

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**KHOA CÔNG NGHỆ THÔNG TIN**

**BỘ MÔN BIỂU DIỄN THÔNG TIN**

-----🙞🙜🕮🙞🙜-----



**BÁO CÁO**

**ONLINE RETAIL**

Giảng viên hướng dẫn*:*

Lê Thị Nhàn

Lương Vĩ Minh

Sinh viên thực hiện: Nhóm 12

Thành phố Hồ Chí Minh, ngày 14 tháng 1 năm 2018

Mục lục

[I. Thông tin nhóm 3](#_Toc503743523)

[II. Giới thiệu đề tài 3](#_Toc503743524)

[III. Bài toán trực quan hóa 4](#_Toc503743525)

[IV. Các công trình liên quan 4](#_Toc503743526)

[IV. Xử lí dữ liệu 8](#_Toc503743527)

[V. Giải pháp trực quan hóa của nhóm 11](#_Toc503743528)

[VI. Nhận xét 11](#_Toc503743529)

[VII. Phân công công việc 12](#_Toc503743530)

[VIII. Tự đánh giá 12](#_Toc503743531)

# Thông tin nhóm

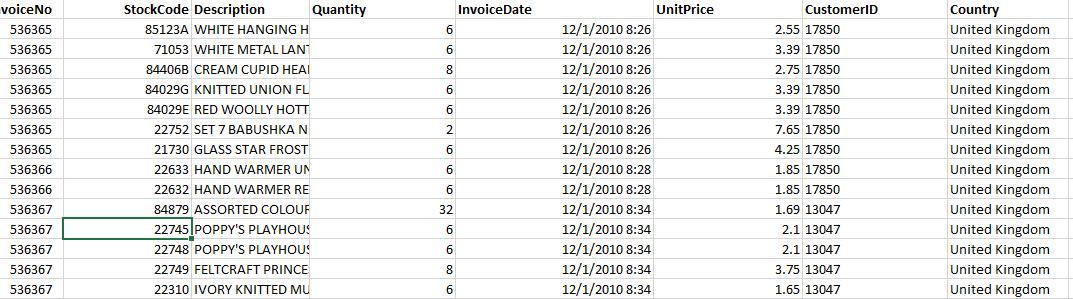
|  |  |  |  |
| --- | --- | --- | --- |
| **MSSV** | **Họ tên** | **Email** | **Chức vụ** |
| **1412304** | Trương Hữu Luân | huuluantruong@gmail.com | Trưởng nhóm |
| **1412356** | Nguyễn Đào Thảo Nguyên | ndtnguyen9596@gmail.com | Điều phối |
| **1412405** | Nguyễn Hồng Phúc | nhphuc030796@gmail.com | Thư ký |
| **1412564** | Trần Thùy Bích Trâm | bichtramtran.khtn@gmail.com | Quản lý thời gian |

# Giới thiệu đề tài

Online Retail gồm một bộ dữ liệu ghi lại toàn bộ giao dịch diễn ra trong khoảng thời gian 1/12/2010 đến 9/12/2011 của một cửa hàng online tọa lạc tại nước Anh. Công ty này chủ yếu bán mặt hàng quà lưu niệm dành cho mọi dịp lễ. Người mua chủ yếu là người bán sỉ.

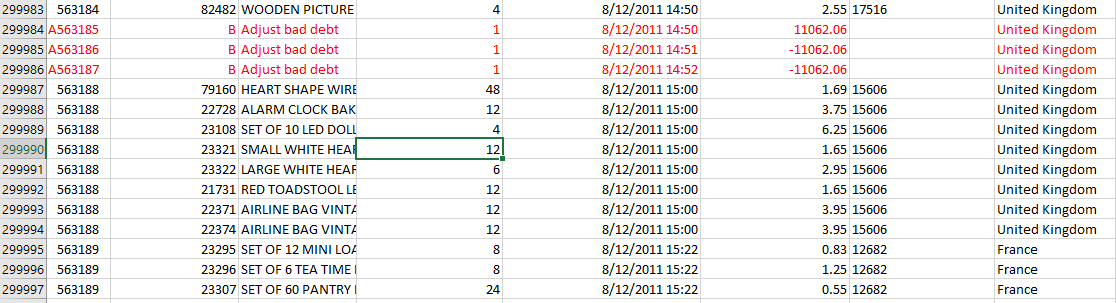
Dữ liệu cho trước bao gồm 8 thuộc tính và 541909 dòng dữ liệu. Thông tin thuộc tính:

1. InvoiceNo: Mã đơn hàng, kiểu nominal, gồm 6 kí tự số. Mỗi mã đơn hàng xác định một giao dịch duy nhất. Nếu mã đơn hàng bắt đầu bằng chữ ‘c’ thì đơn hàng đó bị hủy.
2. StockCode: Mã sản phẩm, kiểu nominal, gồm 5 kí tự số. Mỗi mã sản phẩm xác định duy nhất một sản phẩm.
3. Description: Tên sản phẩm, kiểu nominal.
4. Quantity: Số lượng của mỗi sản phẩm trong một giao dịch. Kiểu dữ liệu số (Numeric)
5. InvoiceDate : Ngày thực hiện giao dịch. Kiểu dữ liệu numeric
6. UnitPrice : Đơn giá của một sản phẩm. Kiểu dữ liệu numeric
7. CustomerID: Mã khách hàng, kiểu nominal, gồm 5 kí tự số. Mỗi mã khách hàng xác định duy nhất một khách hàng
8. Country: Tên quốc gia, kiểu nominal. Tên quốc gia cho biết nơi cư trú của khách hàng. Các quốc gia trong đề bài bao gồm 28 nước nhưng phần lớn là các nước châu Âu.



Dữ liệu rác, không hợp lí:

Chứa nhiều dòng dữ liệu có các thuộc tính không hợp lí như tên sản phẩm rỗng, đơn giá bằng 0



# Bài toán trực quan hóa

* Các câu hỏi:

1. Thị trường trọng điểm ở đâu?

* Một thị trường được đánh giá là trọng điểm hay không dựa trên tiêu chí lợi nhuận.
* So sánh lợi nhuận thu được giữa các nước
* So sánh chi phí mất mát giữa các nước

1. Thời điểm nào đạt được doanh thu cao/ thấp nhất?

* Cần biết được doanh thu theo thời gian và so sánh giữa các đoạn thời gian

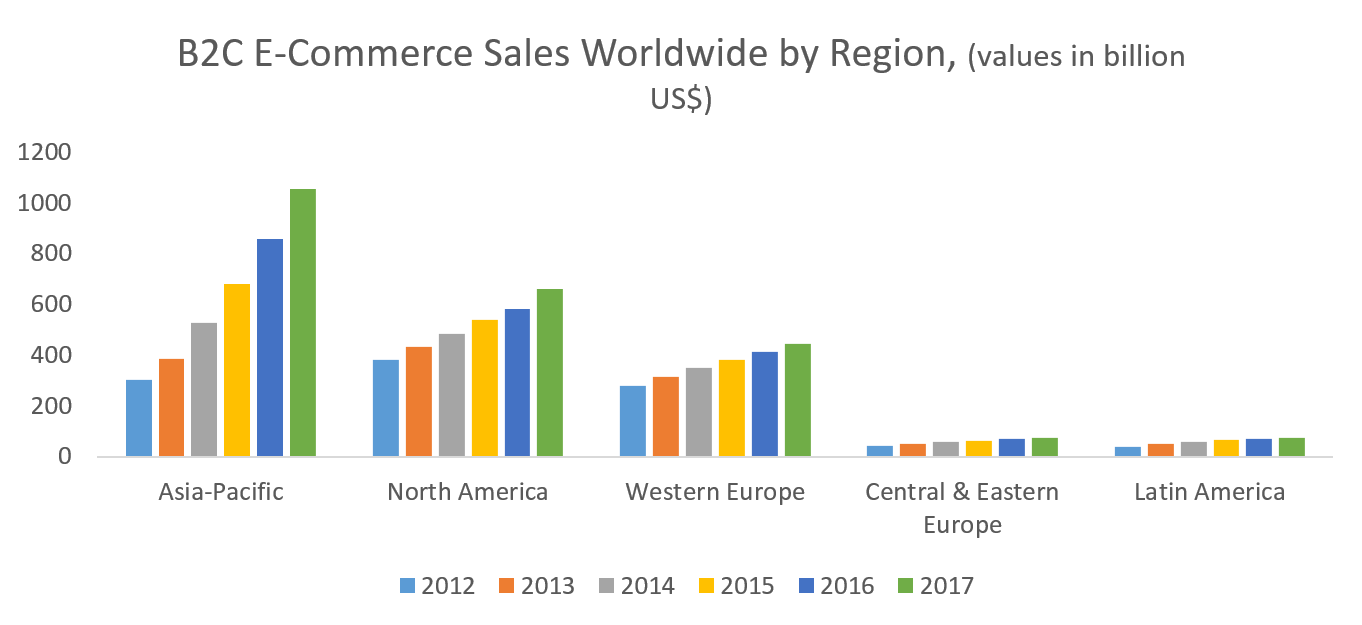
1. Sản phẩm được yêu thích nhất tại mỗi nước?

