

«Talento Tech»

Business Intelligence

Clase 03



Clase N° 3 | Fuentes de datos - Obtención de datos - Buenas Prácticas

Temario:

- Fuentes de datos
 - Métodos de recopilación de datos.
 - Google Forms.
 - Datos sintéticos.
 - Buenas prácticas.
-

Objetivos de la clase:

El objetivo de esta clase es explorar las distintas fuentes y tipos de datos, comprendiendo sus características y aplicaciones. Aprenderemos a recopilar información mediante Google Forms y a vincularla con Google Sheets para su procesamiento. Además, abordaremos la generación de datos sintéticos y se presentarán buenas prácticas en la gestión y manejo de datos, asegurando su calidad, integridad y uso responsable.

Fuentes de datos

En el ámbito de la consultoría en **Business Intelligence (BI)**, comprender de manera efectiva las fuentes de datos y los métodos de recopilación es fundamental para tomar decisiones informadas. En esta clase, se realizará un análisis detallado de las fuentes de datos, los métodos para recopilarlos, la herramienta Google Forms y el concepto de datos sintéticos.

Fuentes de Datos: Internas y Externas

Fuentes de Datos Internas

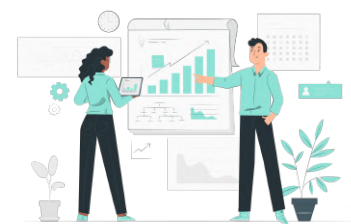
Las fuentes de datos internas provienen de los sistemas y procesos dentro de una organización. Estos datos son fundamentales para el análisis de Business Intelligence (BI), ya que **reflejan las operaciones, la productividad, y la interacción con los clientes**. Algunas de las principales fuentes de datos internas incluyen:



- **Sistemas de Gestión:** Los sistemas ERP (Enterprise Resource Planning) y CRM (Customer Relationship Management) recopilan datos sobre las operaciones del negocio, la gestión de relaciones con los clientes, ventas, inventarios y finanzas. Estos datos son esenciales para entender el rendimiento del negocio y para la planificación estratégica.
- **Bases de Datos Transaccionales:** A medida que las empresas llevan a cabo sus actividades diarias, generan grandes volúmenes de datos transaccionales. Esto incluye información sobre ventas, compras, procesamiento de órdenes y atención al cliente. La creación de informes a partir de esta base de datos permite identificar patrones y tendencias.
- **Análisis de Redes Sociales Internas:** Si la empresa utiliza plataformas internas para la colaboración entre empleados, como sistemas de gestión de proyectos, puede recopilar datos sobre la comunicación y colaboración entre los equipos. Esto puede proporcionar valiosos insights sobre la cultura organizacional y la dinámica del equipo.

Fuentes de Datos Externas

Las fuentes de datos externas son aquellas que provienen de fuera de la organización y pueden influir en sus operaciones y estrategia. Estas fuentes son cruciales para obtener una **visión más amplia del entorno de negocio y**



de las tendencias del mercado. Algunas categorías de fuentes externas son:

- **Datos del Mercado:** Estas fuentes incluyen informes de analistas, estudios de mercado y datos de la industria. Proporcionan información sobre las tendencias del sector, el comportamiento del consumidor y el posicionamiento de la competencia. Esto es útil para desarrollar estrategias de marketing y ventas.
- **Datos Gubernamentales:** Los organismos gubernamentales publican una gran cantidad de datos que pueden ser utilizados para análisis de BI. Estos datos pueden abarcar desde estadísticas demográficas hasta información económica, regulaciones y políticas que afectan a la industria.
- **Redes Sociales y Opiniones en Línea:** Las plataformas sociales ofrecen un vasto reservorio de datos sobre la percepción de los clientes y la reputación de la marca. Analizar estos datos puede ayudar a las empresas a comprender mejor las necesidades y preferencias de sus clientes.
- **Datos de Terceros y APIs:** Muchas empresas utilizan servicios de datos proporcionados por terceros que ofrecen información relevante para su sector. Esto puede incluir datos sobre comportamiento del consumidor, tendencias de búsqueda y análisis de sentimiento.

