

Explorar la visualización de datos

Las visualizaciones de datos proporcionan perspectivas prácticas de lo que pueden ser conjuntos complejos de datos. En este curso, va a conocer cómo comunicarse claramente con visualizaciones de datos y hacer llegar su idea rápidamente.

Tabla de contenido

1. [Explorar la visualización de datos](#)
2. [Introducción a la visualización de datos](#)
3. [Seleccionar software de visualización de datos](#)
4. [Comunicarse claramente con visualizaciones de datos](#)
5. [Decidir el tipo de visualización de datos que se usará](#)
6. [Diseñar mejores prácticas para visualizaciones de datos](#)
7. [Evitar visualizaciones de datos engañosas](#)

Explorar la visualización de datos

El mundo de los negocios está repleto de datos. Usarlos estratégicamente ofrece a una empresa ventajas competitivas importantes. Lo difícil está en dar sentido a los datos y comunicar lo que significan a los interesados de la empresa.

En este curso, conocerá las consideraciones para crear visualizaciones de datos y cómo transmitir su mensaje usando herramientas de visualización. También ganará perspectiva sobre cómo usar mejores prácticas clave de diseño para asegurarse de que sus visualizaciones sean claras y precisas al contar la historia que desea.

Introducción a la visualización de datos

Puede que su organización, como muchas otras, tenga un equipo de analistas que recopilen, gestionen y den sentido a los datos. Aunque hay algo más. Deben comunicarse de una forma significativa. No es fácil explicar datos y presentarlos en un formato que sea fácil de entender. Aquí es cuando entra en juego la visualización de datos. Considere el ejemplo de Boost Labs, una empresa de visualización de datos que pidió ayuda a un cliente para dar sentido a sus datos.

Como la mayoría de las organizaciones, el cliente tenía su propio equipo de investigadores competentes, por lo que estaban bastante seguros de sus conclusiones. No obstante, una vez que Boost Labs creó una ilustración de los datos recopilados, el cliente se dio cuenta de que sus supuestos en relación con lo que estaban transmitiendo realmente los datos eran incorrectos. Con los números plasmados en un papel, las relaciones entre los puntos de datos no estaban demasiado claras. Las herramientas de visualización de datos convierten datos en imágenes integrales, lo que permite a los responsables de la toma de decisiones clave comprender fácilmente conceptos complejos y analizar patrones de datos.

Suponga que a su equipo directivo le dicen que el Net Promoter Score de la organización ha caído algunos puntos en una región específica. Los datos aislados no muestran claramente el motivo. Con herramientas de visualización de datos, su equipo podría profundizar más para descubrir que la resolución de llamadas es un problema en esa región, lo que explica por qué no es probable que los clientes recomienden los productos o servicios de su organización. Esa es información sobre la que puede actuarse para mejorar la puntuación NPS.

Las visualizaciones de datos pueden adoptar muchas formas. Por ejemplo, la forma de mapa interactivo en tiempo real de varios sistemas de transporte para un área específica. Permite que los usuarios en tránsito filtren información para que solo se muestren datos relevantes para ellos. Un beneficio adicional de la visualización de datos es que las personas pueden comprenderlos con más rapidez que el texto. Considere un "infoscape" que usa toques de color para mostrar una gama de síntomas de salud y cómo se relacionan entre sí. Ahora compare eso con escudriñar puntuaciones de texto médico con el fin de encontrar correlaciones entre el mismo conjunto de síntomas de salud.

La diferencia resulta bastante evidente. De esta manera, el siguiente beneficio resulta igualmente evidente. Las organizaciones pueden usar la información derivada de la visualización de datos para tomar mejores decisiones de negocio. Considere las redes sociales, donde un gráfico de datos de redes sociales no estructurado podría representar redes de amigos e información de su perfil en redes sociales. Sin análisis de datos, esta información sería inútil. Aunque cuando se transforma en un gráfico de intereses, muestra redes de personas que tienen intereses similares en términos de marcas o productos. Usados juntos, estos gráficos tienen el poder de dirigir futuras tendencias de compra.

Los beneficios de la visualización de datos superan con creces cualquier riesgo. No obstante, para que cualquier organización coseche todas las recompensas de este método de presentación de datos, debe comprender las consideraciones a la hora de seleccionar software de visualización de datos. Una vez que tenga el software, hay mejores prácticas que deben seguirse para comunicarse claramente con visualizaciones de datos. Es importante saber cuándo usar el mejor tipo de visualización y cómo crear visualizaciones claras y precisas.

