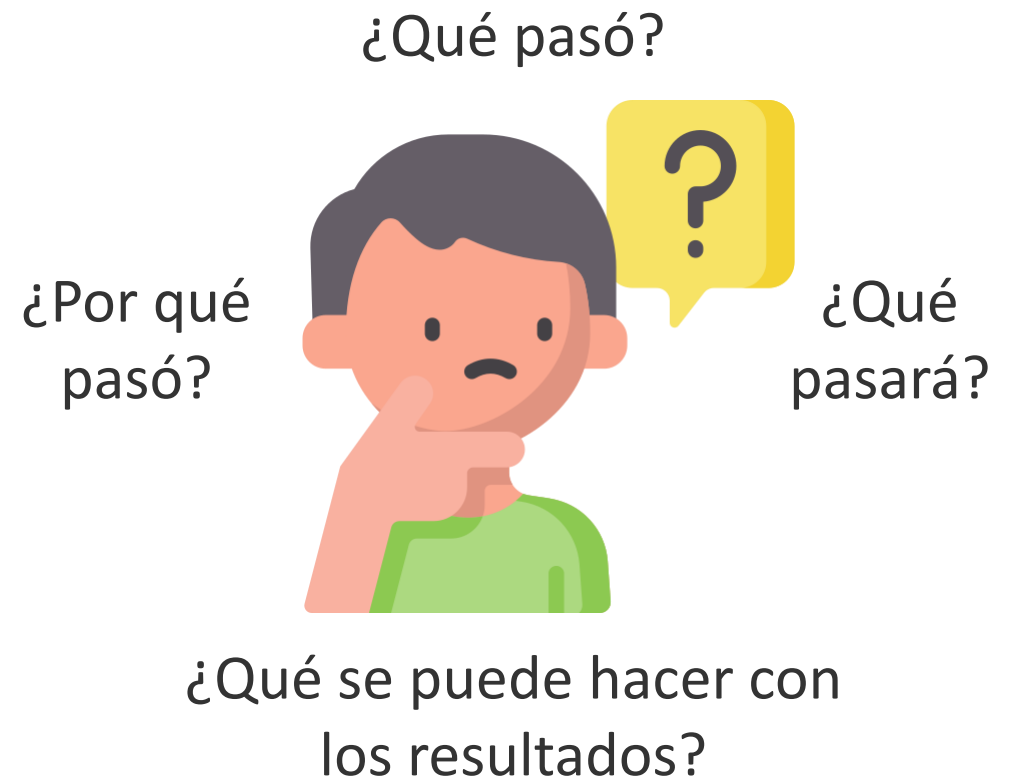


# Introducción a la Ciencia de Datos

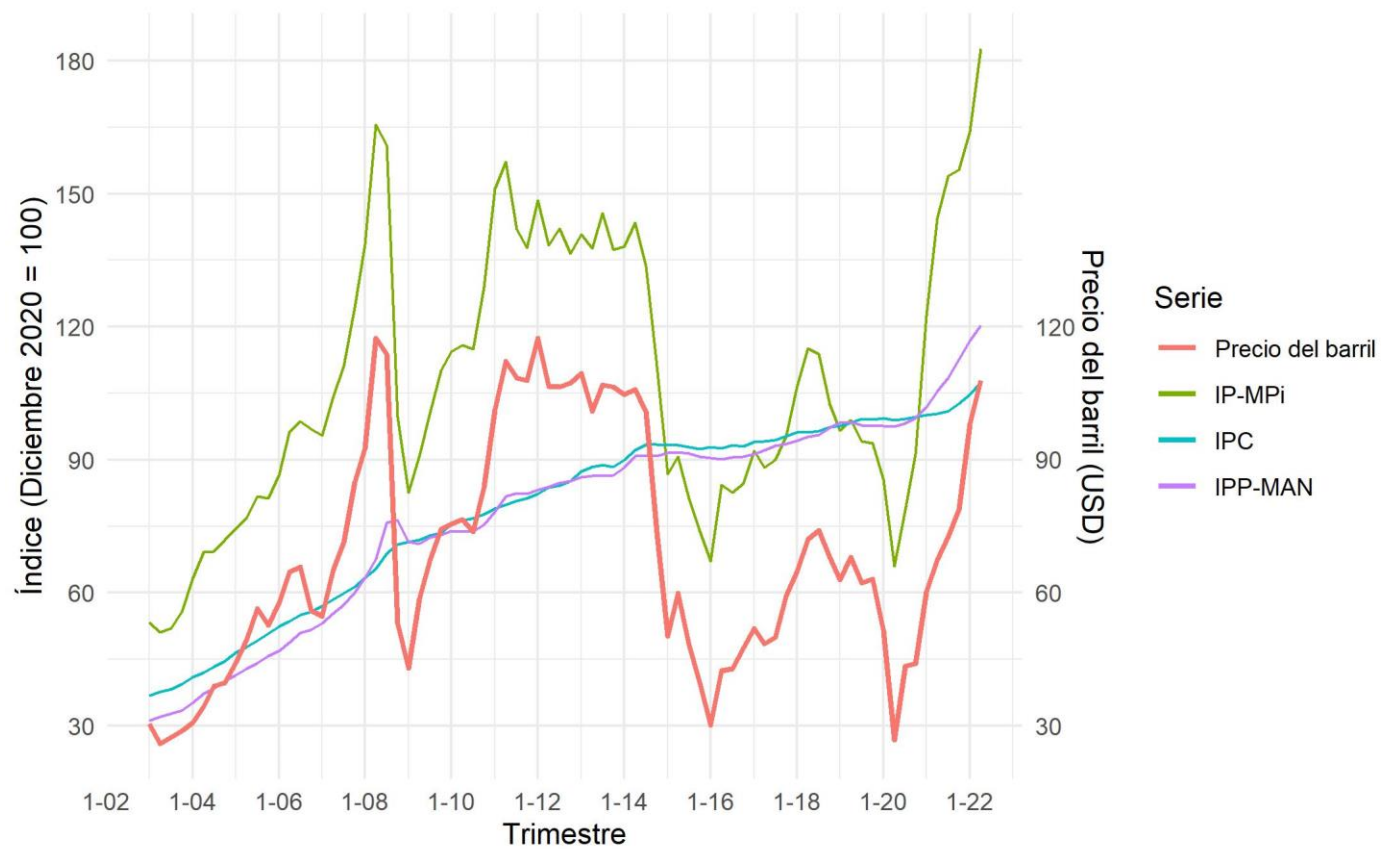


# Ciencia de datos

- Estudio de **datos** con el fin de **extraer información** significativa para empresas.
- Es un **enfoque multidisciplinario** que combina principios y prácticas del campo de las *matemáticas*, la *estadística*, la *inteligencia artificial* y la *ingeniería de computación* para analizar grandes cantidades de datos.



## Costa Rica: Índices de precios y precio internacional del petróleo por trimestre, 2002-2022



Fuente: BCCR (2022) y OPEP (2022).



# Historia



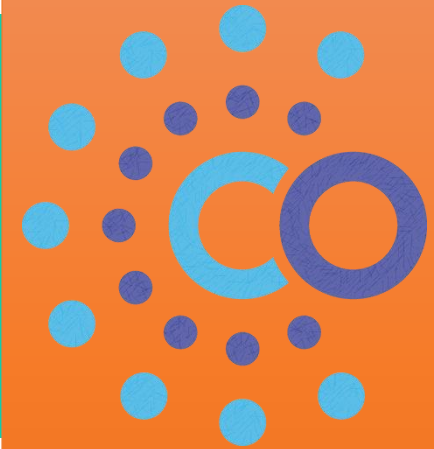
John Tukey  
*1963*



Peter Naur  
*1974*



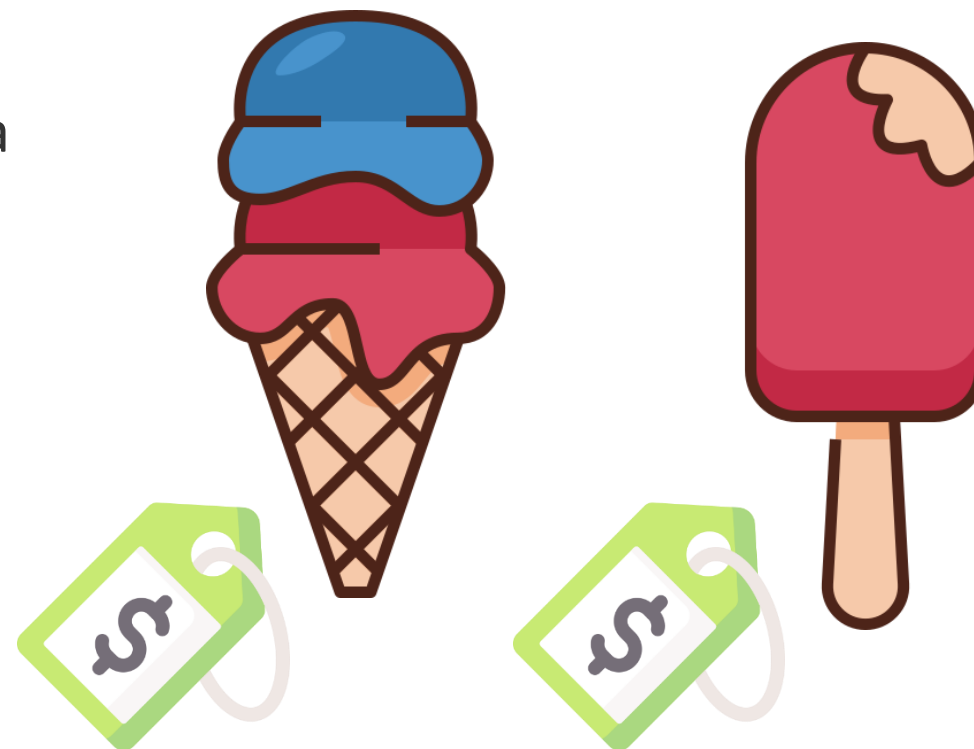
IFCS  
*1996*



Data Science  
Journal  
*2002*

# Dato

- Conjunto discreto, de factores objetivos sobre un hecho real.
- Todas las organizaciones necesitan datos. ¿Para qué?
  - Bancos.
  - Compañías de seguros.
  - Agencias gubernamentales.
  - Pulpería.
- ¿Hay que almacenar todos los datos?
- Los datos son únicamente descriptivos.



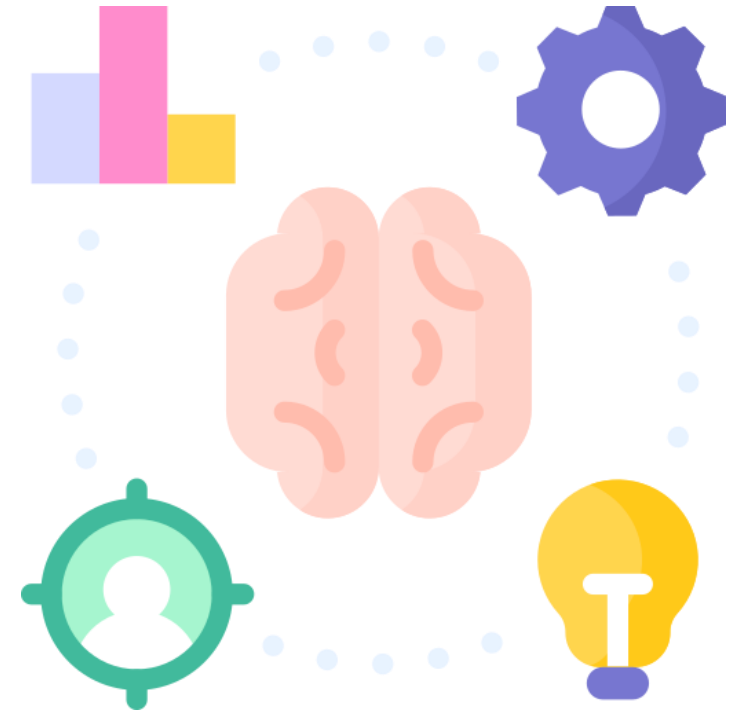
# Información

- Mensaje.
- Informar = dar forma.
- Emisor y receptor.
- Tiene significado (relevancia y propósito).
- Transformamos los datos:
  - Contextualizando.
  - Categorizando.
  - Calculando.
  - Corrigiendo.
  - Condensando.



# Conocimiento

**Davenport y Prusak (1999):** el conocimiento es una mezcla de experiencia, valores, información y “saber hacer” que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción.



# Usos

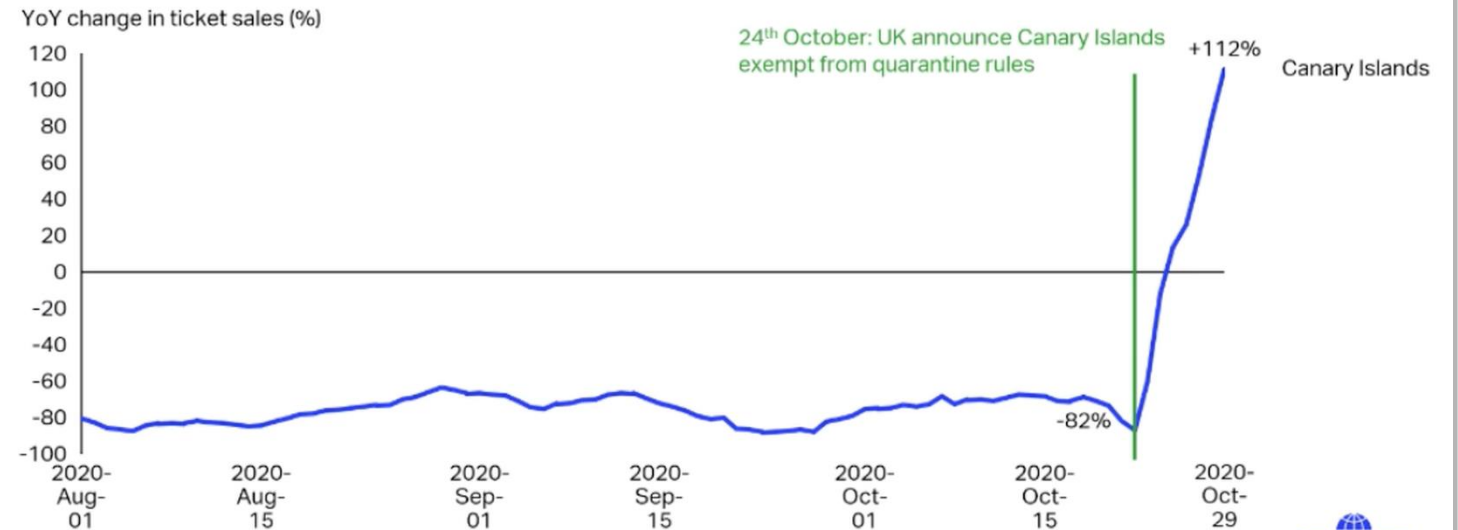
## 1. Análisis descriptivo



## 2. Análisis de diagnóstico

### Clear signs of willingness to fly within Europe Surge in October bookings once UK drops quarantine for Canary Islands

YoY change in ticket sales between the UK and the Canary Islands, Aug-Oct 2020



IATA Economics analysis based on data from DDS





### 3. Análisis predictivo

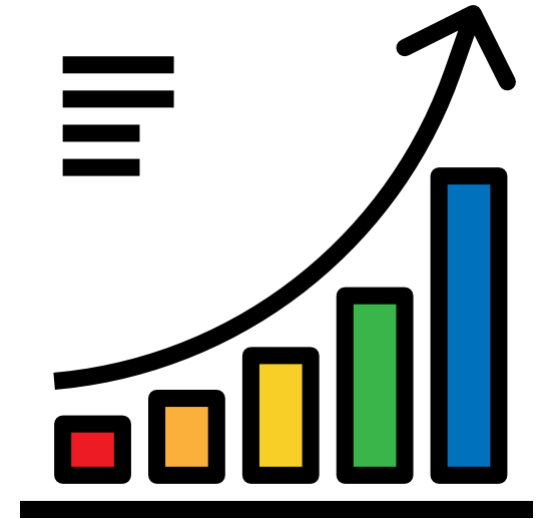
Job information   **Results**   JSON   Execution details

