

1.- Primeros pasos con el análisis de datos.

Definición de análisis de datos y analista de datos.

Los datos son información objetiva (como indicadores o cifras estadísticas) utilizada como base de un argumento, un debate o un cálculo. Otra definición: Información en formato digital que puede transmitirse o procesarse.

El análisis es un examen detallado de algo complejo, para conocer su naturaleza o determinar sus elementos fundamentales. Analizar será estudiar o determinar la naturaleza y la relación de las partes (de algo) mediante análisis.

Un analista de datos es alguien que estudiará o determinará la naturaleza y las relaciones de información objetiva utilizada como base del razonamiento, un debate o un cálculo llevando a cabo un examen detallado de información en formato digital para conocer su naturaleza o determinar su esencia.

Descubrir si eres un analista de datos.

Las personas recopilan y utilizan datos en su vida diaria, incluso sin darse cuenta. Utilizan dispositivos para rastrear su actividad física y patrones alimenticios, estableciendo objetivos basados en datos. Los analistas de datos van más allá, realizando tareas complejas. En un estudio, se encontró que muchos trabajadores de datos no tenían "analista de datos" en sus títulos, pero estaban involucrados en tareas y procesos relacionados con datos. Los datos son esenciales para informar a líderes y departamentos. Los analistas de datos limpian, transforman y moldean datos, usando herramientas como funciones VLOOKUP y creando visualizaciones. Su objetivo es responder preguntas y tomar decisiones informadas a partir de los datos.

El rol de los datos en la organización.

El análisis de datos en las organizaciones implica funciones organizativas diversas y adaptables según el tamaño, alcance y necesidades de la organización. Los roles y la estrategia varían, destacando cuatro áreas clave: Investigación, Gobernanza de datos, Tecnología y Datos.

La investigación define las preguntas a responder y establece parámetros para los datos. Implica especialistas que se enfocan en áreas clave del negocio. Dedicar tiempo a métodos de análisis y resultados estadísticos.

La gobernanza de datos establece claridad en la propiedad y el acceso a los datos. Debe abarcar todos los departamentos y tiene una conexión esencial con el análisis de datos.

La tecnología engloba diversas áreas operativas, no solo la producción. Controla la seguridad de los datos y el acceso. Juega un papel crucial en la gobernanza de datos.

Los roles de datos están presentes en distintos niveles de la organización, pueden incluir científicos de datos, ingenieros de datos y analistas. En organizaciones sin departamento de datos dedicado, a menudo son ubicados en departamentos tecnológicos.

El análisis de datos contribuye a mejorar productos, servicios y resultados finales, y nos da la posibilidad de abordar problemas a nivel global.

Comprensión de los tipos de roles de trabajo de datos.

La tecnología abarca diversas habilidades especializadas por lo que no hay una tecnología general en lo que respecta a datos. Los equipos multidisciplinarios de investigación tienen roles específicos, como proporcionar conjuntos de datos estructurados y guiar con conocimiento en áreas como economía, salud pública y educación.

Los arquitectos de datos definen cómo se gestionan y utilizan los datos. Los científicos de datos realizan análisis y estadística con los datos. Los ingenieros de datos transforman y administran datos, y los administradores, gestionan las bases de datos.

Los usuarios comerciales también utilizan herramientas de TI para profundizar en la inteligencia empresarial, y hay roles de datos en los departamentos comerciales y administrativos, como analistas de sistemas y analistas de negocio, alineados con TI y comercio respectivamente. Los roles de analista de inteligencia empresarial se enfocan en visualizar y mejorar datos para el negocio.

Descubrir las habilidades del analista de datos.

Al igual que un carpintero emplea varias herramientas según la necesidad, los profesionales de datos utilizan diversas habilidades para abordar diferentes escenarios. La experiencia acumulada se une a las herramientas disponibles para obtener resultados efectivos.

El análisis de datos se define como una actividad común a todos, y se exploran roles organizativos y laborales en este contexto. En cuanto a las habilidades del analista de datos, abarcan tanto competencias técnicas como habilidades interpersonales. Los conceptos fundamentales en análisis de datos son independientes de las herramientas, lo que permite aplicarlos universalmente.

Una de las habilidades clave es la capacidad de formular preguntas fundamentales y acceder a los datos necesarios para responderlas. La calidad de los datos se valora no solo por los números, sino también por cómo se obtuvieron y se procesaron. Además, es importante identificar qué datos son relevantes para responder preguntas específicas y cómo crear datos válidos a través de cálculos cuando no están directamente disponibles.

La presentación de los resultados es otro aspecto crucial; la visualización eficaz es esencial para transmitir respuestas de manera clara. El aprendizaje continuo es vital en un campo en constante evolución, donde la construcción de una base sólida de habilidades permite adaptarse a nuevas herramientas y cambios.

2.- Fundamentos de la comprensión de datos.

Aprender a identificar los datos.

Es importante mirar con detenimiento los datos y no quedarse solo en conceptos superficiales. Como analista de datos, es esencial considerar detalles como el proceso de fabricación, costes, envíos, registros y patrones. Mirar con atención los datos puede proporcionar información valiosa que en un principio pudimos pasar por alto.

Aprender sobre campos y tipos de datos.

La estructura de los datos está compuesta por: nombre de campo, tipo de datos y valor. Cuando hablamos de datos, normalmente nos referimos a los valores y su manipulación involucra introducir, exportar y reportar estos valores. Los tipos de datos más comunes son texto, número y fecha. También hay que destacar la diferencia entre el valor mostrado y cómo se almacena este valor. Por ejemplo, para excel 20/11/2025 y 45,981 son equivalentes.

Tratar con datos que no tienes.

No todos los datos necesarios están almacenados en nuestros sistemas, ya que las empresas en constante cambio pueden requerir información adicional que no se recabó originalmente. Por esto es posible que falten algunos de los datos necesarios para la toma de decisiones o para elaborar informes. Una solución sería el uso de funciones, para obtener los datos de manera eficiente.

Es importante comprender y utilizar funciones para obtener datos que no estén presentes. Por ejemplo la función lógica IF, que permite tomar decisiones basadas en condiciones.

No hay un conjunto de fórmulas y funciones estándar, ya que las necesidades varían según los datos y los requisitos. El analista debe aprender las sintaxis y saber cuándo aplicar cada función para obtener los datos faltantes y satisfacer las necesidades específicas.

Aprender la sintaxis.