La combinación de datos internos y externos permite a las organizaciones obtener una visión más completa y precisa del mercado y sus operaciones.

Métodos de recopilación

A. Fuentes Primarias de Datos

Las fuentes primarias se refieren a los **datos que se recogen directamente de primera mano**, lo que significa que el analista es responsable de la recopilación. Estas fuentes son valiosas porque ofrecen información fresca y específica, adaptada a las necesidades del experto en análisis. A continuación, se detallan algunas de las fuentes primarias más comunes.

		Aplicación	Ejemplo
Fuentes primarias de datos	Encuestas	Utilizadas para estudios de mercado, análisis de satisfacción del cliente, o investigaciones sociales.	Una empresa que quiere evaluar la satisfacción del cliente puede enviar una encuesta a sus consumidores para entender mejor sus experiencias y áreas de mejora.
	Entrevistas	Se utilizan para explorar en profundidad	Un investigador que estudia el impacto de un nuevo producto en los usuarios

		percepciones, experiencias y motivaciones individuales.	podría entrevistar a personas para obtener información detallada sobre sus experiencias.
	Experimentos	Se utilizan para determinar relaciones causales y efectos de variables independientes sobre dependientes.	Una empresa de alimentos puede realizar un experimento A/B para ver cómo diferentes empaques afectan las decisiones de compra de los consumidores.
	Observación	Utilizada en estudios de campo, ciencias del comportamiento y sociología.	Un analista de comercio minorista puede observar cómo los consumidores interactúan con diferentes disposiciones de productos en una tienda.

B. Fuentes Secundarias de Datos

Las fuentes secundarias consisten en **datos que han sido recopilados, procesados y analizados por otros**. Este tipo de datos puede ser útil para complementar la información obtenida de fuentes primarias. Aquí se describen algunas fuentes secundarias relevantes:

Fuentes secundarias de datos	Bases de datos públicas	Se utilizan para investigaciones demográficas, análisis económicos y sociales.	El Instituto Nacional de Estadística y Censos proporciona datos sobre la población, ingresos, y educación que pueden ser utilizados para análisis de tendencias.
	Informes de investigación	Utilizados para presentar revisiones de literatura, análisis de mercado y estudios de caso.	Un estudio de mercado publicado por una firma consultora sobre tendencias de consumo puede ser valioso para una empresa que busca nuevas oportunidades de negocio.
	Medios de comunicación	Ayudan a realizar análisis cualitativos y a identificar tendencias emergentes en diversas áreas.	Un analista puede revisar artículos sobre las tendencias tecnológicas actuales para evaluar oportunidades en el sector de la tecnología.
	Redes sociales	Análisis de marca, estudios de percepción pública y segmentación de mercado.	Una empresa puede analizar los comentarios de los clientes en plataformas de redes sociales para entender la percepción de su marca y los problemas que enfrentan los consumidores.

Tanto las fuentes primarias como las secundarias tienen su lugar en la práctica de la analítica de datos. Se recomienda estar familiarizado con diferentes tipos de fuentes, que se comprendan las ventajas y desventajas de cada una, y se aprenda a seleccionar las que mejor se adapten a los objetivos del análisis.

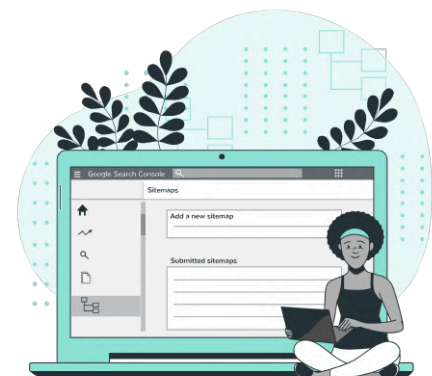
Google Forms: Herramienta de Recopilación de Datos



Una de las herramientas más útiles para la recopilación de datos en línea es **Google Forms**. Esta plataforma permite crear encuestas y formularios de manera simple y efectiva. Su interfaz intuitiva facilita la creación de formularios personalizados, sin la necesidad de contar con habilidades técnicas avanzadas. Entre sus características más destacadas se encuentra la variedad de tipos de preguntas que se pueden incluir, ya sea opción múltiple, texto o casillas de verificación, lo que permite adaptarse mejor a las necesidades del recopilador.

