1. **데이터 수집 및 EDA , 전처리**

0 Survived 891 non-null int64

1 Pclass 891 non-null int64

2 Name 891 non-null object

3 Sex 891 non-null object

4 Age 714 non-null float64

5 SibSp 891 non-null int64

6 Parch 891 non-null int64

7 Ticket 891 non-null object

8 Fare 891 non-null float64

9 Cabin 204 non-null object

10 Embarked 889 non-null object

EDA 및 전처리

Sex – male = 0, Female = 1

Embarked – Q, S, C = oneHotEncoding

SibSp, Parch => SibSp + Parch = Companion

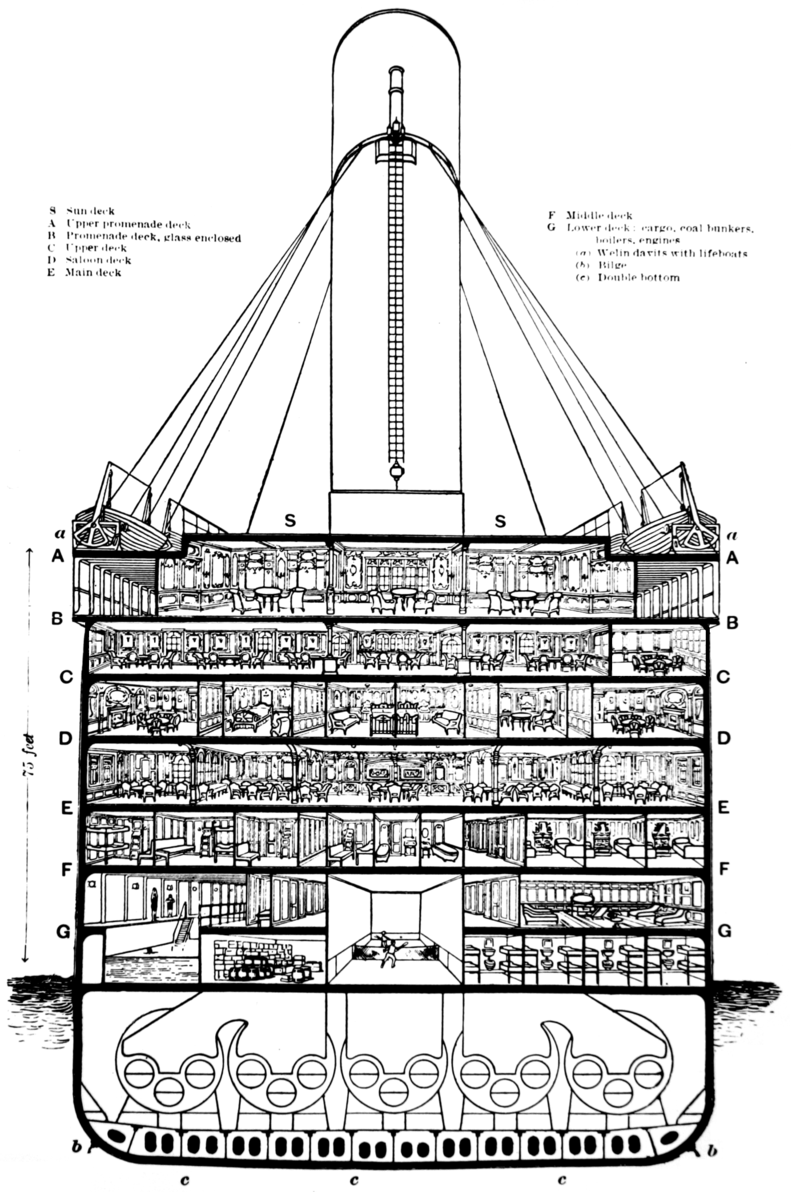
Fare => Fare / Companions = TicketPer1

결측치 예측

Age = Name에 호칭에서 예측

Embarked = 최빈값으로 예측

Cabin = 머신러닝으로 예측



전처리 과정에서 결측치를 채워줘야 하는데

<class 'pandas.core.frame.DataFrame'>

Int64Index: 891 entries, 1 to 891

Data columns (total 11 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Survived 891 non-null int64

1 Pclass 891 non-null int64

2 Name 891 non-null object

3 Sex 891 non-null object

4 Age 714 non-null float64

5 SibSp 891 non-null int64

6 Parch 891 non-null int64

7 Ticket 891 non-null object

8 Fare 891 non-null float64

9 Cabin 204 non-null object

10 Embarked 889 non-null object

dtypes: float64(2), int64(4), object(5)

memory usage: 83.5+ KB

1. 모델 학습 및 평가
2. 모델 적용