

5.- Transformación de datos:

Cómo transformar datos en excel con Power Query.

Queremos ver cuánto tiempo tardan las transacciones desde la fecha de transacción hasta la fecha de finalización, para ver si algún proveedor tarda un poco más en una determinada transacción. Esperamos que la mayoría de nuestras transacciones estén por debajo de tres días.

Desde Datos - Consultas y conexiones - Seleccionamos la consulta de proveedores y le damos a Editar.

Primero necesitamos que todos los nombres de nuestros proveedores estén en mayúsculas. Para ello hacemos clic derecho en la columna - Transformar - Mayúsculas.

Necesitamos la fecha y el año de la transacción, para ello duplicamos la columna TransactionDate con clic derecho - Duplicar columna. En el duplicado cambiamos el nombre a TransactionYear y la transformamos con clic derecho - Transformar - Año - Año.

En la hoja tenemos dos importes, el importe sin impuestos y el importe de los impuestos. Necesitaremos el importe total: Agregar columna - Columna personalizada y la nombramos como TotalAmount. Seleccionamos la columna AmountExcludingTaxes, signo (+) y TaxAmount. Aceptar. Y cambiamos el tipo de dato a Moneda. Eliminamos las dos columnas anteriores.

Ahora queremos ver el número de días que han transcurrido entre la fecha de la transacción y la fecha de finalización. Añadir columna personalizada y la nombramos como Days. Seleccionamos TransactionDate (-) FinalizationDate. Aceptar. Cambiamos el tipo a un número entero. Clic derecho - Transformar - Valor absoluto.

Ahora vamos a agregar una columna condicional que nos muestre un texto y nos indique si tarda más o menos de tres días. Vamos a Agregar columna condicional y la nombraremos como OverUnder. Seleccionamos Si (columna Days) es mayor o igual que 3 entonces 3 days or more. De lo contrario 2 days or less. Aceptar. Inicio - Cerrar y cargar.

Ahora que los datos están preparados, podemos responder algunas preguntas sobre los días de producción. Empezamos insertando una tabla dinámica, Insertar - Tabla dinámica. Arrastramos OverUnder a filas y SupplierTransactionID a valores. En valores abrimos el desplegable - Configuración de campo de valor... - Resumir el campo de valor por - Recuento. Podemos transformar estos valores a porcentajes con clic derecho - Mostrar valores como - % del total general.

Transformar datos puede abarcar muchas pequeñas técnicas diferentes aplicadas a los datos mientras trabajas para llegar a tu análisis.

Transformación de datos en SQL.

SQL es un lenguaje informático que trabaja con datos y las relaciones entre ellos. Microsoft SQL Server es un sistema de gestión de bases de datos relacionales, con la función de almacenar y recuperar datos.

Una consulta básica permite seleccionar datos de la base de datos. Hay dos instrucciones necesarias para un Select; debes saber qué quieres y de dónde lo quieres. Se trata de las instrucciones Select y From. Select listará todos los campos de la tabla y From listará el nombre de la tabla. Si quiero filtrar datos entonces usaré la sentencia Where y para ordenar los datos se usa Order by. Debes filtrar los datos antes de ordenarlos.

Usaremos SQL Server Express y Microsoft SQL Server Management Studio para trabajar con la base de datos Wide World Importers.

Con clic derecho en la tabla PurchasingSupplierTransactions - Seleccionar las 1000 primeras filas. Esto genera una sentencia SQL básica que selecciona todos los campos de Wide World Importers y aplica un filtro para los primeros 1000 registros. Podemos eliminar la sentencia del filtro y ejecutarla de nuevo lo que nos dará todos los registros.

Para añadir más significado a estos datos, necesitamos añadir otra tabla, lo que nos lleva a trabajar con combinaciones. Cuando tienes datos en varias tablas, puedes utilizar combinaciones para controlar qué datos se muestran y los resultados. Para ello seleccionamos la sentencia Select y con clic derecho accedemos al Diseñador de consultas del editor. Hacemos clic derecho - Agregar tabla y agregamos proveedores. Debido a que estas tablas tienen una relación establecida en el diseño de la base de datos, se unen automáticamente, están unidas por SupplierID, que está en ambas tablas. También podemos ver que es una relación uno a muchos, es decir tenemos un proveedor en la lista y puede estar unido a varias transacciones. Vamos a añadir SupplierName, para ello marcamos el checkbox y automáticamente colocará el campo en la parte final de la sentencia Select. Si queremos ponerlo al principio se puede arrastrar sin más. Clic en Aceptar y ejecutamos la sentencia. Una combinación interna funciona buscando en ambas tablas para encontrar una coincidencia. Esto significa que, en estas dos tablas, si tengo un nombre de proveedor y este proveedor tiene un registro de transacciones, aparecerá en los resultados.