Además, Google Forms ofrece la ventaja de la recopilación automática de datos. Todos **los datos ingresados por los encuestados se almacenan en tiempo real en una hoja de cálculo de Google Sheets**, lo que facilita su análisis posterior. La accesibilidad de los formularios es también notable; se pueden compartir fácilmente mediante enlaces, correos electrónicos o redes sociales, lo cual amplía el alcance de las encuestas.

Los beneficios de Google Forms son significativos. Como es una herramienta gratuita, está al alcance de cualquier persona con una cuenta de Google. **La plataforma permite ver los datos recopilados en tiempo real, generando gráficos y análisis básicos integrados que facilitan la interpretación de los resultados.** También ofrece integraciones con otras herramientas de Google y aplicaciones de terceros, lo que mejora la recopilación y análisis de datos.



Datos Sintéticos: Una Alternativa en la Analítica

El concepto de datos sintéticos se refiere a aquellos datos que son generados artificialmente y que imitan las características estadísticas de datos reales, pero sin corresponder a individuos específicos. Este tipo de datos ha tomado relevancia en el ámbito de BI y el análisis de datos por diversas razones.

Una de las principales ventajas de usar datos sintéticos es la privacidad y seguridad que ofrecen. Permiten realizar análisis y pruebas de modelos sin comprometer la información personal de los individuos, lo cual es vital para cumplir con regulaciones como el GDPR. Además, también son útiles en situaciones en las que los datos reales son escasos o difíciles de obtener, brindando así una alternativa valiosa.

Los datos sintéticos se utilizan en el desarrollo de software y en la formación de modelos de inteligencia artificial, permitiendo probar algoritmos sin los riesgos asociados al uso de datos reales. Esto resulta especialmente beneficioso en entornos donde la protección de datos sensibles es una prioridad.

Obtención de datos en Google Sheets

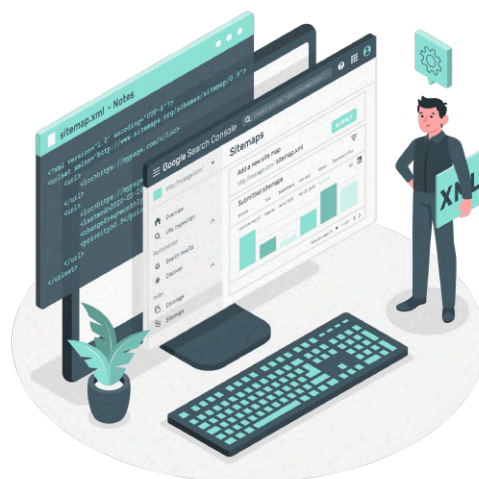
Google Sheets ofrece una variedad de métodos y funciones para obtener datos. Aquí se enumeran algunos de los más comunes, algunos de los cuales ya hemos visto en clases anteriores:

- **Importar datos externos:** desde un archivo CSV u otros formatos de planillas convertidos a formato de planillas de cálculo Google.
- **Vínculo con Google Forms.**
- **IMPORTRANGE:** esta función permite importar un rango de celdas de otra hoja de cálculo de Google Sheets. Se utiliza para combinar datos de diferentes hojas en una sola. La sintaxis es `IMPORTRANGE("URL de la hoja", "rango")`, donde la URL es la dirección de la hoja de cálculo fuente y el rango indica qué celdas se desean importar.
- **IMPORTXML:** extrae datos de una página web utilizando consultas XPath. La sintaxis es `IMPORTXML("URL", "consulta_xpath")`. Esto es útil para obtener información estructurada de sitios web, como tablas o listas.
- **QUERY:** realiza consultas SQL-like sobre un conjunto de datos. Esto permite realizar operaciones complejas como agrupaciones, filtros y ordenamientos de manera eficiente.
- **BUSCARX / BUSCARV / BUSCARH:** estas funciones buscan un valor y devuelven otro valor en la misma fila desde otra columna especificada. Es una función esencial para buscar datos relacionados.

