

Dentro de las acciones que se pueden realizar para la transformación de datos están:

- Identificación de encabezados y nombres de columna
- Promoción de encabezados
- Cambio de nombre de las columnas
- Quitar filas superiores
- Eliminación de columnas
- Anulación de dinamización de columnas
- Cambio de nombre de una consulta
- Reemplazo de valores, entre otras.
- Todas estas acciones y muchas más se encuentran en la cinta de opciones de Power Query.

Archivo Inicio Transformar Agregar columna Vista Herramientas Ayuda

Cerrar y aplicar ▾ Nuevo origen ▾ Orígenes recientes ▾ Especificar datos Configuración de origen de datos Administrar parámetros ▾ Actualizar vista previa ▾ Propiedades Editor avanzado Administrar ▾ Elegir columnas ▾ Quitar columnas ▾ Administrar columnas Cerrar Nueva consulta Orígenes de datos Parámetros Consulta Administrar Consulta Administrar columnas

Conservar filas ▾ Quitar filas ▾ Dividir columna ▾ Agrupar por Tipo de datos: Número entero ▾ Usar la primera fila como encabezado ▾ Reemplazar los valores Combinar consultas ▾ Anexar consultas ▾ Combinar archivos Combinar Text Analytics Visión Azure Machine Learning Conclusiones de IA Reducir filas Ordenar Transformar

Referencias bibliográficas



- <https://www.alteryx.com/es-419/glossary/data-preparation#:~:text=Limpieza%20de%20datos%3A%20mejora%20de,para%20producir%20insights%20m%C3%A1s%20significativos.>
- <https://learn.microsoft.com/es-es/training/modules/clean-data-power-bi/2-shape-data>