

Google Merchandise Sales Data

Người thực hiện: Trần Quốc Khang

Mục lục

1

Data dictionary

2

Data modeling

3

Data preprocessing

4

Data visualization

5

Generative AI

6

Take action

1. Data dictionary

💡 Mô tả dữ liệu

- Dataset is subset of anonymized Google Analytics data from the Google Merchandise Store
- Dữ liệu được thu thập trong 3 tháng 11/2020, 12/2020, 1/2021



1. Data dictionary



Data dictionary

#	Table	Variables/Features/Columns	Description	Data types 1	Data types 2	Keys/Value	Notes
1	events1	user_id	User Id	int	qualitative	[5115 10904 29457 ... 260725 18261 3772]	
2		ga_session_id	Session Id of Google Analytics	int	qualitative	[17001 16401 17113 ... 18001 17969 17918]	
3		country	Country	chr	qualitative	['US' 'TR' nan ... 'MM' 'KE' 'OM']	108 values - Tên viết tắt của country
4		device	Device access	chr	qualitative	['mobile' 'desktop' 'tablet']	3 values
5		type	Type of user action	chr	qualitative	['purchase' 'add_to_cart' 'begin_checkout']	3 values
6		item_id	Item Id	int	qualitative	[0 1 2 ... 1378 1379 1380]	
7		date	Date of session	datetime	timeseries	['2020-11-02 12:05:14' ... '2020-12-30 14:42:10']	
8	items	id	item Id	int	qualitative	[0 1 2 ... 1378 1379 1380]	
9		name	Name of items	chr	qualitative	['Google Land & Sea Cotton Cap' 'Google KeepCup' ... Mountain View Campus Bottle']	
10		brand	Brand of items	chr	qualitative	['Google' 'Android' 'YouTube' '#IamRemarkable' 'Google Cloud']	5 values
11		variant	Variant of items	chr	qualitative	['Single Option Only' ' LG' ... '2XL' 'No options available']	44 values
12		category	Category of items	chr	qualitative	['Apparel' 'New' 'Drinkware' ... 'Eco-Friendly' 'Gift Cards']	21 values
13		price_in_usd	Price in USD	int	quantitative	[14 28 20 ... 56 31 313]	
14	users	id	User Id	int	qualitative	[5115 10904 29457 ... 260725 18261 3772]	
15		ltv	Lifetime value	int	quantitative	[85 40 33 ... 352 523 1200]	
16		date	User creation date	datetime	timeseries	['2020-11-02 11:53:43' ... '2020-12-28 06:34:48']	

Hình 1.1. Data dictionary

1. Data dictionary

Data quality assessment

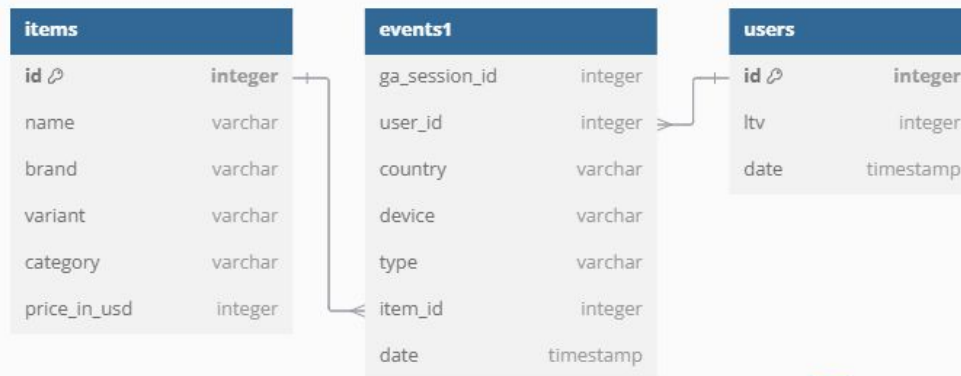
No.	Table	Important Features	Group Name	# records	Distribution						% Missing values	Notes
					Min	Q1/p25	Q2/ p50	Q3/ p75	Max	Outlier		
1	test	user_id		719386								
2		ga_session_id		719386								
3		country		715095							0.6%	
4		device		719386								
5		type		719386								
6		item_id		719386								
7		date_session		719386								
8		item_name		719386								
9		brand		719386								
10		variant		84283							88.3%	
11		category		719386								
12		price_in_usd		719386	1	12	22	30	313	[57, 313]		
13		ltv		719386	0	0	0	76	1530	[190, 1530]		
14		user_create_date		719386								

