MÓDULO 2 - UNIDAD 2

Introducción

La importancia de la abstracción y la manipulación de los datos, se basa en que podamos definir una serie de términos genéricos que permitan a todas las audiencias, sin importar su rama o especialización, poder entender cuál es la base de los datos presentados, así como el significado de cada uno de los elementos presentes en el dataset, ya que, para dos audiencias diferentes, los objetivos del estudio puedan ser totalmente diferentes. Por ejemplo, si encontramos un dataset clínico, es posible que para el área de biología, se quiera estudiar el comportamiento y la evolución de ciertas enfermedades o virus o bacterias bajo ciertas condiciones o cierto tipo de pacientes, y a su vez, para el área médica, se quiere valorar la efectividad de un medicamento en un paciente de cierto tipo de condiciones, lo cuales, aunque partan del mismo principio y puede que se basen en el mismo dataset, sus objetivos de estudio son totalmente diferentes. Es por ello que, a lo largo de esta unidad, se presentarán diferentes técnicas que permitan abstraer y manipular los datos, de modo que, puedan derivarse en nuevos datos o transformar los mismos, con el fin de vislumbrar el objetivo de la visualización.

Estrategias para búsqueda de los datos

Muchas veces se requiere que el usuario o quién está realizando el análisis de los datos, busque por los elementos de interés o identifique el mensaje que quiere transmitir a lo largo de la visualización. Para ello, lo que se usan son 4 alternativas para identificar y localizar el objetivo, que puede ser o no conocido.

Búsqueda

La búsqueda hace referencia a que conoces con exactitud dónde y qué tipo de datos quieres estudiar. Por ejemplo, quieres buscar la correlación entre el rendimiento de un empleado en el top 4 de organizaciones más grandes basado en la autoestima de las personas, con el fin de identificar si el tamaño de la organización incide directamente en la felicidad de un empleado. En este caso, conoces exactamente qué deseas buscar, dónde lo puedes encontrar y qué datos necesitarías para realizar el análisis.

1. **Búsqueda:** La búsqueda hace referencia a que conoces con exactitud dónde y qué tipo de datos quieres estudiar. Por ejemplo, quieres buscar la correlación entre el rendimiento de un empleado en el top 4 de organizaciones más grandes basado en la autoestima de las personas, con el fin de identificar si el tamaño de la organización incide directamente en la felicidad de un empleado. En este caso, conoces exactamente qué deseas buscar, dónde lo puedes encontrar y qué datos necesitarías para realizar el análisis.
2. **Localización:** Cuando se debe encontrar un objetivo en una locación desconocida, es posible que debamos buscar en diferentes conjuntos de datos para lograr encontrar nuestro dato objetivo. Una analogía que se realiza en este punto, es que, cuando estás buscando por ejemplo conejos, es posible que a lo largo de la búsqueda te topes con diferentes tipos de roedores, antes de encontrar los conejos, o en su defecto, con familias o variantes de conejo. Mismo caso podría ser cuando quieres encontrar algo demasiado específico: “Conocer la cifra total de niños de 5-7 años que fueron reportados como caso positivo de Covid - 19 en el 2021 en Bogotá D.C.”. Para ello, seguramente primero encontrarás los datos de los menores de edad que fueron reportados como positivo en Covid - 19, luego, tendrás que hacer la búsqueda entre esos datos de las personas que estuvieron en Colombia, luego, las personas de Bogotá y finalmente, buscar los datos de los niños de 5-7 años, lo cual, te permite llegar y localizar tu set de datos objetivo.
3. **Navegación:** La navegación por el contrario, puede basarse en varias características que se buscarían analizar, pero que no se conoce su locación o su objetivo específico. Para ello, se puede citar como ejemplo el análisis del comportamiento de las acciones en la bolsa de EE.UU de varias empresas y se “navega” un día en específico el comportamiento de las mismas, por ejemplo, día en que inició la caída del mercado de valores por efectos de pandemia (20 de Febrero del 2020), allí, aunque la locación es desconocida, hay varias características agrupadas que permiten ir a buscar el dataset o estudio o mensaje objetivo de la visualización.
4. **Exploración:** En muchos casos, el objetivo no es conocido, y tampoco se conoce el posible lugar donde puedan estar los datos que queremos analizar, esto se conoce como exploración, la cual, principalmente, se caracteriza por “explorar” a lo largo de un set grande de datos, cuál es nuestro dataset o dato objetivo. Para ello, seguramente podrías analizar múltiples opciones hasta encontrar la que quieras analizar a profundidad. Un ejemplo de ello, podría ser, analizar la tendencia de precios de diferentes líneas de productos. Para lograrlo, te basas en una o varias compañías que compartan este tipo de líneas de producto y realizar una comparativa, en la cual, posiblemente puedas encontrar situaciones objetivo a analizar, cómo la tendencia de los productos cosméticos o la tendencia del consumo de cierta línea en particular que ha presentado cambios abruptos a lo largo del tiempo o ha sufrido alguna regulación de precios y es allí, donde a través de una búsqueda, obtienes tu dataset o tipo de estudio o mensaje objetivo.

