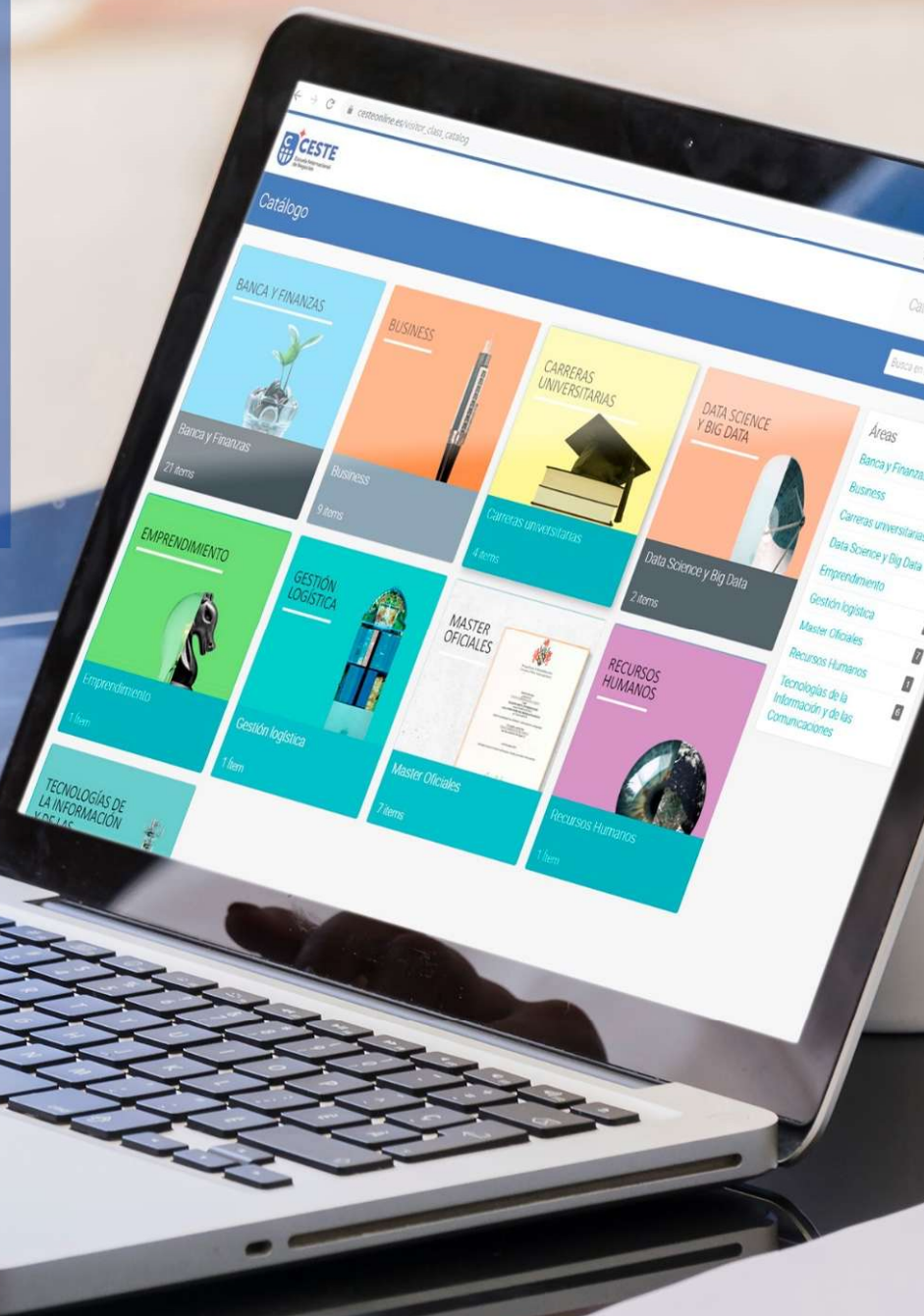


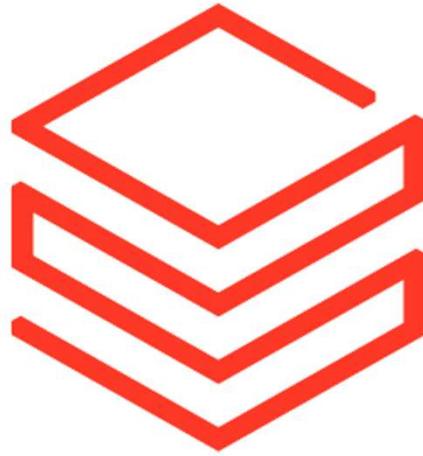
# CESTE

## Escuela Internacional de Negocios

Zaragoza (España)



## Análisis Exploratorio e Introducción a Unity Catalog



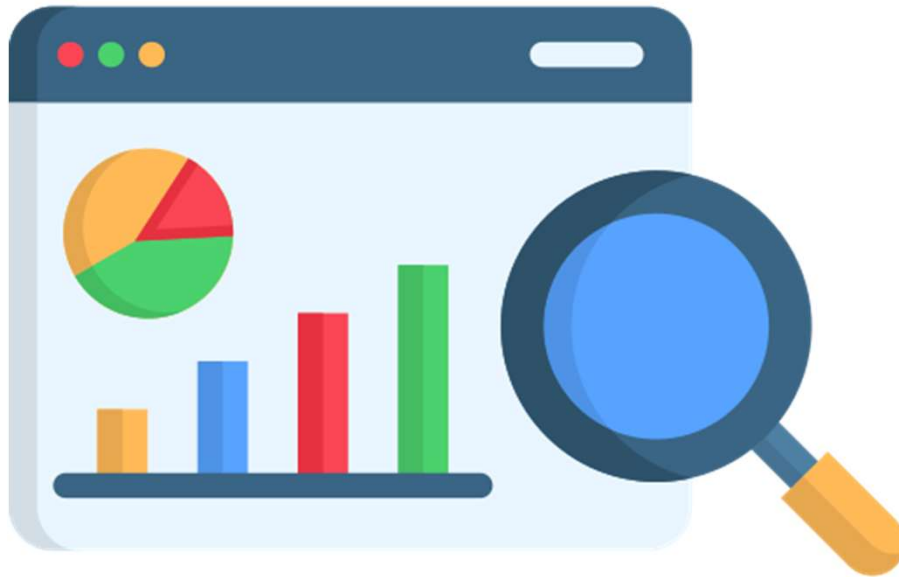
# databricks

## *Objetivos de la sesión*



- Exploración de datos con Spark y SQL (EDA)
- Construcción de Dashboards en Databricks
- Actividad práctica de dashboard
- Unity Catalog

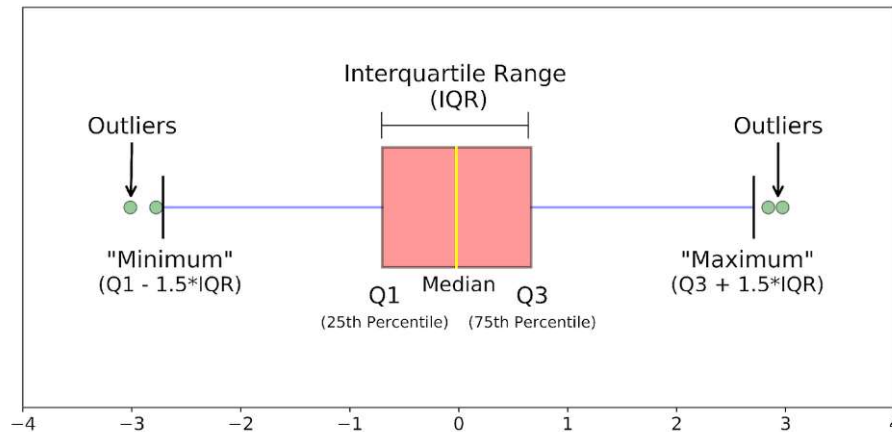
## *¿Qué es el análisis exploratorio?*



- Investigar características
- Ajuste de características.
- Detección e Identificación de valores atípicos (outliers)
- Investigación de la correlación.

