

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

**BÁO CÁO CUỐI KỲ MÔN TƯƠNG TÁC DỮ LIỆU TRỰC QUAN**

**ĐỀ TÀI**

**PHÂN TÍCH THẢM HỌA TÀU TITANIC**

******

**GVHD: ThS Lê Minh Tân**

**Lớp: Thứ 4 ( tiết 1-4 )**

**Sinh viên thực hiện: Nhóm**

**Nguyễn Hiếu Gia Cường MSSV: 20133027**

**Huỳnh Công Hậu MSSV: 20133039**

**Phan Hoàng Việt MSSV: 17133072**

[**LỜI CẢM ƠN** 4](#_Toc134387377)

[**CHƯƠNG 1: LÝ DO CHỌN DATASET VÀ GIỚI THIỆU TỔNG QUAN DATASET** 6](#_Toc134387378)

[**1. Lời mở đầu** 7](#_Toc134387379)

[**1.1.** **Vấn đề nhận thấy** 7](#_Toc134387380)

[**1.2.** **Giải pháp** 7](#_Toc134387381)

[**1.3.** **Mục tiêu và ý nghĩa của dự án** 7](#_Toc134387382)

[**1.4.** **Giới thiệu tổng quan Dataset** 9](#_Toc134387383)

[**1.4.1.** **Nguồn dữ liệu sử dụng** 9](#_Toc134387384)

[**1.4.2.** **Giới thiệu nơi cấp dữ liệu** 9](#_Toc134387385)

[**1.4.3.** **Hướng dẫn tải dataset thực hiện trong đồ án và các dataset khác của nhà cung cấp** 10](#_Toc134387388)

[**1.5.** **Mô tả chi tiết dữ liệu** 10](#_Toc134387389)

[**1.6.** **Thông số dataset** 10](#_Toc134387390)

[**1.7.** **Dữ liệu sau khi trích xuất** 11](#_Toc134387391)

[**1.8.** **Mô tả chi tiết các thuộc tính trong dataset** 12](#_Toc134387392)

[**1.9.** **Giới thiệu các công cụ được sử dụng trong đồ án** 13](#_Toc134387393)

[**1.9.1.** **Tổng quan Zeppelin** 13](#_Toc134387394)

[**1.9.2.** **Giới thiệu Python3** 13](#_Toc134387395)

[**CHƯƠNG 2: THIẾT KẾ XÂY DỰNG DASHBOARD** 15](#_Toc134387396)

[**2.1.** **Quá trình nạp dữ liệu vào zeppeline** 15](#_Toc134387398)

[**2.2.** **Kết hợp các bảng** 16](#_Toc134387399)

[**2.3.** **Quá trình EDA** 17](#_Toc134387401)

[**2.3.1.** **Xem số lượng các dòng bị thiếu trong DataFrame** 17](#_Toc134387402)

[**2.3.2.** **Xóa các dòng có giá trị null** 17](#_Toc134387404)

[**2.3.3.** **Xem mô tả của từng cột** 18](#_Toc134387406)

[**2.3.4.** **Tạo biểu đồ histogram để thấy phân phối các biến số** 18](#_Toc134387408)

[**2.3.5.** **Tạo biểu đồ countplot để xem lượng giá trị của các biến phân loại**  19](#_Toc134387410)

[**2.3.6.** **Tạo biểu đồ violinplot để xem phân bố của các biến số với các biến phân loại** 19](#_Toc134387412)

[**2.3.7.** **Xóa cột SibSp Parch** 20](#_Toc134387414)

[**2.3.8.** **Sort theo độ tuổi** 20](#_Toc134387416)

[**2.3.9.** **Tính toán trung bình tuổi** 20](#_Toc134387418)

[**2.3.10.** **Tạo cột giá trị tuổi trung bình với những giá trị null được thay thế bằng mean(age)** 21](#_Toc134387420)

[**2.3.11.** **Thay thế cột Age** 21](#_Toc134387422)

[**2.4.** **Biểu đồ tròn thể hiện phần trăm sống sót trên tàu** 22](#_Toc134387424)

[**2.5.** **Biểu đồ tròn thể hiện số hành khách nam sống/chết trên tàu** 23](#_Toc134387427)

[**2.6.** **Biểu đồ tròn thể hiện số hành khách nữ sống/chết trên tàu** 24](#_Toc134387430)

[**2.7.** **Biểu đồ cột thể hiện phân phối độ tuổi trên tàu** 25](#_Toc134387432)

