

Clasificación de Datos: Numéricos y Categóricos para Analistas de Datos

Introducción

En el análisis de datos, la clasificación adecuada de los tipos de datos es fundamental para seleccionar los métodos de análisis y modelos adecuados. En este documento, describiremos dos categorías principales de datos: **numéricos** y **categóricos**. Cada uno tiene sus propias características y aplicaciones en el análisis.

1. Datos Numéricos

Definición

Los datos numéricos son aquellos que representan cantidades o mediciones que pueden expresarse mediante números. Son datos cuantitativos que se pueden contar o medir, lo que permite realizar operaciones matemáticas con ellos.

Clasificación de Datos Numéricos

Existen dos subcategorías principales de datos numéricos:

1. **Datos Discretos:** Son aquellos que pueden tomar un número limitado o contable de valores. Generalmente, estos datos representan conteos o valores enteros.

- **Ejemplos:**

- Número de empleados en una empresa.
- Cantidad de productos vendidos en un día.
- Número de cursos completados por un estudiante.

2. **Datos Continuos:** Son datos que pueden tomar un número infinito de valores dentro de un rango determinado. Generalmente representan mediciones.

- **Ejemplos:**

- Altura de una persona (175.5 cm).
- Tiempo empleado para completar una tarea (2.45 horas).
- Temperatura registrada en un día (23.7 °C).

Ejemplos en el Ambiente de un Analista de Datos

- **Ventas:** El monto total de ventas generadas en un mes.
- **Producción:** La cantidad de productos fabricados en una línea de producción.
- **Métricas de rendimiento:** La tasa de conversión de clientes en una campaña de marketing.

Análisis Común de Datos Numéricos

Los datos numéricos suelen analizarse usando estadística descriptiva (promedio, mediana, moda) y visualizaciones como histogramas, gráficos de dispersión y gráficos de líneas.

2. Datos Categóricos

Definición

Los datos categóricos son aquellos que representan categorías o grupos, y no pueden medirse o realizarse operaciones aritméticas con ellos. Son cualitativos y se utilizan para clasificar o etiquetar elementos.

Clasificación de Datos Categóricos

Existen dos tipos principales de datos categóricos:

1. **Datos Nominales:** Son categorías que no tienen un orden inherente.

- **Ejemplos:**

- Colores de productos (rojo, azul, verde).
- Género de las personas (masculino, femenino).

- Marcas de automóviles (Toyota, Ford, BMW).
- 2. **Datos Ordinales:** Son categorías que tienen un orden o una secuencia lógica, aunque la diferencia entre ellas no sea medible de manera exacta.
 - **Ejemplos:**
 - Niveles de satisfacción (bajo, medio, alto).
 - Grados académicos (licenciatura, maestría, doctorado).
 - Tamaños de camiseta (pequeño, mediano, grande).

Ejemplos en el Ambiente de un Analista de Datos

- **Segmentación de Clientes:** Clasificación de clientes según su nivel de lealtad (nuevo, frecuente, VIP).
- **Encuestas:** Respuestas a una encuesta de satisfacción (insatisfecho, neutral, satisfecho).
- **Demografía:** Estado civil (soltero, casado, divorciado).

Análisis Común de Datos Categóricos

Los datos categóricos se analizan usando tablas de frecuencia, gráficos de barras o gráficos de sectores. En modelos predictivos, estos datos suelen transformarse en variables dummy o codificarse usando técnicas como "one-hot encoding".

3. Diferencias Clave entre Datos Numéricos y Categóricos

Característica	Datos Numéricos	Datos Categóricos
Tipo de Información	Cantidades o mediciones.	Categorías o grupos.
Operaciones	Se pueden realizar operaciones aritméticas.	No se pueden realizar operaciones aritméticas.
Ejemplos	Altura, ingresos, edad.	País de origen, nivel educativo, color favorito.
Visualización	Gráficos de líneas, histogramas, gráficos de dispersión.	Gráficos de barras, gráficos de sectores.

Conclusión

Entender la diferencia entre los datos numéricos y categóricos es esencial para cualquier analista de datos, ya que la selección de las técnicas de análisis depende de la correcta clasificación de los datos. Cada tipo de dato tiene sus propios métodos de tratamiento y visualización que permiten obtener información útil para la toma de decisiones.