Seleccionar software de visualización de datos

En el pasado, las organizaciones usaban inteligencia de negocios tradicional, o BI, para mover datos estructurados de una fuente externa individual a un almacén de datos. No obstante, esto supone una presión considerable para los departamentos de TI. El sistema era inflexible, la creación de informes requería mucho tiempo y generalmente solo podían acceder a ellos miembros del personal de TI.

Con el crecimiento exponencial actual de los datos, los métodos tradicionales no funcionarán, especialmente para organizaciones que necesitan acceso a información en tiempo real. Aquí es donde entra en juego Inteligencia de negocios de autoservicio, o SSBI. SSBI faculta a usuarios autorizados, incluso sin minería de datos o sin formación en

análisis, a trabajar con datos. Esto evita departamentos de TI, ya que los usuarios pueden crear fácilmente sus propios informes y realizar consultas para tomar decisiones de negocio basadas en los datos.

Aunque con el tipo de flexibilidad que ofrece SSBI se tiene que aplicar alguna forma de gobernanza de datos para evitar confusión. Por ejemplo, políticas de datos deben definir métricas clave del éxito, describir procesos para crear y compartir informes, asignar privilegios para el acceso a datos y describir procesos de mantenimiento para fines de privacidad, seguridad y calidad. La mayoría de herramientas SSBI incluyen capacidades de visualización de datos.

No obstante, las organizaciones deben considerar detenidamente qué herramienta o combinación de herramientas se ajusta mejor a sus necesidades. Por ejemplo, algunas herramientas pueden diseccionar datos perfectamente, aunque no pueden proyectarlos. Esto podría requerir una combinación de herramientas para analizar y mostrar visualmente datos específicos. Existen algunas consideraciones clave para seleccionar herramientas de software de visualización de datos.

La primera consiste en los tipos de visualizaciones admitidas por varias herramientas, comparándolos con el tipo de datos con los que trabaja su organización. Aunque muchos se deciden por solo una forma de analizar datos, no subestime el poder de intentar nuevos métodos. Suscríbase a tres pruebas y experimente con nuevas herramientas de visualización para encontrar una que sea adecuada. A continuación, considere los formatos de datos permitidos por la herramienta. Las bases de datos SQL y NoSQL son solo dos de los formatos permitidos que deben incluirse. También podría haber formatos de datos para CRM, o gestión de relaciones con los clientes, aplicaciones, SAP Financials u Oracle, herramientas de ventas y cualquier otra plataforma empresarial que use la organización.

Hadoop, un marco de software libre para almacenar datos y ejecutar aplicaciones, resulta fundamental para organizaciones que se preparan para pasar a procesamiento de big data. Considere también el grado con el que la herramienta puede profundizar en los datos fuente. ¿La organización necesita diferentes herramientas y sintaxis lingüística para excavar datos más allá de la primera capa? ¿La herramienta puede manipular elementos visuales de datos en vivo? ¿La vista puede ajustarse de forma rápida y sencilla? Busque una característica de auditoría.

Esto es fundamental para que las organizaciones hagan el seguimiento de los responsables del análisis de la visualización de datos. Considere también capacidades de exportación. Debe haber múltiples formas de que una organización comparta sus visualizaciones de datos con los demás.

Las opciones no deben limitarse a formatos gráficos planos, como PDF y JPEG. También podría haber una forma de dejar caer fragmentos de código en páginas web e incorporarlos en varias aplicaciones a través de API para dispositivos móviles y equipos de escritorio. La consideración final consiste en capacidades avanzadas de procesamiento de la herramienta para organizaciones que recopilan big data. La herramienta debe ser capaz de manejar la carga de consultas. Ejecute el mayor número de consultas posible a través de la herramienta

durante su prueba gratuita con el fin de probar sus capacidades de rendimiento. La selección del software correcto de visualización de datos es clave a la hora de aprovechar al máximo las oportunidades de BI.

Comunicarse claramente con visualizaciones de datos

La comunicación con visualización de datos requiere el uso de buenos datos. Esto requiere comprender que los buenos datos equivalen a la selección y comparación de grupos. Esta es solo una mejor práctica para usar visualizaciones de datos de forma efectiva. Pongamos que un grupo de personas ve un anuncio; ¿cómo calificaría una organización la efectividad de ese anuncio?

Si el grupo estuviera formado por personas ya interesadas en el producto antes de ver el anuncio, estos serán malos datos ya que existe sesgo. No obstante, si se seleccionaran a las personas de forma aleatoria, los datos resultantes podrían considerarse buenos. A continuación, considere si el propósito de la visualización de datos es declarativo o exploratorio. Los datos declarativos afirman información, como por ejemplo el presupuesto de un departamento.