Buenas prácticas de trabajo

A continuación se enumeran algunas buenas prácticas para trabajar con Google Sheets, teniendo en cuenta los entornos colaborativos:

- No modificar las fuentes de datos originales. **Siempre se debe preservar una versión intacta.**
- **Evitar romper el vínculo dinámico:** No se deben descargar las hojas de cálculo para trabajarlas en forma local, copiar y pegar desde otras hojas o cualquier otro método de obtención de datos no dinámico. Estas prácticas producen errores, pérdida de información y no permiten las actualizaciones en tiempo real.
- **No agregar columnas intermedias:** Si es necesario agregar columnas, **debe hacerse al final, luego de la última columna.** Esto evita que se generen errores en otras planillas asociadas o cualquier aplicación que utilice a la hoja como fuente de datos.
- **Organización y nomenclatura clara:** Utilizar nombres descriptivos para las hojas y archivos, así como un formato de nomenclatura consistente. Esto facilita la identificación y el acceso a los documentos por parte de todos los colaboradores.
- **Uso de comentarios y notas:** Añadir comentarios o notas en celdas específicas para proporcionar contexto o explicar el propósito de ciertas fórmulas o datos. Esto ayuda a los colaboradores a entender el trabajo y reduce malentendidos.
- **Control de versiones:** Aprovechar la función de historial de versiones de Google Sheets para rastrear cambios y recuperar versiones anteriores si es necesario. Esto es útil para identificar quién hizo cambios y cuándo.
- **Evitar modificaciones simultáneas:** Si se requiere un trabajo más minucioso, deben establecerse momentos específicos para que cada colaborador trabaje en secciones determinadas, evitando así conflictos y sobrescrituras de datos.
- **Uso de funciones y fórmulas consistentes:** Aplicar fórmulas estándar y funciones conocidas para facilitar la comprensión del análisis. Evitar el uso de fórmulas complejas que pueden confundir a otros colaboradores.
- **Utilizar funciones y fórmulas apropiadas para el problema:** Familiarizarse con la sintaxis y aplicaciones de las funciones, para aplicarlas de la manera más eficaz y eficiente, evitando el uso de fórmulas genéricas generadas por IA.



- **Registro de cambios significativos:** Llevar un resumen de cambios importantes en una hoja específica para que todos los colaboradores estén al tanto de las modificaciones críticas que puedan afectar el análisis.
- **Uso de gráficos y visualizaciones:** Incorporar gráficos y visualizaciones para ayudar a comunicar los hallazgos de manera más efectiva, facilitando la interpretación de los datos por parte de todos los involucrados.
- **Validación de datos:** Implementar validaciones de datos para asegurarse que ingresen valores apropiados y consistentes, evitando errores que pueden distorsionar el análisis.
- **Protección de celdas y hojas:** Si es necesario, utilizar la protección de celdas y hojas para restringir el acceso a ciertas partes del documento. Esto es útil para evitar que los colaboradores modifiquen datos sensibles o estructurales.
- **Seguridad y permisos:** Configurar permisos de acceso según el rol de cada colaborador. Determinar quién puede editar, comentar o solo ver el documento, garantizando la seguridad de los datos.

Reflexión final

Al combinar fuentes internas, externas, primarias y secundarias, los analistas pueden obtener una **visión más completa y precisa** del fenómeno en estudio, lo que permite tomar **decisiones informadas** basadas en datos concretos y relevantes.

Siguiendo las buenas prácticas, se puede mejorar la colaboración y la eficiencia en el análisis de datos en Google Sheets.



Material de referencia

- [Lista de funciones de Google Sheets](#)

Próximos pasos

- Definición de calidad de datos
- Técnicas para asegurar la calidad de los datos. Concepto de normalización de tablas.
- Prácticas de seguridad de datos y acceso controlado.

