**Giải thích các cột trong dữ liệu Titanic**

| **Tên cột** | **Ý nghĩa** |
| --- | --- |
| PassengerId | Mã số định danh duy nhất cho mỗi hành khách |
| Pclass | **Hạng vé** của hành khách (1 = hạng nhất, 2 = hạng nhì, 3 = hạng ba) |
| Name | Họ và tên đầy đủ của hành khách |
| Sex | Giới tính (male = nam, female = nữ) |
| Age | Tuổi của hành khách (có thể bị thiếu) |
| SibSp | Số lượng **anh chị em/vợ chồng** đi cùng hành khách |
| Parch | Số lượng **cha mẹ/con cái** đi cùng hành khách |
| Ticket | Số vé tàu (không mang nhiều ý nghĩa phân tích) |
| Fare | Giá vé mà hành khách đã trả (số thực, đơn vị: bảng Anh) |
| Cabin | Mã phòng/lều trên tàu (rất nhiều giá trị bị thiếu) |
| Embarked | Cảng mà hành khách **lên tàu**:  - `C` = Cherbourg  - `Q` = Queenstown  - `S` = Southampton (phổ biến nhất) |
| Survived | **Kết quả sống sót**:  - 0 = Không sống sót - 1 = Sống sót (chỉ có trong train.csv) |

| **PassengerId** | **Pclass** | **Name** | **Sex** | **Age** | **SibSp** | **Parch** | **Fare** | **Embarked** | **Survived** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 3 | Braund, Mr. Owen Harris | male | 22 | 1 | 0 | 7.25 | S | 0 |
| 2 | 1 | Cumings, Mrs. John Bradley | female | 38 | 1 | 0 | 71.3 | C | 1 |

**Một số lưu ý:**

* **Pclass** có thể phản ánh mức sống => người hạng 1 dễ sống sót hơn.
* **Sex** là yếu tố ảnh hưởng mạnh đến khả năng sống sót (nữ ưu tiên hơn).
* **SibSp** và **Parch** giúp xác định hành khách có đi theo gia đình hay không.
* **Cabin** nhiều giá trị bị thiếu, thường sẽ bỏ hoặc xử lý đặc biệt.
* **Embarked** có thể ảnh hưởng đến nhóm người sống sót do lên tàu từ cảng khác nhau.

