

Desafío Fix Partners Consulting

1. Poniéndose en la posición de un cliente en la industria retail que debe realizar una compra de sus productos en el extranjero e importarlo al país, su problema es tener que tomar la decisión de compra basado en la experiencia de su personal y la intuición que generan las diferentes tendencias en los clientes, lo cual provoca perdidas de eficiencia por quedarse corto o largo de stock en sus diferentes productos.

El cliente quiere poder tomar mejores decisiones para las compras futuras, por lo que se proponen a cambiar su proceso de compra y ahora desean utilizar los datos históricos que disponen (transacciones, ventas, tiendas, etc..) y generar modelos de predicción de demanda.

Con lo expuesto anteriormente, ¿Cuáles serían sus pasos a seguir? ¿Qué datos pediría/necesita? ¿Cómo elaboraría un modelo que satisfaga la necesidad del cliente?

Tome los supuestos que estime pertinentes dejándolos explícitos y explique su procedimiento en máximo una plana.

Respuesta: En síntesis, estamos en presencia de un problema de *forecasting* para la predicción de demanda, para el cual necesitamos contar con un dataset con las siguientes características:

- a) Número de registros suficientes para cubrir un intervalo de tiempo determinado, y en caso de ser necesario, tantos ciclos como sea necesario para el intervalo de tiempo necesario
- b) Los datos deben constituir una sucesión de registros temporales, ordenados y equidistantes:

$$Y = \{Y_{t_i} : t_i \in T, t_i - t_{i-1} = \lambda \in \mathbb{R}, \forall i \in \mathbb{N}\}$$

- c) En el análisis inicial de la serie de tiempo, se deberían observar las siguientes componentes:
 - Tendencia o comportamiento a largo plazo (crecimiento o decrecimiento, comportamiento de la media y varianza).
 - Estacionalidad o comportamiento en un período de tiempo dado (búsqueda de patrones)
 - Ciclos o desviaciones de tendencia.
 - Aleatoriedad o fluctuaciones impredecibles en el tiempo.
- d) Incorporar al dataset, aquellas variables contextuales externas, que incidan en el comportamiento comercial de los clientes, particularmente variables micro y macro económicas.
- e) Posteriormente, previo al proceso de búsqueda de un modelo para el pronóstico deseado, se debe tener en consideración que se pretende extender los valores históricos hacia el futuro, para esto se debe definir el nivel de agregación (horas, días, semanas, meses) y el horizonte esperado (uno o más períodos de tiempo, dados por el nivel de agregación de entrada)
- f) A partir del análisis inicial de los datos, se debería elegir alguna de las siguientes estrategias:
 - Métodos de pronóstico y suavizamiento simple: extrapolación simple a partir de las características de los datos

- Métodos de análisis de correlaciones y modelo ARIMA: a partir de los patrones existentes, se utilizan funciones de diferenciación, autocorrelación y autocorrelación parcial.
 - LSTM: a partir de las características de memoria de corto plazo de las redes neuronales recurrentes (RNN), se puede aplicar para el pronóstico de series temporales.
- g) Finalmente, construiría un modelo a partir de una red LSTM, ajustaría hiperparámetros y evaluaría desempeño final del modelo, y a partir de estos resultados una recomendación ad-hoc.

2. Se tiene una base de datos con las siguientes tablas de las transacciones de una empresa:

POS_TRX	POS_TRX_PAGO	POS_TRX_CLIENTE
ID_TRX INT	ID_TRX INT	ID_TRX INT
Tienda VARCHAR(255)	Monto_pago INT	Nombre VARCHAR(45)
Fecha_transaccion DATE	Medio_de_pago VARCHAR(45)	Rut VARCHAR(45)
Tipo_transaccion INT		
Indexes	Indexes	Indexes

- a) Escriba una consulta en SQL que retorne la cantidad de clientes distintos por tienda que realizaron una transacción del tipo 15.
- b) Escriba una consulta en SQL que entregue las ventas de cada tienda por año en orden de mayor a menor.
- c) Escriba una consulta en SQL que entregue el valor de la transacción promedio de cada tienda por año en orden de mayor a menor.
- d) Escriba una consulta en SQL que responda: ¿Cuál es el principal medio de pago?

Respuesta: Observemos que la base descrita en el esquema, carece de relaciones entre las entidades, por tanto utilizamos las llaves de cada tabla como clave de indexación implícita, bajo la hipótesis de que la llave de cada tabla es: POS_TRX.ID_TRX, POS_TRX_PAGO.ID_TRX, POS_TRX_CLIENTE.ID_TRX.

- a)

```
SELECT COUNT(DISTINCT(POS_TRX_CLIENTE.ID_TRX))
FROM POS_TRX_CLIENTE
WHERE POS_TRX_CLIENTE.ID_TRX=POS_TRX.ID_TRX
AND POS_TRX.Tipo_transaccion = '15';
```
- b)

```
SELECT POS_TRX.Tienda AS tienda ,
YEAR(POS_TRX.Fecha_transaccion) AS year ,
SUM(POS_TRX_PAGO.Monto_pago) AS total
FROM POS_TRX, POS_TRX_PAGO
WHERE POS_TRX.ID_TRX = POS_TRX_PAGO.ID_TRX
GROUP BY YEAR(POS_TRX.Fecha_transaccion)
ORDER BY year DESC;
```

c) `SELECT POS_TRX.Tienda AS tienda ,
YEAR(POS_TRX.Fecha_transaccion) AS year ,
AVG(POS_TRX_PAGO.Monto_pago) AS total
FROM POS_TRX, POS_TRX_PAGO
WHERE POS_TRX.ID_TRX = POS_TRX_PAGO.ID_TRX
GROUP BY YEAR(POS_TRX.Fecha_transaccion)
ORDER BY year DESC;`

d) `SELECT medio
FROM(
SELECT medio , MAX(count)
FROM(
SELECT POS_TRX_PAGO.Medio_de_pago AS medio ,
COUNT(POS_TRX_PAGO.Medio_de_pago) AS count
FROM POS_TRX_PAGO))`