La sintaxis es el lenguaje de programación de un programa. Permite interactuar con programas para ejecutar comandos y funciones. Esta puede variar entre los diferentes programas, pero aprender uno facilita la comprensión de los demás. Si bien algunos programas cuentan con una interfaz gráfica, conocer la sintaxis da más control sobre el mismo. Aprender la sintaxis es clave para ampliar las habilidades en programación y análisis de datos.

Aprender sentencias SQL básicas.

SQL es el lenguaje estructurado de consulta para manipular bases de datos. Las herramientas de datos usan sintaxis SQL, bien escribiendo las consultas o usando una interfaz gráfica. Las sentencias básicas en SQL son:

SELECT: Utilizada para recuperar datos de una base de datos. Permite especificar qué columnas deseas seleccionar y de qué tabla o tablas deseas obtener los datos.

FROM: Indica la tabla o tablas desde las cuales se obtendrán los datos en la consulta SELECT.

WHERE: Permite aplicar condiciones de filtro a los datos seleccionados. Solo se seleccionarán los registros que cumplan con la condición especificada.

ORDER BY: Utilizada para ordenar los resultados de la consulta en función de una o más columnas. Puedes ordenar en orden ascendente (ASC) o descendente (DESC).

INSERT INTO: Utilizada para agregar nuevos registros a una tabla.

UPDATE: Permite modificar los registros existentes en una tabla.

DELETE FROM: Se utiliza para eliminar registros de una tabla que cumplan con ciertas condiciones.

3.- Elementos clave a tener en cuenta al iniciar el análisis de datos.

Aprende a interpretar los datos existentes.

Las empresas emplean informes para la toma de decisiones, ya estén actualizados o no. Comprender los datos en los informes actuales y las reglas comerciales es esencial. Esto se asemeja a leer un menú de un restaurante: contiene categorías como entrantes, bebidas, platos principales y postres. Los platos se crean siguiendo recetas con ingredientes, medidas y reglas de preparación específicas. En el análisis de datos, el informe es el menú, los ingredientes son los datos y las preparación son las reglas comerciales.

Al evaluar un informe, es crucial identificar los puntos de datos recopilados y seguir reglas específicas, como mostrar solo los datos de los últimos 30 días. La analogía de

olvidar una lista en el supermercado enfatiza la importancia de la organización en el análisis de datos. Los “pasillos” que recorre un chef en un supermercado se relacionan con las tablas de datos y bases de datos, que equivalen a las “tiendas”. Al comenzar un proyecto, analizar los informes utilizados, categorías, puntos de datos y reglas comerciales es fundamental para crear una lista de datos necesarios.

Encontrar los datos existentes.

Las personas que toman decisiones requieren un acceso más amplio y eficiente a los datos. Los informes y tablas dinámicas se originan en archivos, sistemas o bases de datos, siendo este el punto de partida. Muchos informes están basados en tablas dinámicas o contienen gráficos generados manualmente por los usuarios. Hacer preguntas desempeña un papel vital en este proceso. Preguntar sobre la ubicación, obtención y exportación de los datos es crucial. Una vez se tienen estas respuestas se pueden realizar cuestionamientos adicionales, como la posibilidad de exportar datos o acceder a las tablas originales. Saber a quién preguntar también es importante. Sin embargo, el acceso puede estar limitado por reglas organizativas y políticas de IT.

En un escenario positivo, podrías configurar la conexión y obtener los permisos necesarios. No obstante, si eres nuevo o los datos son confidenciales, es posible que el acceso sea limitado según las normas de la organización. La extracción de datos inicia si se tiene acceso, permitiendo generar los informes necesarios.

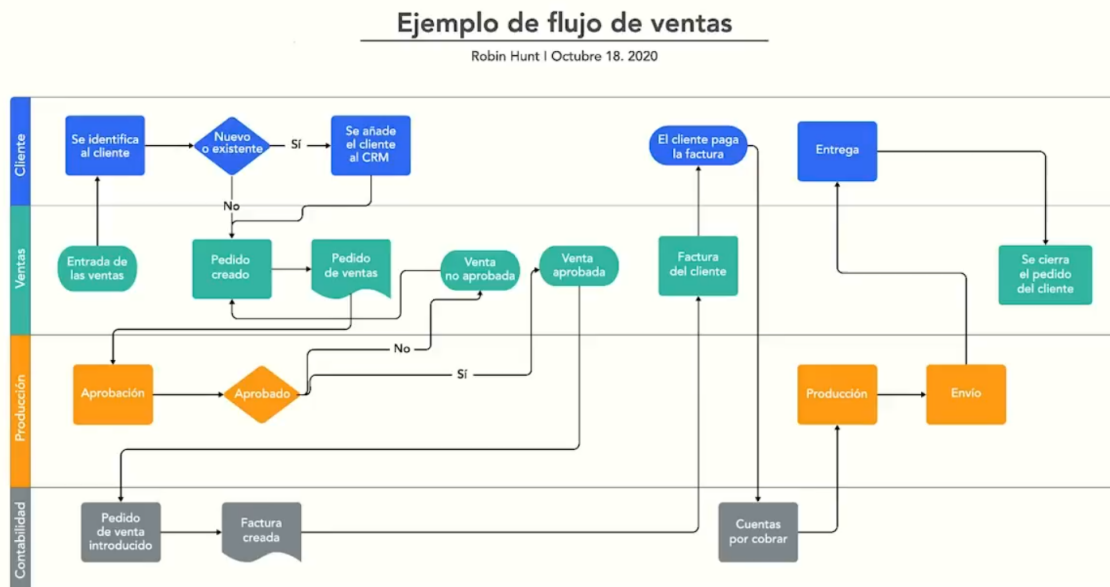
Limpieza de datos.

Cuando manejas datos para tus informes, parte del proceso es explorar y limpiarlos. La limpieza de datos es el proceso de estandarizar y hacer significativos los datos. En excel, es común limpiar los datos al filtrar columnas innecesarias o eliminar espacios adicionales con la función “Espacios”. También puedes abreviar y formatear valores en mayúscula. Puede ser necesario dividir información en columnas separadas, como las direcciones. Eliminar filas inválidas es esencial, por ejemplo, excluyendo ventas incompletas de un informe de ventas completadas. Eliminar datos duplicados es otra transformación clave, creando un conjunto apropiado para tu negocio. Estas acciones definen la limpieza de datos, buscando un conjunto de alta calidad que pueda ser confiable. Tomar notas detalladas es esencial para recordar las acciones aplicadas.