Los datos exploratorios confirman o revelan información. Por ejemplo, si una gerente de ventas simplemente presentara información de ventas trimestrales, esto sería una visualización declarativa. Si la visualización mostrara un retraso en las ventas y el consejo preguntara el porqué del retraso, se requerirá una exploración adicional de los datos. La gerente de ventas puede tener una idea de por qué existe un retraso, y puede usar la visualización exploratoria para confirmarla o refutarla. No obstante, si no tuviera idea de lo que está buscando, puede realizar minería de datos para revelar tendencias y patrones, aunque no siempre se sepa lo que se está buscando. En esos casos, la exploración visual es más útil.

Las visualizaciones de datos abiertas son beneficiosas porque incluyen y examinan más datos que una visualización basada en una hipótesis, y son más interactivas. Suponga que un estratega de marketing sospecha que hay un incremento en las compras del cliente mediante otros dispositivos inteligentes en ciertos momentos del día. Los programas de marketing existentes de la organización no están configurados para beneficiarse de esto, por lo que introduce datos relevantes en Datawrapper, una plataforma de software libre de visualización de datos para probar y confirmar su hipótesis visualmente. No puede confirmar nada usando un gráfico de barras.

No obstante, usando diferentes iteraciones y cambiando las variables y tipos de gráficos, puede confirmar sus observaciones. Para visualizaciones de datos diarias, unos sencillos gráficos de barras y líneas suelen ser suficientemente claros. Para mostrar una reducción en el crecimiento anual del gasto en atención médica, comparado con otras áreas, un sencillo gráfico de líneas transmite el mensaje con más claridad que un gráfico que detalle 20 años de datos históricos. Y considere siempre quién es la audiencia.

Ya sean responsables de la toma de decisiones, personal técnico o el público, cada una de estas audiencias puede requerir un formato visual diferente. Independientemente del formato, una relación de 80% de gráficos y de 20% de texto resulta ideal. Cuando haya que ilustrar muchos datos, use múltiples pequeños, el mismo gráfico básico para mostrar diferentes porciones de un conjunto de datos o añada interactividad.

Un gráfico con demasiadas mediciones resulta difícil de leer; en su lugar, desglose varios puntos en gráficos más concisos y pequeños. De forma similar, la inclusión de demasiados colores puede confundir a los observadores. Resaltar el rendimiento de alguien en comparación con el de otra persona es un ejemplo de interactividad de la visualización. Independientemente de dónde se encuentre el rendimiento de alguien en relación con el de los demás, los resultados resaltados destacan. Otro tipo de interactividad es el filtrado, que ayuda a trabajar con múltiples niveles de datos. Estas prácticas permiten aclaración y comprensión cuando se presentan datos a alguna audiencia.

Decidir el tipo de visualización de datos que se usará

El diseño de una visualización de datos requiere revisar los datos para determinar la forma más efectiva de transmitir la historia que cuentan. Existen muchos gráficos y diagramas disponibles, desde muy sencillos hasta complejos. El gráfico de líneas representa puntos de datos conectados por líneas rectas. Se usa para ilustrar cambios específicos como por ejemplo cifras de ventas con el tiempo. En el caso de muchas variables, puede utilizarse un gráfico de líneas apiladas. De forma similar, los gráficos de barras resultan útiles para hacer el seguimiento de cambios con el tiempo para realizar análisis comparativos entre un número limitado de variables.

La longitud de cada barra en el gráfico se dimensiona en proporción con la sección que representa. Y puede ser horizontal o vertical. Se recomienda mostrar datos en orden, usando barras verticales, como por ejemplo una progresión de años. Y barras horizontales para mostrar datos por categoría, como por ejemplo marcas de vehículos más vendidas. Para varias categorías de datos, puede usarse un gráfico de barras agrupadas, por ejemplo para mostrar millas por galón para clases de automóviles.

El uso de diferentes colores para diferentes variables facilita su comparación. Los gráficos de barras apiladas se usan para enfatizar una comparación de grupos de datos en relación con otros grupos. El enfoque de los gráficos de barras apiladas se encuentra en la imagen general, no en los detalles más sutiles. Por ejemplo, si una empresa quiere comparar tres de sus mejores vendedores durante un periodo de cuatro meses, un gráfico de barras apiladas resulta ideal, ya que proporciona una visión general de las ventas de cada elemento en comparación con los demás. Los gráficos circulares se usan para mostrar proporciones.

