

Realizado por:
Juan Camilo Restrepo Velez
William Leonardo Andrade Collazos
Wilder Valencia Ocampo

PRÁCTICA DE BODEGAS DE DATOS 10%

Bank Marketing

Los datos están relacionados con campañas de marketing directo de una institución bancaria portuguesa. Las campañas de marketing se basaron en llamadas telefónicas. A menudo, se requería más de un contacto con el mismo cliente, para poder acceder a si el producto (depósito bancario a plazo) sería ('sí') o no ('no') suscrito.

Información de atributos

Información Bancaria de los clientes

Age - Edad

Job - Trabajo: tipo de trabajo

Marital - Estado civil: estado civil

Education - Educación: Nivel educativo

Default - Incumplimiento: ¿tiene el crédito en mora?

Housing - Vivienda: ¿tiene un préstamo de vivienda?

Loan - Préstamo: ¿tiene préstamo personal?

Relacionado con la última llamada de la actual campaña

Contact - Contacto: tipo de comunicación

Month - Mes: último mes de contacto del año

DayofWeek - Día de la semana: último día de contacto de la semana

Duration - Duración: duración del último contacto, en segundos (numérico). Nota importante: este atributo afecta en gran medida al objetivo de salida (por ejemplo, si la duración = 0, entonces y = "no"). Sin embargo, no se conoce la duración antes de una llamada se realiza. Además, después del final de la llamada se conoce obviamente y. Por lo tanto, esta entrada sólo debe incluirse a efectos de referencia y debe descartarse si se pretende tener un modelo predictivo realista.

Otros

Campaign - Campaña: número de contactos realizados durante esta campaña y para este cliente

Pdays - pDías: número de días que pasaron después de que el cliente fue contactado por última vez en una campaña anterior. Nota, 999 significa que el cliente no fue contactado anteriormente

Previous - Anterior: número de contactos realizados antes de esta campaña y para este cliente

Poutcome: resultado de la anterior campaña de marketing

Atributos del contexto social y económico

Emp.var.rate - Tasa de variación del empleo - indicador trimestral

Cons.price.idx: Índice de Precios al Consumidor - Indicador mensual; el Índice de Precios al Consumidor o IPC mide los cambios en los precios pagados por los consumidores por una cesta de bienes y servicios cada mes.

Cons.conf.idx: Índice de confianza del consumidor - Indicador mensual; En Portugal, el índice de confianza del consumidor se basa en entrevistas con los consumidores sobre sus percepciones de la situación económica actual y futura del país y sus tendencias de compra. Se estima utilizando la diferencia entre la proporción de respuestas de evaluación positivas y las respuestas de evaluación negativas, pero no incluye la proporción de respuestas neutras

Euribor3m: euribor 3 meses - Euribor es la abreviatura de Euro Interbank Offered Rate. es un índice de referencia publicado diariamente que indica el tipo de interés promedio al que un gran número de bancos europeos dicen concederse préstamos a corto plazo entre ellos para prestárselo a terceros.

Nr.employed - Número de empleados: Número de empleados - Indicador trimestral; Número de personas empleadas para el trimestre.

Variable objetivo

y - ¿el cliente ha suscrito un depósito a plazo?


*Tomado de <https://www.kaggle.com/henriqueyamahata/bank-marketing>

INSTRUCCIONES

1. Determinar los datos (hechos) que se convertirán en el objetivo de análisis del negocio. Una ayuda es buscar los datos susceptibles de cambiar constantemente. Diseñar una tabla de la base de datos con dicha información, la llamaremos la tabla de hechos.

Hechos

- Contact
- Month
- DayofWeek
- Duration
- Campaign
- Pdays
- Previous
- Poutcome

Hechos		Cliente	Social	Bancaria
	Nombre del campo	Tipo de datos		
	Id_Hechos	Autonumeración		
	Id_Cliente	Número		
	Id_Social	Número		
	Id_Bancaria	Número		
	Contact	Texto corto		
	Month	Texto corto		
	DayofWeek	Texto corto		
	Duration	Número		
	Campaign	Número		
	Pdays	Número		
	Previous	Número		
	Poutcome	Texto corto		

2. Determinar la periodicidad de los hechos, incluir la periodicidad en la tabla de hechos (opcionalmente puede ser una dimensión).
La periodicidad de los hechos es de **un mes** esto se debe a que esta es la duración de cada una de las campañas con la que despliega la entidad bancaria portuguesa.
3. Determinar las entidades que influyen en los hechos. Cada entidad es una nueva tabla que llamaremos dimensiones.

Dimensiones

- **Cliente** (age, job, marital, education)

Hechos		Cliente		Social		Bancaria	
		Nombre del campo		Tipo de datos			
		Id_Cliente		Autonumeración			
		Age		Número			
		Job		Texto corto			
		Marital		Texto corto			
		Education		Texto corto			

- **Social** (Emp.var.rate, Cons.price.idx, Cons.conf.idx, Euribor3m, Nr.employed)

Hechos		Cliente		Social		Bancaria	
		Nombre del campo		Tipo de datos			
		Id_Social		Autonumeración			
		Emp_var_rate		Número			
		Cons_price_idx		Número			
		Cons_conf_idx		Número			
		Euribor3m		Número			
		Nr_Employed		Número			

- **Bancaria** (Default, Housing, Loan, y)

Hechos		Cliente		Social		Bancaria	
		Nombre del campo		Tipo de datos			
		Id_Bancaria		Autonumeración			
		Default		Sí/No			
		Housing		Sí/No			
		Loan		Sí/No			
		Y		Sí/No			

4. Normalizar las dimensiones.
No se hace necesario utilizar normalización, ya que las dimensiones ya se encuentran normalizadas.
5. Incluir las llaves foráneas en la tabla de hechos.

	Hechos	Cliente	Social	Bancaria
	Nombre del campo	Tipo de datos		
🔑	Id_Hechos	Autonumeración		
	Id_Cliente	Número		
	Id_Social	Número		
	Id_Bancaria	Número		
	Contact	Texto corto		
	Month	Texto corto		
	DayofWeek	Texto corto		
	Duration	Número		
	Campaign	Número		
	Pdays	Número		
	Previous	Número		
	Poutcome	Texto corto		

6. Construir el cubo multidimensional en un motor de base de datos.

