

Prueba técnica BI Analyst

Contexto

Como Business Intelligence Analyst eres responsable del conocimiento de los datos y de la información de Holaluz. Eres capaz de proporcionar los insights necesarios para mejorar procesos y conocer el estado actual de ellos.

Tu y tu equipo centralizáis todos los datos en una base de datos relacional **MySQL** que llamáis *Data Warehouse*. En ella se guardan y se actualizan las tablas de diferentes procesos de negocio, desde información de facturación hasta información de llamadas de clientes.

Dos de las tablas principales del *Data Warehouse* consisten en la tabla de clientes “*client_dim*” y la tabla de contratos “*contracts_dim*”.

La tabla de clientes tiene como columnas el identificador de fila (*client_id*), el identificador web (*client_web_id*), el nombre del cliente (*cl_name*) y el email.

Por otro lado la tabla de contratos tiene el identificador de contrato (*contract_id*), el identificador de cliente (*client_id*), el estado del contrato (*co_status*), la tarifa del contrato (*contract_tariff*), el producto del contrato (*contract_product*), el tipo de canal de venta (*contract_sales_channel*), la fecha de inicio del contrato (*contract_start_date*) y la fecha fin del contrato (*contract_end_date*). Tanto *contract_start_date* como *contract_end_date* tienen formato fecha dentro de la base de datos.

Una extracción simulada de estas dos tablas las puedes encontrar en el archivo .xlsx junto con una tercera tabla llamada “*tickets_ft*” donde puedes encontrar las incidencias (tickets) de cada cliente. Esta tabla tienen el identificador de ticket (*ticket_id*), el identificador de cliente (*client_id*), el estado de la incidencia (*ticket_status*), la fecha de apertura de la incidencia (*ticket_created_at*) y la fecha de cierre de la incidencia (*ticket_closed_at*). Tanto *ticket_created_at* como *ticket_closed_at* tienen formato fecha dentro de la base de datos.

Se tiene que tener en cuenta que un cliente puede tener más de un ticket, es decir la relación entre la tabla de clientes con la de tickets es 1 a n respectivamente. Además, la tabla de clientes está relacionada con la de contratos con relación 1:N respectivamente, donde un cliente puede tener más de un contrato activo en el mismo momento (más de un punto de suministro con nosotros).

Primera parte

Suponiendo que estas tablas del xlsx son las mismas que te encuentras en el Data Warehouse, escribe las queries necesarias para saber los puntos a continuación. Si lo consideras oportuno puedes escribir comentarios :)

- 1) Cuántos clientes con estado 'Alta' tenemos con cada tarifa y producto.
- 2) Cuántos clientes en estado Alta tienen o han tenido una incidencia.
- 3) Una lista de los client_id y el número de incidencias abiertas de los clientes que tengan más de 2 incidencias con estado 'Abierta'.
- 4) Una lista de todos los clientes (client_id) con su número de incidencias total y el número de incidencias dividido el tiempo que lleva activo el cliente con Holaluz (incidencias por mes).

Segunda parte

Una parte importante de tu trabajo conseguir insights para ayudar a los procesos de Holaluz entre otros. Es por esto que has decidido investigar porque se abren las incidencias. Te propones descubrir si hay algún factor que influya más a que los clientes tengan una incidencia (no entras a valorar el número de incidencias, sólo si **almenos** se abre una). Hay alguna característica del contrato que puedes afirmar que afecta al número de clientes con incidencia?

Utiliza las tablas del archivo .xlsx para este análisis. Puedes usar Excel, el lenguaje de programación que prefieras (en Holaluz trabajamos con Python y R) o utilizar la herramienta que creas más conveniente. Justifica tu respuesta.

Para simplificar el problema, puedes hacer el análisis sólo con la información del primer contrato de cada cliente (contract_id más pequeño) y descartar al resto.

Tercera parte

Un canal importante canal de contacto de Holaluz es el emailing. Por esta razón es importante que los emails de la base de datos sean válidos. Para asegurarte de este data quality decides crear un validador de email. Genera un validador de email que dado un email nos indique si es válido o no.

Por un lado debes explicar qué condiciones lógicas pondrías y por otro lado debes implementarlo con la herramienta que prefieras (no importa cuál, sino que funcione), como por ejemplo Excel mismo o un lenguaje de programación.

Si es en Excel simplemente puede consistir en que si se rellena una celda con un email, otra nos indique si es válido o no. Si es en un lenguaje de programación debe ser una función.

Cuarta parte

Te reunes con el responsable de atención al cliente y te comenta que no son capaces de saber cómo está funcionando su equipo porque no tienen ninguna métrica para hacerlo. Tu sabes que en tu Data Warehouse tienes datos de tickets recibidos y llamadas de los clientes, aparte de las tablas del archivo .xlsx.

Haz una propuesta de un Dashboard tamaño aprox A4 horizontal para hacer el seguimiento de métricas que creas que sean útiles para un equipo de atención al cliente. Ten en cuenta que el mejor dashboard no es el que más cosas tiene, sino el que transmite mejor un estado.

Puedes incorporar métricas relacionadas con incidencias o llamadas de cliente que no estén en el .xlsx, como por ejemplo el tiempo de respuesta. La propuesta de dashboard lo puedes hacer con cualquier tecnología (por ejemplo un croquis con excel) o a mano, aunque una buena herramienta (gratuita) para hacer dahsboards puede ser el Power BI. En el caso que quieras usar datos o indicadores para hacer el dashboard te los puedes inventar.