

«Talento Tech»

Business Intelligence

Clase 07



Clase N° 07 | Estadística Descriptiva Básica

Temario:

- Conceptos básicos: media, mediana, moda, desviación estándar. Correlaciones.
- Análisis exploratorio de datos con Google Sheets
- Data-viz básicas en estadística: barras, pies y líneas.

Objetivos de la clase:

- Definir y calcular la media, mediana, moda y desviación estándar.
- Explorar cómo estas medidas resumen conjuntos de datos.
- Comprender el concepto de correlación y cómo se mide.
- Identificar ejemplos de correlaciones en datos reales.
- Utilizar funciones básicas de Google Sheets para realizar análisis descriptivos.
- Comprender la importancia de la visualización de datos en la estadística.
- Crear gráficos de barras, sectores y líneas en Google Sheets.
- Interpretar y presentar datos utilizando diferentes tipos de visualizaciones.

Conceptos Básicos de Estadística Descriptiva

En esta clase, se abordará el tema de la **estadística descriptiva** y cómo realizar un análisis exploratorio de datos utilizando Google Sheets. Los objetivos de esta clase incluyen comprender los conceptos básicos de la estadística descriptiva, realizar análisis exploratorios de datos y aprender a visualizar información a través de gráficos.



Definición de Estadística Descriptiva

La estadística descriptiva es una rama fundamental de la estadística que se ocupa de resumir y describir las características de un conjunto de datos. A continuación se explorarán varias **medidas de tendencia central y dispersión** que nos ayudarán a comprender mejor los datos que estamos analizando.

Medidas de Tendencia Central

Las medidas de tendencia central son parámetros estadísticos que indican el valor central de un conjunto de datos. Su función es resumir la información y simplificar el análisis de datos.

Las medidas de tendencia central más utilizadas son la media, la mediana y la moda.

Media

La media, también conocida como **promedio**, es una de las medidas más comunes utilizadas para resumir un conjunto de datos. Se calcula sumando todos los valores del conjunto y dividiendo esta suma entre la cantidad total de valores. Por ejemplo, si tenemos los datos de las edades de un grupo de amigos: 20, 22, 21, 19, 24, podemos calcular la media sumando estas edades ($20 + 22 + 21 + 19 + 24 = 106$) y dividiendo entre la cantidad de amigos (5). Así, la media sería $106 / 5 = 21.2$ años.



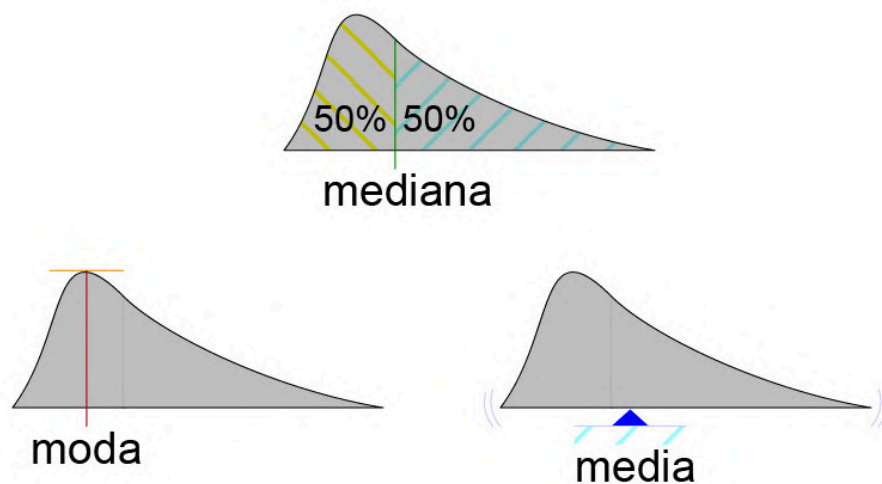
Esta cifra nos da una idea del "promedio" de edad en el grupo.

Mediana

La mediana, por otro lado, es el **valor que se encuentra en el medio de un conjunto de datos ordenados**. Para calcularla, debemos primero organizar los datos de menor a mayor. Si hay un número impar de observaciones, la mediana será el valor central. En cambio, si el número es par, la mediana será el promedio de los dos valores centrales. Por ejemplo, si las edades son 19, 20, 21, 22, 24, la mediana es 21, ya que es el tercer valor en la lista. Si las edades fueran 19, 21, 22, 24, la mediana se calcularía como $(21 + 22) / 2 = 21.5$.

