

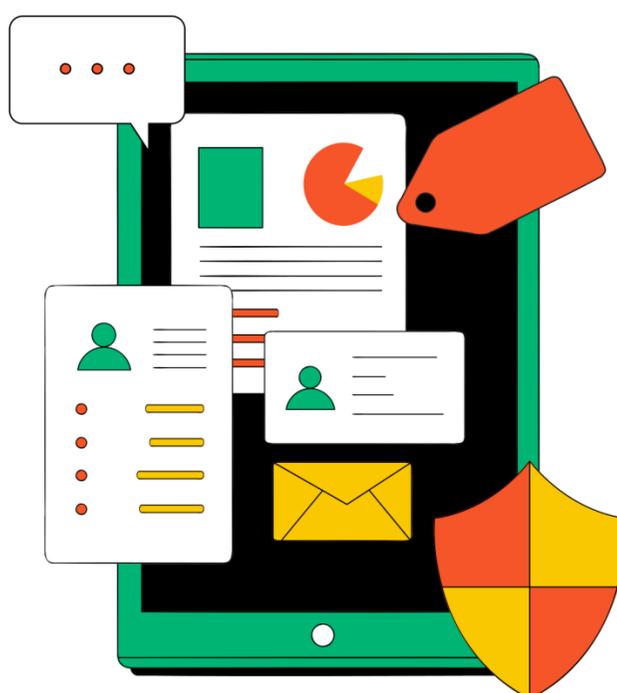
Lección 2: Gráficos y diagramas



La visualización de datos es un rol importante en el análisis de los datos. A través de los gráficos se puede resumir grandes cantidades de información de forma comprensible para las personas, también es posible determinar qué tendencias tienen los datos y cómo se comportan con solo observar un gráfico. Utilizando Python y algunas de sus bibliotecas es posible crear muchos tipos de visualizaciones para representar la información de manera clara.

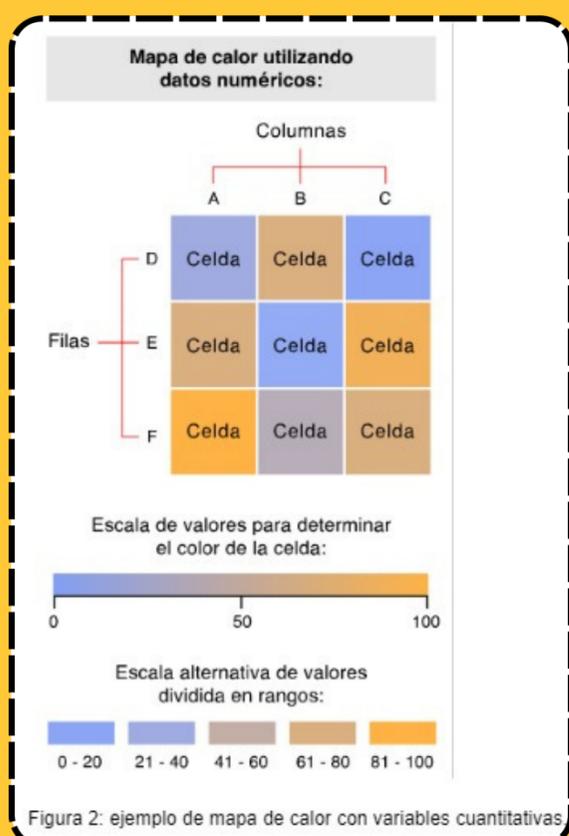
Gráficos de matrices de puntos

El gráfico de matriz de puntos muestra datos discretos en unidades de puntos, cada uno coloreado para representar una categoría particular y agrupado en una matriz. Se utiliza para dar una visión general rápida de la distribución y las proporciones de cada categoría en un conjunto de datos y también para comparar la distribución y la proporción entre otros conjuntos de datos, con el fin de descubrir patrones.



Mapas de calor

Los mapas de calor visualizan los datos a través de variaciones en el color. Cuando se aplican a un formato tabular, los mapas de calor son útiles para el examen cruzado de datos multivariados, mediante la colocación de variables en las filas y columnas, y coloreando las celdas dentro de la tabla. Los mapas de calor son buenos para mostrar la varianza a través de múltiples variables, revelando cualquier patrón, mostrando si las variables son similares entre sí y para detectar si existen correlaciones entre ellas. Se requiere una leyenda junto a un mapa de calor para que se pueda leer correctamente. Los datos categóricos están codificados por colores, mientras que los datos numéricos requieren una escala de color que se mezcla de un color a otro para representar la diferencia en valores altos y bajos. Puede utilizarse una selección de colores sólidos para representar un rango múltiple de valores (0-10, 11-20, 21-30, etc.) o una escala de gradiente para un solo rango (por ejemplo 0 - 100) mezclando dos o más colores juntos. Debido a su dependencia del color para comunicar valores, los mapas de calor son un gráfico más adecuado para mostrar una visión más generalizada de los datos numéricos, ya que es más difícil distinguir con precisión las diferencias entre tonos de color y extraer puntos de datos específicos (a menos que incluyan los datos sin procesar en las celdas). Los mapas de calor también se pueden utilizar para mostrar los cambios en los datos con el tiempo, si una de las filas o columnas se establecen en intervalos de tiempo. Un ejemplo de esto sería utilizar un mapa de calor para comparar los cambios de temperatura a lo largo del año, en varias ciudades, para ver dónde están los lugares más calientes o más fríos para vivir. Así que las filas podrían enumerar las ciudades a comparar, las columnas contienen cada mes y las celdas contendrían los valores de la temperatura.



