

Etsy User Engagement Pipeline

Team members:

Surabhi Soni

Thy Anh Bui

Venetis Pallikaras

Monica Radhakrishnan

Shanmukh Sanka

Problem Statement

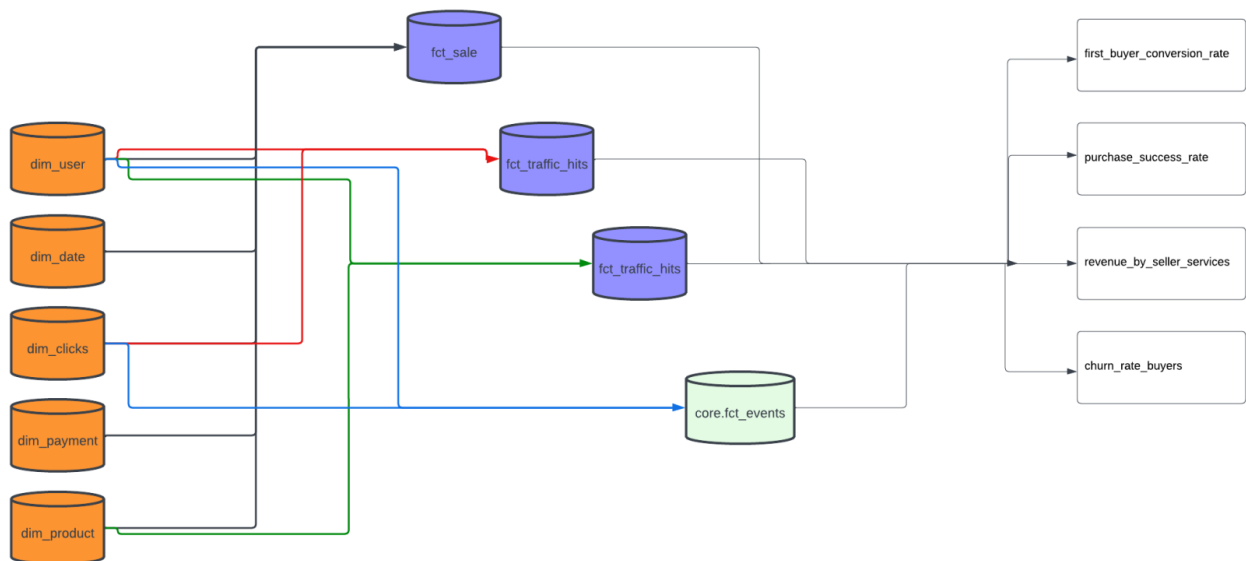
Etsy.com, a renowned global marketplace for unique and creative goods, looking to access its user engagement and business goals. The platform seeks to

- attract new users (sellers) and get them to take actions
- track the ratio of new visitors/sellers towards the number of visitors/inactive accounts/sellers
- measure the conversation rate between add to cart clicks and orders
- Breakdown of revenue generated from various seller services like promoted listings, Etsy Ads (via pinterest, instagram etc)

Business Metrics

Metric Name	Definition	Is Guardrail
first_buyer_conversion_rate	$\text{COUNT}(\text{purchase}) / \text{COUNT}(\text{first impression})$	
purchase_success_rate	$\text{COUNT}(\text{completed_orders}) / \text{COUNT}(\text{add_to_cart_clicks})$	
revenue_by_seller_services	$\text{COUNT}(\text{revenue from specific service}) / \text{COUNT}(\text{total revenue})$	
churn_rate_buyers	$(\text{COUNT}(\text{buyers at the beginning of the period}) - \text{COUNT}(\text{buyers at the end of the period})) / \text{COUNT}(\text{buyers at the beginning of the period})$	

Flow Diagram



Schemas

Dimension tables

1. dim_user:

Column Name	Column Type	Column Comment
user_id	INT	
user_type	STRING	
registration_date	DATETIME	
location_id	INT	
age_group	STRING	

Quality Checks:

- Ensure user_id is unique and not null.
- Validate that user_type is either "Buyer" or "Seller."
- Check that registration_date is within a reasonable date range.
- Verify that location_id exists in dim_location.
- Age_group values should align with predefined categories.

Assumptions:

- Each user has a unique user_id.
- Users are categorized as either buyers or sellers.
- Location_id refers to a specific location in dim_location.
- Age_group segments users into predefined age ranges.

2. dim_date:

Column Name	Column Type	Column Comment
date_time	DATETIME	
is_weekend	BOOLEAN	Boolean that indicates if date is within the weekend
is_holiday	BOOLEAN	
quarter	INT	

Quality Checks:

- Validate that date is within a reasonable date range.
- Quarter values should range from 1 to 4.

Assumptions:

- Is_weekend and is_holiday are derived based on predefined criteria.