* Số lượng sản phẩm trong các đơn hàng thành công nhiều nhất trong một tháng.
* Bài toán đặt ra

Nhìn vào trực quan hóa, ta có thể thấy được thị trường trọng điểm. Làm sao thể hiện được …Đây là vấn đề quan trọng nhất mà nhóm muốn nhấn mạnh.

# Các công trình liên quan

1. Biểu đồ doanh số thương mại điện tử B2C toàn thế giới từ năm 2012 – 2017



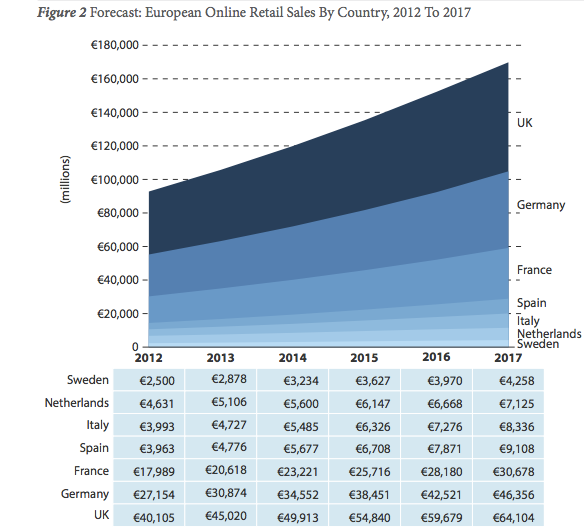
***Link:*** [*https://www.firstcircle.ph/blog/philippine-e-commerce-to-reach-200bn-by-2020?r=/blog&rt=The%20Blog*](https://www.firstcircle.ph/blog/philippine-e-commerce-to-reach-200bn-by-2020?r=/blog&rt=The%20Blog)

***Nhận xét:***

    Khu vực châu Á Thái Bình Dương đang bùng nổ về lĩnh vực thương mại điện tử trong những năm gần đây và đã vượt qua Bắc Mỹ và Tây Âu về doanh số bán hàng tiêu dùng (B2C). Biểu đồ trên cho thấy mức tăng trưởng theo cấp số nhân mà khu vực Châu Á Thái Bình Dương đang đạt được so với con số tăng trưởng tương ứng của các khu vực khác trên thế giới: Bắc Mỹ, Tây Âu, Trung và Đông Âu, Mỹ La-tinh từ năm 2012 đến 2016, bao gồm cả dự báo năm 2017.

    Theo báo cáo mới của quỹ Thương mại điện tử Châu Âu, doanh số bán lẻ thương mại điện tử khu vực Bắc Mỹ cao nhất trong năm 2012 với gần 400 tỉ USD. Tuy nhiên, Châu Á - Thái Bình Dương mới chính là khu vực có tốc độ tăng trưởng thương mại điện tử nhanh nhất trong năm khu vực chính, với tổng doanh thu là 350 tỉ USD (năm 2012) và dự kiến vượt mốc 1000 tỉ USD trong năm 2017. Doanh số bán hàng trực tuyến ở Bắc Mỹ và Tây Âu tăng trưởng ở mức ổn định từ năm 2013 đến năm 2017, tương ứng đạt 600 tỉ USD và 400 tỉ USD trong năm 2017. Bên cạnh đó, doanh thu thương mại điện tử Trung - Đông Âu và Mỹ La-tinh lại khá thấp và không có dấu hiệu tăng cao, với khoảng 10 tỉ USD năm 2012 và đạt khoảng 50 tỉ USD năm 2017.