Comprender los datos y el flujo de trabajo.

Cada organización es única, aunque compartan similitudes en muchos aspectos. El flujo de trabajo guía la actividad organizativa, y los datos se mueven según ese flujo. La presencia de manuales y documentación de políticas es valiosa para entender los procedimientos. Preguntar sobre la existencia de políticas, documentación y flujos de

trabajo ayuda a aprovechar la información disponible o crear nuevos diagramas de proceso. Un ejemplo es el proceso de ventas, que se descompone en varias etapas.



La entrada de un pedido es el inicio, seguido por la identificación del cliente y la adición del CRM. Se indaga sobre cómo se obtiene el pedido y cómo llega al sistema. El equipo de ventas ajusta la propuesta inicial hasta convertirla en un pedido de ventas. El proceso se examina detenidamente, identificando donde residen los datos utilizados por los vendedores y si se comparte sistema con la orden de venta. Luego, la orden se envía al equipo de producción, donde surgen más preguntas sobre la aprobación y como se envían los datos. La orden aprobada regresa a ventas y luego se envía a contabilidad para realizar la factura. Los datos de contabilidad son confidenciales y específicos. Después, la factura se envía al cliente, quien paga y lo notifica a contabilidad. El proceso culmina con la entrega al cliente, cerrando el ciclo.

Este análisis proporciona un flujo de trabajo y puntos clave para los datos. Puede usarse para identificar el tiempo entre etapas y comprender las interacciones entre sistemas. En última instancia, se busca entender el flujo de trabajo organizativo y cómo los datos fluyen a lo largo del proceso.

Comprender las uniones.

Uno de los retos más difíciles para los analistas novatos es combinar datos de múltiples tablas. No hay una solución única que funcione en todas las situaciones. Idealmente, trabajar con conjuntos de datos ya relacionados es preferible. Sin embargo, si estás tratando con datos de tablas distintas, es probable que necesites unirlos, a menudo utilizando consultas. Programas que permiten manejar varios conjuntos de datos ofrecen formas de hacerlo, como BUSCARV en Excel o consultas en bases de datos. Para

relacionar tablas, se requiere una columna común, a menudo un campo clave que identifica de manera única cada fila.

Existen varios tipos de uniones en consultas: internas, externas y cruzadas. El tipo de unión afecta el resultado final. A medida que adquieras experiencia en proyectos reales, te será más natural decidir qué unión aplicar según tus objetivos. Muchos emplean BUSCARV en Excel para unir datos basados en un campo clave.

Los tipos de uniones incluyen la cruzada, que vincula todos los registros entre dos tablas. La unión interna, o inner join, es predeterminada cuando dos tablas se unen por un campo común, mostrando solo las filas coincidentes según las columnas de unión. Las combinaciones externas izquierda y derecha devuelven todos los registros de una tabla y los registros coincidentes de la otra. La elección entre izquierda y derecha depende del contexto.

El mundo real ofrece experiencia en la utilización de uniones con datos recibidos. Cada programa proporciona métodos para unir datos y guiará en el proceso para comprender el resultado al combinar conjuntos de datos.

Trabajar con uniones y validación.

Trabajar con uniones y validación implica seleccionar columnas comunes entre tablas para unir las. Las relaciones deben tener el mismo tipo de datos. Las propiedades de la relación se pueden ajustar en la ventana de Propiedades o en el menú contextual de la consulta. Corregir errores, como relaciones incorrectas, es esencial para obtener resultados precisos.

La ausencia de relaciones entre tablas provoca una combinación cruzada, generando filas para cada combinación de registros. Verificar si hay relaciones y comprender el efecto en la cantidad de datos es fundamental para evitar resultados no deseados.

Al unir tablas, se deben comprender los tipos de uniones. La unión interna (INNER JOIN) muestra solo coincidencias. Las combinaciones externas izquierda y derecha priorizan una tabla sobre la otra, mostrando todos los registros de una tabla y los coincidentes de la otra. Elegir el tipo de unión correcto se basa en el objetivo del análisis.

Se demuestra cómo filtrar resultados usando criterios como valores nulos. La elección de las columnas y el tipo de unión dependen del informe que se desea generar. La habilidad para trabajar con uniones y relaciones entre tablas es esencial para los analistas de datos.

4.- Primeros pasos con un proyecto de datos.

Primeros pasos.

A la hora de trabajar con datos es importante evitar las conjeturas y admitir cuando no se sabe la respuesta, así como comprometerse a investigar, en lugar de inventarse una respuesta. Nunca debemos publicar nada sin la verificación precisa. Consultar con el equipo y realizar las revisiones por duplicado ayuda a garantizar la exactitud de la información.

También debemos preparar bien las reuniones. Es fundamental estar en la agenda, realizar intervenciones y tener respuestas claras, así como respetar el tiempo de los demás. Practicar la escucha activa y la toma de notas. La información de las reuniones es esencial y debe documentarse para futuras referencias. Hacer las preguntas pertinentes y bien preparadas proporcionando contexto sin ambigüedades, es importante para la comunicación efectiva.

La presentación de datos debe ser organizada y detallada. Proporcionar información relevante sobre el origen de los datos, reglas aplicadas y definiciones ayuda a comprender los datos en profundidad. En la exposición debemos destacar lo esencial y permitir tiempo para que las personas comprendan los datos presentados, repitiendo los puntos clave si fuera necesario. También debemos preguntar a otros profesionales sobre sus mejores prácticas para enriquecer la propia comprensión y enfoque en proyectos de datos.

Descubrir errores comunes de principiantes.

Es importante aprender sobre los datos antes de intentar impresionar desde el desconocimiento. Es recomendable invertir tiempo en comprender la estructura y el significado de los datos. Se aconseja prestar atención a los detalles como: la estructura de los datos, los encabezados de las columnas, los tipos de datos y los posibles valores. Así como detectar duplicados y eliminarlos, ya que esto puede afectar a la precisión del análisis.

Otro punto clave es la verificación de los cálculos antes de presentarlos, sin importar cuán simple pueda parecer la operación matemática. Se sugiere conocer la cantidad de filas en las tablas de datos, lo que es útil para evaluar la calidad de los resultados en procesos de combinación y filtrado. Sin olvidarnos de generar una documentación de las preguntas y respuestas para evitar olvidos y pérdidas de tiempo en futuras consultas.

