

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO LAB 01
TẬP DỮ LIỆU TITANIC

Môn học: Trực quan hóa dữ liệu – 21KHDL

Thành viên nhóm:

- Lê Nguyễn Kiều Oanh 21127129
- Đoàn Ngọc Mai 21127104

Giảng viên hướng dẫn:

- Thầy Bùi Tiến Lên
- Thầy Lê Ngọc Thành
- Thầy Lê Nguyễn Nhựt Tường

Thành phố Hồ Chí Minh, ngày 7 tháng 4 năm 2024

Mục lục

| | |
|--|---|
| I. Thông tin nhóm | 3 |
| II. Bảng phân công công việc và đánh giá mức độ hoàn thành | 3 |
| III. Chia sẻ những phát hiện từ tập dữ liệu | 4 |

I. Thông tin nhóm

| STT | Họ và tên | MSSV |
|-----|---------------------|----------|
| 1 | Lê Nguyễn Kiều Oanh | 21127129 |
| 2 | Đoàn Ngọc Mai | 21127104 |

II. Bảng phân công công việc và đánh giá mức độ hoàn thành

| Họ và tên | MSSV | Công việc | Mức độ hoàn thành |
|---------------------|----------|--|-------------------|
| Lê Nguyễn Kiều Oanh | 21127129 | Hoàn thành phần Phân tích và khám phá dữ liệu 2D | 100% |
| | | Hoàn thành phần Phân tích và khám phá dữ liệu 3D | 100% |
| | | Hoàn thành phần Phân tích và khám phá chi tiết đặc trưng | 100% |
| | | Hoàn thành phần Kết luận | 100% |
| | | Hoàn thành Báo cáo | 100% |
| Đoàn Ngọc Mai | 21127104 | Hoàn thành phần Mô tả dữ liệu | 100% |
| | | Hoàn thành phần Tìm hiểu về tập dữ liệu | 100% |
| | | Hoàn thành phần Phân tích và khám phá dữ liệu 1D | 100% |
| | | Hoàn thành phần Đặt câu hỏi và trả lời | 100% |

III. Chia sẻ những phát hiện từ tập dữ liệu

- Trong cùng một hạng vé nhưng hành khách có thể mua với những mức giá khác nhau, có người mua được với giá cao và cũng có người mua được với giá thấp.
- Vé tàu ở cảng Cherbourg được bán với giá cao nhất và ở cảng Queenstown có giá vé thấp nhất.
- Trên tàu, gần 70% hành khách lên từ cảng Southampton và khoảng 47% trong số họ là nam giới và trên tàu nam giới chiếm số đông hơn so với nữ giới.
- Hành khách mua vé hạng 1 chủ yếu đến từ cảng Cherbourg, trong khi hành khách mua vé hạng 3 thường đến từ cảng Southampton. Đặc biệt, những hành khách từ cảng Queenstown không chỉ mua vé với giá ưu đãi hơn so với các cảng khác cho cả ba hạng vé, mà cũng chiếm tỷ lệ ít hơn so với các cảng khác.
- Những người phụ nữ, đặc biệt là những phụ nữ chưa kết hôn có xác suất sống sót cao nhất.
- Phụ nữ kết hôn thích hạng vé 1 hơn hạng vé 3 cho thấy một xu hướng tìm kiếm sự thoải mái và tiện nghi cao hơn trong khi đi du lịch. Sự ưa thích này có thể liên quan đến việc tìm kiếm một trải nghiệm cao cấp và an toàn hơn, có thể là do họ đi cùng gia đình nên họ chọn điều tốt nhất.
- Những người đàn ông chi ít tiền mua vé nhất trong khi những người phụ nữ đã kết hôn lại chi rất nhiều tiền để mua vé .
- Những hành khách trên tàu phần lớn là đi một mình hoặc là đi với gia đình từ 2-4 người.
- Nhóm hành khách đi một mình (Solo) thường có tỷ lệ sống sót thấp, trong khi nhóm hành khách đi với gia đình từ 2-4 người thường có tỷ lệ sống sót cao nhất.
- Nhóm hành khách đi một mình (Solo) mua vé với giá rẻ nhất trong khi những hành khách đi với gia đình 5-7 người thì mua vé với giá đắt nhất.
- Hạng vé 3 được ưa chuộng nhất ở hai điểm lên tàu S và Q, trong khi đó ở C, hạng vé 1 là phổ biến nhất.
- Phân bố tuổi của hành khách trên Titanic chủ yếu tập trung ở độ tuổi từ 20 đến 30 và rất ít hành khách trên 70 tuổi.