Ejercicios prácticos



DATAWISE CONSULTING
INSIGHTS ESTRATÉGICOS

Actividad 1: Fuentes de Datos

Contexto



Estás en las oficinas de DataWise Consulting, donde estás cursando tu pasantía. Silvia, la Project Manager y Especialista en Datos, te asignó una tarea esencial. La satisfacción del cliente es clave para el crecimiento de la empresa, y necesitas identificar fuentes que puedan proporcionar datos valiosos sobre este aspecto.

Objetivos

El objetivo de esta actividad es buscar y recopilar tres fuentes públicas de datos que te permitan comprender mejor la satisfacción del cliente. A través de esta tarea, aprenderás a evaluar la relevancia de diferentes fuentes y cómo estructurar información para un análisis efectivo.

Ejercicio Práctico

Utilizá herramientas de búsqueda en línea para identificar tres fuentes públicas que proporcionen datos sobre la satisfacción del cliente. Pueden ser encuestas de satisfacción, estudios de mercado o cualquier otra fuente relevante. Una vez que las hayas encontrado, creá una hoja de Google Sheets donde resumas cada fuente incluyendo enlaces, tipos de datos disponibles y el potencial que tienen para el análisis posterior.

Sets de Datos

- Fuentes públicas relacionadas con la satisfacción del cliente.
- Links a encuestas de satisfacción y estudios de mercado relevantes.

¿Por qué importa esto en DataWise Consulting?

DataWise Consulting se fundamenta en datos precisos y relevantes para ofrecer soluciones efectivas a sus clientes. Al identificar y analizar estas fuentes, contribuís

a crear una base sólida para futuras estrategias y decisiones, lo que a su vez impacta directamente en la satisfacción del cliente y el éxito de la empresa.

Actividad 2: Importación de Datos

Contexto



Después de completar la búsqueda de fuentes, Luis, el Analista de BI, te invita a profundizar en la manipulación de datos. La habilidad para importar datos es crucial en el análisis y en la presentación de informes que faciliten la toma de decisiones.

Objetivos

En esta actividad, aprenderás a importar datos desde una fuente externa, en este caso, Wikipedia. Se busca fomentar tu capacidad para gestionar datos y hacer uso de herramientas que transformen información en insights.

Ejercicio Práctico

Accedé a Wikipedia y buscá la tabla titulada "Comparación de medias comunes de valores [1, 2, 2, 3, 4, 7, 9]". Una vez que la localices, importá esta tabla directamente a Google Sheets. Asegurate de que todos los datos se transcriban correctamente y descubrí qué análisis inicial podés realizar a partir de esta información.

Sets de Datos

- [Comparación de medias comunes de valores \[1, 2, 2, 3, 4, 7, 9\]](#).

Comparación de medias comunes de valores [1, 2, 2, 3, 4, 7, 9]

Tipo	Descripción	Ejemplo	Resultado
Rango medio	Punto medio entre el mínimo y el máximo de un conjunto de datos	1, 2, 2, 3, 4, 7, 9	5
Media aritmética	Suma de los valores de un conjunto de datos dividida por el número de valores: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	$(1 + 2 + 2 + 3 + 4 + 7 + 9) / 7$	4
Mediana	Valor intermedio que separa la mitad de los valores mayores y la mitad de los valores menores de un conjunto de datos.	1, 2, 2, 3, 4, 7, 9	3
Rango	Diferencia entre el valor máximo y el mínimo	1, 2, 2, 3, 4, 7, 9	8
Moda	Valor más frecuente en un conjunto de datos	1, 2, 2, 3, 4, 7, 9	2

¿Por qué importa esto en DataWise Consulting?

La capacidad de importar datos eficientemente permite a DataWise Consulting operar con agilidad y exactitud. Cada dato es una pieza del rompecabezas que ayuda a entender mejor las tendencias y comportamientos de los clientes, facilitando la formulación de estrategias más efectivas y basadas en evidencia.

— 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 Habilidades para el Futuro

BA Buenos
Aires
Ciudad