También es importante mostrar los datos precisos y sin manipularlos para hacerlos parecer mejores de lo que son. Así como evitar errores lógicos que pueden dar lugar a resultados incorrectos. Se sugiere formular preguntas bien elaboradas y buscar documentación cuando sea necesario. Y aprovechar los informes ya existentes para ganar conocimientos y ahorrar tiempo.

Por último, hemos de destacar la importancia de compartir informes y conocimientos con otros, lo que puede hacer el trabajo más eficiente y fiable.

Extender los conjuntos de datos de excel.

A la hora de comprender los datos de un documento de excel es importante combinar el uso de habilidades técnicas con la exploración de datos. Al analizar el contenido de las columnas y su disposición, se deduce que exponen detalles de pedidos, productos vendidos y clientes compradores. Se nota que las columnas numéricas están alineadas a la derecha para resaltar su naturaleza numérica y relevancia.

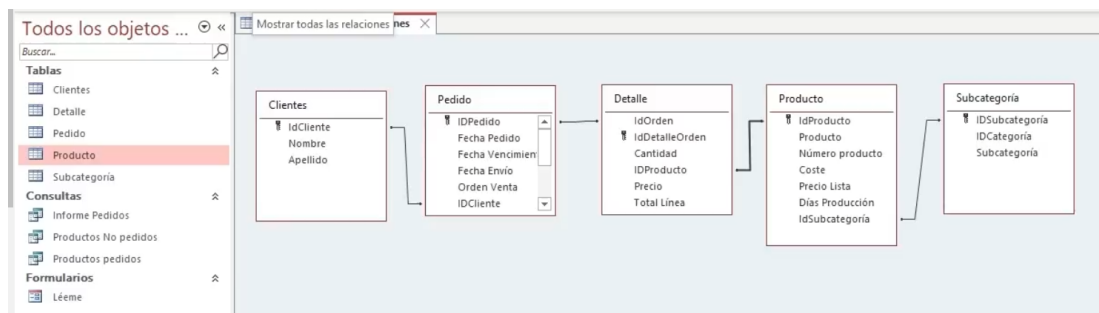
El análisis se extiende a las celdas de datos, realizando una exploración para identificar huecos o interrupciones en los rangos de datos. Mediante la inspección, se verifica la integridad de los datos y se ajusta el zoom para una revisión detallada.

Se examina una columna específica, para obtener su valor agregado, calculando promedio, recuento y suma. Estas medidas preliminares son útiles para comprender la distribución y tendencias de los datos. Se señala que es común que el identificador de orden se repita si hay múltiples productos en una orden, pero no debe ocurrir repetición en el identificador de la línea detalle.

Se explora la detección de duplicados en los identificadores de orden mediante formato condicional y reglas de resaltado. Hay que destacar que podemos conseguir información valiosa sobre los datos sin necesidad de escribir fórmulas, lo que subraya la utilidad de la exploración y análisis inicial para comprender la naturaleza y calidad de los conjuntos de datos.

Explorar la base de datos.

El video trata sobre la exploración de una base de datos usando access.



Se empieza revisando los nombres significativos de las tablas en pantalla, lo que facilita la comprensión de su contenido. Luego, se procede a explorar los datos en una tabla específica, identificando valores relevantes y notando filas con valores cero. Se observa que la columna "IdProducto" está relacionada con otras tablas y presenta también la columna "IdSubcategoría".

Se explica cómo verificar el tipo de datos de cada columna mediante la vista de diseño, donde se encuentran detalles sobre los campos y sus propiedades. El siguiente paso es explorar las relaciones entre las tablas. Esto se hace a través de la opción "Relaciones" en el grupo de herramientas de base de datos. Se resalta la importancia de las

relaciones y se muestran las claves primarias en las tablas, que son útiles para consultas de datos.

Se menciona que la información obtenida es valiosa, pero que no siempre es tan clara. Puede haber casos en los que los elementos no estén bien definidos, los campos tengan tipos incorrectos o falten relaciones. Para descubrir las dependencias entre los objetos, se utiliza la función "Dependencias del objeto" en el menú Herramientas de base de datos. Se muestra cómo ver los objetos que dependen de la tabla actual y los objetos de los que depende.

Se describen dos consultas específicas en la base de datos: "Informe Pedidos", que combina datos de pedidos, productos y clientes, y "Productos Pedidos", que muestra información de pedidos y datos base de productos, utilizando la columna "IdOrden" como unión. Se señala la importancia de entender la estructura de la base de datos, incluyendo tablas, columnas, relaciones, dependencias y consultas, como punto de partida fundamental.

Mantenimiento de los datos originales.

Es importante el mantenimiento de los datos originales en el proceso de trabajo con conjuntos de datos. Se destaca que en ocasiones, se reciben datos en formatos como hojas de cálculo o archivos CSV, y es crucial conservar los datos originales por varias razones. Mantener los datos originales permite reiniciar sin empezar de cero y proporciona una traza de auditoría. Una opción sería crear una carpeta "Datos originales", donde guardarlos y crear una estructura coherente de carpetas para los siguientes guardados. El siguiente punto de control se realizaría una vez hayamos limpiado nuestros datos, generando una copia de "Encuesta_Uso_Limpio". Mantener y organizar estas transformaciones ayuda a no tener que comenzar de cero en caso de necesitar cambios.

Comprender las verdades.

Hay tres verdades esenciales en el contexto de los datos: las verdades de las estadísticas, la verdad de los datos y la verdad empresarial.

La verdad de las estadísticas se enfoca en la importancia de los resultados estadísticos. En proyectos de ciencia de datos, se busca presentar resultados estadísticos relevantes y la importancia del informe. Se busca determinar si los resultados son genuinos o simplemente coincidencias.

La verdad de los datos puede revelar otras verdades. Aunque se hagan las transformaciones y cálculos correctamente, es posible presentar informes con datos incorrectos, lo que puede resultar frustrante. A veces, los datos incorrectos no son el resultado de un error en el análisis, sino de que los datos mismos podrían ser inexactos desde la fuente. La calidad de los datos y la forma en que se recopilan puede afectar a la verdad de los datos.

La verdad empresarial es la realidad del negocio. Para una empresa, su verdad empresarial es una medida de su producción y su desempeño. La verdad de los datos y la verdad de las estadísticas se entrelazan para exponer la verdad del negocio. Estas verdades colaboran para informar, crear procesos mejorados y establecer indicadores de rendimiento clave (KPI) que ayudan a evaluar el éxito.

