

«Talento Tech»

Business Intelligence

Clase 06



Clase N° 06 | Preparación y Procesamiento de Datos II

Temario:

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

Objetivos de la clase:

Proporcionar las herramientas y técnicas necesarias para la preparación efectiva de datos.

- Aprender a gestionar y optimizar sus conjuntos de datos.
- Mejorar la calidad y precisión de los análisis.

Uso de Google Sheets para Combinar y Agrupar Datos

Una de las funcionalidades más útiles en Google Sheets es la **capacidad de combinar y agrupar datos**. Esto se puede lograr mediante varias **funciones y técnicas**. Repasemos algunas de las funciones ya aprendidas y cómo se aplican en este contexto.



1. Función IMPORTRANGE

La función IMPORTRANGE permite importar un rango de celdas de una hoja de cálculo a otra. Esto es especialmente útil si trabajamos con múltiples fuentes de datos.

Ejemplo	<code>=IMPORTRANGE("URL_de_la_hoja", "Hoja1!A1:B10")</code>
----------------	---

2. Función QUERY

La función QUERY permite realizar consultas similares a SQL sobre un rango de datos. Esto facilita la combinación y el agrupamiento de datos.

Ejemplo	<code>=QUERY(A1:C10, "SELECT A, SUM(B) WHERE C='Valor' GROUP BY A", 1)</code>
----------------	---



3. Uso de UNIQUE y FILTER

La combinación de UNIQUE y FILTER puede ayudar a agrupar datos sin duplicados y aplicar filtros a conjuntos específicos.

Ejemplo	<code>=UNIQUE (FILTER (A1:A10, B1:B10="Condición"))</code>
----------------	--

4. Combinación de Tablas usando BUSCARX

La función BUSCARX es una herramienta potente para buscar y combinar datos de diferentes tablas. Permite realizar búsquedas en rangos específicos y devolver resultados de manera efectiva. Es especialmente útil si estás trabajando con datos estructurados en dos tablas diferentes.

Ejemplo:

Supongamos que tenemos dos tablas:

Tabla 1: (A1:B5)

ID	Nombre
1	Juan
2	Ana
3	Luis
4	María
5	Pablo

Tabla 2: (D1:E5)

ID	Edad
1	30
2	25
3	28
4	32
5	29

Para combinar las dos tablas y obtener la edad de cada nombre, usaríamos `BUSCARX()` de la siguiente manera:

`=BUSCARX(A2, D:D, E:E)`

En este caso, colocamos esta fórmula en la celda C2 y la arrastramos hacia abajo para llenar las edades correspondientes a cada nombre, buscando en la columna ID de la Tabla 2.

Casos Prácticos de Limpieza de Datos

La **limpieza de datos** es un paso crítico en la preparación de datos. Aquí se presentan algunos casos prácticos:

1. Obtener valores únicos

Para obtener los valores únicos en un rango de datos, podemos usar la función `UNIQUE()`.

Ejemplo:

`=UNIQUE(A1:A100)`



2. Manejo de Valores Nulos

Para reemplazar valores nulos o vacíos, se puede usar `SI()` junto con `ESBLANCO()`.

Ejemplo:

`=IF(ESBLANCO(A1), "Valor por defecto", A1)`

3. Normalización de Datos

A veces, los datos pueden venir en diferentes formatos, como mayúsculas y minúsculas. Usar `NOMPROPIO`, `MAYUSC` o `MINUS` puede ayudar a normalizarlos:

Ejemplo:

`=MAYUSC(A1)`

Transformación de Datos para Análisis

La **transformación de datos** es fundamental para facilitar el análisis posterior. Algunas funciones útiles incluyen:

1. Funciones Matemáticas Básicas

Para realizar operaciones matemáticas, como sumas o promedios:

Ejemplo:

`=SUM(A1:A10)`

`=PROMEDIO(A1:A10)`

Estas funciones básicas pueden combinarse con condicionales.



2. Concatenación de Datos

Para combinar texto de diferentes celdas, podemos usar & o la función `CONCATENAR()`.

Ejemplo:

```
=CONCATENAR(A1; " "; B1)
=A1 & " " & B1
```

Recordemos que la operación inversa a la concatenación se puede ejecutar con la función `SPLIT()`.

3. Crear Columnas Derivadas

Si necesitamos crear una columna que dependa de otras, como calcular el precio después de impuestos:

Ejemplo:

```
=A1 * (1 + Tasa_impuesto)
```

Fórmulas Condicionales

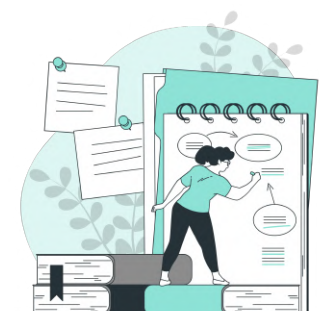
Las fórmulas condicionales permiten aplicar lógicas específicas en la manipulación de datos.

1. Uso de IF

La función `SI()` permite ejecutar diferentes acciones basadas en condiciones.