[**2.8.** **Tạo biểu đồ đường so sánh tỷ lệ sống còn theo PCLASS** 25](#_Toc134387434)

[**2.9.** **Tạo ComboBox hiển thị 3 biểu đồ tròn** 26](#_Toc134387436)

[**CHƯƠNG 3: KẾT LUẬN** 27](#_Toc134387439)

[**3.1.** **Kết quả đạt được** 28](#_Toc134387440)

[**3.2.** **Kết luận rút ra được từ dashboard** 28](#_Toc134387441)

[**3.3.** **Những hạn chế** 29](#_Toc134387442)

[**3.4.** **Bảng phân công nhiệm vụ trong nhóm** 30](#_Toc134387443)

[**3.5.** **Tài liệu tham khảo** 31](#_Toc134387505)

# LỜI CẢM ƠN

Lời đầu tiên, nhóm em xin được gởi lời cảm ơn đặc biệt đến Thầy -Lê Minh Tân - Giảng viên phụ trách môn Tương Tác Dữ Liệu Trực Quan – trường đại học Sư Phạm Kỹ Thuật Tp Hồ Chí Minh .

Trong thời gian nhóm em làm đồ án , tụi em đã nhận được nhiều sự giúp đỡ từ thầy. Thầy đã cung cấp đầy đủ kiến thức, chỉ bảo và đóng góp những ý kiến quý báu giúp tụi em có thể hoàn thành được đồ án môn học của mình một cách tốt nhất.

Xuất phát từ mục đích học tập, tìm hiểu sâu hơn các kiến thức về dữ liệu và các thuật toán, cũng như tìm hiểu kỹ về quy trình lên ý tưởng, xây dựng dashboard. Nhóm chúng em đã thực hiện đồ án “Xây dựng dashboard để phân tích sự kiện tàu Titanic.”. Trong quá trình thực hiện đồ án, dựa trên kiến thức được Thầy cung cấp qua các buổi học lý thuyết cũng như thực hành trên lớp, kết hợp với việc tự tìm hiểu những công cụ và kiến thức mới , nhóm đã cố gắng thực hiện đồ án một cách tốt nhất .Tuy nhiên, đồ án còn chưa được hoàn hiện và có nhiều sai sót.

Nhóm rất mong nhận được sự góp ý từ thầy nhằm rút ra những kinh nghiệm quý báu và hoàn thiện vốn kiến thức để nhóm có thể hoàn thành những đồ án, dự án khác trong tương lai .

**Nhóm chúng em xin chân thành cảm ơn thầy!**

**Lời Nhận Xét Của Giảng Viên**

**CHƯƠNG 1: LÝ DO CHỌN DATASET VÀ GIỚI THIỆU TỔNG QUAN DATASET**

*Giới thiệu tổng quan về dataset, lý do hình thành dự án, nguồn dữ liệu thực hiện. Khảo sát, nghiên cứu và phân tích các báo cáo nghiệp vụ cần phục vụ cho dự án.*

* 1. **Lời mở đầu**
  2. **Vấn đề nhận thấy**

Dự án được hình thành từ tập dữ liệu của chuyến tàu Titanic bởi vì đây là một trong những sự kiện lịch sử nổi tiếng và đầy cảm xúc, khiến cho nhiều người quan tâm và muốn tìm hiểu.

Dữ liệu được thu thập từ hành khách và phi hành đoàn trên tàu, bao gồm thông tin về tuổi, giới tính, hạng ghế, điểm đến và liệu họ đã sống sót hay không trong vụ đắm tàu. Bằng cách phân tích tập dữ liệu này, ta có thể đưa ra được những thông tin và kiến thức mới về sự kiện đắm tàu Titanic và phương pháp xây dựng mô hình dự đoán khả năng sống sót của một hành khách.

* 1. **Giải pháp**

Dựa trên nhu cầu thống kê, phân tích và khai thác dữ liệu các hành khách trên tàu. Giải pháp là xây dựng dashboard phục vụ mục đích phân tích, khai thác, và tạo báo cáo tổng. Đưa ra các kết quả phù hợp.

* 1. **Mục tiêu và ý nghĩa của dự án**

Thứ nhất là nghiên cứu học tập xây dựng một dashboard phân tích dữ liệu.

Xây dựng và phát triển ứng dụng nhằm phục vụ việc phân tích, khai thác, tạo cáo nhằm nắm rỏ xu hướng công nghệ đang được ưa chuộng, sự phân câp giữa các lâp trình viên, mức lương của họ theo từng quốc gia. Việc này giúp dễ dàng nhận biết xu hướng công nghệ, so sánh mức lương giữa các quốc gia.