El análisis y la interpretación de datos no solo se basan en los datos reales, sino también en comprender la importancia de los datos y cómo informan la realidad del negocio. Las tres verdades trabajan en conjunto para mejorar los resultados y tomar decisiones informadas en función de los datos y el contexto empresarial.

5.- Importación, exportación y conexiones de datos.

Aprender sobre la gobernanza de datos.

Aunque las organizaciones varíen, todas deben cumplir con regulaciones internas o externas relacionadas con los datos. El gobierno de datos va más allá de los datos en sí e involucra procesos, tecnologías y personas. Los objetivos son datos fiables, comprensibles, correctos y seguros,

La propiedad de los datos: Las diferentes áreas de una organización recopilan datos de diferentes maneras y los utilizan en distintas aplicaciones. La propiedad de los datos está ligada a quien los recopila y utiliza.

Acceso y permisos: El acceso a datos y los permisos están controlados por el plan de gobernanza de datos. En algunas organizaciones, solo el equipo de TI tiene acceso a los datos originales. Es importante tomar notas sobre el origen de los datos y cómo se obtuvo acceso.

Automatización y Protección: Se puede automatizar el acceso a los datos para protegerlos y asegurarse de que las personas sólo tengan acceso a lo que necesitan. A medida que una información se vuelve más relevante para la toma de decisiones, es probable que se requieran más acceso, herramientas y habilidades.

Adopción de la gobernanza de datos: Algunas organizaciones recién están adoptando la gobernanza de datos. La seguridad y protección se logran al solicitar acceso solo a los datos necesarios y limitando el acceso a la infraestructura completa.

Comprender el origen de los datos.

Es muy importante comprender y conocer el origen de los datos con los que se trabaja.

Es crucial conocer el origen de los datos y cómo se recibieron. Idealmente, se trabaja con bases de datos como punto de partida. Sin embargo, en la realidad, acceder a la fuente original puede ser un desafío debido a la gobernanza de datos y la estructura del sistema.

También se puede dar el caso en el que trabajemos con múltiples fuentes de datos. Los datos pueden ser recibidos de personas con acceso a la fuente o exportados de sistemas. El analista debe ser capaz de trabajar con diversas fuentes y conectarlas para un análisis más completo. Por ejemplo, una empresa con sistemas de recursos humanos y nóminas separados. Los datos se comparten entre estos sistemas, pero tienen diferentes propósitos y niveles de detalle.

Hay que tener en cuenta la restricción de acceso a ciertas fuentes de datos, debido a la sensibilidad de la información.

Saber de dónde provienen los datos es crucial para explicar su origen y resolver problemas en caso de errores. Los sistemas se actualizan y cambios en las fuentes pueden afectar a los datos. Tener en cuenta si los datos provienen directamente de la fuente o son transferidos hacia ti es fundamental.

Trabajar con archivos planos.

Si no puedes conectarte al origen de datos de un sistema, quizás recibas un archivo en formato excel o CSV. Son archivos planos y están desconectados de los datos, lo que significa que cuando se actualizan los datos en el sistema, no se actualizan a la vez en el archivo plano. Por lo que será necesario que te vuelvan a enviar el archivo actualizado.

Los archivos planos más comunes son los CSV, que son archivos con valores separados por comas. Los archivos CSV se abren con excel, por eso muchas personas los transforman directamente a excel. La recomendación aquí es utilizar un excel vacío que se conecte al archivo CSV a través de Power Query.

Power Query facilita la conexión y el procesamiento de archivos planos, permitiendo combinar datos de múltiples archivos, actualizar datos y crear informes dinámicos. Power Query simplifica la actualización de datos cuando se agregan archivos nuevos a la carpeta de archivos.

Trabajar con conexiones.

Aprovechar las herramientas para conexiones con datos externos como Excel y Power BI, nos aporta beneficios como mejoras en la productividad y mantener los datos actualizados. En Excel, las conexiones se encuentran en la pestaña Datos, Grupo, Obtener y Transformar Datos. En Power BI, estas conexiones se encuentran en la pestaña Inicio, Grupo, Datos, y se presentan diversos conectores integrados.

Al conectarse directamente a los datos externos, estás interactuando con ellos a través de tu herramienta. Si se trata de una base de datos, tus permisos en esa base de datos determinarán a qué datos puedes acceder. Las herramientas trabajarán de manera diferente en términos de cómo se manejan los datos actualizados en origen.

En caso de Power BI, se destaca que Power Query es la herramienta interna que se encarga de conectar, preparar y transformar los datos. Al conectarse directamente a los datos externos, se elimina un intermediario y se autorizan las acciones, lo que ayuda a prevenir problemas potenciales que pueden surgir en los procesos manuales.

Crear conjuntos de datos para otras personas.

A la hora de crear conjuntos de datos para compartir con otros cobrará aún más importancia el proceso de documentación y organización de los datos. Es importante proporcionar datos precisos y bien estructurados.

Documentar la información relevante sobre los archivos, como su origen y cualquier manipulación realizada. Esto ayudará a recordar detalles importantes y a comunicar información clara a los demás.

Se recomienda crear una pestaña especial en el archivo de trabajo destinada a documentar las columnas y los contenidos de los datos. Esto facilita la comprensión de los datos para cualquier persona que acceda al archivo.

Dado que no todos tienen el mismo nivel de acceso a los datos de origen, se aconseja proporcionar copias desconectadas a las personas que no tienen acceso directo. Esto asegura que todos puedan trabajar con los datos de manera eficiente. Aunque se comparta una copia desconectada, el archivo de trabajo original permanece conectado para permitir consultas y actualizaciones de las mismas.

También es importante planificar y dedicar tiempo a preparar la información detallada sobre el archivo. Esta práctica no solo es útil para ti, sino también un ahorro de tiempo para quienes trabajan con los datos.

6.- Primeros pasos para la limpieza y el modelado de datos.

Comprender el proceso ETL en los datos.

ETL (Extract, Transform, Load) es el proceso de obtener los datos de una fuente determinada, convertirlos en información con significado y prepararlos para que otras personas puedan usarlos. El ETL es esencial en la creación de almacenes de datos o Data Warehouses. Aunque existen diversos enfoques y términos relacionados con los datos, es importante centrarse en aprender a tratarlos de manera efectiva. La transformación de los datos (T), engloba actividades como limpieza, estandarización y enriquecimiento. Este

proceso es aplicable a cualquier tipo de dato y es esencial para generar informes significativos y útiles.

Limpieza de los datos usando macros de Excel.