Moda

La moda es la medida que nos **indica el valor que aparece con mayor frecuencia en un conjunto de datos**. Por ejemplo, si en una encuesta sobre mascotas preferidas obtuvimos las respuestas: perro, gato, gato, pez, perro, la moda sería "gato", ya que es la opción más mencionada. Es posible tener un conjunto de datos sin moda (si todos los valores ocurren con la misma frecuencia) o con más de una moda (en el caso de que dos o más valores se repitan con la misma frecuencia más alta).

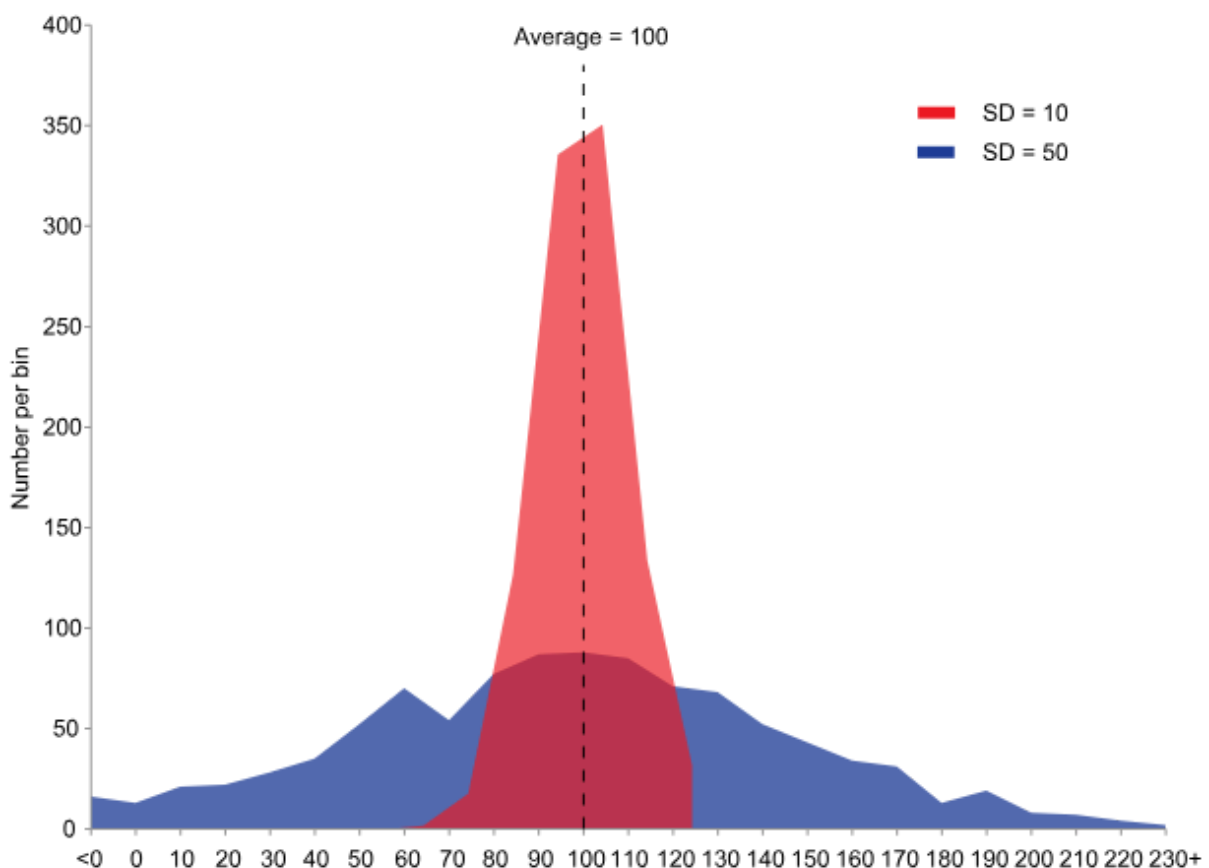


Medidas de dispersión

Las medidas de dispersión son un **conjunto de estadísticas que indican la variabilidad de los datos en relación a la media aritmética**. Permiten entender la distribución de los datos y evitar conclusiones erróneas al comparar grupos.