Row	start_station_name	forecast_timestamp	forecast_value	standard_error	confidence_level	prediction_interval_lower_bound	prediction_interval_upper_bound
1	Central Park North & Adam Clayton Powell Blvd	2016-10-01 00:00:00 UTC	248.92639037485574	28.976986939548496	0.9	201.31513840709755	296.5376423
2	Central Park North & Adam Clayton Powell Blvd	2016-10-02 00:00:00 UTC	363.68298264818185	28.950407995433263	0.9	316.1154017773513	411.250563
3	Central Park North & Adam Clayton Powell Blvd	2016-10-03 00:00:00 UTC	198.88838055443952	37.80946364266388	0.9	136.7647407152758	261.012020
4	Central Park S & 6 Ave	2016-10-01 00:00:00 UTC	315.0966902597872	33.08129588892611	0.9	260.74176656556523	369.4516139
5	Central Park S & 6 Ave	2016-10-02 00:00:00 UTC	372.2015822785193	34.26247838317384	0.9	315.9058916760313	428.4972728
6	Central Park S & 6 Ave	2016-10-03 00:00:00 UTC	171.05749819000278	34.397509635586104	0.9	114.53994162900312	227.5750547
7	Central Park W & W 96 St	2016-10-01 00:00:00 UTC	88.2794448473145	25.851071107600923	0.9	45.80429507732261	130.754594
8	Central Park W & W 96 St	2016-10-02 00:00:00 UTC	142.8865570581737	30.510300562418326	0.9	92.75596167502351	193.017152
9	Central Park W & W 96 St	2016-10-03 00:00:00 UTC	61.79444762054615	31.775255511796885	0.9	9.585441222877144	114.0034540
10	Central Park West & W 100 St	2016-10-01 00:00:00 UTC	67.95383241302838	17.786042545729188	0.9	38.73009859554769	97.1775662
11	Central Park West & W 100 St	2016-10-02 00:00:00 UTC	59.71885135589996	17.786042545729188	0.9	30.49511753841927	88.9425851
12	Central Park West & W 100 St	2016-10-03 00:00:00 UTC	40.904989813854826	17.786042545729188	0.9	11.681255996374137	70.1287236
13	Central Park West & W 102 St	2016-10-01 00:00:00 UTC	51.953703703703695	16.088375294476204	0.9	25.519358010855836	78.3880493
14	Central Park West & W 102 St	2016-10-02 00:00:00 UTC	51.953703703703695	16.088375294476204	0.9	25.519358010855836	78.3880493
15	Central Park West & W 102 St	2016-10-03 00:00:00 UTC	51.953703703703695	16.088375294476204	0.9	25.519358010855836	78.3880493

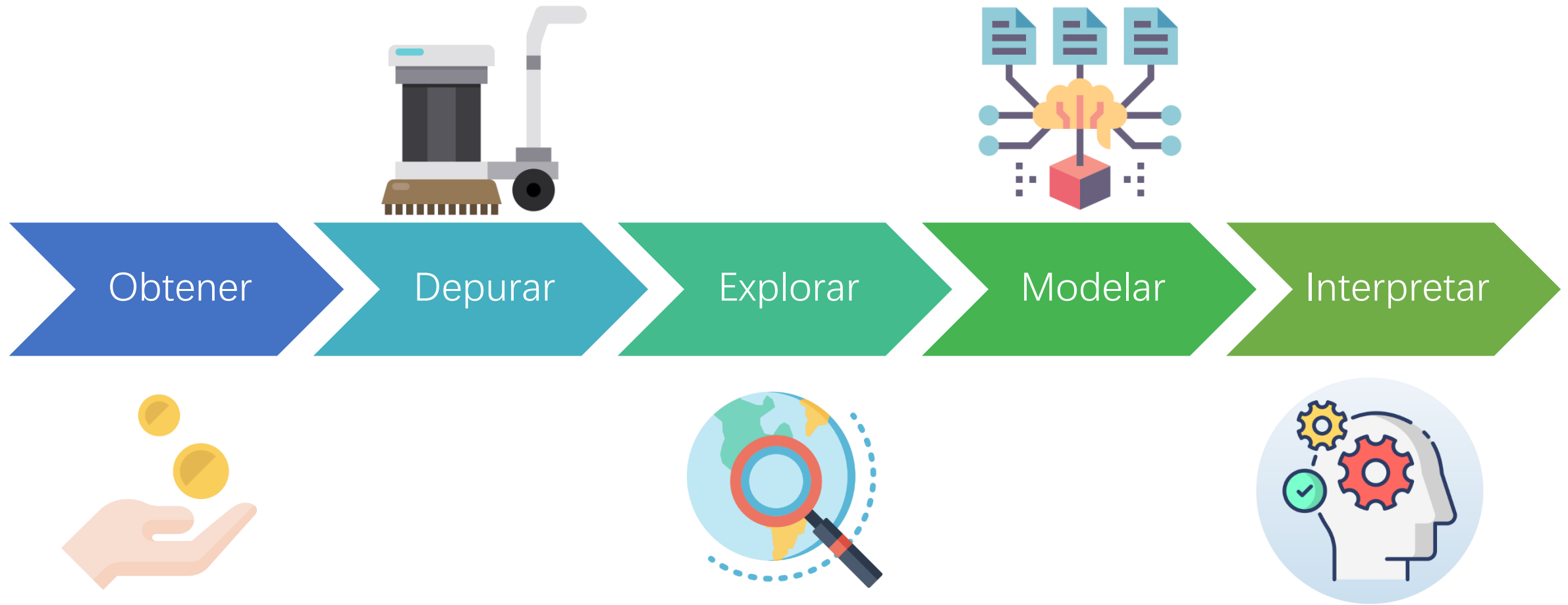
Rows per page: 100 1 - 36 of 36 First page < < > > Last page