Aunque es mejor usarlos con moderación y solo para comparar pequeñas categorías de información, como por ejemplo partidas en el presupuesto de un proyecto. A continuación están los gráficos de dispersión, que ilustran la relación o dependencia entre dos conjuntos de datos, y generalmente se representan en un eje X e Y. Resultan útiles cuando existen muchos puntos de datos con los que trabajar, como a la hora de demostrar la relación entre

dos variables como por ejemplo volumen y coste por unidad. Si un gráfico de dispersión tuviera más de dos variables, se conoce como una matriz de gráficos de dispersión.

La matriz es un grupo de gráficos de dispersión dispuesta en una cuadrícula. La cuadrícula se usa para determinar si existen correlaciones lineales entre numerosas variables. Para disponer de visualizaciones de datos más avanzadas, considere un gráfico de mapa. Como el nombre lo sugiere, el gráfico de mapa representa visualizaciones en función de la geografía. Suelen contener capas de datos y marcadores interactivos que los usuarios pueden filtrar para encontrar información específica. Google Maps es un ejemplo de gráfico de mapa. Los usuarios pueden encontrar diferentes rutas del punto A al B, en función de su modo de transporte.

Un gráfico de coordenadas paralelas proporciona a cada variable su propio acceso y cada acceso está en paralelo con el siguiente. Comparan una selección de productos con atributos idénticos, como por ejemplo especificaciones de vehículos a través de modelos. Las coordenadas paralelas podrían representar las diferentes variables para vehículos en función de características como ahorro de combustible, peso, régimen de aceleración y similares. Los gráficos de burbujas se usan para representar relaciones entre tres o cuatro rangos de datos en los ejes X e Y, donde el valor de la tercera serie de datos determina el tamaño de una burbuja. Con texto explicativo, estos gráficos pueden interpretarse rápidamente.

Los gráficos radiales no se suelen utilizar en el negocio, aunque si se usan es mejor usar de dos a tres grupos dispares para evitar confusión. Un gráfico de caja compara variaciones de conjuntos de datos y es efectivo para análisis estadístico. Por ejemplo, empresas inmobiliarias pueden usar un gráfico de caja para ilustrar y comparar los valores de venta de varias casas a través de varias ciudades. Técnicamente, cualquier dato puede mostrarse en cualquier tipo de gráfico o diagrama. No obstante, algunos son más adecuados que otros para un mensaje específico.

El uso de un gráfico circular para mostrar el cambio de las ventas con el tiempo no resulta efectivo. Un sencillo gráfico de líneas es suficiente. Además, no importa lo sencillo que sea, la inclusión de texto explicativo ayuda a aclarar el mensaje y resalta los datos relevantes.

Diseñar mejores prácticas para visualizaciones de datos

Utilizada adecuadamente, la visualización de datos aplica los principios de la teoría Gestalt, que hace referencia a cómo interpreta el cerebro lo que hay que visualizar para comunicar de forma efectiva información estadística. Considere iconos y símbolos que muchos pueden encontrar diariamente, como el icono de papelería en la computadora. Aunque pueda parecer abstracto, representa la tarea muy real de eliminar algo.

Al tratar con datos, es importante tomar conceptos abstractos representados por los datos y convertirlos en visualizaciones que sean más tangibles e inteligibles. La teoría Gestalt se centra en la ley de Pragnanz, que significa ser sustancial aunque conciso. Símbolos como la papelería de la pantalla deben tener un sentido inmediato sin requerir consideración ni

explicación adicionales. Deben ser suficientemente específicos para transmitir significado, sin permitir confusión.

De acuerdo a la ley de pragnanz, el cerebro humano favorece la simplicidad. Procesamos patrones visuales básicos con más rapidez que los complejos. Para identificar a las personas que ofrecen el mejor rendimiento del año mediante un gráfico de barras, por ejemplo, si se ordenan en orden alfabético, los datos relevantes no serían inmediatamente evidentes. No obstante, si se ordenan de la puntuación mayor a la menor, los datos son fáciles de interpretar. En lo que se refiere a la visualización de datos, los datos deben disponerse de una forma lógica.

A continuación está la ley Gestalt de continuidad, que plantea que el cerebro humano agrupa de forma natural elementos que están en consonancia entre sí. Tomemos el gráfico de barras de las personas que ofrecen mejor rendimiento. En un gráfico dispuesto alfabéticamente, las barras crean una línea irregular con subidas y bajadas repentinas. Aunque en un gráfico organizado por puntuación, existe una línea continua que hace que sea comprensible de forma más inmediata. La ley Gestalt de similitud sugiere que agrupamos automáticamente objetos de forma, color, tamaño, etc., similares. La asignación de un color diferente a cada barra de un gráfico de barras, por ejemplo, hace que los datos sean más sencillos de seguir. Y también interpretamos objetos situados cerca entre sí como un grupo de acuerdo con la ley Gestalt de proximidad.