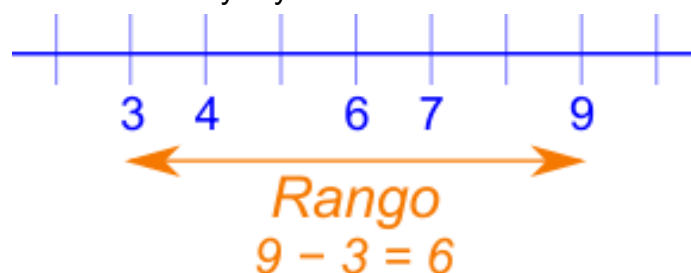
Las medidas de dispersión son complementarias a las medidas de tendencia central, que ofrecen una idea general del comportamiento de los datos.

Algunas medidas de dispersión son el rango, la desviación media, la varianza y la desviación standard.



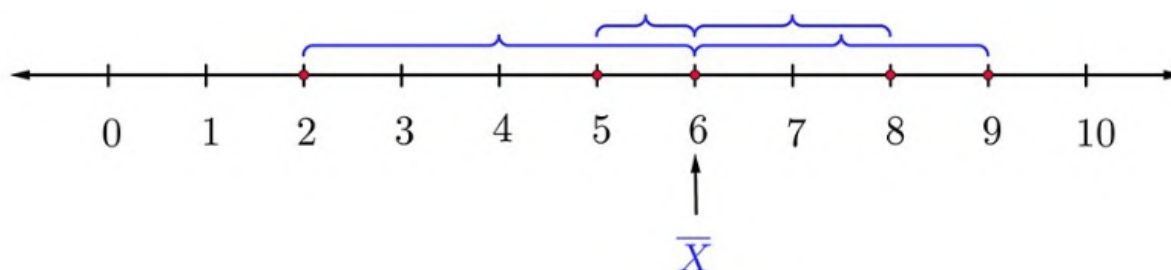
Rango

Es la diferencia entre el valor mayor y el menor de una serie de datos



Desviación media

Se trata de calcular la media aritmética de los valores absolutos de la desviación de cada valor de la variable con respecto a la media



$$DM = \frac{\sum_{i=1}^n |x_i - \bar{X}|}{n}$$

$$DM = (|2 - 6| + |5 - 6| + |8 - 6| + |9 - 6|) / 5$$

$$DM = (4 + 1 + 2 + 3) / 5 = 10 / 5 = 2$$

Varianza

Es el cociente entre la suma de los cuadrados de las desviaciones de los valores de la variable y el número de datos del estudio.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

$$S^2 = (16 + 1 + 4 + 9) / 4 = 7.5$$

Desviación Standard

La desviación estándar es una medida que nos **indica la cantidad de variabilidad o dispersión existente en un conjunto de datos**. Se trata de un indicador relevante para entender qué tan dispersos están los valores respecto a la media. Para calcularla, primero debemos encontrar la varianza y luego tomamos la raíz cuadrada de ese resultado. Siguiendo con nuestro ejemplo de las edades, si la media es 21.2, restamos este valor de cada edad, elevamos al cuadrado y calculamos la media de esos cuadrados. La desviación estándar nos dará una idea clara de cuán alejados están, en promedio, los valores respecto al promedio.