Hình 1.2. Data quality assessment

2. Data modeling



Star schema

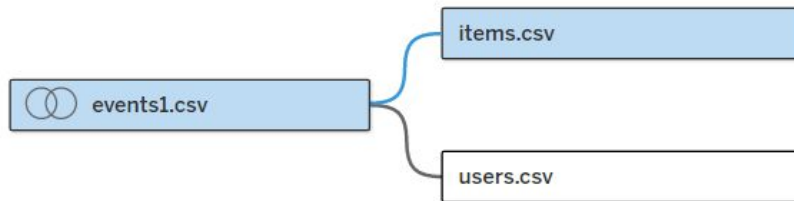


Hình 2.1. Star schema use dbdiagram
([link: star schema - dbdiagram.io](https://dbdiagram.io))

2. Data modeling



Star schema



Relationship		Related Fields	
events1.csv	* - 1	items.csv	item_id = Id
events1.csv	* - 1	users.csv	user_id = Id (Users.Csv)

Hình 2.2. Star schema in Tableau

2. Data modeling



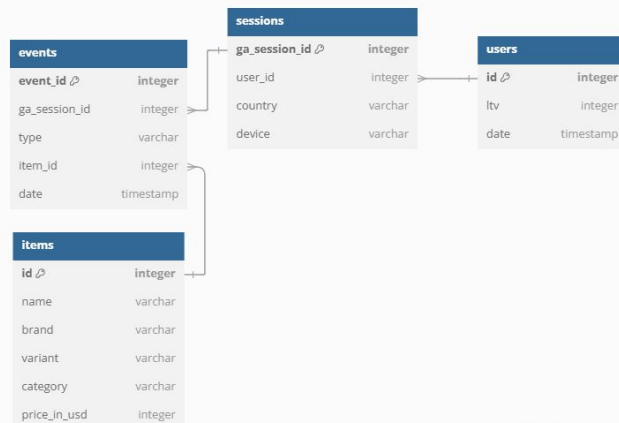
Snowflake schema - Phân tích NF based on primary keys

- Chia table events1 thành 2 table: sessions{ga_session_id (PK), user_id, country, device} và events{event_id (PK), ga_session_id, date, type, item_id}. Lí do:
 - sessions:
 - Atomic attributes: không có nhóm lặp hay quan hệ lồng (1NF)
 - ga_session_id (PK) → user_id, country, device: phụ thuộc trực tiếp. Chứng minh: Các thuộc tính không khóa không xác định lẫn nhau(2NF), không có phụ thuộc bắc cầu (3NF)
 - events: tương tự đạt 3NF
- Table items và users: phân tích tương tự đều đạt 3NF

2. Data modeling



Snowflake schema



Hình 2.3. Snowflake schema use dbdiagram
(link: [snowflake schema - dbdiagram.io](https://dbdiagram.io))

3. Data preprocessing



Use Colab for data preprocessing



Hình 3.1. Colab

(link: https://drive.google.com/file/d/1JtXAfPWC0_DyAEUH-qMaspAby9biNdL6/view?usp=sharing)

4. Data visualization



Key metrics



GMV

GMV (Gross Merchandise Volume) đo lường tổng giá trị hàng hóa hay tổng giá trị giao dịch



Total Users

Total Users: total new users, total current users,...