Considere la forma en la que un gráfico de barras muestra las ventas de cuatro regiones en cuatro trimestres. Cuando se agrupan por trimestre, determinar las ventas por región a través de los cuatro trimestres resulta difícil. Pero cuando se agrupan por región, resulta mucho más fácil. Al usar visualización de datos, las agrupaciones deben apoyar el objetivo de un gráfico. Finalmente, use siempre una leyenda en un gráfico que contenga múltiples variables. La leyenda actúa como una explicación de las variables usadas en el gráfico cuando aparecen. Proporciona una guía rápida sobre cómo identificar los datos de cada variable, aunque considere la claridad de la posición y detalle de la leyenda. Cuando una leyenda se coloca o etiqueta incorrectamente, puede provocar confusión, lo que hace que se malinterpreten los datos. Y el propósito de todas las decisiones tomadas en la creación de visualizaciones de datos es la claridad, lo que hace que la historia que cuentan los datos sea lo más precisa y concisa posible.

Evitar visualizaciones de datos engañosas

La idea de la comunicación de datos consiste en hacer que la información que represente sea coherente y rápidamente comprensible. La visualización de datos es una forma efectiva para hacerlo. Las visualizaciones engañosas o confusas no solo fallan a la hora de transmitir su mensaje, sino que tienen el riesgo de que los datos se malinterpreten o no se actúe al respecto de forma significativa. Para empezar, los cerebros humanos no se adaptan bien en relación con la escala. Resulta difícil comprender completamente grandes números. Por lo tanto, empiece la escala en cero.

Los cambios en la escala están abiertos a diferentes interpretaciones de los datos, y el valor de inicio de un eje tiene un gran impacto en el valor relativo de los mismos. Suponga que tenemos dos gráficos que muestran información de la tasa de interés. En uno, los valores de la izquierda o del eje Y indican valores pequeños y progresivos, lo que hace que parezca que la tasa de interés aumenta drásticamente. En realidad, hay apenas una diferencia de un punto. Un gráfico diferente, con escala desde 0 en el eje Y, hace que esto resulte mucho más obvio. Evite también usar gráficos acumulativos. Muchas organizaciones los usan para exagerar el rendimiento de ventas de productos o ingresos generales.

Siempre que haya una pendiente continua hacia arriba en un gráfico, se asume automáticamente que las cosas se están haciendo bien. Sin embargo, la presentación de la misma información en un formato no acumulativo puede revelar que se cumple lo opuesto. Asegúrese de seguir convenciones gráficas estándar. Por ejemplo, no use ejes Y invertidos, ya que al hacerlo se sugiere un resultado diferente del real.

Colocar datos en formatos no estándar puede ser engañoso en el mejor de los casos, y deshonesto en el peor. Finalmente, no omita datos innecesariamente. Cuando se ignoran puntos de datos específicos, genera una tendencia o patrón que no refleja los datos reales. Esto puede generar conclusiones incorrectas. Supongamos que tenemos dos gráficos de tendencias del mercado bursátil. Las empresas pueden engañar fácilmente a inversores al hacerles pensar que el mercado está estable mostrándoles un gráfico que omita años alternos para mostrar un incremento estable.

La precisión y honestidad requieren mostrar un gráfico que incluya puntos de datos para cada año con el fin de reflejar la volatilidad real del mercado. Considere una empresa que usa un gráfico de rosquilla, similar a un gráfico circular. Para ilustrar listados de puestos vacantes, este es un ejemplo de demasiados datos en un gráfico. La imagen resultante es confusa y poco clara en cuanto a la idea principal que se está transmitiendo y lo que el observador está intentando obtener de ella. El anillo exterior se descompone posteriormente en subcategorías con porciones demasiado estrechas para etiquetas, lo que hace que el observador esté confuso y el mensaje sea poco claro. Al igual que un gráfico circular, los gráficos de rosquilla se adaptan mejor para mostrar porcentajes de dos a cuatro componentes con variaciones drásticas. Aunque no tantos como cinco o seis, o una docena.

La idea es que presentar datos visualmente, con gráficos o diagramas, requiere la decisión consciente de determinar el tipo de visualización que represente mejor los datos de una forma comprensiva, coherente y precisa. La presentación de datos para explicar, persuadir, informar e ilustrar significa hacer que sean claros y comprensibles para su audiencia. Las visualizaciones de datos son herramientas de comunicación para facilitar esa tarea.