Mục tiêu của dự án là phân tích và dự đoán khả năng sống sót của hành khách trên chuyến tàu Titanic dựa trên các thông tin về đặc điểm cá nhân, hành lý và hạng ghế của họ. Dự án này có ý nghĩa quan trọng trong việc nghiên cứu về sự kiện Titanic, giúp chúng ta hiểu rõ hơn về nguyên nhân, tác động và hậu quả của thảm họa này đối với con người và xã hội. Ngoài ra, kết quả của dự án cũng có thể áp dụng vào việc dự đoán rủi ro và đưa ra các biện pháp phòng ngừa cho các tình huống khẩn cấp tương tự trong tương lai.

Hướng tới đối tượng sử dụng là các nhà phát triển muốn cập nhật thông tin nhằm phát triển nên các biện pháp đối phó với thảm họa tự nhiên, các doanh nghiệp có nhu cầu sản xuất các thiết bị vận tải, …

* 1. **Giới thiệu tổng quan Dataset**
     1. **Nguồn dữ liệu sử dụng**

Nguồn dữ liệu được thu thập từ kaggle.com, dataset kaggle.com **Titanic – Machine Learning from Disaster.**

* + 1. **Giới thiệu nơi cấp dữ liệu**

Kaggle là một trang web chuyên về kho dữ liệu và cuộc thi phân tích dữ liệu trực tuyến. Trang web này cung cấp các tập dữ liệu từ nhiều lĩnh vực khác nhau, từ kinh doanh đến khoa học và công nghệ. Ngoài ra, Kaggle còn tổ chức các cuộc thi về phân tích dữ liệu và học máy, cho phép các nhà khoa học dữ liệu, chuyên gia phân tích và lập trình viên trên toàn thế giới tham gia để giải quyết các thách thức về dữ liệu. Việc tham gia các cuộc thi này giúp các chuyên gia phát triển kỹ năng phân tích dữ liệu và học máy, cũng như giúp các công ty và tổ chức giải quyết các vấn đề phức tạp liên quan đến dữ liệu của mình.

Kaggle có:

* Hơn 4 triệu truy cập mỗi tháng (theo SimilarWeb)
* Hơn 22.000 câu hỏi (tính đến tháng 4 năm 2021)
* Hơn 35.000 câu hỏi từ nhiều lĩnh vực khác nhau như tài chính, y tế, thể thao, v.v
* Hơn 120.000 câu trả lời
* Số cuộc thi: Kaggle đã tổ chức hơn 300 cuôc thi về khoa học dữ liệu và AI từ 2010
  + 1. **Hướng dẫn tải dataset thực hiện trong đồ án và các dataset khác của nhà cung cấp**

Link tải dataset: https://www.kaggle.com/competitions/titanic/data?select=train.csv

<https://www.kaggle.com/competitions/titanic/data?select=test.csv>

https://www.kaggle.com/competitions/titanic/data?select=gender\_submission.csv

* 1. **Mô tả chi tiết dữ liệu**

Tập dữ liệu Titanic là một trong những tập dữ liệu phổ biến nhất trong lĩnh vực khoa học dữ liệu. Nó chứa thông tin về hành khách trên chuyến tàu Titanic bao gồm tên, giới tính, tuổi, hạng ghế, số lượng người thân, địa chỉ, số phiếu đặt chỗ, giá tiền vé, khu vực lên tàu, thông tin về hành trang và sống hay chết trong thảm họa đắm tàu. Tập dữ liệu này được sử dụng rộng rãi để phân tích về các yếu tố có ảnh hưởng đến khả năng sống sót của hành khách trên tàu.

Ngoài ra, tập dữ liệu này còn được sử dụng để huấn luyện và đánh giá mô hình học máy, xây dựng mô hình dự đoán khả năng sống sót của hành khách trên tàu dựa trên các thông tin có sẵn và nhiều ứng dụng khác trong lĩnh vực khoa học dữ liệu.