Diagramas de árbol

Un **diagrama de árbol** es una forma de representar visualmente la jerarquía en una estructura similar a un árbol. Normalmente, la estructura de un diagrama de árbol consta de elementos como nodo raíz, un miembro que no tiene superior/padre. Luego están los nodos, que están conectados entre sí con conexiones de línea llamadas ramas que representan las relaciones y conexiones entre los miembros. Finalmente, los nodos de hojas (o nodos finales) son miembros que no tienen hijos o nodos secundarios. Los diagramas de árbol se utilizan a menudo:

- Para mostrar las relaciones familiares y el
- descenso. En la taxonomía, la práctica y la ciencia
- de la clasificación. En la ciencia evolutiva, para
- mostrar el origen de las especies. En informática y
- matemáticas. En las empresas y organizaciones con fines gerenciales.

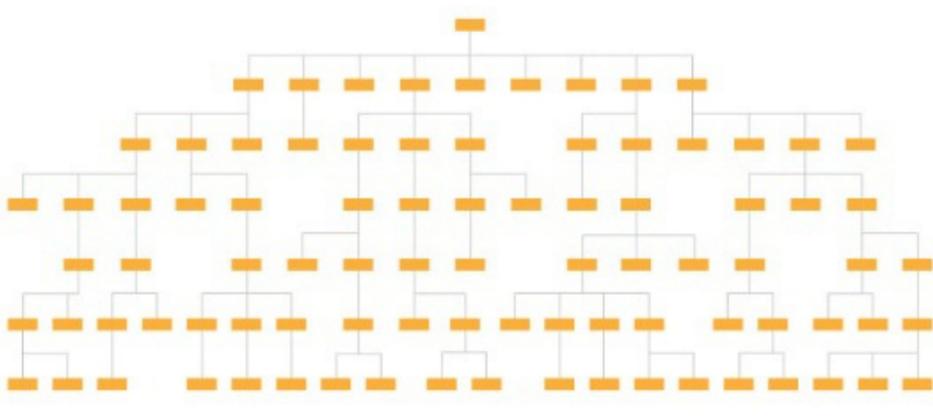
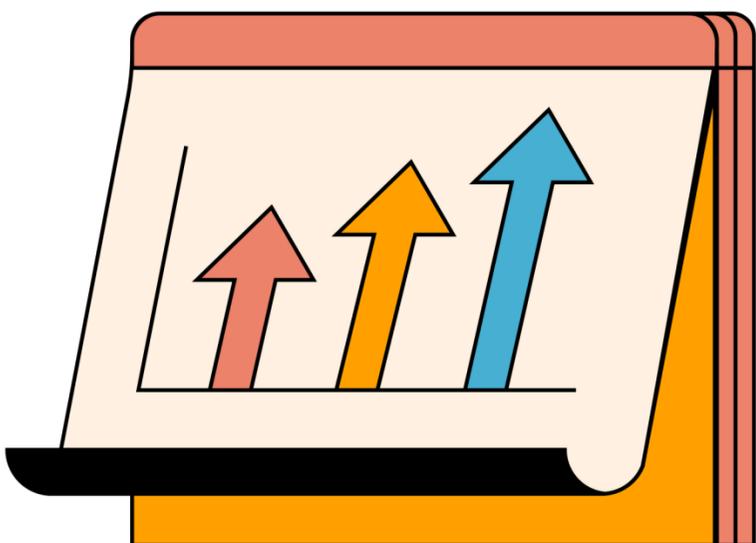


Figura 3: ejemplo de diagrama de árbol.



