

Goals

1. How many people are going to visit the website on a daily basis
 - a. What is the geographical and device break down
 - b. Where are the people coming from ? LinkedIn, X, Substack, Email etc.
2. How many people are logging in with an account on a daily basis
 - a. What percentage of the traffic is converting to signing up ?
3. How many people are purchasing the boot camps and courses
 - a. What percentage of signups convert to paying customers ?

Metric Name	Definition	Is Guadrail
signup_conversion_rate	COUNT(signups)/ COUNT(website_hits)	Yes
purchase_conversion_rate	COUNT(purchases)/ COUNT(signups)	Yes
traffic_breakdown_by_channel	COUNT(website_visits) GROUP BY channel	No

Note : Guadrail here means a metric that is business critical and any decline in number on it triggers an alert, and the text that is doing this should not be pushed to production.

Schema

core.fct_website_events : List of all events from the website, including IP and user agent enrichment for country and device information

Primary Key is the **logged_out_user_id**

Column	Column Type	Comments
user_id	BIGINT	Nullable. Indicates the user who generated this event
logged_out_user_id	BIGINT	Hash of IP address and device information (Murmur3 hash)
dim_hostname	STRING	Host associated with the event
dim_country	STRING	Country associated with the IP address of the request
dim_device_brand	STRING	Device associated with the request

action_type	ENUM	List of actions that the user can make on the website (Log-in, watch video, page visit etc.)
event_timestamp	TIMESTAMP	UTC timestamp logging when the event occurs
other_properties	MAP{String:String}	Any other valid properties that are part of this request
ds	STRING	Partition column for the table

Quality Checks: -

1. Not null checks on (dim_hostname, dim_action_type, event_timestamp, dim_country, logged_out_user_id)
2. Make sure no duplicates on Primary Key
3. Dim_hostname is well formatted (something like this www.xyz.com)
4. Row count checks - GROUP ON dim_hostname and check week-over-week counts for the hosts
5. Enumeration check on dim_action_type (should be signup, login,purchase etc.)

core.agg_website_events : Daily aggregates of the events on the website

Primary Key is the **logged_out_user_id**

Column	Column Type	Comments
action_type	ENUM	Enumerated action_type
logged_out_user_id	BIGINT	Hash of IP address and device information (Murmur3 hash)
dim_hostname	STRING	Host associated with the event
dim_country	STRING	Country associated with the IP address of the request
dim_device_brand	STRING	Device associated with the request
event_hour	INTEGER	Hour of day the event took place
m_total_events	BIGINT	Total number of events for this slice
aggregation_level	STRING	Granularity of the aggregation for this table
ds	STRING	Partition column for the table

Quality Checks: -

1. Row count checks
 - a. (overall) rollup should have more data than any other rollup
 - b. Event_hour should look like it's old seasonal pattern
2. M_total_events should be > some minimum number