* 1. **Thông số dataset**

Dữ liệu gồm có 2 bảng:

**Train.csv** có: 892 (dòng) \* 12(cột), với mỗi dòng thể hiện một hành khách trên tàu.

**Test.csv** có: 419(dòng) \* 11(cột), với mỗi dòng thể hiện một hành khách trên tàu.

**Gender\_submission.csv** có: 419(dòng) \* 2(cột), với mỗi dòng thể hiện một id của hành khách trên tàu.

* 1. **Dữ liệu sau khi trích xuất**

Thực hiện lấy dữ liệu từ bảng **Train.csv, Test.csv** và **Gender\_submission.csv:**

**A screenshot of a computer

Description automatically generated with medium confidence**

**A screenshot of a computer

Description automatically generated with medium confidence**

**A screenshot of a computer

Description automatically generated with medium confidence**

* 1. **Mô tả chi tiết các thuộc tính trong dataset**

|  |  |
| --- | --- |
| **Tên thuộc tính** | **Mô tả** |
| **PassengerId** | ID hành khách trên tàu |
| **Survived** | 1 – còn sống, 0 – chết |
| **Pclass** | Lớp ghế mà hành khách đang ngồi trên tàu (1 = hạng nhất, 2 = hạng 2, 3 = hạng 3) |
| **Name** | Họ tên hành khách |
| **Sex** | Giới tính của hành khách (nam = male, nữ = female) |
| **Age** | Độ tuổi của hành khách |
| **SibSp** | Số lượng anh/chị/em của hành khách cùng đi trong chuyến đi |
| **Parch** | Số lượng cha/mẹ/con của hành khách cùng đi trong chuyến đi |
| **Ticket** | Số vé của hành khách |
| **Fare** | Giá vé của hành khách |
| **Cabin** | Số hiệu phòng của hành khách |
| **Embarked** | Cảng lên tàu của hành khách ( C = Cherbourg, Q = Queenstown, S = Southampton) |

* 1. **Giới thiệu các công cụ được sử dụng trong đồ án**

Công cụ được sử dụng trong đồ án này là: Zeppelin

Ngôn ngữ lập trình: Python

* + 1. **Tổng quan Zeppelin**

Apache Zeppelin là một ứng dụng web mã nguồn mở để thực thi, quản lý và chia sẻ các notebook tương tác cho phép thực hiện phân tích dữ liệu, truy xuất cơ sở dữ liệu, thực thi các thuật toán và trình bày các kết quả dưới dạng tài liệu tương tác. Zeppelin hỗ trợ nhiều ngôn ngữ lập trình và các công cụ phân tích dữ liệu phổ biến như Python, R, Spark, SQL, Flink, Cassandra và nhiều hơn nữa. Nó cung cấp cho người dùng một cách tiếp cận tương tác để phân tích dữ liệu và có khả năng tạo ra tài liệu tương tác với các biểu đồ, bảng và hình ảnh.

Zeppelin là một công cụ mạnh mẽ và phổ biến trong cộng đồng phân tích dữ liệu và được sử dụng rộng rãi trong các dự án phát triển và nghiên cứu khoa học.

* + 1. **Giới thiệu Python3**

Python là một ngôn ngữ lập trình bậc cao cho các mục đích lập trình đa năng, do Guido van Rossum tạo ra và lần đầu ra mắt vào năm 1991. Python được thiết kế với ưu điểm mạnh là dễ đọc, dễ học và dễ nhớ. Python là ngôn ngữ có hình thức rất đơn giản, cấu trúc rõ ràng, thuận tiện cho người mới học lập trình. Cấu trúc của Python còn cho phép người sử dụng viết mã lệnh với số lần gõ phím tối thiểu. Vào tháng 7 năm 2018, Van Rossum đã từ chức Leader trong cộng đồng ngôn ngữ Python sau 30 năm lãnh đạo.

Python hoàn toàn tạo kiểu động và dùng cơ chế cấp phát bộ nhớ tự động, do vậy nó tương tự như Perl, Ruby, Scheme, Smalltalk, và Tcl. Python được phát triển trong một dự án mã mở, do tổ chức phi lợi nhuận Python Software Foundation quản lý.

Ban đầu, Python được phát triển để chạy trên nền Unix. Nhưng rồi theo thời gian, nó đã "bành trướng" sang mọi hệ điều hành từ MS-DOS đến Mac OS, OS/2, Windows, Linux và các hệ điều hành khác thuộc họ Unix. Mặc dù sự phát triển của Python có sự đóng góp của rất nhiều cá nhân, nhưng Guido van Rossum hiện nay vẫn là tác giả chủ yếu của Python. Ông giữ vai trò chủ chốt trong việc quyết định hướng phát triển của Python.

