

# **Notas de clase Business Analytics**

Daniel Parra

Invalid Date

# **Tabla de contenidos**

# Bienvenidos

Este libro ha sido creado con el propósito de ofrecer un resumen claro y conciso de los conceptos clave de las clases de Business Analytics para los posgrados de administración de la Pontificia Universidad Javeriana. Es un complemento diseñado para reforzar el aprendizaje en clase, pero no pretende reemplazar la experiencia educativa que estas ofrecen.

Parte del texto en estas notas de clase ha sido elaborado con ayuda de inteligencia artificial (ChatGPT), combinando apuntes propios y referencias de diversos libros. Las ilustraciones, inspiradas en el libro “Math with Bad Drawings” de Ben Orlin, también han sido generadas utilizando inteligencia artificial.

© 2025 Pontificia Universidad Javeriana. Este libro está licenciado bajo una licencia CC BY 4.0.

# 1 Introducción: Descubriendo la Analítica de Datos

Imagina que trabajas en una empresa que debe decidir qué producto lanzar o qué precio poner a sus artículos. ¿Cómo tomas estas decisiones? Antiguamente, las decisiones se tomaban según la intuición del jefe (**HiPPO**, la persona mejor pagada en la organización), pero actualmente contamos con algo mejor: la analítica de datos.

## 1.1. ¿Qué es la Analítica de Datos para Negocios?

La analítica de datos es la aplicación de herramientas tecnológicas y estadísticas para analizar información relevante y apoyar la toma de decisiones en una organización.

Los datos, sin embargo, no son suficientes por sí solos. Necesitamos interpretarlos de manera efectiva para contar historias convincentes que permitan tomar decisiones informadas.

## 1.2. ¿Qué tipo de preguntas podemos responder con la analítica?

La analítica de negocios nos permite responder preguntas prácticas, como:

- ¿Quiénes son mis clientes?
- ¿Qué campaña de marketing es más efectiva?
- ¿Cuál sería el efecto de cambiar los precios?
- ¿Cómo predecir la demanda de un producto?

## 1.3. Tipos de Analítica

Hay tres tipos principales de analítica:

### **1.3.1. Analítica Descriptiva**

Describe la situación actual mediante visualizaciones y estadísticas básicas.

#### **Ejemplos:**

- ¿Han crecido las ventas este año?
- ¿Qué región vende menos?

### **1.3.2. Analítica Predictiva**

Predice qué sucederá, aprovechando correlaciones entre variables. Aquí no es necesaria la causalidad, sino una relación estadística sólida.

#### **Ejemplos:**

- ¿Cuál es la probabilidad de que un cliente deje de comprar?
- ¿Qué tan probable es que una transacción sea fraudulenta?

### **1.3.3. Analítica Prescriptiva**

Sugiere decisiones óptimas, enfocándose en relaciones causales y, frecuentemente, utilizando experimentos.

#### **Ejemplos:**

- ¿Subir el precio incrementará las ganancias?
- ¿Qué diseño de página web aumentará las ventas?

## **1.4. La cadena de valor en Business Analytics**

La analítica de negocios incluye:

- **Recolección y organización de datos:** Crear bases de datos (SQL, Power BI).
- **Análisis de datos:** Usar herramientas como R, Excel o Python para análisis estratégicos.
- **Comunicación de resultados:** Presentar hallazgos en reportes y visualizaciones (PowerPoint, Tableau, Quarto).

## **2 Capítulo 2**

### **2.1. Introducción**

Este es el nuevo capítulo 2.

### **2.2. Contenido**

Contenido del capítulo...

## 3 Explorando tus Datos con Estadísticos Descriptivos

Imagina que eres gerente de un restaurante que acaba de lanzar un nuevo menú y quieres evaluar la satisfacción de tus clientes a partir de las calificaciones que te han dado. Tienes cientos de puntuaciones y necesitas entender rápidamente cuál es la opinión general, qué tan variadas son las experiencias y si hay patrones o valores atípicos que merecen atención.