1. Biểu đồ danh thu bán hàng trực tuyến tại các nước châu Âu



***Link:*** [*https://techcrunch.com/2013/03/13/forrester-2012-2017-ecommerce-forecast/*](https://techcrunch.com/2013/03/13/forrester-2012-2017-ecommerce-forecast/)

***Nhận xét:***

    Theo nguồn dữ liệu cung cấp bởi Forrester - thông qua The Next Web - doanh số bán lẻ trực tuyến của châu Âu dự kiến sẽ đạt 170 tỷ Euro trong năm 2017, tăng so với mức 95 tỷ Euro năm 2012. Dự báo này là kết quả của nghiên cứu trên 7 quốc gia châu Âu. Theo báo cáo, thị trường châu Âu hiện đang có sự chênh lệch đáng kể giữa các nước trong lĩnh vực mua sắm trực tuyến. Ví dụ, Forrester lưu ý rằng Thụy Điển có lượng doanh thu bán lẻ trực tuyến thấp nhất trong các nước và mức tăng trưởng này qua các năm cũng không tăng đáng kể. Trong khi đó, công ty nghiên cứu dự đoán Ý và Tây Ban Nha lại là thị trường tiềm năng khi từ 4 tỉ năm 2013 Euro đến 9 tỷ Euro năm 2017. Bên cạnh đó, Anh luôn duy trì mức doanh thu hàng đầu châu Âu.

1. Biểu đồ doanh thu bán lẻ trực tuyến ở khu vực châu Á Thái Bình Dương



***Link:*** [*https://www.insideretail.com.au/blog/2017/03/06/apac-online-sales-to-hit-trillion-dollar-mark/*](https://www.insideretail.com.au/blog/2017/03/06/apac-online-sales-to-hit-trillion-dollar-mark/)

***Nhận xét:***

    Châu Á Thái Bình Dương tiếp tục là khu vực bán lẻ trực tuyến lớn nhất, bao gồm cả thị trường lớn nhất (Trung Quốc) và thị trường phát triển nhanh nhất (Ấn Độ), theo dữ liệu mới nhất từ ​​Forrester. Với 681 tỷ USD doanh số bán lẻ trực tuyến vào năm 2016, Trung Quốc là thị trường bán lẻ trực tuyến lớn nhất trên toàn cầu. Trung Quốc chiếm gần 80% doanh số bán lẻ trực tuyến tại Châu Á Thái Bình Dương và Forrester dự đoán nó sẽ  trở thành thị trường đầu tiên đạt doanh thu bán lẻ trực tuyến lên tới hơn 1.000 tỷ USD vào năm 2021. Ấn Độ vẫn là thị trường phát triển nhanh nhất trên thế giới, với doanh thu khoảng 16 tỉ USD năm 2016 và dự kiến ​​đạt 64 tỷ USD vào năm 2021. Bên cạnh đó, việc cung cấp nhiều lựa chọn cho sản phẩm trực tuyến với giá cả hợp lý cũng giúp các thị trường Nhật Bản, Hàn Quốc và Úc đạt doanh thu khả quan, nâng tổng doanh thu bán lẻ trực tuyến của khu vực Châu Á Thái Bình Dương đạt 861 tỉ USD năm 2016 và sẽ chạm mốc 1426 tỉ USD năm 2021.

* **Nhận xét chung:**

    Từ năm 2012 đến năm 2013, Bắc Mỹ là thị trường bán lẻ trực tuyến đạt doanh thu cao nhất. Tuy nhiên, từ năm 2014 đến nay, danh hiệu này lại thuộc về Châu Á Thái Bình Dương, đồng thời đây cũng là thị trường có tốc độ tăng trưởng cao nhất, với khoảng 861 tỉ USD năm 2016. Trong đó, không thể không kể đến Trung Quốc -  thị trường bán lẻ trực tuyến lớn nhất trên toàn cầu, chiếm khoảng 80% doanh số bán lẻ trực tuyến tại khu vực Châu Á Thái Bình Dương và Ấn Độ - thị trường phát triển nhanh nhất khu vực khi được dự đoán doanh thu tăng từ 16 tỉ USD năm 2016 đến 64 tỉ USD năm 2021. Bên cạnh Châu Á Thái Bình Dương và Bắc Mỹ, Châu Âu cũng là một thị trường được đánh giá là lý tưởng và có mức doanh thu bán lẻ trực tuyến ổn định từ năm 2012 đến năm nay. Anh là thị trường trọng điểm của khu vực châu Âu, với doanh thu khoảng 69,2 tỷ USD ( 59, 679 tỉ Euro) vào năm 2016, tuy là thị trường bán lẻ trực tuyến lớn nhất khu vực châu Âu, nhưng doanh thu của Anh chỉ chỉ bằng 1/12 doanh thu khu vực Trung Quốc - thị trường bán lẻ trực tuyến lớn nhất trên toàn cầu vào cùng thời điểm.