# Detección de Outliers (I): Métodos numéricos

## Regla de IQR



## Z-Score

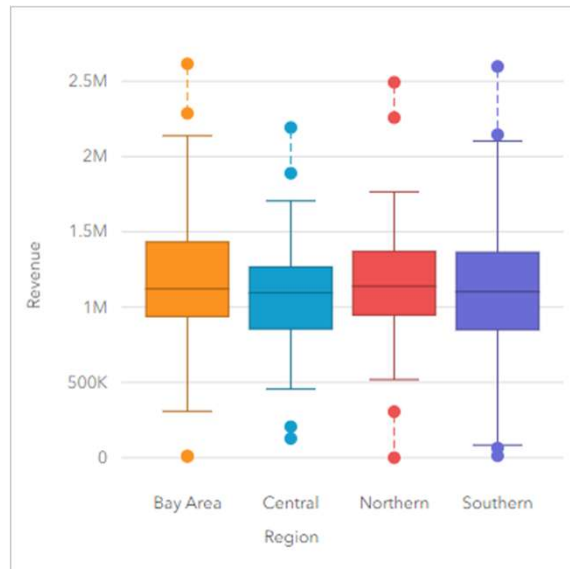
$$z = \frac{x - \mu}{\sigma}$$

$\mu$  = Mean

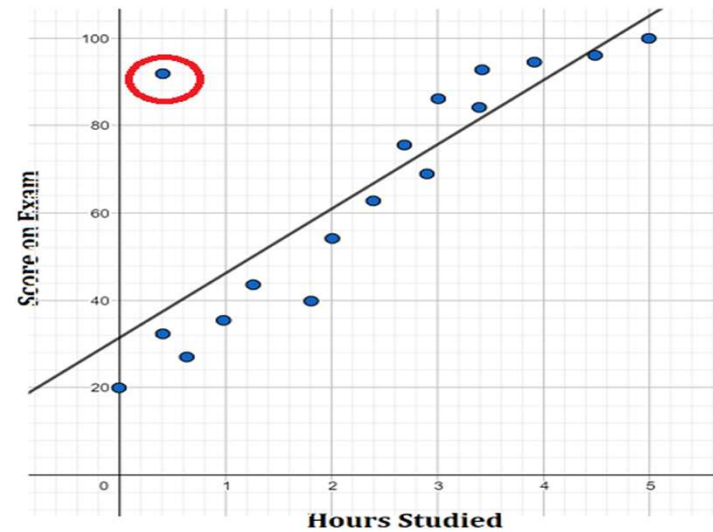
$\sigma$  = Standard Deviation

# *Detección de Outliers (II): Métodos Visuales*

Boxplot



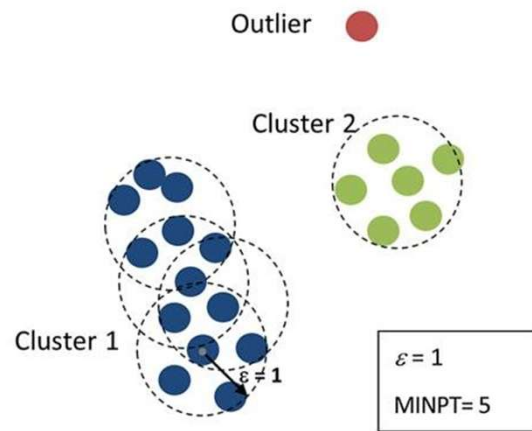
ScatterPlot



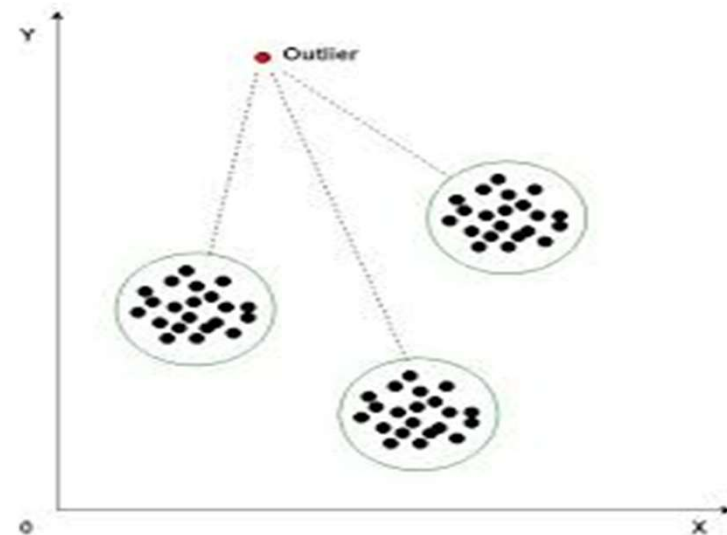


# *Detección de Outliers (III): Métodos no supervisados*

DBSCAN



KMeans



## ***¿Cómo tratar outliers?***



Eliminar outliers



Imputación de Outliers



Investigar más afondo valores



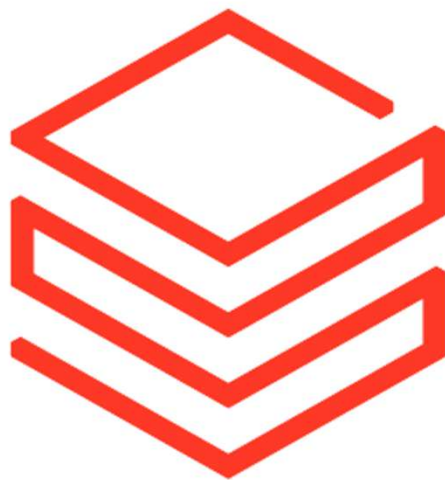
Marcarlos sin modificar dataset



Comparar resultados con/sin outliers



## *Práctica con Notebook*



# **databricks**

# ***Preguntas / Sugerencias***



[www.ceste.es](http://www.ceste.es)