Estrategias para manipular los datos

Selección de Elementos

Esta estrategia hace referencia a que se seleccionen uno o más elementos de interés de la visualización. Generalmente la salida de esta selección debería ser la entrada de la subsecuente operación. Y este tipo de selección permite o genera dependencia de la estrategia de cambio, ya que normalmente se requiere previamente hacer una selección.

¿Qué se debe tener en cuenta a la hora de escoger o seleccionar los elementos?

Una de las sugerencias fundamentales es seleccionar cuáles serán los elementos que son objetivo del estudio, para ello, normalmente se usa el elemento que tiene el valor cuantitativo de la categoría que se estudia y un identificador, que permite correlacionar este elemento con otras características a través de esta llave. En algunas herramientas de visualización también es posible escoger tipos de atributos específicos, generando que sea posible presentar en más de una vista o visualización, toda la perspectiva analizada de los datos. Otra de las reglas comúnmente usadas, se basa en el uso del canal del color, en el cual, apoyando en la iluminación, el color o la saturación a nivel de contraste, permite identificar o transicionar entre los elementos seleccionados (asignándoles un color en específico diferenciable de los datos originales) y los datos no seleccionados (generalmente opacados). En muchos casos, apoyándose de esta última técnica, se preservan los colores a lo largo de diferentes vistas con el fin de mantener o relacionar más fácilmente el mensaje entre las vistas, sin necesidad de conocer el id del objeto en estudio.

Navegación: Cambiando entre Puntos de Vista

Cuando se tienen dataset con grandes cantidades de datos, generalmente éstos no pueden ser condensados o sintetizados en un sólo punto de vista. Es por ello que, se usa la metáfora de navegación para cambiar entre diferentes puntos de vistas, con el fin de sintetizar la vista total (la visualización final objetivo, el mensaje a transmitir, la sumatoria de los puntos de vistas que comunican los mensajes, entre otras ideas referentes a “la vista total”).

Navegación: Reduciendo atributos

Cuando se tienen dataset con grandes cantidades de Así como se indica en la selección, el objetivo al ir navegando entre las vistas es que cada una contenga sólo los atributos de los elementos que se encuentran en estudio, permitiendo que, cada uno de los puntos de vista exprese el mensaje que específicamente se requiere, ya que sino se realiza esta correcta segmentación, es posible que los diferentes puntos de vistas terminan pareciéndose entre sí y expresando el mismo mensaje pero con visualizaciones diferentes.

**Estrategias para separar en múltiples vistas**

Estas estrategias fueron creadas para dividir o mostrar la información a través de múltiples vistas o puntos de vistas, de modo que, para la audiencia o stakeholder objetivo, sea más fácil engranar toda la información presentada y que pueda llevarse el mensaje completo al armar “las piezas del rompecabezas”. En estas estrategias se debe tener cuidado de cuántas regiones usar, cómo agruparlas, cuando dividir, cuando detenerse en la división, como usarlas, entre otras características relevantes a tener en cuenta. Entre las estrategías existen:

* Yuxtaposición y Coordinación entre Vistas

Usar una yuxtaposición implica como coordinar y crear relaciones entre las vistas (vistas vinculadas). Para ello, a su vez, existen cuatro alternativas para realizarlo, entre las que encontramos:

* Compartir la Codificación

Este tipo de opción de diseño hace énfasis en que se use la misma codificación visual entre múltiples vistas, generando que, la audiencia o stakeholder interesado, pueda comprender cómo se encuentran estas vistas interrelacionadas. Esto significa a su vez, que todos los canales son manejados de la misma manera a lo largo de las vistas. Sin embargo, es posible que existan variaciones entre las vistas y se mantenga sólo uno de los canales codificados de igual manera, por ejemplo, en una vista indicar que el elemento A corresponde al color verde con una marca circular, pero en la siguiente vista, se use una marca diferente pero se mantenga el canal del color (verde), pero, ésto no sólo se limita al canal del color, también es posible orientar las vistas con el canal de posición, ángulo, tamaño, entre otros.