Sau đây là các đặc điểm của **Python**:

* Ngữ pháp đơn giản, dễ đọc.
* Vừa hướng thủ tục (procedural-oriented), vừa hướng đối tượng (object-oriented)
* Hỗ trợ module và hỗ trợ gói (package)
* Xử lý lỗi bằng ngoại lệ (Exception)
* Kiểu dữ liệu động ở mức cao.
* Có các bộ thư viện chuẩn và các module ngoài, đáp ứng tất cả các nhu cầu lập trình.
* Có thể nhúng vào ứng dụng như một giao tiếp kịch bản (scripting interface).

**CHƯƠNG 2: THIẾT KẾ XÂY DỰNG DASHBOARD**

Trình bày chi tiết các bước thực hiện trong đồ án

* 1. **Quá trình nạp dữ liệu vào zeppeline**

Import tập dữ liệu **train.csv**, **test.csv** và **gender\_submission.csv**

A picture containing text, screenshot, line, number

Description automatically generated

* 1. **Kết hợp các bảng**

Kết hợp 2 tập dữ liệu **test.csv** và **gender\_submission.csv**

A picture containing text, line, screenshot

Description automatically generated

Kết hợp 2 tập dữ liệu **train.csv và merge.df12** để tạo thành tập dữ liệu hoàn chỉnh

A picture containing text, screenshot, line

Description automatically generated

* 1. **Quá trình EDA**
     1. **Xem số lượng các dòng bị thiếu trong DataFrame**

A picture containing text, screenshot

Description automatically generated

* + 1. **Xóa các dòng có giá trị null**

A screenshot of a computer

Description automatically generated with low confidence

* + 1. **Xem mô tả của từng cột**

A screenshot of a computer

Description automatically generated with low confidence

* + 1. **Tạo biểu đồ histogram để thấy phân phối các biến số**

A picture containing diagram, text, plot, screenshot

Description automatically generated

* + 1. **Tạo biểu đồ countplot để xem lượng giá trị của các biến phân loại**

A screenshot of a computer

Description automatically generated with low confidence

* + 1. **Tạo biểu đồ violinplot để xem phân bố của các biến số với các biến phân loại**

A screenshot of a computer

Description automatically generated with medium confidence

* + 1. **Xóa cột SibSp Parch**

A picture containing text, screenshot, font, number

Description automatically generated

* + 1. **Sort theo độ tuổi**

A screenshot of a computer

Description automatically generated with medium confidence

* + 1. **Tính toán trung bình tuổi**

A picture containing text, font, screenshot, algebra

Description automatically generated

* + 1. **Tạo cột giá trị tuổi trung bình với những giá trị null được thay thế bằng mean(age)**

A picture containing text, screenshot

Description automatically generated

* + 1. **Thay thế cột Age**

A screenshot of a computer

Description automatically generated with medium confidence

* 1. **Biểu đồ tròn thể hiện phần trăm sống sót trên tàu**

A screenshot of a computer

Description automatically generated with medium confidence

A blue and red pie chart

Description automatically generated with medium confidence

* 1. **Biểu đồ tròn thể hiện số hành khách nam sống/chết trên tàu**

A screenshot of a computer program

Description automatically generated with low confidence

A blue and red pie chart

Description automatically generated with low confidence

* 1. **Biểu đồ tròn thể hiện số hành khách nữ sống/chết trên tàu**

A screenshot of a computer program

Description automatically generated with low confidenceA blue and red pie chart

Description automatically generated with medium confidence

* 1. **Biểu đồ cột thể hiện phân phối độ tuổi trên tàu**

A picture containing text, screenshot, diagram, plot

Description automatically generated

* 1. **Tạo biểu đồ đường so sánh tỷ lệ sống còn theo PCLASS**

A screenshot of a computer

Description automatically generated with medium confidence

* 1. **Tạo ComboBox hiển thị 3 biểu đồ tròn**

A screenshot of a computer

Description automatically generated with medium confidence

A blue and red pie chart

Description automatically generated with low confidence

**CHƯƠNG 3: KẾT LUẬN**

*Báo cáo kết quả được sau khi hoàn thành đồ án, nêu ra những hạn chế còn tồn tại, định hướng phát triển trong trương lai, các tài liệu tham khảo. Bảng phân công công việc cụ thể và mức độ hoàn thành công việc.*

* 1. **Kết quả đạt được**

Trong đề tài này, nhóm đã tìm hiểu và vận dụng kiến thức đạt được kết quả như sau :