¿Cómo puedes resumir y comprender toda esa información de forma clara y sencilla? Para eso utilizamos los estadísticos descriptivos, herramientas fundamentales que nos permiten condensar grandes cantidades de datos en medidas simples y significativas.

Estas medidas serán esenciales para tomar decisiones informadas que mejoren tu decisión.

### 3.1. Estadísticos Descriptivos

Los estadísticos descriptivos son herramientas clave que nos permiten resumir y comprender mejor la información contenida en nuestros datos. Para facilitar su estudio, los podemos clasificar en cuatro categorías principales:

- **Medidas de tendencia central:** Estas nos muestran el valor típico o representativo en un conjunto de datos, es decir, alrededor de qué número se agrupan la mayoría de las observaciones.
- **Medidas de variación:** Nos indican qué tan dispersos o concentrados están los datos respecto a esa tendencia central, ayudándonos a entender la consistencia o diversidad dentro de la información.
- **Medidas de forma:** Describen la distribución general de los datos, revelando si están simétricos, sesgados hacia un lado o presentan picos o colas particulares.
- **Medidas de relación:** Evalúan cómo dos variables numéricas se relacionan entre sí, especialmente si existe una conexión lineal que pueda ser útil para análisis más avanzados.

## 3.2. Medidas de Tendencia Central

Cuando queremos entender las calificaciones que los clientes dan a tu restaurante, buscamos encontrar un valor que represente lo que la mayoría piensa. Las medidas de tendencia central nos ayudan a esto.

### 3.2.1. Media (Promedio)

La media o promedio es una forma simple pero útil de resumir un conjunto de datos. La media es el valor que obtienes al sumar todas las calificaciones y dividirlas entre el número total de clientes.

💡 Matemáticamente se escribiría así:

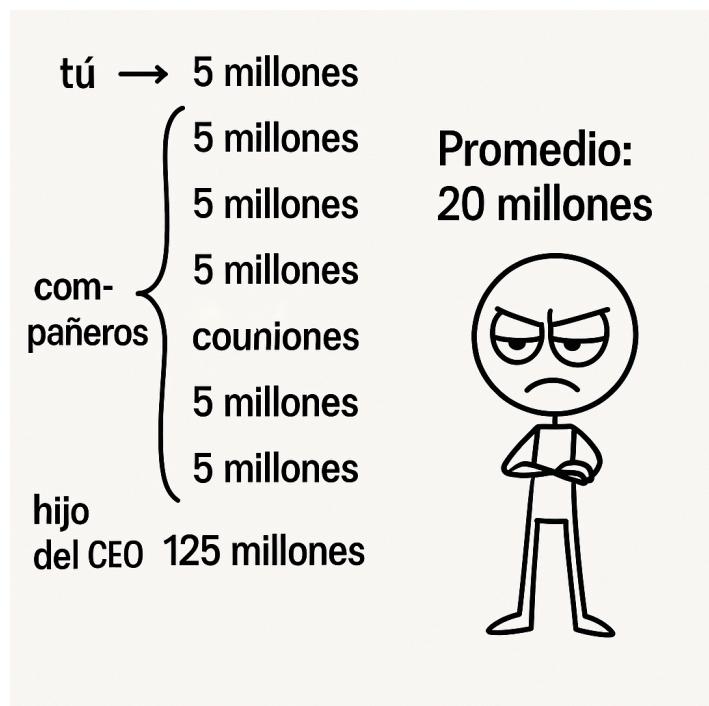
$$\text{Media} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Por ejemplo, si cinco clientes dieron las calificaciones: 4, 5, 3, 4 y 5, la media será:

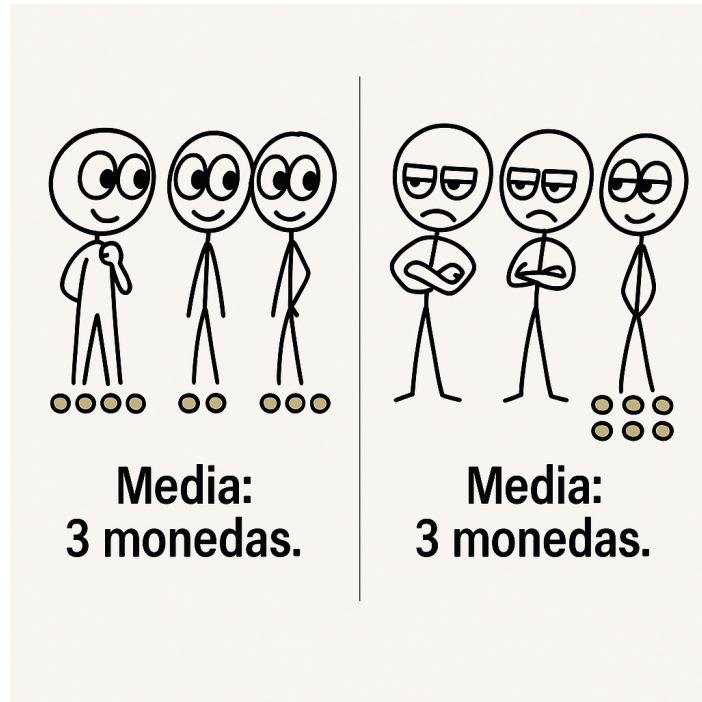
💡

$$\bar{x} = \frac{4 + 5 + 3 + 4 + 5}{5} = \frac{21}{5} = 4.2$$

Sin embargo la media, tiene limitaciones: no muestra cómo están distribuidos los datos y puede ser engañosa si hay grandes diferencias entre los valores. Por ejemplo:



Aquí hay otro ejemplo donde la media es la misma en ambas situaciones pero el contexto individual que esconde es muy diferente:



### 3.2.2. Mediana

Volviendo a las calificaciones de tu restaurante, la **mediana** es el valor que está justo en el medio cuando ordenas todas las opiniones de menor a mayor. Esto significa que la mitad de los clientes dieron una calificación igual o menor que la mediana, y la otra mitad dio una calificación igual o mayor.

Por ejemplo, si tus clientes calificaron así: 3, 4, 4, 5, 5, la mediana es 4 — porque es el valor que divide el grupo en dos partes iguales.



Mediana = valor central en datos ordenados

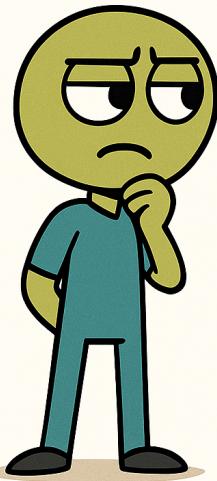
Una gran ventaja de la mediana es que **no se ve afectada por calificaciones muy bajas o muy altas** que podrían distorsionar la media. Por ejemplo, si alguien puso un 1 o un 10, la mediana sigue mostrando el punto medio real de la mayoría.

Sin embargo, la mediana **no nos dice qué tan dispersas están las calificaciones a cada lado**. Por eso, para entender mejor la variabilidad de las opiniones, necesitaremos otras medidas que veremos más adelante.

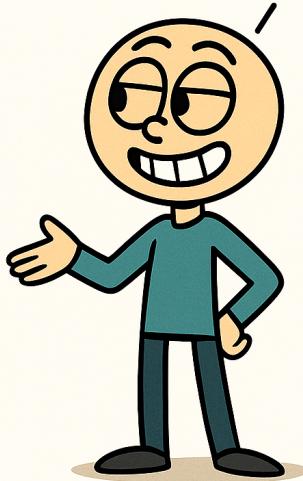
A continuación podemos ver un ejemplo donde la mediana es usada para entregar un mensaje erróneo:

Entonces,  
por qué  
debería  
invertir  
contigo?

