

## Cápsula 3: Tipos de atributos

Hola, bienvenidxs a una cápsula del curso Visualización de Información. En esta hablaré sobre tipos de atributos.

Como recordarán, los atributos son propiedades que se pueden medir. Entre ellos podemos diferenciar dos tipos intrínsecamente distintos de atributos, los categóricos versus los ordenados.

Primero, los **atributos categóricos** son aquellos que no tiene un orden intrínseco, como puede ser nombres de frutas, el código postal, el género de personas, razas de perros, nombres de ciudades, géneros de películas, etc...

Las comparaciones entre valores que son posibles son igualdad o diferencia, permiten entonces separar a los datos en **categorías**. Claramente uno puede siempre ordenar valores categóricos según un orden externo como el alfabético, pero ese orden no es necesariamente intrínseco a la semántica del dato como lo es en el caso de los atributos ordenados.

Entonces los **atributos ordenados** son todos aquellos con un orden intrínseco, como la edad, el peso de tortugas y tallas de ropa. Pero este grupo se puede subdividir más aún en atributos ordinales o atributos cuantitativos.

La diferencia esencial entre estos dos grupos es **si existe o no comparación aritmética exacta entre los valores posibles del atributo**. Si hay un orden pero no una comparación aritmética, entonces es **ordinal**. Un ejemplo son las tallas de ropa que usan letras para diferenciarse: "pequeño", "mediano" o "grande". Es claro que "pequeño" es menor que "mediano", pero ¿por cuánto? ¿Es dos o tres veces más pequeño?

Lo mismo ocurre con los rankings o listas de mejor a peor, estos definen un orden entre ítems, como los 10 mejores jugadores de tenis, pero no es directo cuanto "mejor" es la primera posición en relación al resto de las posiciones.

El resto de los atributos ordenados entonces son **cuantitativos**, ya que estos presentan definiciones aritméticas que permiten comparar cantidades de mucho mejor forma. La altura de personas es un atributo cuantitativo, y es posible comparar cuántas veces más alta es una persona adulta de un infante. Otros ejemplos son el precio de productos, la temperatura en ubicaciones en el mundo, el número de cafés vendidos en una cafetería, etc... Suelen corresponder a atributos que se representan como números enteros o reales.

Otra forma de distinguir atributos ordenados es si son secuenciales o divergentes. Un atributo es **secuencial** si sus valores posibles se distribuyen a lo largo de un rango homogéneo con un valor mínimo y un valor máximo. Mientras que el rango de un atributo

**divergente** puede separarse en dos rangos secuenciales que apuntan en direcciones opuestas que se encuentran en un cero o valor nulo común.

Ejemplos de atributos secuenciales son la altura de montañas, que van desde valores pequeños sobre el nivel del mar hasta la altura del Everest, y por separado las distancias de profundidad marítima, que van desde poco profundas hasta lo más profundo conocido por la humanidad. Pero se pueden considerar ambas escalas juntas como una gran escala divergente donde el nivel del mar sirve de cero común, las montañas aumentan hacia arriba y las profundidades marítimas aumentan hacia abajo.

Otra característica a considerar en atributos ordenados es si son **cíclicos o no**. Es decir, si sus valores siguen un orden que los devuelve al valor mínimo eventualmente en vez de aumentar indefinidamente o llegar a un máximo. Muchas mediciones **temporales** son cíclicas gracias a sus semánticas, como los días de la semana, las horas en los días, o los meses en el año. Los años en contraste no son cíclicos ya que aumentan de forma indefinida. Por otro lado, un porcentaje usualmente no es cíclico, ya que al aumentar puede llegar a un valor máximo de 100%, que bajo sus semánticas no tiene siguiente valor.

Finalmente, pero no menos importante, también existe la diferenciación entre atributos llave y atributos de valor. Esto puede referirse tanto a atributos ordenados como a categóricos, y se refiere a que si un atributo en un *dataset* actúa como índice para buscar y determinar ítems u otros atributos. Si es capaz de actuar como índice o identificador, se dice que es un atributo de **llave**, mientras que si no, es un atributo de **valor**.

Esto generalmente tiene como implicancia que los valores posibles de un atributo de llave no tiene duplicados en el *dataset*, y cada valor identifica un ítem específico. Las llaves pueden ser implícitas o explícitas en el *dataset*, y por lo general corresponden a atributos categóricos u ordinales y no a cuantitativos. Un ejemplo de atributo llave es el RUT identificador de cada persona en Chile, no hay dos personas con el mismo. ¿Se te ocurren ejemplos de atributos llaves? ¡Déjalos en los comentarios!

También, grupos de atributos pueden comportarse como llave en conjunto. Por ejemplo, en un *dataset* sobre la temperatura en distintas ciudades de Chile durante todo el 2020, no es posible determinar un valor de temperatura específico solo conociendo la ciudad o solo conociendo la fecha de interés. Pero al conocer ambas la ciudad y fecha, si es posible. Casos con grupos de atributos que se comportan como índice se consideran **tablas multi-índice**, mientras que aquellas tablas con índices individuales, serían **tablas planas**.

Con eso termina el contenido de esta cápsula. Recuerda que si tienes preguntas, puedes dejarlas en los comentarios del video para responderlas en la sesión en vivo de esta temática. ¡Chao!