* Áp dụng được biểu đồ có thể tương tác bằng cách trỏ chuột hiện chú thích và đổi màu.
* Năm vững kiến thức và có thể vận dụng , xây dựng một dashboard hoàn chỉnh dùng để khai thác dữ liệu.
* Áp dụng được biểu đồ có sử dụng combobox.
* Có sử dụng nhiều bảng dữ liệu kết hợp với nhau.
* Áp dụng các thao tác xóa cột, nhóm (group), nối (concat), kết hợp (join), lọc (filter), thay thế cột bằng tính toán dữ liệu.
* Tạo ra được biểu đồ thể hiện đường mục tiêu.
  1. **Kết luận rút ra được từ dashboard**

Dựa trên phân tích dữ liệu từ dự án Titanic trên Kaggle, chúng ta thu được các kết quả sau:

* Tỷ lệ sống còn: Tỷ lệ sống còn trên tàu Titanic là khoảng 38%, điều này cho thấy thảm họa Titanic đã gây ra nhiều thiệt hại và gây mất mát lớn.
* Giới tính: Tỷ lệ sống còn của phụ nữ là khoảng 74%, trong khi tỷ lệ sống còn của nam giới chỉ là khoảng 19%. Điều này cho thấy có sự ưu tiên trong việc cứu hộ cho phụ nữ trong thảm họa này.
* Lớp hành khách: Hành khách ở lớp hạng nhất có tỷ lệ sống còn cao hơn so với hành khách ở lớp hạng thứ hai và thứ ba. Điều này cho thấy sự ưu tiên trong việc cứu hộ cho nhóm hành khách giàu có và có địa vị xã hội cao.
* Tuổi: Tỷ lệ sống còn của trẻ em (dưới 18 tuổi) cao hơn so với người lớn. Tuy nhiên, người cao tuổi (trên 65 tuổi) có tỷ lệ sống còn thấp nhất. Điều này có thể cho thấy việc ưu tiên cứu hộ cho nhóm yếu thế và khả năng di chuyển hạn chế của người cao tuổi.

Dựa trên các kết quả trên, chúng ta có thể rút ra một số nhận định và khuyến nghị như sau:

* Nâng cao quy định và chuẩn bị cho các biện pháp cứu hộ để tăng cường khả năng sống sót trong các tình huống thảm họa tương tự.
* Tăng cường sự chú trọng đến an toàn và chuẩn bị sẵn sàng cho nhóm yếu thế như trẻ em và người cao tuổi trong các kế hoạch cứu hộ.
* Đảm bảo sự công bằng và không phân biệt đối xử dựa trên giới tính trong quá trình cứu hộ và phân chia tài nguyên trong tình huống khẩn cấp.

Tuy nhiên, cần lưu ý rằng kết quả phân tích chỉ dựa trên dữ liệu từ dự án Titanic và không thể áp dụng trực tiếp vào các tình huống khác. Để có kết quả phân tích chi tiết và đáng tin cậy hơn.

* 1. **Những hạn chế**
* Chưa hoàn chỉnh ComboBox (định dạng layout, hiển thị title của đồ thị).
* Quá trình EDA rườm rà và có thể chưa đầy đủ.
* Biểu đồ tương tác còn khá đơn giản.
* Chưa xử lý thay đổi dữ liệu.
  1. **Bảng phân công nhiệm vụ trong nhóm**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Nhiệm vụ** | **Người phụ trách** | **Ghi chú** |
| 1 | EDA, Word | Cường, Hậu, Việt | Hoàn thành 100% |
| 2 | Tìm dataset | Hậu, Cường, Việt | Hoàn thành 100% |
| 3 | Vẽ biểu đồ | Việt, Hậu, Cường | Hoàn thành 100% |
| 4 |  |  |  |
| 5 |  |  |  |
| 6 |  |  |  |
| 7 |  |  |  |
| 8 |  |  |  |
| 9 |  |  |  |
| 10 |  |  |  |
| 11 |  |  |  |
| 12 |  |  |  |

* 1. **Tài liệu tham khảo**
* Slide bài giảng của thầy Lê Minh Tân
* Công cụ sử dụng:
  + Pham Thi Hong Anh, Giới thiệu Visualization với ploty với tập dataset Titanic, 16/01/2018, <https://viblo.asia/p/gioi-thieu-visualization-data-voi-plotly-voi-tap-dataset-titanic-gAm5y8bqldb>, truy cập ngày: 5/5/2023.
  + Vietnambiz, Biểu đồ hộp (Box Plot) là gì ? Đặc trưng và ví dụ, 12/11/2019, <https://vietnambiz.vn/bieu-do-hop-box-plot-la-gi-dac-trung-va-vi-du-20191112102052212.htm>, truy cập ngày 6/5/2023.