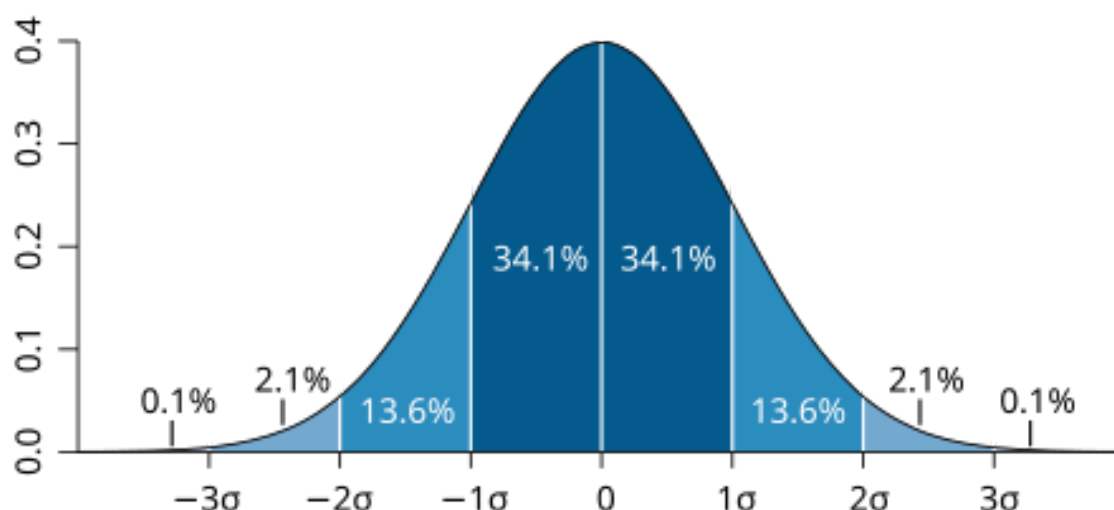
$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

$$\sigma = \sqrt{S^2}$$

$$\sigma = \sqrt{7.5}$$

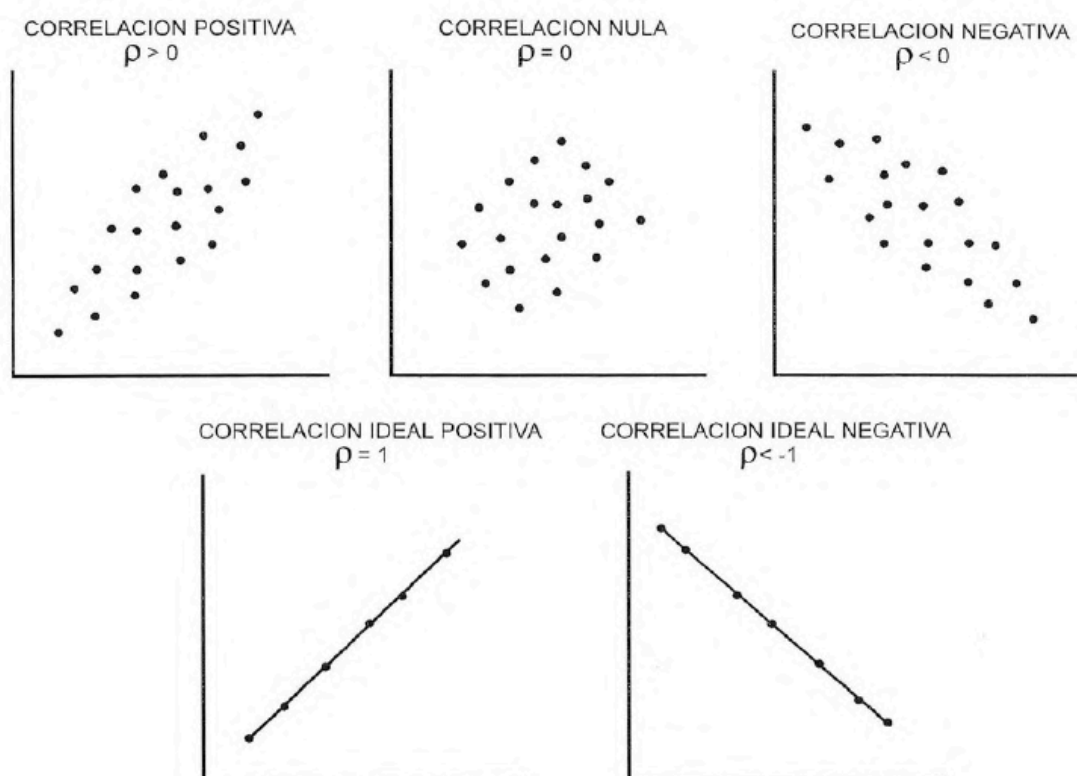
$$\sigma = 2,738612788 \approx 2,74$$

Interpretación gráfica y significado de la Desviación Standard



Correlaciones

La correlación es un concepto que **describe la relación entre dos variables**. Un coeficiente de correlación de Pearson, que varía entre -1 y 1, nos permite medir la fuerza y dirección de esa relación. Un valor de 1 implica una relación positiva perfecta, donde, al aumentar una variable, también aumenta la otra. Un valor de -1 indica una relación negativa perfecta, donde al aumentar una, la otra disminuye. Un coeficiente de 0 sugiere que no hay relación entre las variables. Por ejemplo, si estudiamos la relación entre la cantidad de horas que un estudiante estudia y sus resultados en un examen, podemos esperar que haya una correlación positiva: a medida que aumentan las horas de estudio, también lo hace el rendimiento en el examen.