La limpieza de datos mediante macros de Excel es una técnica para gestionar la herencia de archivos y estilos. Algunas organizaciones emplean macros para la limpieza, esta técnica puede resultar antigua pero es efectiva. La grabación de macros es una manera de automatizar pasos repetitivos. Para usarlas, se accede a la pestaña Programador y se graban acciones como duplicar hojas, eliminar columnas y filas o cambiar nombres de cabeceras. Las macros pueden ejecutarse en un solo paso y son útiles para simplificar tareas de limpieza repetitivas. Esta metodología es particularmente relevante cuando se heredan archivos de otras personas o se retoman trabajos anteriores, ya que permite entender y modificar las rutinas de limpieza existentes.

Limpieza de datos con Power Query.

La limpieza de datos utilizando Power Query es una metodología más eficiente que grabar macros o realizar cambios manuales repetitivos. Para llevar a cabo la limpieza y transformación de datos con Power Query, se comienza desde un archivo Excel vacío y se sigue la ruta Datos > Obtener y transformar datos > Obtener datos > De un archivo > De un libro Excel. Se conecta con la fuente de datos, en este caso, una tabla llamada Datos en la hoja Encuestas del libro Datos Encuestas.

Power Query permite realizar transformaciones en los datos de manera visual y organizada. Se pueden eliminar columnas innecesarias utilizando opciones como "Elegir columnas". También es posible ajustar los tipos de datos, como cambiar una columna de tipo Fecha/Hora a solo Fecha para reducir la carga de datos innecesarios.

Se pueden mejorar los encabezados de las columnas y eliminar redundancias en los datos. Esto incluye mejorar los nombres de las columnas y reemplazar valores nulos con texto descriptivo, como "No contesta". Es recomendable cambiar los nombres de los pasos realizados para facilitar la comprensión.

Además, Power Query permite combinar columnas utilizando diferentes separadores para crear columnas compuestas. Estas transformaciones se basan en los requisitos del negocio y ayudan a tener datos limpios y estructurados.

Aprender Power Query en Excel es beneficioso, ya que estas habilidades son transferibles a herramientas como Power BI. Power Query simplifica y agiliza el proceso de limpieza y transformación de datos, permitiendo cargar datos limpios y transformados para su análisis posterior.

Trabajar con datos reutilizables.

Trabajar con datos reutilizables implica crear conjuntos de datos que son necesarios de manera frecuente, como tablas de calendario. Estas tablas son comunes en proyectos y pueden ser creadas como plantillas para su reutilización. Para crear una tabla calendario en Excel, se utilizan fórmulas.

Se empieza escribiendo la fecha de inicio deseada y se arrastra para llenar la serie. Luego, se pueden usar diferentes funciones para calcular atributos de las fechas. Por ejemplo, la función "VALOR" convierte la fecha en un número de serie, "DIASEM" calcula el día de la semana, "TEXTO" forma el nombre del día, "MES" y "NOMPROPIO" calculan el número y nombre del mes, "AÑO" calcula el año y "FINDEMES" da el último día del mes.

Estas fórmulas se pueden aplicar a las celdas y luego rellenarlas para obtener los resultados deseados. El calendario resultante puede ser utilizado en herramientas como Power Query, Excel, Power BI u otras, según las necesidades del proyecto. En Power Query, se puede crear un calendario usando el lenguaje M, mientras que en Power BI u otros entornos de lenguaje tabular, se puede usar DAX.

Este enfoque de trabajar con datos reutilizables se puede aplicar a otros conjuntos de datos estáticos, como códigos postales o coordenadas geográficas. Es importante identificar elementos que se utilizan con frecuencia y crear procesos eficientes para mantener y utilizar estos conjuntos de datos.

Modelado de datos con consultas.

El modelado de datos mediante consultas es esencial para enriquecer conjuntos de datos y prepararlos para el análisis. En el contexto de análisis por fechas, es importante tener un calendario que contenga días con columnas descriptivas como día de la semana, mes y año. Esta tabla calendario puede ser creada en Excel u otras herramientas.

Si se trabaja con bases de datos, las consultas son fundamentales para modelar los datos. En el ejemplo, se utiliza Microsoft Access y se muestra cómo importar datos y realizar consultas para enriquecer el modelo. Se conecta una tabla calendario con los datos de encuestas utilizando la fecha como clave de relación. Las consultas se crean para cumplir con requisitos específicos:

Enriquecimiento por día de la semana: Se crea una consulta que relaciona la tabla calendario con las encuestas por la fecha. Esto permite enriquecer las encuestas con atributos como el nombre del día de la semana. La consulta resultante puede ser utilizada para análisis y creación de informes.

Análisis de fechas faltantes: Se crea una consulta que selecciona un período específico del calendario y lo cruza con las encuestas. Esto permite identificar días en los que no hubo encuestas, enriqueciendo los datos con información sobre la ausencia de respuestas.

Estas técnicas de modelado de datos son poderosas para enriquecer conjuntos de datos, generar atributos adicionales y realizar análisis más profundos. Las consultas y uniones son herramientas valiosas para cumplir con los requisitos del proyecto y mejorar la calidad del análisis de datos.

Modelado de datos en Power Query.

El modelado de datos en Power Query es esencial para unir y transformar conjuntos de datos de manera efectiva. En el contexto de la combinación de consultas, se utilizan dos fuentes de datos: una que contiene información de encuestas y otra que contiene datos de poblaciones, con un enfoque en la columna de Código Postal. En el editor de consultas de Power Query en Excel, se realizan las siguientes acciones:

Asegurarse de mantener nombres de columna coherentes: Aunque se pueda mover una columna, es una mala práctica que afecta al rendimiento de la consulta, por lo que se debe evitar.

Asegurarse de la consistencia del tipo de datos: Al combinar consultas, es esencial asegurarse de que las columnas que se van a unir tengan el mismo tipo de datos. Esto es especialmente crucial en columnas clave, como los códigos postales.

Combinar consultas: Se utiliza la función "Combinar consultas" para crear una nueva consulta que combine los datos de las encuestas y los datos de poblaciones. Se eligen las tablas, las columnas relevantes y el tipo de combinación (en este caso, Externa izquierda) para obtener los resultados deseados.

Expandir columnas: Después de la combinación, se expanden las columnas relevantes desde la consulta combinada. Esto permite obtener los datos de poblaciones relacionados con los códigos postales en las encuestas.