# Usos

## 4. Análisis prospectivo

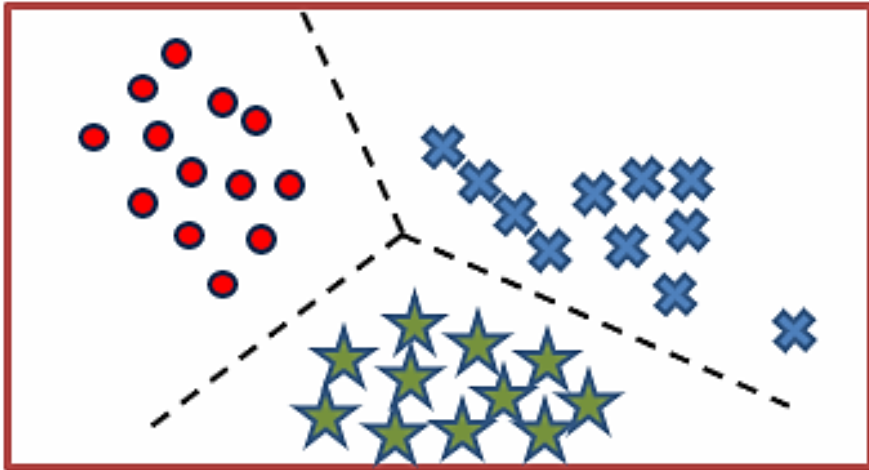


# Proceso



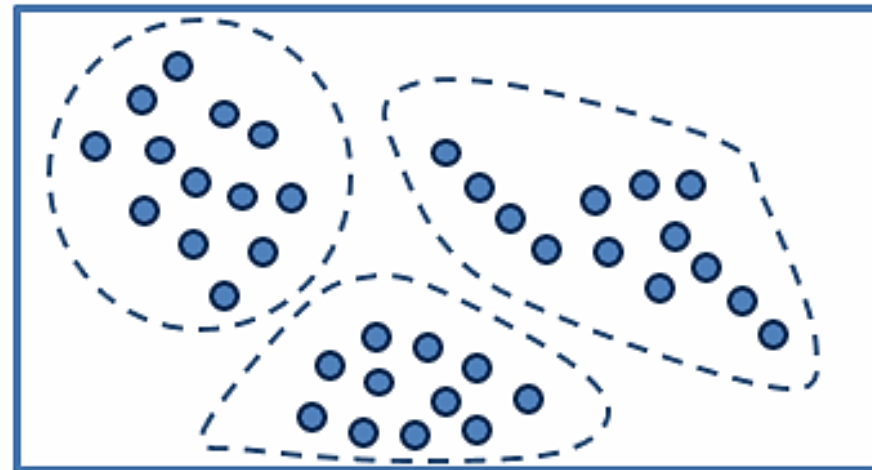
# Técnicas: clasificación

Classification



Supervised learning

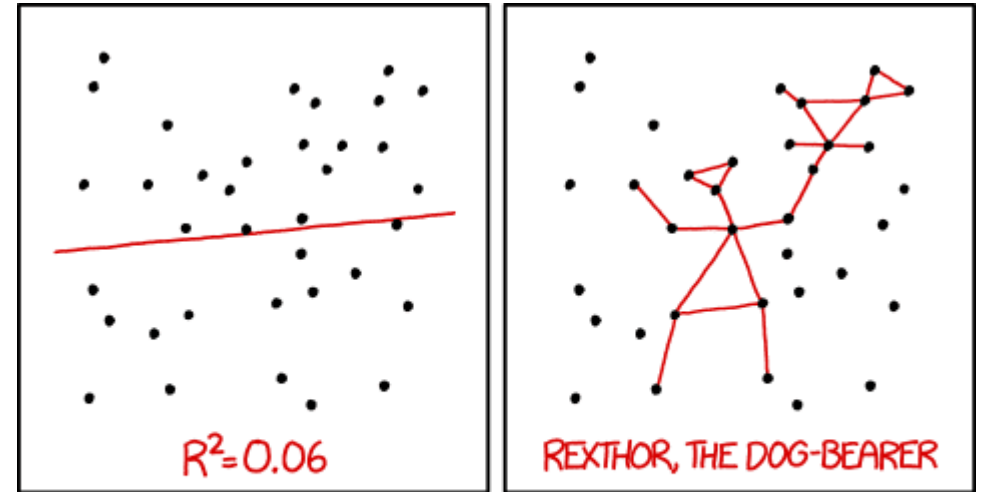
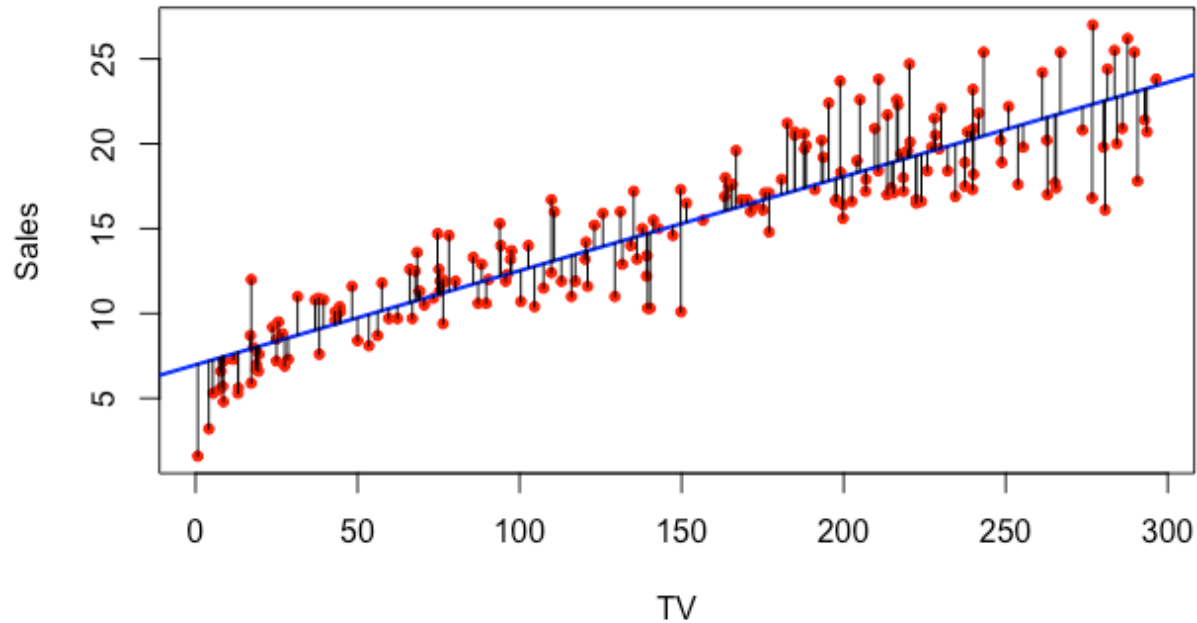
Clustering



Unsupervised learning

# Técnicas: regresión

Ventas vs. gasto en TV



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

# Big Data



# Retos



Obtener valor



Integrar código abierto



Entender el problema de la empresa



Aspectos éticos



Varios orígenes de datos



Rendimiento



Eliminación del sesgo

# Referencias

- Amazon. (s.f.). ¿Qué es la ciencia de datos? <https://aws.amazon.com/es/what-is/data-science/>
- Carrión, J. (s.f.) Diferencia entre dato, información y conocimiento. <https://iibi.unam.mx/voutssasmt/documentos/dato%20informacion%20conocimiento.pdf>
- IT Digital Media Group (2020). Los 4 desafíos a los que se enfrenta la ciencia de datos en 2020. <https://discoverthenew.ituser.es/predictive-analytics/2020/07/los-4-desafios-a-los-que-se-enfrenta-la-ciencia-de-datos-en-2020>
- Santamaría Muñoz, R. (2021). *El origen y evolución de la ciencia de datos*. <https://isdfundacion.org/2021/07/02/el-origen-y-evolucion-de-la-ciencia-de-datos-data-science/>
- Universidad Pontificia Católica de Chile. ¿Qué es la ciencia de datos? <https://datascience.uc.cl/que-es-ciencia-de-datos>