! Importante: La correlación sólo se refiere a cómo dos o más variables tienden a variar juntas. Sin embargo, observar que dos variables se mueven conjuntamente no significa necesariamente que una sea la causa de la otra.

La correlación no implica causalidad.

Análisis Exploratorio de Datos con Google Sheets

Google Sheets nos permite llevar a cabo análisis exploratorios de datos de forma sencilla. Una de las primeras tareas que realizaremos será importar y organizar nuestros datos.

Entre las funciones estadísticas más comunes están:

- **Media:** `=PROMEDIO (A1:A10)` calcula el promedio de los valores en el rango especificado.
- **Mediana:** `=MEDIANA (A1:A10)` nos da el valor central de los datos ordenados.
- **Moda:** `=MODA (A1:A10)` permite identificar el valor que más se repite.
- **Desviación standard:** `=DESVEST (A1:A10)` nos indica la variabilidad de los datos.
- **Correlación entre dos variables:** `COEF.DE.CORREL (A1:A10;B1:B10)` calcula r , el coeficiente de correlación de un conjunto de datos.

Además, Google Sheets cuenta con **herramientas de filtrado y ordenamiento**, que facilitan la visualización de subconjuntos específicos de datos. Por ejemplo, podemos filtrar las calificaciones de un grupo de estudiantes para ver sólo aquellos que hayan aprobado.

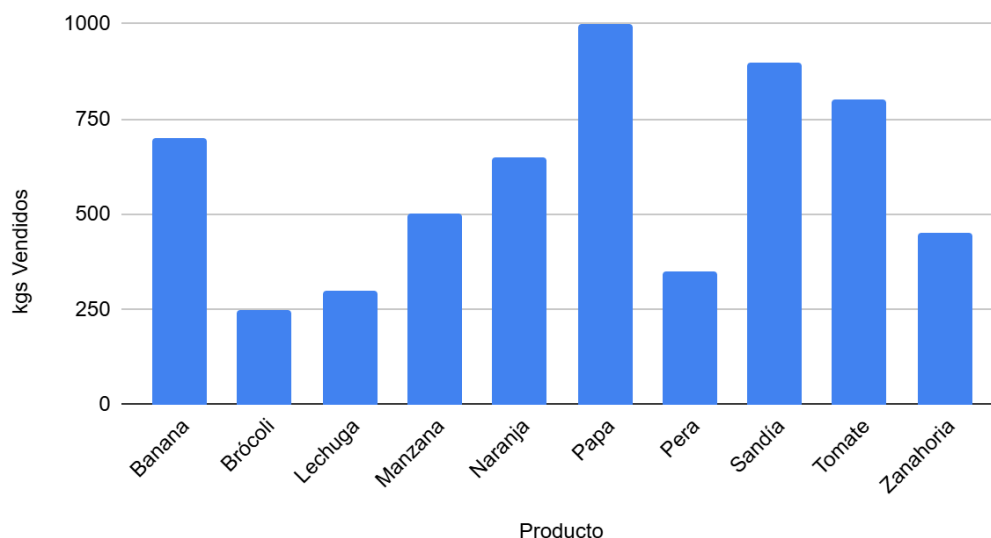
Visualización de Datos

Ventas					
Nombre	Categoría	#	kgs Vendidos	Mes	
Manzana	Frutas		500	Enero	
Lechuga	Verduras		300	Febrero	
Tomate	Verduras		800	Marzo	
Naranja	Frutas		650	Abril	
Zanahoria	Verduras		450	Enero	
Banana	Frutas		700	Febrero	
Brócoli	Verduras		250	Marzo	
Pera	Frutas		350	Abril	
Papa	Verduras		1000	Enero	
Sandía	Frutas		900	Febrero	