Realizar tareas de limpieza y transformación: Con las columnas combinadas y expandidas, se pueden realizar tareas de limpieza, transformación y filtrado según las necesidades del análisis.

El proceso de modelado de datos en Power Query permite crear conjuntos de datos enriquecidos y transformados, que son luego cargados en Excel u otras herramientas para análisis adicionales. Es fundamental seguir buenas prácticas para asegurar la coherencia de los tipos de datos y obtener resultados precisos y eficientes.

7.- Aplicación de técnicas comunes para todos los analistas de datos.

Convertir datos en Power Query.

Power Query es una herramienta poderosa para limpiar y transformar datos en Excel. Cambió la forma en que se trabaja con datos, simplificando la limpieza y conversión de datos. A continuación se presentan los pasos clave para convertir datos utilizando Power Query:

1.- Conexión a los datos: Desde Excel, acceder a Power Query a través del menú "Datos" > "Obtener y transformar datos" y elegir "De una tabla o rango". Seleccionar el rango de datos y confirmar si hay encabezados.

2.- Ventana de Power Query: Al conectarse a los datos, se abre la ventana de Power Query en Excel. Los dos primeros pasos se crean automáticamente: "Origen", que establece la conexión, y "Tipo cambiado", que detecta el tipo de datos de las columnas basándose en las primeras filas.

3.- Cambio de tipos de datos: Si es necesario ajustar los tipos de datos, se puede hacer seleccionando una columna y utilizando el menú para cambiar el tipo de datos. En algunos casos, esto puede reemplazar el paso existente, mientras que en otros casos, se agrega un nuevo paso.

4.- Ajuste de formatos: Power Query permite ajustar los formatos de las columnas, como cambiar números decimales a formatos monetarios. Esto se realiza desde el menú contextual de la columna y seleccionando "Cambiar tipo".

5.- Explorar y limpiar datos: Power Query ofrece numerosas opciones para limpiar y enriquecer datos. Es una herramienta poderosa para aquellos que exploran sus capacidades y la utilizan para mejorar la calidad de los datos.

6.- Cerrar y cargar: Una vez que se han realizado las transformaciones necesarias en Power Query, se puede regresar a Excel seleccionando "Cerrar y cargar" desde el menú "Inicio". Los datos limpios y transformados se cargarán en una nueva hoja de Excel.

Power Query simplifica la tarea de limpiar y transformar datos, ahorrando tiempo y mejorando la calidad de los análisis. Es una herramienta valiosa para aquellos que trabajan con datos en Excel y buscan optimizar su flujo de trabajo.

Encontrar y eliminar duplicados.

Para encontrar y eliminar duplicados en tus datos utilizando Power Query, sigue estos pasos:

1.- Identificar duplicados: Cuando trabajas con datos externos, es importante explorarlos en busca de posibles errores, como duplicados. Puedes usar el formato condicional en Excel para resaltar celdas con valores duplicados. Desde "Inicio" > "Estilos" >

"Formato condicional" > "Reglas para resaltar celdas" > "Valores duplicados", podrás ver las celdas duplicadas resaltadas en rojo.

2.- Editar la consulta: Si descubres que tienes duplicados, puedes utilizar Power Query para limpiarlos. Abre la ventana de Power Query haciendo clic en "Datos" > "Obtener y transformar datos" > "Desde una tabla o rango". Una vez en Power Query, selecciona la opción de "Editar".

3.- Duplicar la consulta: Antes de eliminar duplicados, duplica la consulta original para preservar los datos. Para ello, elige "Duplicar" desde el menú de la consulta. Esto te dará una copia idéntica de la consulta original.

4.- Crear una consulta de cabecera: En la consulta duplicada, elimina las columnas que corresponden a los detalles repetitivos de los pedidos. Esta consulta se convertirá en tu "cabecera", ya que contendrá una única fila para cada pedido.

5.- Eliminar duplicados: En la consulta de cabecera, busca el botón con un ícono de tabla a la izquierda de la primera columna. Despliega las opciones y selecciona "Quitar duplicados". Esto eliminará las filas duplicadas basadas en las columnas que elijas.

6.- Cerrar y cargar: Una vez que hayas eliminado los duplicados y preparado tus consultas, selecciona "Cerrar y cargar" para cargar los resultados en hojas de Excel.

Al realizar estos pasos, habrás eliminado los registros duplicados y organizado tus datos en consultas separadas para cabecera y detalle. Esto mejora la calidad de los datos y te permite trabajar con información más precisa y coherente.

Cambiar mayúsculas y minúsculas y reemplazar valores.

Para cambiar mayúsculas y minúsculas en los datos y reemplazar valores utilizando Power Query, sigue estos pasos:

1.- Cambiar mayúsculas y minúsculas: Abre Power Query seleccionando "Datos" > "Obtener y transformar datos" > "Desde una tabla o rango". En la ventana de Power Query, selecciona la columna que deseas modificar (por ejemplo, "Nombre"). Haz clic con el botón derecho en la columna seleccionada y elige "Transformar". Dentro del menú de transformación, encontrarás opciones para cambiar mayúsculas y minúsculas. Puedes seleccionar "Mayúsculas" o "Minúsculas" según tus necesidades.

2.- Modificar varias columnas: Si deseas aplicar la misma transformación a varias columnas, puedes hacerlo en un solo paso. Mantén presionada la tecla "Control" en tu teclado y selecciona las columnas que deseas transformar. Luego, haz clic con el botón derecho en una de las columnas seleccionadas y elige "Transformar". Aplica la transformación deseada, como "Mayúsculas" o "Minúsculas".

3.- Reemplazar valores: Si deseas reemplazar valores en una columna, selecciona la columna relevante (por ejemplo, "IdPedido"). Haz clic con el botón derecho en la columna

y elige "Transformar". Utiliza la opción "Reemplazar los valores" para buscar y reemplazar valores específicos en la columna. Por ejemplo, reemplaza "NP" por "Pedido".

4.- Ver cambios y aplicar: Después de realizar las transformaciones, puedes cerrar y cargar los resultados en la hoja de Excel. Selecciona "Inicio" > "Cerrar y cargar" para aplicar las transformaciones a los datos en Excel.

Recuerda que trabajar con Power Query te permite realizar cambios en los datos de manera eficiente y consistente, evitando la degradación manual de la calidad de los datos. Además, si necesitas modificar o mejorar los cambios realizados, puedes volver a la ventana de Power Query y ajustar las transformaciones en cualquier momento.

Combinar los datos con columnas combinadas.