El único problema con estos datos es que si queremos informar sobre los proveedores que tenemos en el sistema, independientemente de sus transacciones, tenemos que ajustar el tipo de combinación. Seleccionamos la sentencia y con clic derecho - Diseñar consulta en el editor. Hacemos clic derecho en el rombo e indicamos que muestre todas las filas de proveedores, lo que creará una combinación externa.

Debido a que los tipos de combinación tienen un impacto en los datos de nuestro conjunto de resultados, debes pensar de forma crítica qué quieres lograr con tus datos y saber que podrías necesitar ajustar el tipo de combinación.

Transformación de datos en Power Bi.

Power BI ofrece dos funciones básicas para el análisis de datos; transformar los datos y presentarlos.

Queremos analizar las ventas de nuestros productos para anotar nuestros diez mejores productos. La pantalla de apertura del escritorio de Power BI es una página en blanco lista para la visualización. Esto sucede una vez que te hayas conectado a los datos. Las tablas que necesitaremos para analizar los mejores productos son los detalles del pedido y los productos. Haremos las transformaciones en las tablas de productos y los pedidos en Power Query. Vamos a ver la función Agrupar para poder totalizar los pedidos, así como la función Consulta, donde combinaremos los pedidos y productos con la opción Combinar consultas.

Vamos a la función Transformar datos, para empezar a limpiar datos. Empezaremos con los productos. La única información que necesitamos de los productos es el ID y el nombre de los productos. Hacemos clic en ProductID y manteniendo pulsado Ctrl seleccionamos ProductName. Clic derecho - Eliminar otras columnas.

Ahora pasaremos a los detalles del pedido. Una de las primeras cosas que tenemos que hacer es crear la función que nos da el importe total después del descuento aplicado. Seleccionamos Agregar columna - Columna personalizada. La nombraremos como TotalOrderAmount. En esta sentencia vamos a crear el subtotal, calcular el descuento y luego deducirlos uno del otro. (Cantidad por precio unitario menos el importe del descuento) Cambiamos el tipo de dato a un número decimal fijo. A continuación, utilizamos la función combinar para crear la consulta que combina la tabla de productos con la tabla de detalles del pedido. Empezamos haciendo clic en los productos, y en la pestaña inicio seleccionamos Combinar consultas.

Se abrirá la pantalla Combinar y los productos. Queremos combinarlos con los detalles del pedido, y el campo común entre los dos es ProductID. Por lo tanto, tenemos que asegurarnos de que los hemos resaltado. Los tipos de combinación, al igual que cualquier otro conjunto de datos, están aquí en la combinación. En la parte inferior de la pantalla dice tipo de combinación. Tenemos 75 registros que coinciden con las 77 filas de la primera tabla. Eso quiere decir que tenemos dos productos sin registros, lo que significa que no se han pedido. Estamos viendo los mejores productos, así que es evidente que todos ellos se han pedido. Hacemos clic en el menú desplegable. Tenemos Externa izquierda, que mostraría todos los productos y si hay un pedido. Externa derecha, que mostraría todos los detalles del pedido, independientemente de la coincidencia con el producto. Externa completa, si tenemos productos y detalles de pedidos que no tienen registros que coincidan, mostraría todas las filas de ambos. Interna, que es la que necesitamos aquí: me mostraría solo los productos con los pedidos. También tenemos Anti izquierda y Anti derecha, que muestran solo los valores nulos. Aquí, si seleccionamos Anti izquierda, solo mostraría los dos productos que no tienen detalles de pedidos. Para el análisis, necesitamos Interna.