4. Data visualization



Use Tableau for Dashboard



Hình 4.1. Tableau
(*link: finalproject / Tableau Public*)

5. Generative AI



Use NotebookLM for automatic report

 NotebookLM



Có trích dẫn nguồn



Hạn chế hallucination

5. Generative AI



Use NotebookLM for automatic report

Prompt

Giả sử bạn là một nhân viên data analyst và cần viết một bản báo cáo để trình bày nội dung bạn phân tích được đến CEO. Đầu tiên bạn hãy viết bản báo cáo đó. Sau khi viết xong báo cáo, bạn nghĩ công ty cần thu thập thêm dữ liệu/tập dữ liệu nào và tại sao?

Generative AI

Báo Cáo Phân Tích Dữ Liệu Bán Hàng Google Merchandise Cho CEO

Kính gửi CEO,

Bản báo cáo này trình bày phân tích dữ liệu bán hàng Google Merchandise từ tháng 11/2020 đến tháng 1/2021, nhằm cung cấp cái nhìn tổng quan về hiệu quả kinh doanh và đề xuất các chiến lược phát triển trong tương lai.

I. Tổng Quan Thị Trường

- Tổng doanh thu (GMV) đạt **17.922 triệu USD**, trong đó thị trường Mỹ chiếm tỷ trọng lớn nhất với **8.026 triệu USD** ¹.
- Các quốc gia có doanh thu cao tiếp theo là Ấn Độ (1.682 triệu USD) và Hà Lan (183 nghìn USD) ¹.
- **5 sản phẩm có doanh thu cao nhất** bao gồm Google Zip Hoodie F/C, Google Crewneck, Google Badge, Google Men's Tech Fleece và một sản phẩm Google Crewneck khác ¹.
- **5 danh mục sản phẩm có doanh thu cao nhất** lần lượt là Apparel, New Shop by Brand, Accessories, Uncategorized Items và một danh mục không được hiển thị tên ².
- Doanh thu theo **thiết bị** cho thấy **máy tính để bàn** chiếm ưu thế với 58%, tiếp theo là điện thoại di động (40%) ¹.

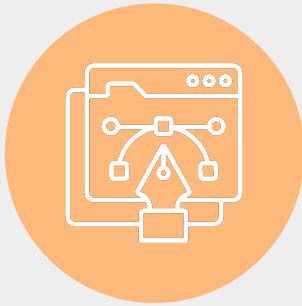
II. Phân Tích Người Dùng

- Tổng số người dùng là **14.594** ² ³.