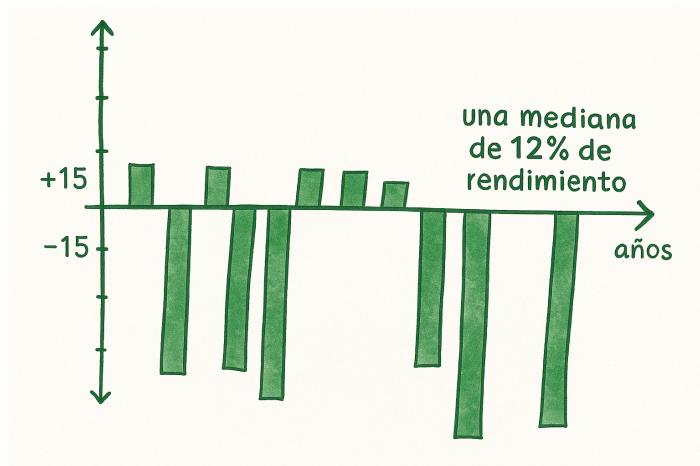
I



No es por  
presumir, pero  
mi fondo tiene  
una **mediana**  
de **12%** de  
rendimiento al año



Pero, los rendimientos anuales del fondo:



Miremos otro ejemplo donde la mediana similar no implica datos similares. Siempre hay tener una combinación de datos para tomar decisiones correctas



### 3.2.3. Moda

La **moda** es la calificación que más se repite entre tus clientes. Es como la opinión más común o popular sobre tu restaurante.

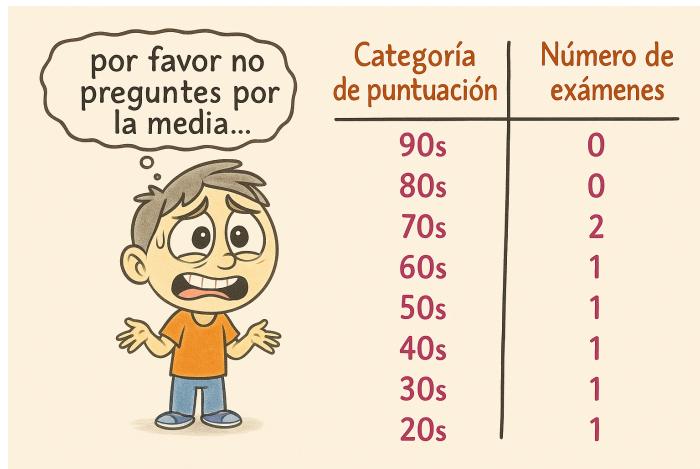
Por ejemplo, si las calificaciones de tus clientes fueron: 3, 4, 4, 5, 5, tanto 4 como 5 se repiten dos veces, por lo que hay dos modas: 4 y 5.

Si no hay repeticiones exactas, se pueden agrupar en categorías y tomar como moda la más común. Es útil especialmente con datos no numéricos, como colores o preferencias políticas, donde no tiene sentido calcular promedios.



Moda = valor que aparece con mayor frecuencia

Su limitación: no considera la totalidad ni la distribución de los datos, y lo más común no siempre es lo más representativo. Miremos este ejemplo:



### 3.3. Medidas de Variación

Las medidas de variación nos ayudan a entender qué tan diferentes o dispersos están los datos entre sí. Es decir, nos dicen si las opiniones o valores están muy juntos o muy separados.

### 3.3.1. Rango

Las medidas de variación nos ayudan a entender qué tan diferentes o dispersos están los datos entre sí. Es decir, nos dicen si las opiniones o valores están muy juntos o muy separados.

Por ejemplo, si las calificaciones en tu restaurante van desde 2 hasta 5, el rango sería:



$$\text{Rango} = 5 - 2 = 3$$

Esto nos dice que las opiniones varían en un rango de 3 puntos, desde una calificación baja hasta una alta.

Su principal ventaja es su simplicidad, da una idea rápida del “ancho” del conjunto de datos.

Pero su debilidad es igual de clara, solo considera los valores extremos, ignorando por completo todos los datos intermedios.

Miremos un ejemplo de un rango que da una impresión incorrecta:

