


# Exploración de datos



Antes de comenzar nuestro  
proceso de Data Analytics  
resultará muy útil **explorar**  
nuestros datos.



# Índice

- 01 ¿Qué es la exploración de datos?
- 02 Tipos de exploraciones
- 03 ¿Qué ventajas tiene la exploración de datos?
- 04 Conceptos importantes

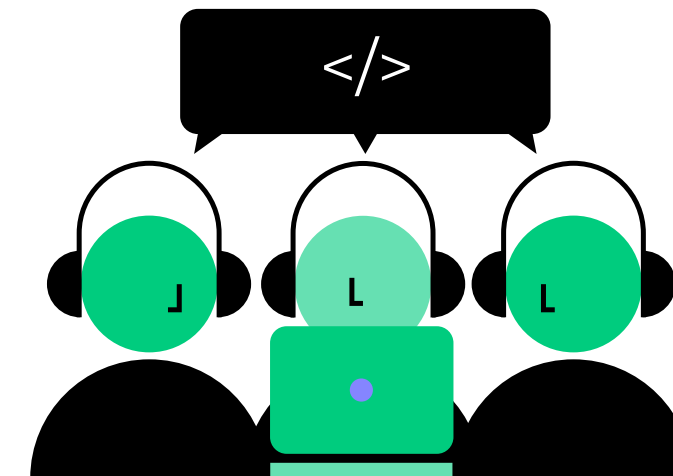



01

# ¿Qué es la exploración de datos?

# Exploración de datos

Es una investigación preliminar de los datos en búsqueda de tener un mejor entendimiento de las bases. Así podremos aplicar las transformaciones necesarias y luego sacar el mayor provecho posible a nuestro análisis.





Adicionalmente puede utilizarse para **analizar los resultados** de los análisis de datos.



También podés  
escucharlo como  
**EDA (Exploratory  
Data Analysis/  
*Análisis Exploratorio  
de Datos*).**



02

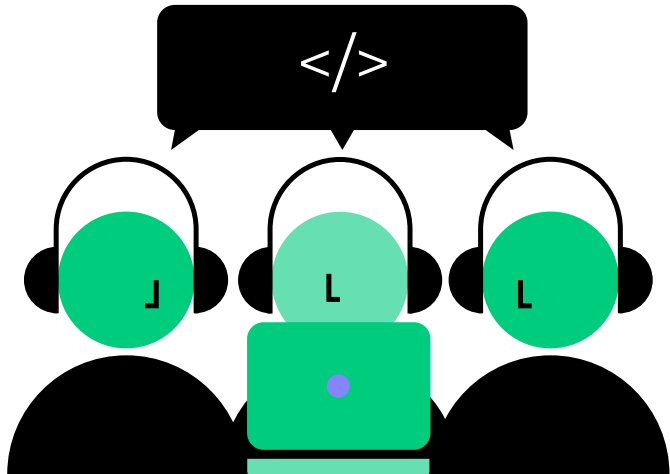
# Tipos de exploraciones



# Sumarización

Es un resumen de los datos a través de medidas específicas que permiten captar características de grandes volúmenes de datos.  
Por ejemplo, una tabla con la edad promedio de los habitantes por ciudad, veamos:

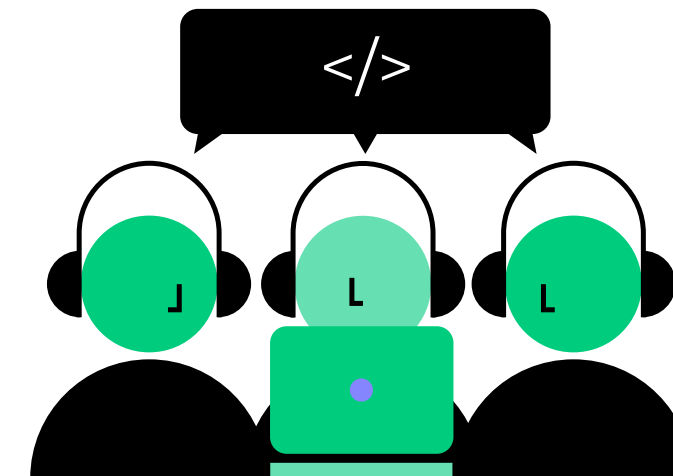
Ciudad	Edad promedio
Buenos Aires	35
Córdoba	42
Mendoza	38



# Visualizaciones

Es la forma de mostrar datos de una forma gráfica o en tablas.

Resulta muy útil para la interpretación de grandes volúmenes de datos por cualquier receptor.



03

¿Qué ventajas tiene la exploración de datos?

# Las visualizaciones son de gran ayuda



Podrás utilizarla para los procesos de Data Manipulation y para analizar los resultados de un análisis.

No reemplaza técnicas de Data Analytics, pero sí contribuirá a la calidad de tus análisis.



Hace que la interpretación de los datos sea **accesible** para cualquier persona involucrada.



Y nos ayudará a alcanzar  
conclusiones **rápidas** y  
**precisas.**



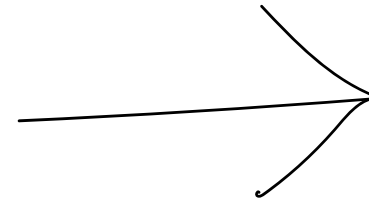
04

# Conceptos importantes

# Frecuencia

Es la cantidad de ocurrencias dentro de la base. Es decir, cuántas veces se repite un determinado valor en la base.

Ciudad	Nacimientos
Buenos Aires	100
Córdoba	120
Santa Fe	100
Mendoza	100
Jujuy	200
Salta	300



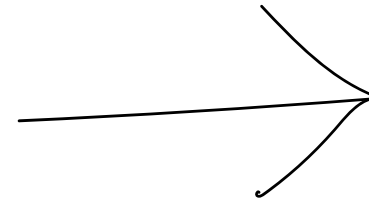
Nacimientos	Frecuencia
100	3
120	1
200	1
300	1



# Moda

Es el valor que más se repite en nuestra base, será aquel que tenga mayor frecuencia.

Ciudad	Nacimientos
Buenos Aires	100
Córdoba	120
Santa Fe	100
Mendoza	100
Jujuy	200
Salta	300

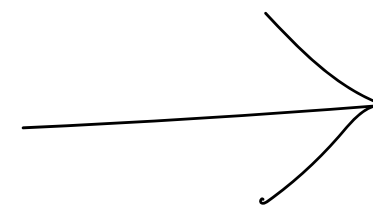


En este caso la moda será el valor de **100**, ya que se repite tres veces.

# Promedio

Es un número representativo que busca ilustrar los valores de una base. También podrán escucharlo como Media o Media aritmética.

Ciudad	Nacimientos
Buenos Aires	100
Córdoba	120
Santa Fe	100
Mendoza	100
Jujuy	200
Salta	300

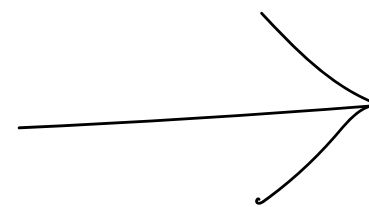


En este caso el promedio será el valor de **153** y surge de la suma de todos los valores dividido la cantidad de registros.

# Percentil

Es el valor máximo de un determinado porcentaje de la base. Nos dirá hasta qué valor está comprendido un X % de la base de datos.

Ciudad	Nacimientos
Buenos Aires	100
Córdoba	120
Santa Fe	100
Mendoza	100
Jujuy	200
Salta	300

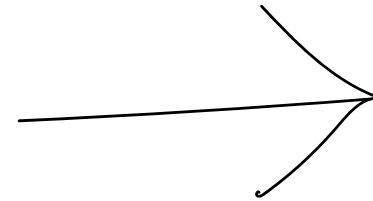


En este caso el percentil 85 será el valor de **225** ya que el 85 % de los datos (5 de 6) tienen un valor inferior a 225.

# Media

Es el valor del medio de nuestra base, en otras palabras es el percentil 50.

Ciudad	Nacimientos
Buenos Aires	100
Córdoba	120
Santa Fe	100
Mendoza	100
Jujuy	200
Salta	300



En este caso la media será el valor de **110** ya que el 50 % de los datos (3 de 6) tienen un valor inferior a 110.

¡Muchas gracias!