3. dim_clicks:

Column Name	Column Type	Column Comment
session_id	INT	Unique session
impression	INT	
click		
click_referrer	ENUM(pinterest, instagram, google_search, etsy)	referrer of where click is from
session_start_time	DATETIME	date and time when session started
session_end_time	DATETIME	date and time when session

		ended
page_type	STRING	Describes the function of the page in the URL
sub_page_type	STRING	Click action within the page

Quality Checks:

- Ensure clicks_id (for dim_clicks) is unique and not null.
- Validate that click_referral (for dim_clicks) is a valid entry.

Assumptions:

- Clicks_id uniquely identify each entry.
- click_referrer represents pre-defined referrer.

4. dim_payment:

Column Name	Column Type	Column Comment
payment_id	INT	
payment_method	STRING	

Quality Checks:

- Ensure payment_id (for dim_payment) is unique and not null.
- Validate that payment_method (for dim_payment) is a valid entry.

Assumptions:

- Payment_id uniquely identifies each entry.
- Payment_method represents valid methods.

5. dim_product:

Column Name	Column Type	Column Comment
product_id	BIG INT (primary key)	Unique identifier for a product
product_group	STRING	Category of the product. Ex: Mobile phones, Tablets
business_group	STRING	Business representing the

		product. Ex: Electronics , Audio books
product_description	STRING	Description of the product
Market_place	STRING	Country / Location
is_live	BOOLEAN	Is the product live on the website ?

Quality Checks:

- Ensure product_id (for dim_product) is unique and not null.
- Validate that is_live (for dim_payment) is 0 or 1

Assumptions:

- business_group has multiple product_group and product_id

Fact Tables

1. fct_sale:

Column Name	Column Type	Column Comment
product_id	BIG INT	Unique identifier for a product
date_time	DATETIME	Time a user completes a purchase
payment_id	BIG INT	Unique identifier for a transaction
click_id	BIG INT	Unique identifier for an ad click
user_id	BIG INT	Unique identifier for a buyer
sale_quantity	INT	Number of product bought by a user
Reg_unit_price	BIG INT	Price of a product
Ad_dollar_amount	BIG INT	Ad spent on a sale
Profit_after_ad	BIG INT	Profit after ad
Profit_by_referrer	BIG INT	Profit by referrer

Churn_sale	BIG INT	Return revenue
session_id	STRING	Unique identifier for a browsing session
click_id	STRING	Unique hit in a customer session

2. fct_traffic_hits :

Column Name	Column Type	Column Comment
request_id	STRING (primary key)	Unique hit in a customer session
session_id	STRING	Unique identifier for a browsing session
user_id	BIG INT	Unique identifier for a given user
device_id	BIG INT	Unique identifier for a given device
is_robot	BOOLEAN	Human or robot traffic
session_start_time	DATETIME	date and time when session started
session_end_time	DATETIME	date and time when session ended
page_type	STRING	Describes the function of the page in the URL
sub_page_type	STRING	Click action within the page

- Quality Checks:
 - Verify that request_id is unique.
 - Verify that session_id is unique.

- Validate that user_id exists in dim_user.
- Assumptions:
 - sub_page_type indicates various user interactions via clicks.

3. fct_order_items :

Column Name	Column Type	Column Comment
order_item_id	BIG INT (primary key)	Unique identifier for an order item within an order
order_id	BIG INT (primary key)	Unique identifier for an order
order_date	DATETIME	Date when an order is placed.
product_id	BIG INT	Unique identifier for a product
user_id	BIG INT	Unique identifier for a given user
business_group	STRING	Business representing the product. Ex: Electronics , Audio books
price	NUMERIC	Total price buyers have to pay for given item(s)
quantity	INT	No times the item was selected
session_id	STRING	Unique identifier for a browsing session
request_id	STRING	Unique hit in a customer session

- Quality Checks:
 - Verify that order_id and order_item_id is unique and not null
 - Verify that session_id is unique.
 - Validate that user_id exists in dim_user.

4. core.fct_events :

Column Name	Column Type	Column Comment
-------------	-------------	----------------

session_id	STRING	Unique identifier for a browsing session
user_id	BIG INT (primary key)	Unique identifier for a given user
event_date	DATETIME	Date when the event is occurred.
session_start_time	DATETIME	date and time when session started
session_end_time	DATETIME	date and time when session ended
request_id	STRING (primary_key)	Unique hit in a customer session
event_name	STRING	Name of the event or action by User
event_sequence_number	BIG INT(primary key)	Sequence of the events within the session.

- Quality Checks:
 - Verify that request_id is unique.
 - Verify that user_id is unique.
 - Verify that event_sequence_number is unique.
 - Validate that user_id exists in dim_user.
- Assumptions:
 - event_sequence_number is local within the session.