Para combinar datos de diferentes columnas en Power Query utilizando columnas combinadas, sigue estos pasos:

1.- Eliminar espacios en blanco: Abre Power Query seleccionando "Datos" > "Obtener y transformar datos" > "Desde una tabla o rango". Si tienes espacios en blanco entre las columnas, puedes combinarlas para eliminar esos espacios y mejorar la legibilidad de los datos.

2.- Combinar columnas: Selecciona la primera columna que deseas combinar (por ejemplo, "Nombre"). Mantén presionada la tecla "Control" en tu teclado y selecciona la segunda columna (por ejemplo, "Apellido"). Haz clic con el botón derecho en las columnas seleccionadas y elige "Transformar". En el menú de transformación, elige "Combinar columnas". Selecciona un separador, como "Espacio", para separar los valores combinados (en este caso, el nombre y el apellido). Cambia el nombre de la columna resultante (por ejemplo, "Nombres y Apellidos"). Acepta la transformación.

3.- Ordenar datos combinados: Si deseas ordenar los datos combinados, inserta un paso antes de la combinación. Haz clic en el paso que contiene las columnas originales ("Valor reemplazado" en este caso). Selecciona la columna por la cual deseas ordenar primero (por ejemplo, "Apellido") y ordena los datos. Inserta el paso de ordenamiento. Repite el proceso para la otra columna (por ejemplo, "Nombre") y también ordena los datos. Inserta el paso de ordenamiento.

4.- Columnas combinadas y ordenadas: Regresa al paso de "Columnas combinadas" que creaste previamente. Como los datos ya están ordenados, la columna combinada mostrará los datos según el orden establecido en los pasos anteriores.

Recuerda que trabajar con columnas combinadas puede facilitar la visualización y el análisis de los datos al eliminar espacios en blanco innecesarios. Además, puedes aprovechar las capacidades de Power Query para ordenar los datos según tus requisitos específicos.

Creación de funciones lógicas.

Para crear funciones lógicas y columnas condicionales en Power Query, sigue estos pasos:

1.- Crear una Referencia (Opcional, para mantener limpieza en los datos): Abre la consulta original en la que deseas trabajar (por ejemplo, la consulta "Cabecera"). Haz clic derecho en la consulta en el panel de Consultas y Conexiones. Selecciona "Crear referencia". Esto crea una copia de la consulta sin afectar los datos originales.

2.- Crear Columna Condicional: Abre la consulta en la que deseas agregar la columna condicional (por ejemplo, la consulta "Marketing"). Haz clic en "Agregar columna" en la barra superior de Power Query. Selecciona "Columna condicional" en el menú desplegable.

3.- Definir la Columna Condicional: En el cuadro de diálogo que aparece, proporciona un nombre para la nueva columna (por ejemplo, "Nota de agradecimiento"). Define la condición utilizando las opciones provistas: Selecciona la columna por la cual deseas establecer la condición (por ejemplo, "Total"). Elige el operador (por ejemplo, "Mayor o igual"). Ingresa el valor (por ejemplo, "100"). Luego, define los valores que tomará la columna condicional en función de la condición: "Resultado sí verdadero" (por ejemplo, "Con nota"). "Resultado si falso" (por ejemplo, "Sin nota"). Acepta la configuración.

4.- Ajustar el Tipo de Dato: Es posible que la columna condicional tenga un tipo de dato predeterminado. Asegúrate de ajustar el tipo de dato según tus necesidades. En este caso, podría ser "Texto".

5.- Cerrar y Cargar: Una vez que hayas creado la columna condicional y ajustado el tipo de dato, cierra el editor de Power Query. Power Query aplicará los cambios y te llevará de vuelta a Excel.

6.- Ver los Resultados en Excel: En la hoja de Excel, verás que se ha creado una nueva hoja con los resultados de la columna condicional (por ejemplo, "Notas para Marketing"). Esta hoja contendrá los valores "Con nota" o "Sin nota" según la condición que estableciste.

Crear columnas condicionales es una manera poderosa de enriquecer tus datos y agregar información útil basada en condiciones lógicas.

Crear un conjunto de datos agregados.

Para crear conjuntos de datos agregados en Power Query, sigue estos pasos:

1.- Crear una Referencia (Opcional, para mantener limpieza en los datos): Abre la consulta original en la que deseas trabajar (por ejemplo, la consulta "Cabecera"). Haz clic derecho en la consulta en el panel de Consultas y Conexiones. Selecciona "Crear referencia". Esto crea una copia de la consulta sin afectar los datos originales.

2.- Crear Conjunto de Datos Agregados: Abre la consulta en la que deseas crear los datos agregados (por ejemplo, la consulta "Total Pedidos"). Haz clic en "Agrupar" en la barra superior de Power Query.

3.- Configurar la Agrupación: En el cuadro de diálogo que aparece, selecciona la columna por la cual deseas agrupar los datos (por ejemplo, "IDCliente"). Si deseas contar las filas, selecciona "Recuento de fila" como una de las operaciones de agregación.

4.- Agregaciones Avanzadas: Si necesitas realizar agregaciones más complejas, como sumar o promediar una columna específica, selecciona "Uso avanzado". Aquí podrás agregar más operaciones y definir las columnas sobre las cuales se aplicarán (por ejemplo, "Suma" para la columna "Total" y "Promedio" para la misma columna).

5.- Filtrar los Resultados: Si necesitas filtrar los resultados según ciertos criterios, puedes aplicar filtros. Por ejemplo, para mostrar solo los clientes con más de una compra, puedes filtrar la columna "Recuento de fila" para valores mayores a 1.

6.- Redondear Valores (Opcional): Si los valores agregados tienen muchos decimales y deseas redondearlos, puedes aplicar esta transformación. Selecciona "Transformar" > "Columnas de número" > "Redondear" y elige el número de posiciones decimales.

7.- Cerrar y Cargar: Una vez que hayas realizado todas las transformaciones y agregaciones necesarias, cierra el editor de Power Query. Power Query aplicará los cambios y te llevará de vuelta a Excel.

8.- Ver los Resultados en Excel: En la hoja de Excel, verás que se ha creado una nueva hoja con los resultados del conjunto de datos agregados (por ejemplo, "Total Pedidos"). Esta hoja contendrá los datos agrupados y agregados según las operaciones que hayas configurado.

Crear conjuntos de datos agregados te permite resumir y analizar tus datos de manera eficiente, proporcionando información útil de manera más concisa y organizada.