

«Talento Tech»

Business Intelligence

Clase 05





Clase N° 05 | Preparación y Procesamiento de Datos

Temario:

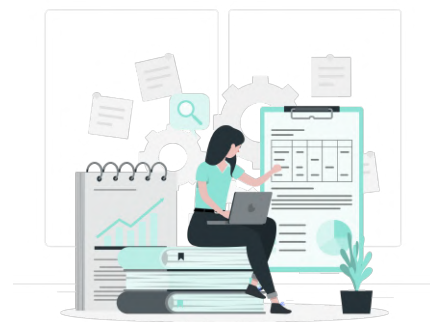
- Introducción a la preparación de datos
 - Limpieza de datos en Google Sheets
 - Uso de funciones básicas para transformar datos
-

Objetivos de la clase:

- Conocer la finalidad y los procesos involucrados en la preparación de datos.
- Familiarizarse con las funciones de Google Sheets.

Introducción a la Preparación de Datos

La preparación y procesamiento de datos es un paso fundamental en la analítica de datos y la inteligencia de negocios (BI). Esta etapa implica **organizar, limpiar y transformar datos brutos en un formato adecuado para su análisis**. La finalidad de este proceso, es obtener una estructura apropiada para su posterior análisis. La calidad de los datos es crucial, ya que cualquier error o inconsistencia puede afectar la precisión de los resultados obtenidos en la posterior toma de decisiones.



Definición

La preparación de datos incluye todas las **acciones necesarias para garantizar que los datos sean precisos, relevantes y estén listos para ser analizados**. Este proceso puede variar desde la recolección de datos hasta la transformación y la integración de diferentes fuentes de información.

Limpieza de Datos en Google Sheets

Una de las herramientas más accesibles para la limpieza y preparación de datos es **Google Sheets**. Esta plataforma permite a los usuarios realizar una variedad de tareas de limpieza con funciones integradas y manualmente.



Principales Actividades de Limpieza de Datos

- **Eliminación de Duplicados:** Identificar y eliminar filas que contengan información repetida. Ejemplo: En un listado de clientes, puede haber entradas duplicadas que se deben remover para evitar contabilizar a un mismo cliente varias veces.

- **Corrección de Errores y Valores Anómalos:** Arreglar errores tipográficos o de formato, corregir valores fuera de rango. Ejemplo: Si un campo "Estado" contiene variantes como "Buenos Aires" y "B.A.", se debe estandarizar.
- **Tratamiento de Valores Faltantes:** Decidir cómo manejar los datos que no se registraron. Ejemplo: Se puede optar por eliminar filas con datos faltantes o imputar valores promediando las columnas de datos presentes.
- **Normalización de Datos:** Ajustar los datos a un formato y rango consistentemente. Ejemplo: Convertir todas las fechas al formato "DD/MM/AAAA".
- **Integración de Diferentes Fuentes de Datos:** Combinar varias fuentes de datos en una sola estructura. Ejemplo: Agregar una columna con los nombres de las localidades a partir de una tabla de códigos postales.

Pasos para Limpiar Datos en Google Sheets

1. **Importar Datos:** Cargar los datos en Google Sheets desde la fuente necesaria.
2. **Eliminar Duplicados:** Utilizar la función "Datos" > "Eliminar Duplicados".
3. **Buscar Errores y Valores Anómalos:** Crear filtros para identificar errores visualmente y corregirlos manualmente. Identificar valores fuera de rango y aplicar una estrategia apropiada para su tratamiento (ignorar, estimar o reemplazar).
4. **Tratar Valores Faltantes:** Aplicar estrategias para trabajar los valores faltantes, por ejemplo, utilizar `PROMEDIO()` para completar con un estimado a partir de otros valores.
5. **Estandarizar y Normalizar:** Usar fórmulas como `MAYUS()`, `MINUS()` o `NOMPROPIO()` para uniformar el texto.
6. **Integrar Diferentes Fuentes de Datos:** Usar fórmulas como `IMPORTRANGE()`, `BUSCARX()` o `QUERY()` para obtener la estructura de la tabla requerida.

Uso de Funciones Básicas para Transformar Datos

Google Sheets ofrece una **variedad de funciones que facilitan la transformación de datos**, lo que permite manipular información para obtener insights más relevantes.



Algunas de las Funciones Comunes

Función CONCATENAR	Permite combinar dos o más celdas en una sola. Ejemplo: =CONCATENAR (A2, " ", B2) fusiona el nombre y apellido en una sola celda.
Función SI	Crea condiciones lógicas que permiten tomar decisiones en función del contenido de las celdas. Ejemplo: =SI (C2>100, "Alto", "Bajo") indica si el valor en C2 es mayor a 100.
Función BUSCARX	Útil para buscar datos en tablas. Ejemplo: =BUSCARX (D2, A2:A, B2:B) busca el valor de D2 en la columna A y devuelve el valor correspondiente de la columna B.
Función TEXTO	Permite dar formato a números y fechas. Ejemplo: =TEXTO (E2, "DD/MM/YYYY") transforma una fecha en otro formato.

