



**UNIVERSIDAD DE GUAYAQUIL**  
**FACULTAD DE INGENIERIA INDUSTRIAL**  
**CARRERA SISTEMAS DE INFORMACIÓN**



**MATERIA**

CIENCIA DE DATOS

**TEMA**

TIPOS DE VARIABLES EN CIENCIA DE DATOS

**DOCENTE**

ING. HECTOR DULCEY ASTUDILLO

**ESTUDIANTE**

ANTHONY FLORES TOMALA

CARLOS GARCIA CEDEÑO

ANTHONY MONCAYO FAJARDO

MADELINE MUÑOZ VILLEGAS

**CURSO**

SIN-S-NO-8-2

**CICLO I**

2023-2024

## **INTRODUCCION**

La ciencia de datos es un campo el cual se centra en extraer conocimientos y obtener información útil, esto a su vez combina la estadística, matemáticas, programación y conocimiento en dominios específicos para extraer conocimientos y revelar patrones ocultos a partir de conjuntos de datos. La ciencia de datos abarca todo el ciclo de vida de los datos, desde la recopilación y el almacenamiento hasta el análisis y visualización de información. Comienza con la identificación de preguntas, o problemas de investigación, seguido de la recolección de datos relevantes.

La ciencia de datos se ha vuelto cada vez más relevante en la era de la información y la tecnología. El crecimiento exponencial de los datos generados por las empresas, las organizaciones y los dispositivos conectados ha creado una necesidad creciente de expertos en ciencia de datos que puedan comprender y aprovechar este vasto conjunto de información.

## **TIPOS DE VARIABLES**

El análisis de datos es el proceso de evaluar los diferentes tipos de variables y su comportamiento utilizando un razonamiento analítico y lógico para examinar cada componente de los datos proporcionados. El análisis de datos es un proceso de inspección, limpieza, transformación y modelado de datos con el objetivo de descubrir información útil, informar conclusiones y respaldar la toma de decisiones y esto se hace a través de las variables que se tiene en la ciencia de datos.

En la ciencia de datos, las variables son elementos fundamentales que representan características o atributos de los datos que estamos analizando. Comprender los diferentes tipos de variables es de suma importancia, ya que afecta como se deben tratar, analizar y modelar los datos. “En el campo de la ciencia de datos, es esencial comprender los diferentes tipos de variables, ya que esto influye en el tipo de análisis y las técnicas que se pueden aplicar” (Brownlee, 2020). Los tipos de variables comunes incluyen variables numéricas, categóricas, ordinales, texto y binarias

Las variables numéricas son aquellas que representan cantidades numéricas y pueden ser continuas o discretas. Los ejemplos de variables numéricas incluyen la edad de una persona, el precio de un producto o la altura de un edificio. Estas variables se pueden someter a operaciones matemáticas y se utilizan con frecuencia en análisis estadísticos.

Las variables categóricas, representan características que se encuentran en categorías o grupos. Estas categorías pueden ser mutuamente excluyentes o superpuestas. Por ejemplo, el género de una persona (masculino o femenino) o el estado civil (soltero, casado, divorciado) son variables categóricas. Estas variables se pueden utilizar en análisis de frecuencia y análisis de asociación.

Las variables ordinales son similares a las variables categóricas, pero tienen un orden o jerarquía inherente. Por ejemplo, una escala de calificación de 1 a 5, donde 1 representa la calificación más baja y 5 la más alta, es una variable ordinal. Estas variables se pueden utilizar en análisis de tendencias y comparaciones.

Las variables de texto contienen texto o cadenas de caracteres. Estas variables se utilizan comúnmente en análisis de procesamiento de lenguaje natural (NLP) y minería de texto. Ejemplos de variables de texto incluyen comentarios de usuarios, tweets o reseñas de productos.

Las variables binarias representan características que solo pueden tomar uno de los dos valores posibles, como si/no, verdadero/falso o 1/0. Estas variables se utilizan a menudo en problemas de clasificación o en la construcción de modelos de predicción.

“Es fundamental comprender los tipos de datos en estadística, ya que esto influye en la elección de las pruebas y técnicas estadísticas adecuadas para analizar los datos” (Chetty, 2018).