6. Take action



System



UI/UX



Desktop & Mobile



High-purchase period

6. Take action



Users



New Users

Tạo campaign marketing
thu hút



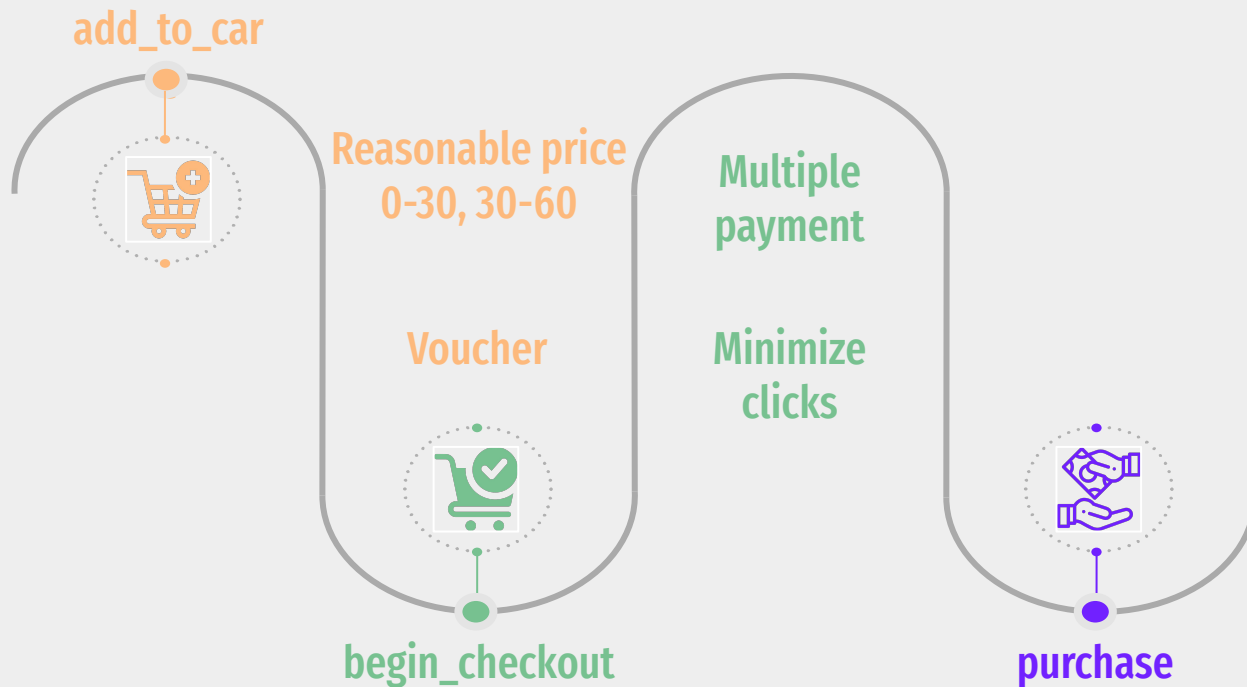
VIP Users

Membership tiers

6. Take action



Conversion rate



6. Take action



Items



Items

Google, Android and Youtube



Ads

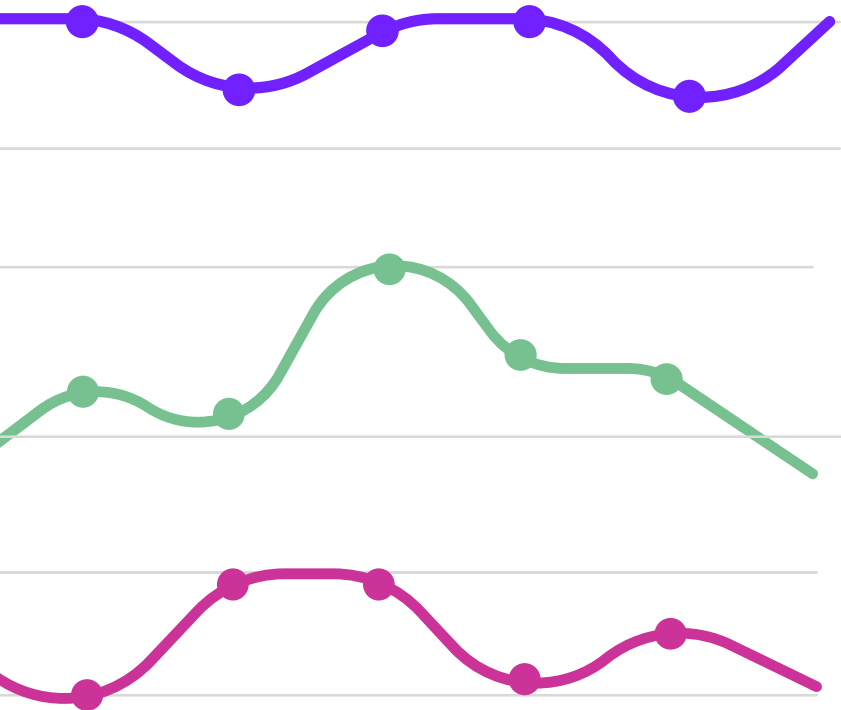
Campaign



Devices

Desktop and Mobile

Trần trọng cảm ơn



Google Merchandise Sales Data

Người thực hiện: Trần Quốc Khang