La visualización de datos es fundamental para entender y presentar de manera efectiva la información analizada. Los **gráficos más utilizados** en estadística descriptiva son:

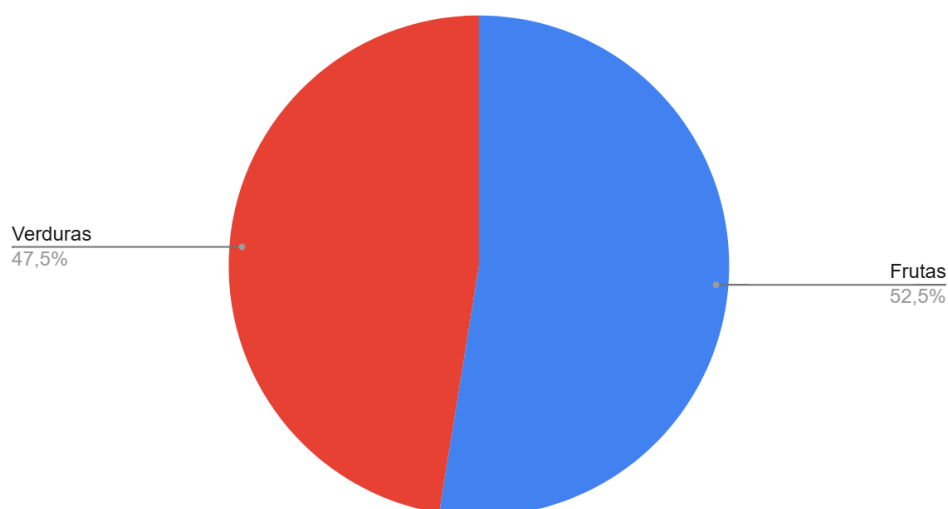
- Los **gráficos de barras** son ideales para comparar diferentes categorías. Por ejemplo, si deseamos comparar las ventas de distintos productos, un gráfico de barras nos mostrará claramente las diferencias entre ellos.

kgs Vendidos

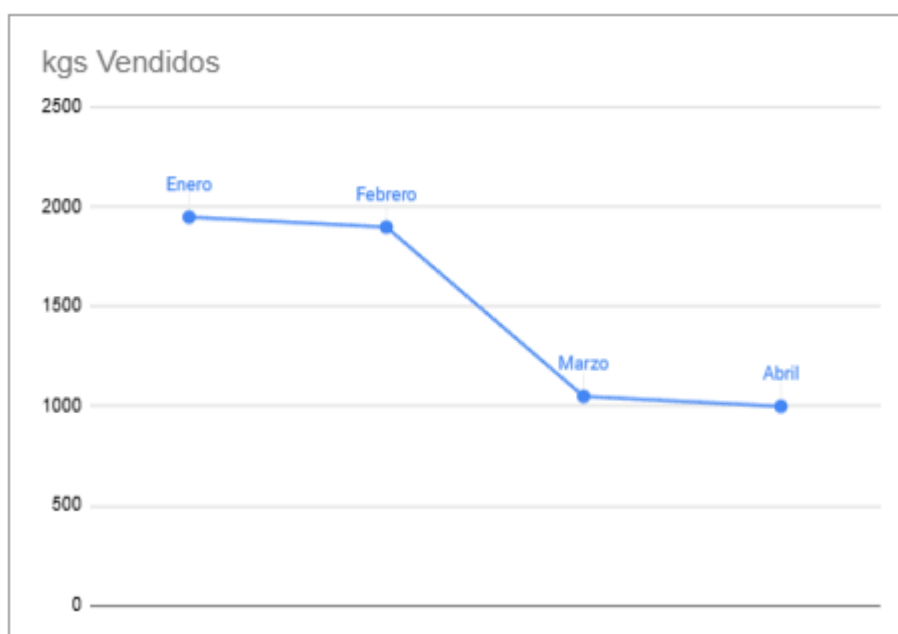


- Los **gráficos de sectores** (o gráficos de torta) son útiles para mostrar la proporción de cada categoría en relación con el total. Imaginemos que tenemos la distribución de las ventas de diferentes categorías de productos; un gráfico de sectores permitirá visualizar cómo se divide el total de ventas entre ellas.

