

VEF ACCADEMY

KHẢO SÁT THỊ TRƯỜNG THƯƠNG MẠI ĐIỆN TỬ TẠI ANH

FINAL PROJECT REPORT - UK E-Commerce Data

TÊN DỰ ÁN	Khảo Sát Thị Trường Thương Mại Điện Tử Tại Anh	DỰ ÁN SỐ	1
THÀNH VIÊN NHÓM DỰ ÁN	THỜI GIAN THỰC HIỆN	NGÀY BẮT ĐẦU DỰ ÁN	NGÀY HOÀN THÀNH DỰ ÁN
Lương Ngọc Lan Hương Nguyễn Ngọc Bích	02/10/2021 – 11/11/2021	11/10/2021	14/11/2021

1. GIỚI THIỆU TỔNG QUAN VỀ MỤC TIÊU DỰ ÁN

- Dự án được thực hiện dựa trên dữ liệu Thương mại điện tử thu được từ một cửa hàng bán lẻ trực tuyến có trụ sở tại Vương quốc Anh để hiểu sâu hơn về hoạt động kinh doanh bán lẻ.
- Khảo sát các đặc điểm của thị trường thương mại điện tử tại Anh, khám phá các mô hình của khách hàng, quốc gia, sản phẩm, dựa trên từng khoảng thời gian. Từ đó làm cơ sở cho các quyết định về các chiến dịch tăng doanh thu như tái định giá sản phẩm, thời gian tạo chương trình khuyến mãi, tung voucher.
- Biết được các khu vực, đất nước mà thị trường thương mại điện tử còn hạn chế. Từ đó có những giải pháp nhằm nâng cao doanh số tại các khu vực này.

2. CÁCH TIẾP CẬN VẤN ĐỀ

- Dataset thiếu nhiều giá trị (missing value), cần phải làm sạch và xử lý các biến ngoại lai trước khi thực hiện phân tích dữ liệu và tìm insights.
- Giả thiết: Các yếu tố ảnh hưởng đến số lượng giao dịch bao gồm: yếu tố địa lý, theo mùa. Thời gian trong ngày, trong tháng hoặc trong năm có ảnh hưởng lớn đến số lượng đơn của khách hàng. Số lượng đơn huỷ phụ thuộc vào các thời gian nhất định của tháng, hoặc năm.

3. TỔNG QUAN VỀ NGUỒN DỮ LIỆU

- Nguồn dữ liệu: Dr Daqing Chen, Director: Public Analytics group. chend '@' lsbu.ac.uk, School of Engineering, London South Bank University, London SE1 0AA, UK. ([Source](#))
- Thời gian lấy dữ liệu: từ 01/12/2010 đến 09/12/2011
- Cách lấy dữ liệu: Dataset được lấy từ Kaggle ([Kaggle](#))

4. GIỚI THIỆU VỀ DỮ LIỆU

- Cửa hàng thương mại điện tử có trụ sở tại Vương quốc Anh
- Sản phẩm bán chủ yếu là quà tặng
- Khách hàng hầu hết là các nhà bán buôn (trong nước và quốc tế)
- Thời gian giao dịch diễn ra trong một năm, từ ngày 1/12/2010 đến ngày 9/12/2011

TÊN CỘT	Ý NGHĨA	MIÊU TẢ	LƯU Ý
InvoiceNo	Số hóa đơn	Biến định danh, dãy kí tự có 6 số nguyên được phân cho mỗi giao dịch	Nếu code bắt đầu bằng ký tự "c", tức là giao dịch bị huỷ
StockCode	Mã sản phẩm	Biến định danh, dãy kí tự có 5 số nguyên được phân cho mỗi loại sản phẩm khác nhau	Nếu mã không phải là dãy số thì không phải là sản phẩm
Description	Tên sản phẩm	Biến định danh	
Quantity	Số lượng	Biến số, số lượng mỗi sản phẩm mỗi giao dịch	
InvoiceDate	Ngày giờ xuất hóa đơn	Biến số, ngày giờ thực hiện giao dịch	
UnitPrice	Đơn giá	Biến số	
CustomerID	Mã số khách hàng	Biến định danh, dãy kí tự có 5 số nguyên được phân cho mỗi khách hàng khác nhau	
Country	Quốc gia	Biến định danh, tên quốc gia mà khách hàng thực hiện giao dịch	

5. TỔNG QUAN VỀ PHÂN TÍCH KHÁM PHÁ (EDA)

Phân phối dữ liệu

Xử lý các giá trị null

- Data có 541909 dòng và 8 cột
- Có 9288 dòng có Invoice bắt đầu bằng ký tự "C" (giao dịch bị huỷ) => xoá các giao dịch bị huỷ
- Trong đó, cột CustomerID, cột Description có null. Với số lượng null của CustomerID là 135080 (24,92%) và tỉ lệ null của Description 1454 (0,268%) => xoá các giao dịch ko có CustomerID
- Sau khi xử lí, data có 397924 dòng và 8 cột

Xử lí outlier

Đặt vấn đề và tìm thêm insights

- Số lượng đơn hàng / chi tiêu (cao nhất vào tháng 11/2011)
 - Thống kê theo customerID
 - Thống kê theo các quốc gia khác nhau
- Khách hàng

- Quốc gia có nhiều khách hàng nhất là UK (cũng là quốc gia xuất xứ của cửa hàng)
- Phân khúc khách hàng bằng cách sử dụng RFM
- Tìm khách hàng định kỳ bằng Cohort
- Sản phẩm
 - Mã sản phẩm đặc biệt là các chi phí phát sinh
 - Sản phẩm có khoảng giá 0-2\$ mang lại lợi nhuận cao nhất
 - Sản phẩm giáng sinh có từ khóa “Christmast” trong miêu tả chiếm khoảng 4% tổng doanh thu
 - Sản phẩm miễn phí được tặng cho một số khách hàng nhất định
- Thời gian khách hàng có số lượng giao dịch nhiều nhất?
 - Thời gian giao dịch tương đối nhiều trong khung giờ từ 10h - 15h hàng ngày. Trong đó đạt đỉnh ở khung giờ từ 12h-13h
 - Ngày trong tuần khách hàng chốt đơn nhiều nhất? => Số lượng đơn trong các ngày thứ 2,3,4,6 tương đối đồng đều. Tuy nhiên, Thứ Năm ghi nhận số lượng đơn nhiều hơn đáng kể, trong khi Chủ nhật lại có số lượng đơn ít nhất (Không ghi nhận có giao dịch vào Thứ 7!)
 - Tính theo tháng, số lượng giao dịch ghi nhận được gần như giảm dần từ đầu đến cuối tháng. Càng về cuối tháng, số lượng giao dịch có xu hướng giảm
 - Số lượng giao dịch trong các tháng tương đối khác nhau trong năm. Ghi nhận sự gia tăng số lượng giao dịch trong các tháng cuối năm (từ tháng 9 đến tháng 12). Điều này có thể là do thời gian cuối năm có nhiều ngày lễ?
- Về việc huỷ đơn
 - Các quốc gia có số lượng đơn huỷ nhiều nhất (United Kingdom, Germany, EIRE, France, USA). Tuy nhiên, United Kingdom, EIRE, France, Singapore, Germany theo thứ tự mới là các nước có giá trị đơn huỷ cao nhất.
 - Các đơn huỷ tăng cao vào các tháng 9 đến tháng 11

6. TRÌNH BÀY KẾT QUẢ

- Project Folder: [Google Drive](#)
- Source Code: [Google Colab](#)
- Storytelling with data: sử dụng Tableau để vẽ biểu đồ thích hợp