Ejemplo:

```
=SI(A1 > 100, "Alto", "Bajo")
```





2. Uso de IFS

Para múltiples condiciones, `IFS()` puede ser más eficaz.

Ejemplo:

```
=IFS(A1 > 100, "Alto", A1 > 50, "Medio", A1 <= 50, "Bajo")
```

3. SWITCH

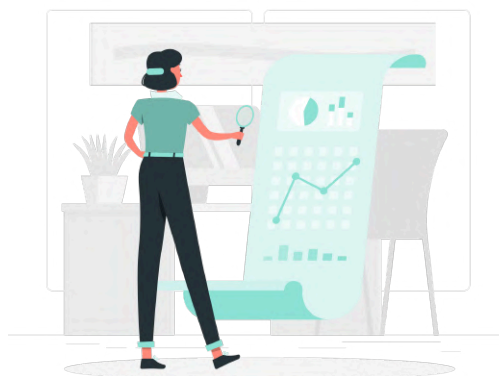
Cuando tenemos un número limitado de opciones, `SWITCH()` puede simplificar el código.

Ejemplo:

```
=SWITCH(A1, "Opción1", "Resultado1", "Opción2", "Resultado2",  
"Opción3", "Resultado3")
```

Reflexión final

Google Sheets es una herramienta versátil para la preparación de datos en BI consulting. Aprender a utilizar funciones como `IMPORTRANGE`, `QUERY`, y `BUSCARX` es esencial para un **manejo eficiente de los datos**. Te recomendamos que profundices en estas prácticas para que estés mejor equipado para realizar tareas de limpieza y transformación de datos con efectividad.





Materiales y recursos adicionales

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

Próximos pasos

- Conceptos básicos: media, mediana, moda, desviación estándar. Correlaciones.
- Análisis exploratorio de datos con Google Sheets
- Visualización de estadísticas básicas: barras, pies y líneas.

Ejercicios prácticos:



DATAWISE CONSULTING
INSIGHTS ESTRATÉGICOS

Actividad 1: Combinación de Datos

Contexto



Al día siguiente de haber trabajado con la normalización de datos, Silvia, la Project Manager y Especialista en datos, te convoca para una nueva tarea. Se trata de combinar varias fuentes de datos relacionadas con el stock y los movimientos de productos. Comprender cómo integrar esta información es fundamental para generar informes precisos.

Objetivos

El objetivo de esta actividad es combinar las fuentes de datos proporcionadas para obtener una tabla que contenga información relevante sobre los productos, su stock y las operaciones realizadas. Aprenderás a manejar y cruzar datos provenientes de diferentes fuentes, una habilidad esencial en el mundo del BI.

Ejercicio Práctico

Debes combinar las siguientes fuentes de datos para obtener una tabla con las siguientes columnas:

- Nombre de producto
- Stock anterior
- Tipo de operación
- Variación

Sets de Datos

- **stock.csv**: último relevamiento [stock.csv](#)
- **movimientos.csv**: movimientos del último período [movimientos.csv](#)
- **productos** (Google Sheets): listado de productos [productos](#)

¿Por qué importa esto en DataWise Consulting?

La combinación de datos es crucial en DataWise Consulting. Al integrar diferentes fuentes de información, Silvia y el equipo pueden obtener una visión completa del inventario y operación de productos. Esto no solo mejora la calidad de los reportes, sino que también permite una mejor toma de decisiones acerca de la gestión de stock, lo que puede tener un gran impacto en la satisfacción del cliente y la eficiencia operativa.

Actividad 2: Preparación de Datos y Preamálisis

Contexto



Tras la exitosa combinación de datos, ahora es el momento de darle un vistazo más profundo a la información. Esta vez, Sabrina, la Consultora Estratégica, te guía en la preparación de los datos y te ayudará a identificar posibles acciones a tomar respecto al stock. La capacidad de analizar datos para prever necesidades es esencial para el éxito en la gestión de inventarios.



Objetivos

El objetivo de esta actividad es crear una tabla que muestre el stock actualizado de los productos y analizar los datos para determinar si es necesario reponer stock desde otro depósito. Desarrollarás habilidades de análisis crítico y toma de decisiones.

Ejercicio Práctico

A partir de los datos obtenidos en la Actividad 1, deberás crear una tabla con las siguientes columnas:

- Nombre de producto
- Stock actualizado

Te informan que un valor de stock negativo implica que se solicitó a otro depósito cumplir con la demanda. Analizá los datos obtenidos y respondé a la siguiente pregunta: ¿Debemos reponer el stock de alguno de los productos que haya sido solicitado a otro depósito?

Sets de Datos

La tabla combinada generada en la Actividad 1.

¿Por qué importa esto en DataWise Consulting?

La preparación de datos y el análisis preliminar son pasos esenciales en DataWise Consulting. Al entender adecuadamente la situación del stock, los miembros del equipo, como Sabrina, pueden anticipar necesidades e implementar acciones que mejoren la eficiencia operativa. Esto no solo ayuda a prevenir faltantes en el inventario, sino que también optimiza la cadena de suministro y mejora la relación con los clientes.

⚠️ **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