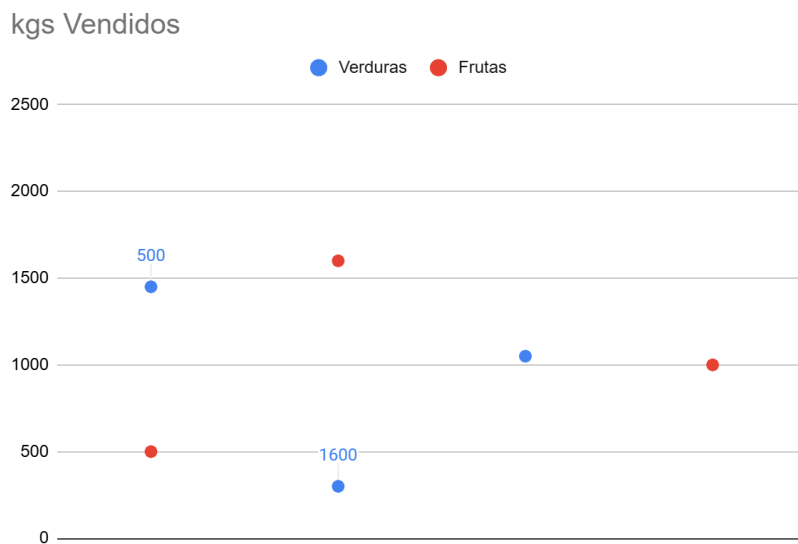
kgs Vendidos



- Los **gráficos de líneas** son ideales para representar cambios a lo largo del tiempo. Si queremos mostrar cómo evolucionaron las ventas mensuales de un producto a lo largo de un año, un gráfico de líneas será perfecto para visualizar esta información.



- Los **gráficos de correlación** muestran el nivel de relación que hay entre dos variables. Por ejemplo, si queremos ver si el mes está relacionado con la venta de alguna categoría de artículos.



Para crear gráficos en Google Sheets, basta con seleccionar los datos relevantes, ir a la opción "Insertar" en el menú y elegir el tipo de gráfico que se desee. Luego, podremos personalizarlo para que se ajuste a nuestras necesidades.

Reflexión final

La **estadística descriptiva** y la **visualización** son **herramientas de BI** que permiten analizar y comprender grandes volúmenes de datos de manera resumida. A través de medidas como la media, mediana y desviación estándar, los consultores pueden identificar patrones y tendencias clave que guían la toma de decisiones estratégicas. Además, la visualización de datos facilita la comunicación de insights, haciendo que la información sea accesible y comprensible para todos los involucrados.





Material y recursos adicionales

- [Guía de visualización de datos. Generalitat de Catalunya.](#)

Próximos pasos

- Creación de informes efectivos en Google Sheets. Dashboards. Formato condicional.
- Introducción a Looker Studio.
- Diseño y estructura de informes.

Ejercicios prácticos



DATAWISE CONSULTING
INSIGHTS ESTRATÉGICOS

Actividad 1: "Analizando Datos de Clientes"

Contexto



Silvia, la Project Manager del equipo de Datawise Consulting, ha recibido un conjunto de datos de clientes que contienen información sobre sus compras y preferencias. Tu tarea es analizar estos datos para identificar patrones que puedan ayudar a la empresa a personalizar ofertas.

Objetivos

- Aplicar medidas de tendencia central (media, mediana, moda) para resumir los datos de compras.
- Comprender cómo estos patrones pueden influir en las estrategias de marketing.

Consigna

1. Importar los datos de clientes a Google Sheets.
2. Utilizar las funciones de Google Sheets para calcular la media, mediana y moda de las compras.
3. Crear un gráfico de barras que visualice el número de compras por categoría de producto.
4. Presentar los hallazgos a Silvia, explicando cómo podrían mejorar la personalización de las ofertas.

Datos

- **Clientes.csv:** Estructura enfocada en las compras de los clientes.

¿Por qué importa esto en DataWise Consulting?

El análisis de los datos de clientes permite a la empresa diseñar estrategias de marketing más efectivas, mejorar la satisfacción del cliente y aumentar las ventas. Comprender la tendencia central de las compras ayuda a identificar qué productos son populares y cuáles no, guiando las decisiones comerciales.