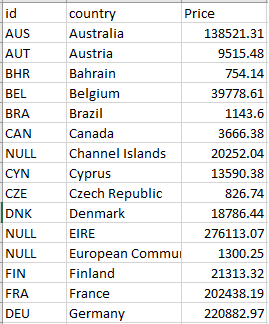
    Qua 3 biểu đồ trên, có thể thấy được tổng doanh thu bán lẻ trực tuyến tại từng khu vực trên thế giới và doanh thu cụ thể của từng nước trong một khu vực trong vài năm cụ thể. Qua đó, cũng dự đoán được tổng doanh thu bán lẻ trực tuyến của từng nước và tổng doanh thu bán lẻ trực tuyến tại một khu vực trong tương lai:

* Biểu đồ 1: Trực quan hóa dữ liệu theo mô hình cột nhóm dữ liệu các năm theo từng khu vực giúp người xem dễ dàng so sánh, đối chiếu nhanh về tổng doanh thu giữa các khu vực trong cùng một năm, cũng như tổng doanh thu giữa các năm trong cùng một khu vực.
* Biểu đồ 2: Việc sử dụng biểu đồ miền giúp người xem nhận ra được sự thay đổi về doanh thu của các nước qua từng năm và tổng doanh thu bán lẻ trực tuyến của khu vực châu Âu, tuy nhiên số liệu doanh thu của từng nước tại một năm lại chưa thể hiện rõ trong biểu đồ.
* Biểu đồ 3: Qua việc biểu diễn doanh thu bán lẻ trực tuyến của từng nước dưới dạng biểu đồ tròn, người xem có thể so sánh được doanh thu từng nước trong khu vực cũng như doanh thu từng nước qua 2 năm khác nhau.

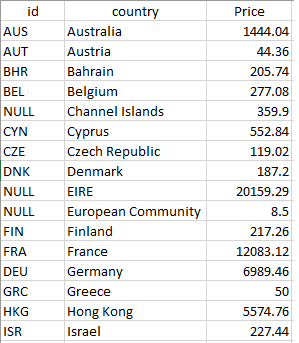
Tuy các biểu đồ trên đều giúp dữ liệu được trực quan hóa rõ ràng, nhưng với bài toán trực quan hóa mà nhóm đã đặt ra, thì việc biểu diễn dữ liệu dưới dạng các biểu đồ trên sẽ gặp nhiều khó khăn để trả lời hết các câu hỏi ấy.

# Xử lí dữ liệu

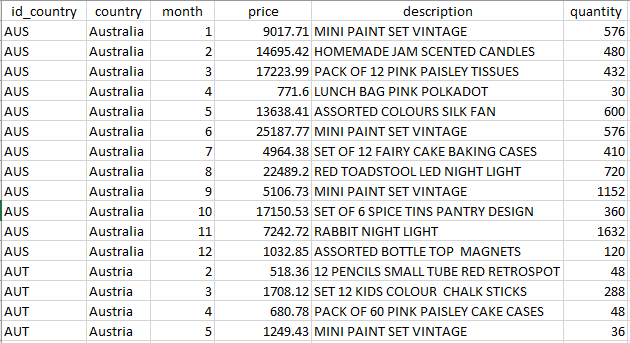
1. Loại bỏ những dòng dữ liệu khác so với mô tả.
2. Tiến hành gộp tổng doanh thu trong năm của mỗi nước



