Users

column_name	description
UUID	Unique identifier of a user
date_sk	Date of registration
country_code	User’s Country
campaign_id	Marketing campaign to which a user's registration was attributed

Sales

column_name	description
UUID	Unique identifier of a user
purchase_id	Unique identifier of a purchase (sale)
date_sk	Date of purchase
campaign_id	Marketing campaign to which a purchase was attributed

Revenue

column_name	description
purchase_id	Unique identifier of a purchase
purchase_transaction_id	Unique identifier of a transaction
start_date_sk	start date of subscription
end_date_sk	end date of subscription

is_first_period	Either purchase first_sale or extension
price_eur	Price of purchases in Euro

Marketing Campaigns

column_name	**description**
campaign_id	Unique identifier of a marketing campaign
campaign_name	Name of marketing campaign
campaign_channel	Name of the marketing channel a campaign belongs to
sub_channel	Name of Sub-Channel the campaign belongs to

Country

column_name	**description**
country_code	A three-letter (ISO 3166-1 alpha-3)
country_name	Name of the country
geo_area	Geographical continent to which a country belongs

Task 1

Questions:

- How many users are there per country, per geo_area and per year of registration?
- How many purchases happened per country, per geo_area and per year of registration?
- How many purchases are there per campaign, per subchannel and per month of purchase?
- Which is the Sales-to-Lead ratio (number of sales per purchase date / number of leads per registration date) per campaign and per date?

- Which are the top 10 new sales revenues per subchannel per month of purchase?
  Hint: New Sales Revenues: purchase_transaction_ID with is_first_period = 'first_sale'

Task 2:

User events are grouped into sessions. We define a session as a sequence of events with the same user ID, ordered by timestamp, such that the time difference between any consecutive pair of events is at most one hour.

Example:

column name	**description**
session_id	Unique identifier of a session
UUID	Unique identifier of a user
session_start	Timestamp of session start
session_end	Timestamp of session end

- How would you calculate the average session duration?
- How to know how many users have several sessions a day? How many sessions do they do?
- Though in theory, a session_id is supposed to be unique, there happen to be duplicates. How would you proceed to find and remove them?

Task 3

Our conversion event looks similar to the structure shown below. In Snowflake, we can use SQL syntax to query the events.

{

"name": 'conversion_event'

"created_at" : 2020-01-01 00:00:00,

"uuid" : 12314512441312312,

"campaign_name": campaign_name1,

"campaign_channel": channel1,

"meta":{

"app_name": iOS,

"os_version": 11,

"country" : ABC,

"latitude": 12.12,

"longitude": 12.12,

},

"type_of_conversion": 'registrations'

}

Questions:

- Write a SQL query to find the number of distinct app_name and os_version.
- If there are two types of conversion like **registrations** and **purchases**, write a SQL to calculate the average number of days between registrations and purchases per country per campaign.
- If there is another event name called "**campaign_touchpoint**"(i.e. different campaign touchpoint which users' contact point before he or she converted), write a SQL to identify the first touchpoint channel, last touchpoint channel per user per session (default session length = 30 mins). Your result must contain UUID, session_Id, first_touchpoint_channel, last_touchpoint_channel.

Task 4: Python

We also use Python to handle our data. Please send us a piece of code you have run in the past and you are proud of, together with a description of its use case.