Actividad 2: "Evaluando el Desempeño del Equipo"



Contexto

Luis, Analista de BI, necesita evaluar el desempeño del equipo de ventas de Datawise en el último trimestre. Tu rol es ayudarlo a analizar los datos de rendimiento de cada miembro del equipo utilizando medidas de dispersión.

Objetivos

- Entender la variabilidad de los resultados de ventas a través de medidas de dispersión.
- Aprender a presentar datos de forma clara y visual.

Consigna

1. Subir los datos a Google Sheets.
2. Calcular el rango, la desviación media y la desviación estándar de las ventas.
3. Crear un gráfico de líneas que muestre las ventas trimestrales de cada miembro del equipo.
4. Elaborar un breve informe para Luis sobre cómo la dispersión en los datos puede reflejar el desempeño y la cohesión del equipo.

Datos

- **Ventas del Equipo.csv:** Conjunto de datos para evaluar el rendimiento del equipo de ventas.

¿Por qué importa esto en DataWise Consulting?

Evaluar el desempeño del equipo con medidas de dispersión permite identificar áreas de mejora y reconocer a los mejores vendedores. Estas acciones son cruciales para fomentar la competitividad y la efectividad del equipo, impactando directamente en los resultados de la empresa.

Actividad 3: "Correlaciones entre Marketing y Ventas"



Contexto

Sabrina, la Consultora Estratégica, está interesada en entender si existe una relación entre el gasto en marketing y las ventas. Tu misión es calcular la correlación entre estos dos conjuntos de datos.

Objetivos

- Aplicar el coeficiente de correlación para determinar la relación entre dos variables.
- Aprender a interpretar los resultados y su relevancia.

Consigna

1. Reunir los datos de gasto en marketing y ventas en Google Sheets.
2. Calcular el coeficiente de correlación utilizando la función `COEF.DE.CORREL`.
3. Crear un gráfico de dispersión para visualizar la relación entre el gasto y las ventas.
4. Presentar a Sabrina tus conclusiones sobre la correlación y lo que podría significar para futuras estrategias de inversión.

Datos

- **Gastos en Marketing y Ventas.csv:** Datos que permiten calcular la correlación entre marketing y ventas.

¿Por qué importa esto en DataWise Consulting?

Comprender la relación entre el gasto en marketing y las ventas ayuda a tomar decisiones más informadas sobre dónde invertir los recursos. Un análisis de correlación efectivo puede guiar a la empresa en sus estrategias comerciales y maximizar el retorno de inversión.

Actividad 4: "Visualizando Tendencias del Mercado"



Contexto

Matias, el Desarrollador de BI, te ha solicitado que ayudes a visualizar las tendencias del mercado utilizando los datos del último año. La visualización efectiva de datos es clave para comunicar insights a los clientes.

Objetivos

- Aprender a crear diferentes tipos de gráficos en Google Sheets para representar tendencias.
- Desarrollar habilidades de comunicación utilizando datos.

Consigna

1. Importar un conjunto de datos del mercado que contenga información mensual sobre las ventas de diferentes productos.
2. Utilizar Google Sheets para crear gráficos de sectores y de líneas que muestren las tendencias de ventas.
3. Compartir tus gráficos y una breve explicación con el equipo sobre lo que indican las tendencias y cómo podrían influir en la toma de decisiones.

Datos

- **Ventas del Mercado.csv:** Estructura para visualizar tendencias de ventas en productos específicos a lo largo de los meses.

¿Por qué importa esto en DataWise Consulting?

Visualizar datos de manera efectiva permite que todos los stakeholders comprendan la información compleja de una manera accesible. Las tendencias del mercado, representadas gráficamente, pueden influir en decisiones estratégicas y dar una ventaja competitiva a la empresa.

⚠️ **Estos ejercicios son una simulación de cómo se podría resolver el problema en este contexto específico. Las soluciones encontradas no aplican de ninguna manera a todos los casos.**

Recuerda que las soluciones dependen de los sets de datos, el contexto y los requerimientos específicos de los stakeholders y las organizaciones.



Buenos Aires
aprende
Agencia de Políticas para el Futuro

BA Buenos
Aires
Ciudad