

Multidimensional modeling should be applied when...

a) We have a relational database that we want to structure

b) We have a linked open data source that we want to analyze

c) We want to organize and create a data model to facilitate our analysis and possibly create an analytic repository

Un modelado multidimensional busca organizar los datos de cierta forma que nos facilite la consulta de ciertas métricas. Este modelado multidimensional puede hacerse desde una base de datos existente por lo que descartamos la opción b y la opción a también la descartamos porque una base de datos relacional está estructurada técnicamente, solo que no está bien estructurada para ciertos análisis o consultas ya que serían demasiado complejas.

What is the main goal of an infographic?

a) Help us understand data better during the analysis

b) Communicate conclusions of an analysis towards a wide audience that does not necessarily have technical or domain knowledge

c) Monitoring business operations, because managers understand better the information if it is presented as an infographic

Vimos un ejemplo en clase que mostraba la deuda estadounidense con billetes de un dólar, esto ayudaba a gente no técnica a entender la magnitud de la deuda a simple visto. Suelen ser gráficos que una persona no técnica o con conocimiento del dominio puede entender, porque la respuesta correcta la b.

From a data science point of view, what are we interested most in from Machine Learning models?

a) Their accuracy, the better performant the model is, the better results we will get no matter how complex it is

b) Their inner logic, weights assigned to attributes and rules followed to obtain the prediction, because they aid us to extract insights

c) The simplicity of the training, we need fast models that are trained quickly in order to be efficient in our data science process

Como vimos la precisión de un modelo no lo es todo. Vimos por ejemplo en clase un modelo que era un clasificador sobre si una persona vive o muere por covid. Tenía una precisión de un 97%. Solo el 3% de los que padecían covid fallecen, por lo que el modelo podía tener una gran precisión diciendo a todos los casos que no. Por lo que tiene sentido que no nos importe tanto la precisión sin importar la lógica interna, entonces descaramos la opción a. Además también vimos que si no definíamos lógicas mas complejas de forma interna podíamos tener problemas como por ejemplo con técnicas de clustering, por lo que la respuesta correcta es la b.