**Variables numéricas:** Representan valores numéricos y se dividen en dos subtipos:

- Variables continuas: Pueden tomar cualquier valor dentro de un rango determinado. Ejemplos: la edad de una persona, la temperatura, el ingreso anual.
- Variables discretas: Solo pueden tomar valores enteros o separados. Ejemplos: el número de hijos, el número de productos vendidos, la puntuación en un examen.

**Variables categóricas:** Representan categorías o etiquetas y se dividen en dos subtipos:

- **Variables nominales:** Las categorías no tienen un orden intrínseco. Ejemplos: el color de los ojos, el estado civil, el país de origen.
- **Variables ordinales:** Las categorías tienen un orden específico. Ejemplos: nivel educativo (primaria, secundaria, universidad), clasificación de la satisfacción del cliente (muy insatisfecho, insatisfecho, neutral, satisfecho, muy satisfecho).

**Variables ordinales:** Representan fechas, horas o ambos. Ejemplos: fecha de nacimiento, fecha de compra, hora de inicio de un evento.

**Variables binarias:** Representan dos categorías mutuamente excluyentes, a menudo codificadas como 0 y 1. Ejemplos: género (masculino/femenino), presencia/ausencia de una característica.

**Variables de texto:** Representan texto o cadenas de caracteres. Ejemplos: comentarios de los clientes, descripciones de productos, nombres de personas.

Además, también existen las variables según su ***Rol de análisis***, las cuales permiten conocer la forma en la que interactúan entre sí y como influyen en los resultados. Al asignar roles específicos a las variables, se facilita la interpretación de los resultados y se aclara su papel en el análisis.

En la ciencia de datos existe dos variables según su rol de análisis:

**Variable predictora:** Es también conocida como variable explicativa o variable independiente, es aquella que se utiliza para predecir o explicar la variable objetiva, es decir esta variable influye o afecta a la variable objetiva.

**Variable objetiva:** Esta es la variable que se pretende predecir o explicar en función de otras variables, en pocas palabras es el resultado o el fenómeno que se quiere entender o predecir.

## CONCLUSIÓN

Los tipos de variables en la ciencia de datos son una parte fundamental del proceso de análisis de datos. Comprender y clasificar correctamente las variables permite seleccionar las técnicas adecuadas y obtener información valiosa de los datos.

La elección del tipo de variable correcto es crucial para seleccionar las técnicas de análisis adecuadas también puede influir en la forma en que se reprocesan los datos.

El entendimiento de los diferentes tipos de variables, como numéricas, categóricas, ordinales, de texto y binarias, es crucial para determinar cómo deben ser tratados, analizados y modelados los datos. Cada tipo de variable tiene características y aplicaciones específicas en el análisis de datos, lo que afecta las técnicas y enfoques que se pueden utilizar. El análisis de datos se basa en un razonamiento analítico y lógico para examinar cada componente de los datos proporcionados, incluyendo la inspección, limpieza, transformación y modelado de los mismos.

## BIBLIOGRAFIA

Brownlee, J. (2020). Understanding Variable Types in Data Science. En J. Brownlee,

*Understanding Variable Types in Data Science.*

Chetty, P. (2018). A Guide to Data Types in Statistics. En P. Chetty, *A Guide to Data*

*Types in Statistics.*

ma, equipo. (2019, September 17). *Tipos de variables en el análisis de datos.* Marketing

Analítico; Marketing-Analítico.com. [https://www.marketing-](https://www.marketing-analitico.com/analitica-web/tipos-de-variables-en-el-analisis-de-datos/)

[analitico.com/analitica-web/tipos-de-variables-en-el-analisis-de-datos/](https://www.marketing-analitico.com/analitica-web/tipos-de-variables-en-el-analisis-de-datos/)

Borrego, Á. (2021, November 2). *Capítulo 1 Introducción / Introducción a la ciencia de*

*datos para estudiantes de Información y Documentación.* Bookdown.org.

[https://bookdown.org/angelborrego/ciencia\\_datos/introduccion.html](https://bookdown.org/angelborrego/ciencia_datos/introduccion.html)