

Análisis Exploratorio de Datos con Google Sheets

Utilizando Google Sheets para EDA

Agenda

1. ¿Qué es el Análisis Exploratorio de Datos (EDA)?
2. ¿Por qué utilizamos EDA?
3. Cómo realizar EDA con Google Sheets
4. Estadísticas y fórmulas esenciales en EDA
5. Tipos de gráficos para EDA

¿Qué es EDA?

Exploratory Data Analysis

- El Análisis Exploratorio de Datos (EDA) es un enfoque para analizar conjuntos de datos para resumir sus características principales.
- Utilizamos EDA para:
 - Descubrir patrones
 - Identificar anomalías
 - Probar hipótesis
 - Comprobar suposiciones

¿Por qué utilizamos EDA?

- EDA nos ayuda a:
 - Comprender la estructura de los datos
 - Detectar errores
 - Determinar relaciones entre variables
 - Seleccionar características importantes para modelos

Cómo realizar EDA con Google Sheets

- Google Sheets ofrece una plataforma accesible y poderosa para realizar EDA.
- Algunas herramientas útiles en Google Sheets incluyen:
 - Filtros
 - Funciones estadísticas
 - Gráficos
 - Tablas dinámicas

Estadísticas y fórmulas esenciales en EDA

Promedio (AVERAGE)

- **Descripción:** El promedio es el valor medio de un conjunto de números. Se calcula sumando todos los valores y dividiendo por la cantidad de valores.
- **Sintaxis:** `=AVERAGE(rango)`
- **Ejemplo:** `=AVERAGE(A1:A10)`

Mediana (MEDIAN)

- **Descripción:** La mediana es el valor central en un conjunto de datos ordenados. Si el número de datos es par, es el promedio de los dos valores centrales.
- **Sintaxis:** =MEDIAN(rango)
- **Ejemplo:** =MEDIAN(A1:A10)

Moda (MODE)

- **Descripción:** La moda es el valor que aparece con más frecuencia en un conjunto de datos. Puede haber más de una moda si varios valores tienen la misma frecuencia máxima.
- **Sintaxis:** =MODE(rango)
- **Ejemplo:** =MODE(A1:A10)

Desviación estándar (STDEV)

- **Descripción:** La desviación estándar mide la dispersión de un conjunto de datos respecto a su promedio. Indica qué tan dispersos están los datos.
- **Sintaxis:** =STDEV(rango)
- **Ejemplo:** =STDEV(A1:A10)

IF en Google Sheets

IF (Si)

- **Descripción:** La función IF evalúa una condición y devuelve un valor si la condición es verdadera y otro valor si es falsa. Es útil para tomar decisiones basadas en condiciones específicas.
- **Sintaxis:** =IF(prueba_lógica, valor_si_verdadero, valor_si_falso)
- **Ejemplo:** =IF(A2 > 50, "Aprobado", "Reprobado")

COUNT en Google Sheets

COUNT (Contar)

- **Descripción:** La función COUNT cuenta el número de celdas que contienen números en un rango. Es útil para determinar la cantidad de entradas numéricas en un conjunto de datos.
- **Sintaxis:** =COUNT(rango)
- **Ejemplo:** =COUNT(A1:A10)

UNIQUE en Google Sheets

UNIQUE (Único)

- **Descripción:** La función UNIQUE devuelve una lista de valores únicos en un rango, eliminando duplicados. Es útil para identificar elementos distintos en un conjunto de datos.
- **Sintaxis:** =UNIQUE(rango)
- **Ejemplo:** =UNIQUE(A1:A10)

SUM en Google Sheets

SUM (Suma)

- **Descripción:** La función SUM suma todos los números en un rango de celdas. Es una de las funciones más básicas y comunes para agregar valores.
- **Sintaxis:** =SUM(rango)
- **Ejemplo:** =SUM(A1:A10)

SUMIF en Google Sheets

SUMIF (Sumar Si)

- **Descripción:** La función SUMIF suma las celdas que cumplen con un criterio específico. Es útil para sumar valores condicionalmente, como sumar ventas mayores a un cierto monto.
- **Sintaxis:** =SUMIF(rango, criterio, [rango_suma])
- **Ejemplo:** =SUMIF(A1:A10, ">50", B1:B10)

Tipos de gráficos para EDA

- **Gráficos de barras:** Comparar cantidades entre diferentes categorías.
- **Gráficos de líneas:** Mostrar tendencias a lo largo del tiempo.
- **Histogramas:** Representar la distribución de datos.
- **Diagramas de dispersión:** Mostrar relaciones entre dos variables.
- **Boxplots:** Resumir la distribución de datos y detectar valores atípicos.

 gráficos

Tipos de gráficos en Google Sheets

Gráfico de líneas

- **Descripción:** Un gráfico de líneas conecta puntos de datos con una línea, ideal para mostrar tendencias o cambios a lo largo del tiempo.
- **Uso común:**
 - **Seguimiento de ventas mensuales:** Puedes observar cómo las ventas aumentan o disminuyen mes a mes.
 - **Evolución de precios:** Muestra cómo los precios de un producto han cambiado con el tiempo.
 - **Monitoreo del clima:** Representar cambios en la temperatura diaria o mensual.

Gráfico de barras

- **Descripción:** Un gráfico de barras compara diferentes categorías usando barras horizontales, donde la longitud de cada barra representa su valor.
- **Uso común:**
 - **Comparar ingresos por diferentes productos:** Visualiza qué productos están generando más ingresos.
 - **Rendimiento de empleados:** Compara el rendimiento entre diferentes empleados o equipos.