Ahora podemos ampliar la tabla, haciendo clic en las flechas al lado del campo desmarcamos todas las opciones menos TotalOrderAmount que será la columna que

necesitamos. Ya solo vemos las columnas ProductID, ProductName y TotalOrderAmount lo que nos permitirá crear una función para agrupar y totalizar cada producto. En la pestaña Transformar - Agrupar por. Seleccionamos la columna ProductName y nombramos la nueva columna como Total, la operación será Suma y la columna TotalOrderAmount. Aceptar. Así solo se nos mostrará el nombre del producto y cuanto se pidió. Seleccionamos la tabla Order_Details y con clic derecho des-seleccionamos Habilitar Carga. Vamos a Inicio - Cerrar y aplicar.

Ahora en la lista de campos vemos los productos. Lo visualizamos en una tabla, así que elegimos una tabla. Arrastramos ProductName y Total al campo de valores. Ajustamos el tamaño para verlo bien. Ahora mismo, esto representa todos los productos. Y estamos intentando obtener los 10 mejores. Vamos a ProductName en Filtros. Le decimos que haga un filtro avanzado Top N, donde top es 10. Y basamos eso en el total. Arrastramos Total a Por valor y aplicamos el filtro. Después de aplicar ese filtro, vemos los 10 mejores productos. Para ordenarlos hacemos clic en el encabezado Total.

Estas técnicas y combinaciones muestran exactamente lo poderoso que pueden ser Power BI y los datos cuando se establecen rutinas de limpieza para las presentaciones básicas de datos.

Tareas habituales de limpieza y transformación.

Al crear la caja de herramientas de limpieza y transformación, hay algunos elementos que utilizaremos habitualmente.

Los espacios son invisibles a la vista, pero en realidad son caracteres. Cuando un campo tiene espacios extra debemos limpiarlo eliminándolos. Hay espacios iniciales al principio del campo y finales al final. Si queremos eliminarlos podemos utilizar funciones como TRIM o CLEAN.

El acto de dividir un texto también se llama analizar el texto. Podemos hacerlo con cualquier delimitador. Los espacios también sirven como delimitadores; los espacios entre palabras son espacios válidos. Por ejemplo; nombre y apellido. Otros delimitadores pueden ser guiones o comas. Para trabajar con el análisis de texto, usamos cosas como texto en columnas, dividir por delimitadores y funciones como izquierda, derecha y centro. No solo dividimos el texto, hay veces que necesitamos combinar campos. Lo que se conoce como concatenar. También podemos sustituir el texto por otro válido. Por ejemplo; si alguien escribe la abreviatura de las provincias podemos sustituir este valor por el nombre completo. Hay varios métodos para sustituir datos no válidos por datos válidos. O cambiar a mayúsculas o minúsculas.

Puede que necesites eliminar duplicados o cambiar el tipo de datos para que se adapte a lo que necesitamos. Por ejemplo puedes tener datos de fecha guardados como texto. O en caso de querer operar con valores el valor del campo deberá ser de tipo numérico.

Uso de funciones integradas.

Uno de los principales beneficios de las herramientas de análisis de datos es que pueden realizar todo tipo de cálculos por nosotros. No tenemos que crear cada función que necesitemos de forma manual.

Por ejemplo en esta tabla de Power BI tenemos Quantity, UnitPrice y Discounts. Añadimos una tabla y expandimos Orders. Queremos incluir OrdersID y ProductName. Expandimos la tabla para visualizarla bien y añadimos Quantity y UnitPrice. Como no tiene sentido mostrar el sumatorio de estos campos los eliminamos haciendo clic derecho en Quantity y le decimos No resumir. Repetimos lo mismo con UnitPrice. Lo que sí nos interesa sería ver el subtotal. Para incorporarlo al modelo vamos a Transformar datos - Agregar columna personalizada. La nombramos como SubTotal y será la multiplicación de UnitPrice por Quantity. Cambiamos el tipo de dato a un decimal fijo. Ahora traemos Subtotal a la tabla y en este caso si queremos mostrar el sumatorio. Esta será la cantidad antes de aplicar el descuento.

A continuación vamos a añadir una matriz mirando los valores basados en Shipping Country. Añadimos Ship Country a las filas y Subtotal a los valores. Esto nos permitirá ver los países y resume automáticamente su subtotal. Si quisiéramos que fuera un promedio, podría hacer clic derecho y elegir Promedio. Si quisiera mostrar el máximo de cualquier subtotal en un país seleccionaremos Máximo.

Otra práctica funcionalidad de Power BI es la posibilidad de utilizar medidas rápidas. Hacemos clic en Medida rápida. Se trata de medidas escritas en DAX. Las medidas rápidas son funciones predefinidas escritas en el lenguaje de fórmula DAX (Data Analysis Expressions). Estas medidas ofrecen una forma rápida y conveniente de realizar cálculos comunes en Power BI sin la necesidad de escribir complejas fórmulas DAX desde cero. Algunos ejemplos de medidas rápidas incluyen promedio, suma, mínimo, máximo, varianza, desviación estándar, entre otras.

Cuando se utiliza una medida rápida, simplemente se selecciona el campo o columna objetivo y se elige la medida rápida deseada desde el menú desplegable. Esto permite obtener rápidamente resultados agregados sin tener que crear manualmente la fórmula DAX correspondiente. Además, Power BI también ofrece medidas rápidas específicas para la inteligencia de tiempo, lo que facilita el análisis y comparación de datos a lo largo del tiempo, como "Total anual hasta la fecha" o "Cambio de año a año".

Algunas consideraciones sobre el uso responsable de datos:

- Entender el propósito antes de realizar cualquier análisis.
- Seleccionar las funciones adecuadas. No todas las métricas son aplicables a todos los conjuntos de datos.
- Verificar y validar los datos, antes de utilizar los resultados del análisis.
- Ser claro y transparente sobre la metodología utilizada al presentar los resultados del análisis.

6.- Modelado de datos:

Bases de datos relacionales.

Las bases de datos relacionales son un tipo de sistema de gestión de bases de datos que organizan y almacenan datos en tablas con estructuras predefinidas y relaciones establecidas entre ellas. Estas relaciones se basan en claves primarias y claves foráneas, lo que permite establecer conexiones entre distintas tablas y garantizar la integridad y consistencia de los datos.

Modelado de datos con Power BI.

Trabajamos con datos en diferentes conjuntos de datos o tablas para hacer el análisis y la representación. Cuando trabajamos con varias tablas, debemos modelar nuestros datos para sacar el máximo partido. Cuando tenemos un diagrama de entidad relación, donde se muestran las tablas y las diferentes relaciones entre ellas, lo que tenemos en realidad es un modelo de datos.

Tenemos diferentes tablas con la información necesaria para nuestros informes. Los conjuntos de datos están destinados a estar unidos. Hay varias formas de unir los datos de un modelo de datos en Power BI. Cuando combinamos consultas, estamos estableciendo una combinación, pero también podemos ir a la sección de modelado y modelar estos datos desde el principio.. Esto permite que los datos se comuniquen a través de combinaciones. Es decir, si hacemos referencia a un pedido, sabemos qué productos y qué detalles estarán relacionados con él.

Power BI trata de combinar los datos automáticamente. Esto se denomina detección automática. Siempre debemos confirmar que las relaciones que se establecen sean correctas. Por ejemplo; tenemos las tablas Orders y OrderDetails, que están unidas, que están unidas por ProductID, con una relación uno a varios. Otros tipos de relaciones cuando nos referimos a cardinalidad serían: varios a uno, uno a uno y varios a varios. Uno a uno implica que en las dos tablas solo hay un registro vinculado a un registro. Uno a varios o varios a uno, significa que tenemos un registro en una tabla que está vinculado a muchos registros en otra tabla. Podemos gestionar todas las relaciones desde la pestaña Administrar relaciones.

Una vez tengamos modelados los datos podemos analizarlos utilizando funciones sin necesidad de crear varias consultas combinadas para obtener las conexiones.

Gestión de datos maestros.

Los datos maestros, como la información de dirección de los clientes, deben ser coherentes y consistentes en toda la organización. Diseñar reglas para los datos maestros ayuda a los analistas a aplicar transformaciones de manera consistente.

Algunas herramientas como Power Query y Power BI facilitan la corrección y estandarización de datos maestros. La gestión de datos maestros busca mantener una lista ordenada, completa y precisa de estos datos en una organización. Tener un plan de gestión de datos maestros es crucial independientemente del tamaño de nuestra organización.

Datos no estructurados.

En el mundo existen muchos más datos no estructurados que estructurados. Muchos de estos datos no estructurados son generados a partir de estructurados.