* Compartir los Datos

Esta opción de diseño habla de qué tantos datos se comparten entre dos vistas. Una de las aproximaciones consiste en compartir todos los datos, en el cual, cada vista trabaja de forma independiente y muestra la información de acuerdo al mensaje objetivo de la misma, normalmente, no se usa con con la aproximación anterior de compartir la codificación, ya que las vistas serían similares y redundantes.

La siguiente aproximación se compone de dos pasos, el “overview” y el “detail”, en la cual, se usa una vista para agregar, resumir y segmentar los datos de forma general, y en la segunda vista, se ve el detalle de los elementos del dataset, presentando la información con todos sus atributos o los atributos seleccionados. Para esta aproximación es posible combinar y compartir la codificación y compartir los datos, dado que, se parte del esquema general y se quiere revisar un comportamiento específico de un elemento en cuestión. En cuanto a la aproximación de cuántas vistas usar, generalmente se recomiendan dos pero, es posible que existan diferentes niveles discretos que requieran segmentar en más vistas detalladas cada uno de los elementos. Un ejemplo de esta aproximación sería:



**Tomado de:**<https://www.google.com/maps/@4.6409379,-74.1913801,11z>

En la cual, se comparten los datos y se comparte la codificación, ya que a partir del overview de Latinoamérica, es posible llegar al detalle de la ciudad de Bogotá con sólo realizar zoom:



**Tomado de:**

<https://www.google.com/maps/@4.6409379,-74.1913801,11z>

Otra aproximación es combinar el overview-detail con el compartir datos, de modo que, cada una de las vistas detalladas tienen diferentes aproximaciones de codificación visual en contraste con el overview. También es posible que, en una de las vistas detalladas, se ofrezca más información que la que se “resume” en el overview, llamando a esta aproximación como detail-on-demand o detalle a demanda.

La tercera alternativa indica que se comparten todos los datos pero se muestra una porción diferente del dataset en cada vista. Esta aproximación se conoce como small múltiples o pequeños múltiples, permitiendo que, se pueda tener un marco de referencia y permita conocer qué porción de los datos se está evaluando. La recomendación para usar este tipo de aproximación consiste en que todas las vistas se muestran al tiempo, de modo que para la audiencia sea fácilmente comparable entre las diferentes vistas el mensaje que expresa cada una, sin generar una carga cognitiva demasiado alta.

* Compartir la Navegación: Sincronizar

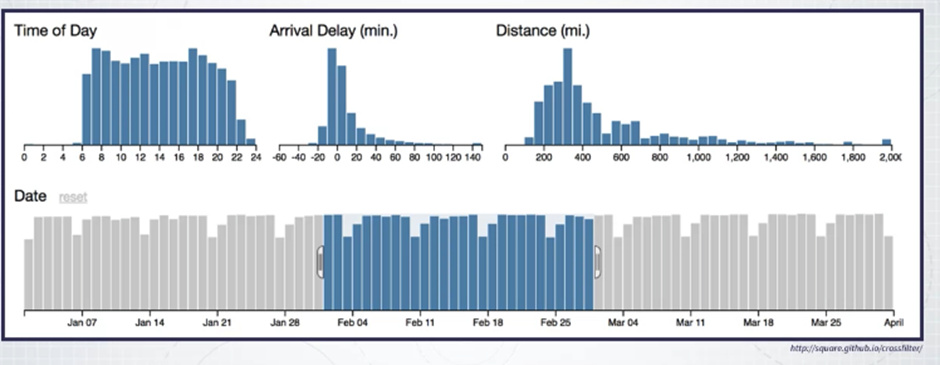
Para esta estrategia, el objetivo es compartir la navegación a través de relaciones, de modo que sea posible moverse entre un punto de vista y otro de manera sincronizada. Un ejemplo nuevamente sería los mapas de navegación, en los cuales, podemos navegar entre el mapa y verlo de forma resumida (vista más alejada) sin perder de referencia el punto que se está buscando.

* Yuxtaposición de las vistas

Esta aproximación básicamente se refiere a que ambas vistas se encuentran “lado a lado” para ser utilizadas por la audiencia, aunque, en muchas herramientas es posible difuminar una vista de otra y usarlas en la misma visualización, permitiendo al usuario conocer la relación entre ambas y la acción de tomar una u otra.