Gráfico de áreas

- **Descripción:** Este gráfico rellena el área debajo de la línea de datos, mostrando la magnitud del cambio a lo largo del tiempo y destacando la contribución total.
- **Uso común:**
 - **Cuota de mercado acumulada:** Representa cómo cada producto contribuye a la cuota de mercado total.
 - **Consumo de recursos:** Visualiza cómo se utiliza un recurso a lo largo del tiempo.

Gráfico de dispersión

- **Descripción:** Muestra la relación entre dos variables mediante puntos dispersos en un eje. Es útil para identificar correlaciones.
- **Uso común:**
 - **Relación entre ventas y marketing:** Analiza cómo el gasto en marketing afecta las ventas.
 - **Análisis de datos científicos:** Comparar variables como peso y altura.

Gráfico circular

- **Descripción:** Representa proporciones de un todo como "rebanadas" de un círculo. Cada rebanada muestra la parte de cada categoría en comparación con el total.
- **Uso común:**
 - **Distribución de presupuesto:** Visualiza cómo se distribuye el presupuesto entre diferentes departamentos.
 - **Encuestas de preferencia:** Muestra las preferencias de los clientes por diferentes productos.
 - **Partes del mercado:** Representa la cuota de mercado de diferentes competidores.

Gráfico de radar

- **Descripción:** Compara múltiples variables entre varias categorías usando un gráfico circular con varias dimensiones.
- **Uso común:**
 - **Evaluación de rendimiento:** Comparar el rendimiento en diferentes áreas como ventas, marketing y desarrollo.
 - **Análisis de habilidades:** Visualizar las habilidades de diferentes empleados o candidatos.
 - **Rendimiento de productos:** Comparar diferentes características de productos.

Gráfico de combinación

- **Descripción:** Combina varios tipos de gráficos (líneas, barras) para representar diferentes series de datos en un solo gráfico.
- **Uso común:**
 - **Ventas totales y tendencias:** Compara ventas totales (barras) con una tendencia de ventas (línea).
 - **Presupuesto vs. gasto real:** Visualiza el presupuesto planificado frente al gasto real.
 - **Rendimiento financiero:** Combinar ingresos y gastos en un solo gráfico.

Histograma

- **Descripción:** Muestra la distribución de datos agrupados en rangos. Útil para entender cómo se distribuyen los datos a través de diferentes intervalos.
- **Uso común:**
 - **Distribución de puntuaciones:** Analiza cómo se distribuyen las puntuaciones en un examen.
 - **Análisis de frecuencia:** Muestra la frecuencia de diferentes rangos de valores en un conjunto de datos.
 - **Control de calidad:** Visualiza la variabilidad de los procesos de producción.

Boxplot

- **Descripción:** Muestra la distribución de datos basada en cinco números resumen: mínimo, primer cuartil, mediana, tercer cuartil y máximo.
- **Uso común:**
 - **Comparar variabilidad:** Compara la variabilidad de diferentes conjuntos de datos.
 - **Identificación de valores atípicos:** Detecta valores que están significativamente fuera del rango esperado.
 - **Análisis de rendimiento:** Evaluar el rendimiento de diferentes grupos o periodos.

Unir conjuntos de datos con Google Sheets

¿Por qué unir conjuntos de datos?

- En análisis de datos, a menudo necesitamos combinar información de diferentes fuentes.
- Unir conjuntos de datos nos permite:
 - Enriquecer el análisis con información adicional.
 - Comparar y contrastar datos de distintas fuentes.
 - Crear informes más completos y detallados.

VLOOKUP en Google Sheets

- **VLOOKUP** (Búsqueda vertical) se usa para buscar un valor en la primera columna de un rango y devolver un valor en la misma fila desde otra columna.
- Sintaxis: `=VLOOKUP(valor_buscado, rango, índice_columna, [ordenado])`
- Ejemplo: `=VLOOKUP(A2, Datos!A:D, 3, FALSE)`

MATCH en Google Sheets

- **MATCH** (Coincidencia) devuelve la posición relativa de un valor en un rango.
- Sintaxis: `=MATCH(valor_buscado, rango_búsqueda, [tipo_de_coincidencia])`
- Ejemplo: `=MATCH("Producto", A1:A10, 0)`

INDEX en Google Sheets

- **INDEX** devuelve el valor de una celda en un rango específico basado en las coordenadas de fila y columna.
- Sintaxis: `=INDEX(rango, num_fila, [num_columna])`
- Ejemplo: `=INDEX(A1:C10, 2, 3)`

Combinando MATCH e INDEX

- MATCH e INDEX se pueden usar juntos para buscar y devolver valores en un rango.
- Ejemplo: `=INDEX(B1:B10, MATCH("Producto", A1:A10, 0))`
- Este método es más flexible que VLOOKUP, ya que no depende de la posición de la columna.

Ventajas de diferentes métodos

- **VLOOKUP:**
 - Fácil de usar para búsquedas simples.
 - Limita las búsquedas a la primera columna del rango.
- **MATCH e INDEX:**
 - Más flexible y potente.
 - Permite búsquedas en cualquier parte del rango.

Ejemplos prácticos

- Uniendo conjuntos de datos de ventas y productos.
- Combinando información de clientes con transacciones.
- Enrichiendo datos de encuestas con detalles demográficos.

class: center, middle, green-background

Preguntas