Los datos estructurados son aquellos que encajan perfectamente en tablas y hojas de cálculo y contrastan con los no estructurados que estarían conformados por fotos, vídeos, documentos PDF... Esta transición se ha intensificado con la proliferación de dispositivos móviles, cámaras y redes sociales.

Los datos no estructurados requieren de una interpretación visual para comprender su contenido. Los datos semiestructurados, por otra parte, representan una fusión entre ambos tipos, representando una mezcla de organización y caos. Un ejemplo ilustrativo es cuando recibes una imagen de un gato en la playa: mientras ves un gato y una playa, un experto en datos puede extraer información adicional, como la hora del día, el clima, la ubicación y más detalles.

Los datos semiestructurados ocupan un lugar intermedio, presentando características de ambos extremos. La interpretación y el análisis de estos datos dependen en gran medida de la naturaleza de la organización y sus objetivos. La comprensión de la existencia y variedad de estos datos puede ser beneficiosa para explorar nuevas funcionalidades y aplicaciones.

7.- Visualización de datos.

Métodos de visualización y mejores prácticas.

Se describe un caso en el que un experto diseña un panel magnífico, pero nadie lo utilizaba, y se plantea el porqué. Aquí cobra importancia la perspectiva del usuario a la hora de hacer el diseño ya que algunas personas pueden requerir más tiempo para comprender y utilizar una nueva herramienta. Se recomienda hacer una documentación y proporcionar formación sobre el material visual creado. Explicar las diferencias y puntos clave que ayuden al usuario a comprender y utilizar el panel de manera efectiva.

Se sugiere ser constante en el uso de colores para elementos similares y presentar los datos con simplicidad. El objetivo del panel será facilitar información valiosa y no tratar de impresionar a nadie con gráficos avanzados.

También se recalca la importancia de usar títulos, etiquetas y proporcionar información contextual destacada, para que los usuarios puedan entender rápidamente el contenido y contexto de la visualización.

Se menciona que las visualizaciones no deben limitarse a los gráficos, también es importante usar matrices, tablas y tablas dinámicas, y ofrecer varias opciones para que el usuario use la que prefiera. Recordemos que el contenido visual debe contribuir a la comprensión y toma de decisiones rápida.

Cómo crear informes para visualizar los datos en las páginas.

No todos los datos se consumen mejor en un panel. Algunos informes son más adecuados para unos u otros formatos debido a la complejidad de los datos. Power BI cuenta con una herramienta para generar informes paginados. Estos informes son útiles cuando se necesita presentar una gran cantidad de datos en varias páginas. Los informes paginados permiten conectarse a los datos y ofrecen una manera más detallada de presentar la información.

Es muy importante mostrar sólo la información relevante de manera práctica para las reuniones. Se destaca el uso de matrices que resulta mucho más útil que la cantidad de páginas original. Una matriz es una representación tabular de datos que permite organizar y mostrar información en filas y columnas de manera similar a las tablas dinámicas de excel. Estas generalmente se utilizan para presentar resúmenes de datos. Es especialmente útil cuando se necesitan generar agregados y totales en una estructura organizada.

Creación de un panel para la elaboración de informes.

Los paneles de control son herramientas valiosas para proporcionar información útil basada en selecciones de datos personalizadas. Estos paneles permiten mostrar indicadores clave de rendimiento, como el rendimiento en ventas por países, mediante las visualizaciones interactivas.

La interacción con las visualizaciones es esencial. Al hacer clic en elementos como países se pueden explorar detalles específicos, como los productos más vendidos en este lugar. Esto ayuda a identificar patrones y tendencias.

El diseño y formato adecuado son cruciales para la efectividad de los paneles. Ordenar y formatear campos, como los de ventas, mejora la presentación y comprensión de la información.

En la creación de paneles, es importante considerar las necesidades del usuario. Por ejemplo, se puede diseñar una página para los directores de ventas, donde pueden interactuar con diversos conjuntos de datos, como detalles de pedidos.

El uso de gráficos, como barras apiladas, permite visualizar relaciones complejas, por ejemplo cómo los vendedores interactúan con los diferentes clientes. Esto proporciona una visión clara de los volúmenes de ventas y patrones de comportamiento.

La inclusión de segmentadores y filtros, agiliza la búsqueda de datos específicos. Se pueden añadir segmentadores para las fechas, clientes y otros criterios, permitiendo una exploración más enfocada.

La capacidad de exportar datos filtrados a hojas de cálculo proporciona una herramienta útil para compartir información. Esto es especialmente beneficioso para los responsables de ventas, ya que pueden usar estos datos para comunicarse con clientes que no tienen acceso al panel.

La clave de los paneles efectivos es el diseño. Al ofrecer una experiencia interactiva con líneas de datos y múltiples páginas, se brinda la oportunidad de explorar diferentes perspectivas y crear situaciones diversas.

Recopilar requisitos para las visualizaciones.

Es muy importante trabajar junto al cliente desde el principio sobre cómo vamos a representar los datos en el panel. Se recomienda la creación de maquetas, que deben incluir elementos visuales como tablas, matrices y gráficos, además de filtros interactivos. El propósito es que todos los involucrados puedan tener una idea clara del diseño y llegar a un consenso sobre cómo se verá nuestro panel.

Hay que destacar los wireframes, que resultan muy útiles en la planificación de la navegación en paneles con múltiples páginas. Los wireframes son representaciones visuales simplificadas y esquemáticas de una interfaz de usuario.

A la hora de diseñar los requisitos y filtros es muy importante identificar las necesidades específicas del panel. Esto incluye determinar tanto los filtros estrictos a nivel de datos como los flexibles y dinámicos para los usuarios finales.

Por último, hay que destacar la importancia de considerar el público al cual están dirigidos los paneles, los permisos de acceso a los datos y las licencias requeridas para su uso.

Presentar los desafíos de datos de manera eficaz.

Aunque la explicación pueda parecer sencilla para el hablante, es posible que nuestros interlocutores no comprendan lo que estamos diciendo. Las habilidades comunicacionales son de vital importancia para un especialista en datos. Siempre resultará útil consultar con la dirección del proyecto para que se entienda la labor del equipo de datos en el proyecto.

A veces al iniciar proyectos puede que aunque tengamos acceso total a los datos y un método de trabajo, nos enfrentemos a problemas iniciales con la consistencia de los datos. En estos casos es importante informar a la persona adecuada sobre los datos faltantes y también destacar lo que se ve en los datos existentes.

Cuando nos enfrentamos al desafío de descubrir que la información va en contra de las expectativas de la organización es importante confirmar la precisión de los datos, hablar con los líderes de la empresa y si es necesario, recurrir a la dirección.

La entrega de resultados y la comunicación de los datos a la empresa puede ser complicada. Por ello se recomienda revisar los resultados, seguir la cadena de mando y mantener la comunicación con el equipo.

Finalizar los paneles.

Uno de los aspectos clave de las funcionalidades de visualización de datos son la generación automática de títulos y sugerencias. No obstante, algunas de estas características pueden resultar confusas si los usuarios no tienen un conocimiento profundo sobre los datos subyacentes. Es importante considerar el cambio de títulos como una parte esencial del proceso de diseño, recomendando hacerlo al final de la configuración del panel. Los títulos deberán ser descriptivos.

A continuación ajustaremos los elementos visuales como gestión de etiquetas de categorías, tamaños y formatos, para mejorar la presentación. Por ejemplo en los segmentadores, agregaremos la palabra clave "Seleccionar" para aclarar las opciones disponibles. También es importante tener en cuenta la uniformidad y precisión.

Nuestra labor en este aspecto consistirá en mejorar tanto la comprensión de la información presentada como la apariencia general de los elementos visuales utilizados.

Adición de filtros a los paneles.

Cuando trabajamos con filtros en un panel, es posible que el usuario quiera eliminar todos los filtros seleccionados, para que este muestre todos los datos. Para permitir a los usuarios volver a la vista original agregaremos un marcador al que llamaremos limpiar. Y para facilitar la navegación lo colocaremos en la interfaz añadiendo un "botón". Ya que no será un botón real, hemos de modificar los estilos para que el usuario lo perciba como tal, agregando un fondo y eliminando el icono de marcador.

Modificar los textos alternativos en los paneles.

En este punto se aborda la refinación y personalización de una visualización de datos centrada en una barra apilada, así como la explotación de la función de información

sobre herramientas integradas. Se enfatiza que, aunque la información sobre herramientas ofrece datos útiles, puede carecer de ciertos detalles deseados.

Los ajustes en la información sobre herramientas pueden aumentar significativamente la cantidad de datos presentados, simplemente modificando nombres y contenido. Power BI permite la modificación del fondo del panel, incluso ofreciendo opciones de diseño oscuro. Así como diseños adaptados a los smartphones. Aunque existen muchas otras opciones para personalizar el panel, los cambios efectuados hasta ese punto constituyen una base sólida para la optimización.

8.- Clasificación de roles en el campo del análisis de datos.

Trabajadores de datos.

Las principales funciones de un trabajador de datos son el uso de manipulación de datos en hojas de cálculo y la creación de información valiosa a través de presentaciones e informes. Los trabajadores de datos utilizan funciones y operaciones agregadas en sus datos, incluyendo funciones lógicas como IF.

Para el crecimiento profesional de un trabajador de datos se recomienda el uso y aprovechamiento de herramientas como Power Query, tablas dinámicas y visualizaciones en Excel e incluso el uso de Power Point para visualizar los datos en presentaciones. Por otro lado, no hay que olvidar algunas habilidades no técnicas, como presentación y comunicación efectiva.

Analistas de datos.

Los analistas de datos tienen un conocimiento más profundo de sistemas y diseños de bases de datos que los trabajadores de datos. Tienen acceso más amplio a tablas y bases de datos, junto con habilidades básicas en consultas SQL.

Un analista de datos comprende el plan de gobernanza de datos y trabaja bajo políticas establecidas. Son hábiles en la formulación de preguntas y en la investigación, además de saber cómo limpiar y transformar datos según los requisitos del proyecto. Pueden crear funciones, incluyendo sentencias condicionales y lógicas. Trabajan con estadísticas, desde conceptos básicos hasta agregados y conexiones de datos para mantener actualizadas visualizaciones e informes. Se recomienda el crecimiento en temas estadísticos, aunque no se requiere un nivel de estadístico completo.

Otras habilidades clave del analista de datos son la escucha activa, la narración de datos y el pensamiento crítico.

Ingenieros de datos.

Hay que entender la diferencia entre depurar y añadir a conjuntos de datos, y la habilidad de crearlos. Las competencias de analista e ingeniero de datos a menudo se solapan. Un ingeniero de datos comprende cómo optimizar conjuntos de datos, hacerlos más manejables y pueden recibir datos de otros para diseñar conjuntos a partir de consultas y proporcionarlos. Un ingeniero de datos suele tener más acceso a datos y entiende su seguridad y privacidad en el contexto de la gobernanza de datos. Puede avanzar hacia roles de arquitectura de datos, cubriendo más sistemas y estrategias de servidor y seguridad.

Para progresar en el rol de ingeniero de datos, es crucial conocer datos estructurados y no estructurados, convertirlos en conjuntos de datos utilizables y comprender las metodologías de diseño de sistemas de bases de datos relacionales. Se requieren habilidades de comunicación, presentación, pensamiento crítico y escucha activa. Estas habilidades son necesarias para transformar numerosas tablas en formatos útiles para procesos, usando ETL (extraer, transformar y cargar) o ELT (extraer, cargar y transformar), que es como los datos se mueven de sistemas de producción a almacenes de datos.

Científico de datos.

Los científicos de datos poseen habilidades de analista e ingeniero de datos, pero además requieren un conocimiento más profundo de programación, matemáticas y estadística. Un científico de datos es esencial para el desarrollo de herramientas y enfoques que generen información valiosa para las empresas. Sin embargo, un científico de datos no puede hacerlo solo y necesita colaborar con otros roles en el equipo. Los científicos de datos interpretan grandes conjuntos de datos, crean modelos de aprendizaje automático, presentan resultados, lideran equipos de datos y contribuyen a la estrategia de gobernanza de datos.

Para progresar en este ámbito, es importante comprender el pensamiento programático y aprender programación, posiblemente comenzando con lenguajes como Python. También es importante mejorar las habilidades en estadísticas, independientemente de la experiencia previa en este campo. Se destacan las diferencias clave entre un científico de datos y otros roles, principalmente en matemáticas, programación y estadísticas avanzadas.