Reflexión final



La **preparación y el procesamiento de datos** son actividades esenciales en el ámbito del BI, ya que aseguran que la información sea válida y útil para el análisis. **Google Sheets** nos permite llevar a cabo estas tareas, permitiendo a los usuarios manipular datos de manera efectiva con funciones básicas y técnicas de limpieza. Con un enfoque sistemático y el uso adecuado de las herramientas disponibles, los analistas de datos pueden optimizar el valor de la información y facilitar la toma de decisiones informadas.

Materiales y recursos adicionales

- [Lista de funciones de Hojas de cálculo de Google](#)

Próximos pasos

- Uso de Google Sheets para combinar y agrupar datos
- Casos prácticos de limpieza de datos
- Transformación de datos para análisis
- Fórmulas condicionales

Ejercicios prácticos:



DATAWISE CONSULTING
INSIGHTS ESTRATÉGICOS

Actividad 1: Limpieza de Datos

Contexto



Estás inmerso en tu pasantía en DataWise Consulting. Luis, el Analista de BI, se acerca a vos con una tarea clave: limpiar un conjunto de datos que contiene información de personas entre 30 y 50 años. La limpieza de datos es fundamental, ya que garantiza que la información que se utilizará más adelante sea precisa y confiable.

Objetivos

El objetivo de esta actividad es practicar la identificación y limpieza de errores en el conjunto de datos. Aprenderás a trabajar con datos desordenados y a aplicar las técnicas necesarias para que la información sea útil.

Ejercicio Práctico

Recibís un conjunto de datos en Google Sheets con información sobre personas entre 30 y 50 años. Tu tarea consiste en realizar las siguientes acciones:

- **Eliminar Duplicados:** Asegurate de que no haya registros duplicados basados en el campo "Email".
- **Corregir Errores Léxicos:** Uniformizá el nombre "Ana" y corregí la variante "Perez" a "Pérez" (con acento).
- **Completar Valores Faltantes:** Usá la función `PROMEDIO()` para completar el campo "Edad" donde falte información.

- Corregir Formatos Incorrectos: Verificá que todas las fechas estén en el formato "DD/MM/AAAA".
- Filtrar Valores Atípicos: Identificá cualquier valor atípico en la columna "Edad".

Sets de Datos

[Datasets Clase 05 - BI Consulting](#) Dataset de personas entre 30 y 50 años,

¿Por qué importa esto en DataWise Consulting?

La limpieza de datos es vital para DataWise Consulting. Un conjunto de datos preciso y ordenado permite que Luis y el resto del equipo realicen análisis fiables y tomen decisiones estratégicas fundamentadas en datos limpios. La calidad de la información que manejas tiene un impacto directo en el éxito de los proyectos y la satisfacción del cliente.

Actividad 2: Normalización y Estandarización

Contexto



Después de realizar la limpieza de datos, ahora es el turno de llevar esa información a un nivel superior. Matías, el Desarrollador de BI, se acerca para explicarte la importancia de la normalización y estandarización de datos. Con esto, vas a garantizar que toda la información esté en un formato coherente y homogéneo, facilitando los análisis posteriores.

Objetivos

El objetivo de esta actividad es ejercitar la normalización y estandarización de datos para mejorar su calidad. Aprenderás a aplicar reglas que hagan que los datos sean más eficaces y fáciles de usar en futuras consultas.

Ejercicio Práctico

Tendrás que trabajar en el mismo conjunto de datos que limpiaste en la actividad anterior y realizar lo siguiente:

- **Estandarizar Nombres:** Asegurate de que los nombres de las categorías estén en el mismo formato (usar mayúsculas solamente en la primera letra).

- **Completar Valores Faltantes:** Completá el campo "Precio" utilizando un valor promedio de otros productos de la misma categoría.
- **Corregir Errores de Entrada:** Revisá que todos los campos de "Fecha de Ingreso" tengan el mismo formato ("DD/MM/AAAA") .
- **Identificar Valores Atípicos:** Filtrá los productos cuyo "Stock" esté fuera de un rango razonable (por ejemplo, menos de 10 o más de 300).
- **Crear Tablas Normalizadas:** Creá tablas separadas para los productos y las categorías, respetando las reglas de normalización. Este punto lo continuaremos en la próxima clase.

Sets de Datos

[Datasets Clase 05 - BI Consulting](#) Dataset de productos que incluye categorías, precios y stock.

¿Por qué importa esto en DataWise Consulting?

La normalización y estandarización de datos son procesos clave en DataWise Consulting. Al trabajar con datos bien estructurados y homogéneos, el equipo puede generar análisis más precisos y útiles. Esto también permite que todos los analistas y desarrolladores, como Luis, Matías y Sabrina, puedan comunicarse más efectivamente y colaborar en proyectos, asegurando que todos trabajen con la misma información confiable.

⊖ **Estos ejercicios son una simulación de cómo se podría resolver el problema en este contexto específico. Las soluciones encontradas no aplican de ninguna manera a todos los casos.**

Recuerda que las soluciones dependen de los sets de datos, el contexto y los requerimientos específicos de los stakeholders y las organizaciones.



Buenos Aires
aprende
Agencia de Políticas para el Futuro

BA Buenos
Aires
Ciudad