* Dividir en múltiples vistas

Para dividir la información en múltiples vistas, es necesario realizar a menudo una jerarquización de atributos o decidir el orden en que se mostrará la información. Muchas veces se realiza una aproximación bottom-up, en la cual, se empiezan a caracterizar todas las partes específicas de los elementos, de modo que, poco a poco se vaya construyendo el “overview” general, que engloba todo el mensaje a transmitir, esto, generalmente se conoce como niveles de abstracción y permiten a su vez, generar la jerarquización y la división a través de dimensiones de los datos. Por ejemplo, podríamos determinar o ir correlacionando cómo es el comportamiento de los vuelos en unas fechas determinadas y encontrar relación entre diferentes atributos de la información presentada así:



**Tomado de:** <https://square.github.io/crossfilter/>

En la cual se presenta a los stakeholders el tiempo por día, luego se presenta el retraso por días y las distancias de cada vuelo por cada día, para al final, condensarla en una vista cómo la que se presentó en el ejemplo anterior.

* Estrategias para reducir ítems y atributos: filtrar, agregar, embeber

La reducción es una de las estrategias más importantes para manejar la complejidad en las visualizaciones, y en muchos de los casos es posible que la estrategia permite reducir o aumentar el número de elementos visibles en cada visualización. Entre las estrategias encontramos:

* Filtrar

La estrategia de diseño de filtrar es directamente la de reducir los números de elementos mostrados, en algunos casos, estos elementos son simplemente eliminados. En algunas visualizaciones también se realiza este filtrado basado en las acciones o dimensiones de interés del usuario o inclusive, el tipo de perfil que ve la misma, por lo cual, se seleccionan basado en ello, los elementos de interés. Sin embargo, la técnica más sencilla para realizar el filtrado de información se basa en eliminar elementos basados en un rango de valores cuantitativos o en su defecto, realizar la reducción de la cantidad de atributos a analizar, esto permite que exista una mayor absorción de la información por parte de la audiencia a la que se presenta la visualización.

* Agrupación / Agregación / Embeber

Otra de las estrategias de reducción de ítems consiste en la agregación o la agrupación de los mismos, de modo que, este grupo de elementos está representado por un nuevo elemento derivado del mismo. Esta estrategia permite no eliminar los elementos y no son mutuamente excluyentes, es decir, puedo realizar un filtrado previo a una agregación o viceversa. Las agregaciones siempre implican el uso de un atributo derivado o creado a partir de otro, como podría ser el promedio, máximos, mínimos, conteos, entre otros. Sin embargo, se debe tener precaución en no eliminar información interesante a través de la agrupación.

**Lectura Complementaria de la estrategia de embeber:** Tamara Muzner Capítulo 14 - Página 323 – 333

**Nota:**

Para acceder a este recurso, deberás ingresar [aquí](https://search-ebscohost-com.bdbib.javerianacali.edu.co/Community.aspx?authtype=ip&stsug=AmE9slWqwy7fZUCW7sYtzdHTNyNg_LScwDPmIoxm-sfE9E7MD6JwFrEnZduM2B5-ALONV3UuBaNSkI267Oga5SNgy8brnOJ1iu40ldf_U_tq6HhqG3bFwgeEkdxwJgbrgSlJJMad_bX9wKDUAx-gndFpbJT67Chphg3-udOpM7jK_nk&IsAdminMobile=N&encid=22D731163C3635973736357632253C973933311375C371C371C375C373C376C370C331&selectServicesToken=A2HXJ_OhUIT8BuLya8oU-bMSskESccih6HuEv3DbHUNyuWVDVySnN70SgBYLdDThzeKQKYivNQMbS-26zrUGi-e_e0p0vOodhMSFfvdoGr7BRmMzuMIZXD4psneQbNlnG9qrxpXlGocMaLW5xIRjh7uzOkCkwX1o4Anix9zYp_3VB55toWeLQWNktnqymF0ISIWbP7nDWQbPSJdL2POTBo8wLKRgsI1bl43-3UJCpVE_oNgWSr0muP_q8l1vW-0waPsVRvUVYrUGSqz-10INpS23WOGvGn-u). Una vez ingreses a la página principal de EBSCO, deberás hacer clic en la primera opción EBSCOhost Research Databases, luego, debes seleccionar la casilla “Seleccionar / anular selección de todo” y hacer clic en “Continuar”. Después de esto estarás en el buscador de EBSCO y en la casilla de búsqueda, deberás ingresar el código 9781466508910 y seleccionar la opción “Texto completo en PDF”.