Mapas de árbol

Los mapas de árbol son una alternativa para visualizar la estructura jerárquica de un diagrama de árbol mientras que también muestra las cantidades para cada categoría a través del tamaño del área. A

cada categoría se le asigna un área rectangular con

sus rectángulos de subcategoría dentro de él. Cuando se asigna una cantidad a una categoría, el tamaño de su área se muestra en proporción a esa cantidad y a otras cantidades dentro de la misma categoría padre en una relación de parte a todo. Además, el tamaño del área de la categoría padre es el total de sus subcategorías. Si no se asigna ninguna cantidad a una subcategoría, su área se divide igualmente entre las otras subcategorías dentro de su categoría padre. La forma en que se dividen y ordenan los rectángulos en subrectángulos depende del algoritmo de mosaico utilizado. Se han desarrollado muchos algoritmos de mosaico, pero el «algoritmo de cuadrado» que mantiene el rectángulo como un cuadrado es uno de los más utilizados.



Figura 4: Ejemplo de mapa de árbol.

Gráficos radiales

Los **gráficos radiales** son una forma de comparar múltiples variables cuantitativas. Esto los hace útiles para ver qué variables tienen valores similares o si hay valores atípicos entre cada variable. Los gráficos radiales también son útiles para ver qué variables son altas o bajas dentro de un conjunto de datos, haciéndolos ideales para mostrar el rendimiento. A cada variable se le proporciona un eje que empieza en el centro. Todos los ejes se disponen radialmente, con distancias iguales entre sí, manteniendo la misma escala entre todos los ejes. Las líneas de cuadrícula que se conectan de eje a eje a menudo se usan como guía. Cada valor de variable se traza a lo largo de su eje individual y todas las variables en un conjunto de datos se conectan para formar un polígono.

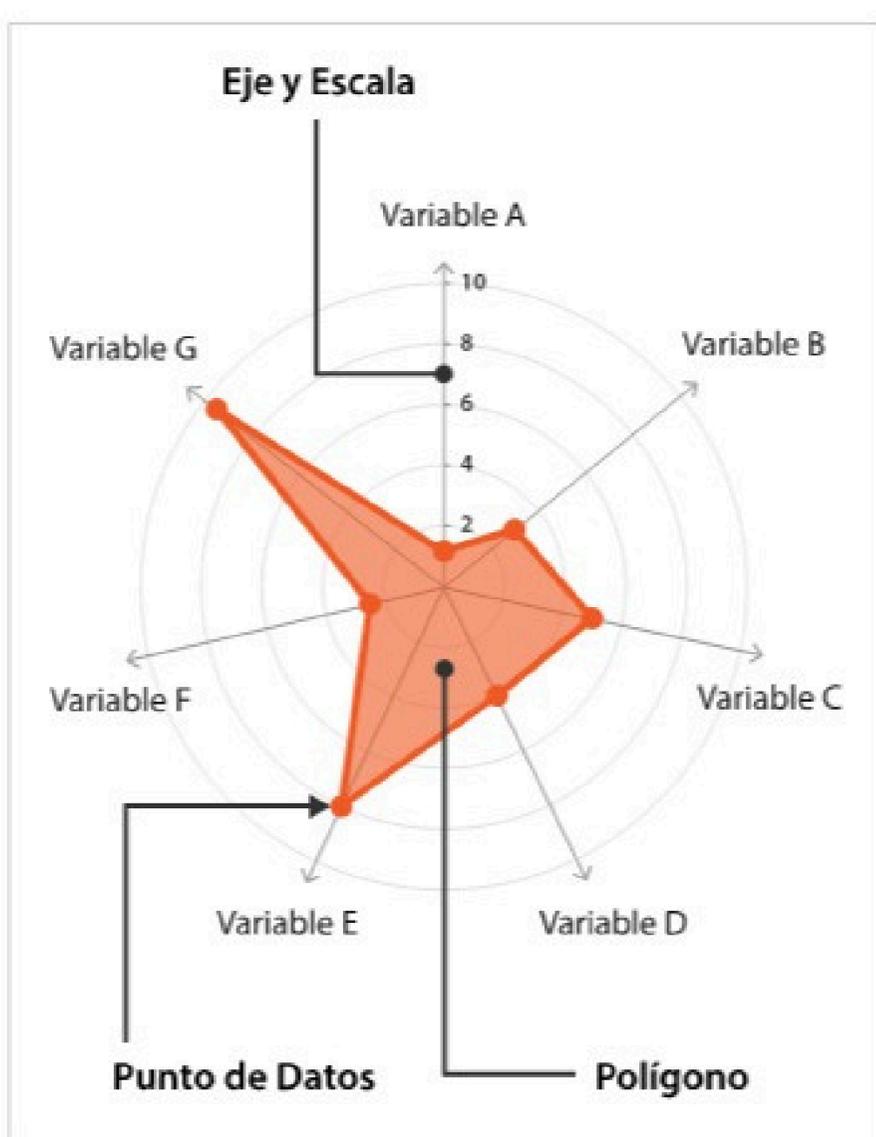


Figura 6: ejemplo de gráfico radial