1. Tiến hành gộp tổng giá trị đơn hàng bị hủy trong năm của mỗi nước



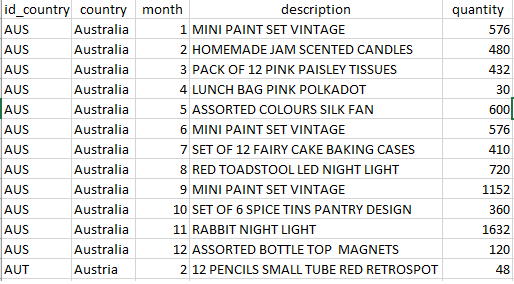
1. Tiến hành gộp tổng doanh thu sản phẩm theo từng tháng trong năm của mỗi nước



1. Tiến hành gộp tổng giá trị sản phẩm bị hủy theo từng tháng trong năm của mỗi nước



1. Tìm sản phẩm bán chạy nhất trong một tháng của mỗi nước



1. Tìm sản phẩm hay bị hủy nhất trong một tháng của mỗi nước



# Giải pháp trực quan hóa của nhóm

# Nhận xét

Tuy ý tưởng biểu diễn trực quan của nhóm khá hay, đơn giản nhưng tiện dụng, có thể giúp các công ty, doanh nghiệp bán hàng rất dễ dàng trong khâu xác định thị trường trọng điểm, thời điểm bán hàng thuận lợi nhất và mặt hàng bán chạy nhất , từ đó vạch ra chiến lược bán hàng, phân phối sản phẩm, marketing,quảng cáo,…Nhưng vì số lượng sản phẩm quá nhiều và đa dạng nên gây khó khăn trong việc truyền tải thông điệp đến người dùng một cách đơn gỉan và dễ dàng. Nhóm cần phải tìm hiểu, học hỏi, sáng tạo nhiều hơn. Tuy nhiên, đây cũng là một biểu đồ khá hợp lý cho bộ dữ liệu của nhóm, nhìn đẹp và có thể quan sát được doanh thu của từng nước, của các khu vực trên thế giới tại nhiều thời điểm trong năm.

# Phân công công việc

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Công việc** | **Ngày bắt đầu** | **Thời gian** | **Ngày kết thúc** | **Người thực hiện** |
| 1. Phân tích vấn đề | 16/10/2017 | 25 | 9/11/2017 | Cả nhóm |
| 1.1. Đọc hiểu dữ liệu + Đặt câu hỏi | 16/10/2017 | 8 | 23/10/2017 | Cả nhóm |
| 1.2. Xử lý dữ liệu | 24/10/2017 | 4 | 27/10/2017 | 1412304 |
| 1.3. Phác họa, vẽ thử hình trực quan hóa | 28/10/2017 | 4 | 31/10/2017 | 1412356 |
| 1.4. Tìm công trình liên quan, phân tích và so sánh | 24/10/2017 | 8 | 31/10/2017 | 1412405 |
| 1.5. Thiết kế slide | 24/10/2017 | 8 | 31/10/2017 | 1412564 |
| 1.6. Tổng hợp nội dung báo cáo, trình bày slide | 1/11/2017 | 3 | 4/11/2017 | Cả nhóm |
| 2. Thiết kế vấn đề | 10/11/2017 | 14 | 24/11/2017 | Cả nhóm |
| 3. Tinh lược dữ liệu | 25/11/2017 | 18 | 13/12/2017 | Cả nhóm |
| 3. Cài đặt | 14/12/2017 | 26 | 05/01/2018 | Cả nhóm |
| 4. Thiết kế Poster, viết báo cáo | 05/01/2018 | 10 | 15/01/2018 | Cả nhóm |

# Tự đánh giá

Mức độ hoàn thiện của nhóm: 8/10

Đánh giá thành viên:

|  |  |  |  |
| --- | --- | --- | --- |
| **MSSV** | **Họ tên** | **Đánh giá** | **Nhận xét** |
| **1412304** | Trương Hữu Luân | 10/10 | Hoàn thành tốt nhiệm vụ, chủ động nhận việc, học hỏi, tìm hiểu công nghệ |
| **1412356** | Nguyễn Đào Thảo Nguyên | 9/10 | Hoàn thành tốt nhiệm vụ, chủ động nhận việc, chịu khó |
| **1412405** | Nguyễn Hồng Phúc | 9/10 | Hoàn thành tốt nhiệm vụ, chủ động nhận việc, chịu khó |
| **1412564** | Trần Thùy Bích Trâm | 9/10 | Hoàn thành tốt nhiệm vụ, phân chia, điều phối công việc hợp lý |