**An eCommerce Company Case Study**

**Bối cảnh**

Bạn là một Data Analyst làm việc cho một công ty thương mại điện tử tên là X. Bạn được giao nhiệm vụ chuẩn bị một bài thuyết trình để trình bày tổng quan tình hình kinh doanh và vận hành của công ty tính đến thời điểm hiện tại cho Giám đốc bán hàng và Giám đốc vận hành. Bài thuyết trình tối thiểu phải bao gồm các thông tin sau: tổng quan tình hình kinh doanh, mức độ hài lòng của khách hàng, và đề xuất 2 đến 3 lĩnh vực (areas) mà công ty có thể cải thiện.   
  
Một số thông tin bổ sung cho case study:

* Vì chỉ có dữ liệu đến năm 2018, nên ta sẽ giả sử hiện tại đang là tháng 9 năm 2018 (các dữ liệu sau tháng 9/2018 bạn có thể bỏ qua)
* Công ty có trụ sở tại Mỹ, tuy nhiên được thành lập ở Brazil (đó là lý do vì sao một số thông tin được viết bằng tiếng Bồ Đào Nha)

**Dữ liệu**

Bộ dữ liệu bao gồm:

* **Orders dataset**: Cung cấp thông tin về các đơn hàng

**order\_id**: unique ID của đơn hàng

**customer\_id**: unique ID của khách hàng

**order\_status**: trạng thái đơn hàng

**order\_purchase\_timestamp**: thời gian đơn hàng được đặt mua

**order\_approved\_at**: thời gian đơn hàng được phê duyệt

**order\_delivered\_carrier\_date**: thời gian hàng được đưa đến cho đơn vị vận chuyển

**order\_delivered\_customer\_date**: thời gian hàng được đưa đến khách hàng

**order\_estimated\_delivery\_date**: thời gian dự kiến đơn hàng được đưa đến khách hàng

* **Order items dataset**: Cung cấp thông tin về từng món hàng trong đơn hàng và chi phí ship

**order\_id**: unique ID của đơn hàng

**order\_item\_id**: ID của món hàng trong đơn hàng (món hàng số 1 có ID là 1, món hàng số 2 có ID là 2, v.v. Dựa vào đây ta cũng biết được mỗi đơn hàng có bao nhiêu món hàng)

**product\_id**: unique ID của sản phẩm nằm trong đơn hàng

**seller\_id**: unique ID của người bán hàng

**price**: giá của món hàng

**freight\_value**: phí ship

* **Order payments dataset**: Cung cấp thông tin về thanh toán của mỗi đơn hàng. Lưu ý là phải cộng tổng tất cả các khoản thanh toán của mỗi đơn hàng lại thì mới ra được giá trị của đơn hàng đó.

**order\_id**: unique ID của đơn hàng

**payment\_sequential**: thứ tự của thanh toán

**payment\_type**: hình thức thanh toán

**payment\_installments**: thanh toán 1 lần (payment\_installments = 1) hay trả góp (payment\_installments > 1, khi đó số tiền thanh toán sẽ được trả dần thành nhiều lần. Tổng giá trị của các lần thanh toán đó bằng payment\_value)

**payment\_value**: giá trị của thanh toán

* **Product dataset**: Cung cấp thông tin về sản phẩm

**product\_id**: unique ID của sản phẩm

**product\_category\_name**: Tên nhóm sản phẩm

**product\_name\_lenght**: Số kí tự (chữ, số) trong tên sản phẩm

**product\_description\_lenght**: Số kí tự trong phần mô tả sản phẩm

**product\_photos\_qty**: Số lượng ảnh mô tả sản phẩm

**product\_weight\_g**: Cân nặng của sản phẩm (g)

**product\_length\_cm**: Chiều dài sản phẩm (cm)

**product\_height\_cm**: Chiều cao sản phẩm (cm)

**product\_width\_cm**: Chiều rộng/sâu sản phẩm (cm)

* **Product category name translated dataset**: Dịch tên các product categories từ tiếng Bồ Đào Nha sang tiếng Anh.
* **Order reviews dataset**: Cung cấp thông tin review của mỗi đơn hàng

**review\_id**: unique ID của review

**order\_id**: unique ID của đơn hàng

**review\_score**: Điểm khách hàng đánh giá

**review\_comment\_title**: Tiêu đề của review

**review\_comment\_message**: Nội dung của review

**review\_creation\_date**: Ngày tạo review

**review\_answer\_timestamp**: Ngày giờ review được trả lời

* **Customers dataset dataset**: Cung cấp thông tin về khách hàng.

**customer\_id**: unique ID của khách hàng. Trường này dùng để link với trường customer\_id ở bảng orders\_dataset.

**customer\_unique\_id**: mã unique ID của khách hàng trong hệ thống quản lý thông tin khách hàng

**customer\_zip\_code\_prefix**: zip code của khách hàng

**customer\_city**: thành phố nơi khách hàng sống

**customer\_state**: bang nơi khách hàng sống