

## Tecnologías Business Intelligence (ORACLE)

Práctica I 24/01/2024, fecha de entrega máxima 12/02/2024

Lee el contenido del punto 5 del enlace: [Create Data Flows in Oracle Analytics](#) donde se documenta el uso de una herramienta moderna de ETL en concepto de autoservicio.

[https://docs.oracle.com/middleware/bidvhelp/desktop/es\\_WW/BIDVD/BIDVD.pdf](https://docs.oracle.com/middleware/bidvhelp/desktop/es_WW/BIDVD/BIDVD.pdf)

Visualiza enlace: [Self-service data modeling for business users](#) en donde encontrarás un ejemplo completo de los conceptos vistos aplicados a la plataforma tecnología del fabricante Oracle

<https://www.youtube.com/watch?v=8fvSLV1LXn0>

Practica en OAD con los ejemplos del tutorial: [Explore Data Flows in Oracle Analytics](#)

[https://docs.oracle.com/en/cloud/paas/analytics-cloud/tutorial-explore-data-flows/#before\\_you\\_begin](https://docs.oracle.com/en/cloud/paas/analytics-cloud/tutorial-explore-data-flows/#before_you_begin)

Repite y practica en Oracle Analytics Desktops (OAD) el ejemplo de flujo del enlace: [Create Data Flows in Oracle Analytics](#)

<https://www.youtube.com/watch?v=ig5Lbx6sn5U>

**Elabora el entregable: *Self Service Data Preparation*, en el que te enfrentarás a un problema de preparación de datos para un análisis posterior. El documento debe justificar en todo momento que se pretende hacer, como se hace y que se obtiene. Incorpora todas las capturas de pantalla necesarias para explicar y justificar el trabajo realizado.**

### Preparación de los datos:

**Aprovecha los diferentes elementos que proporciona Oracle Analytics Desktop (OAD) que incluya como mínimo dos ejemplos de corrección de problemas que puedan tener los datos, adecuación de datos para un análisis posterior y generación de una nueva variable indicador a partir de las existentes en los datos.**

**Será valorable la resolución de otros tipos de problemáticas en función de los datos escogidos.**

**Los datos para preparar son de libre elección. Deben ser como mínimo dos ficheros que deberán importarse y relacionarse. Es valorable la diversidad de tipologías de variables en los datos: carácter, numérico, fecha/hora, lógicos, cartográficos... Si el conjunto de registros es muy elevado se puede utilizar un subconjunto de datos escogiendo una muestra.**

Conocer el dominio de los datos ayuda a que el trabajo sea más provechoso e interesante.

Posibles fuentes de datos, puedes escoger de otras fuentes si lo deseas.

Datos abiertos

Google Dataset Search

<https://datasetsearch.research.google.com/>

Datos abiertos España

[https://datos.gob.es/es/catalogo?q=&frequency=%7B%22type%22%3A+%22months%22%2C+%22value%22%3A+%221%22%7D&sort=score+desc%2C+metadata\\_modified+desc](https://datos.gob.es/es/catalogo?q=&frequency=%7B%22type%22%3A+%22months%22%2C+%22value%22%3A+%221%22%7D&sort=score+desc%2C+metadata_modified+desc)

Datos abiertos Madrid

<https://datos.madrid.es/portal/site/egob/>

Datos abiertos Barcelona

<https://opendata-ajuntament.barcelona.cat/es/>

Datos abiertos Londres

<https://data.london.gov.uk/>

Datos abiertos New York

<https://opendata.cityofnewyork.us/>

Conjuntos de datos para aprendizaje automático e investigación

UCI Machine Learning

<https://archive.ics.uci.edu/ml/datasets.php>

Datasets for machine-learning research (Wikipedia)

[https://en.wikipedia.org/wiki/List\\_of\\_datasets\\_for\\_machine-learning\\_research](https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research)

Kaggle datasets

<https://www.kaggle.com/datasets>