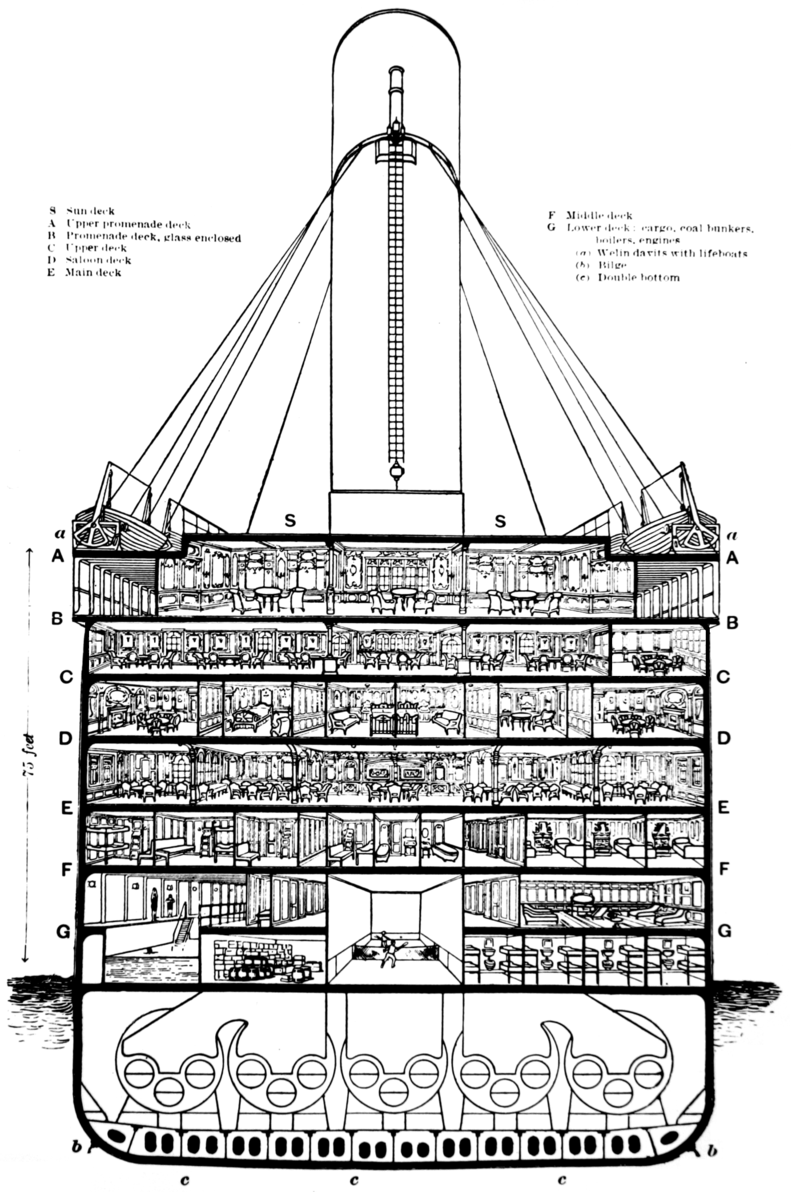
타이타닉 학습하기

타이타닉 머신러닝 과정

1. **데이터 수집 및 EDA , 전처리**

데이터는 어차피 있으니까 전처리 과정이 중요하다 이말이야!

EDA 를 좀 해야겠다.



A 갑판 - 산책 갑판이라고도 불렸다. 거의 모든 영역이 1등실 전용이었고 1등실 객실과 라운지, 흡연실, 독서실, 야자수 코트가 있었다. 산책 갑판이라는 이름답게 기나긴 산책로가 있었다.

B 갑판 - 선교루 갑판이라고도 불렸다. 객실은 1등실이 상당수를 차지하고 있었으며, 어떤 객실에는 개인 산책로가 있었다. 레스토랑과 카페도 있었다. 이 갑판에는 선두와 선미 윗쪽도 포함하고 있는데, 선미는 3등실 산책로로 쓰였다

C 갑판 - 배에서 선수, 선미와 분리되는 지점이자 하얗게 칠한 곳이 시작하는 지점으로, 3등실의 산책로로 쓰이고 선수와 선미로 각각 이어지는 요갑판과 이어져 있었다. 선수는 주로 선원들의 숙소로 쓰였고, 선미는 3등실 전용 휴게실과 흡연실이 있었다. 그 사이에는 1등실 객실과 2등실 도서관이 있었다.

D 갑판 - 3개의 넓은 공공시설이 상당부분을 차지하고 있었는데, 1등실 대합실, 1등실 식당과 2등실 식당이 있었다. 3등실을 위한 열린 공간도 있어서 밤만 되면 신나는 연회가 열렸다. 1등실, 2등실, 3등실 객실과 화부 숙소가 있던 곳이기도 했다. 방수격벽이 위치해 있던 가장 높은 갑판이기도 했다.

E 갑판 - 주로 1등실, 2등실, 3등실 모두의 객실과 선원 숙소가 있었다. "스코틀랜드 로드"라고 불리던 아주 긴 복도가 있어 선원과 승객이 오고갈수 있었다. (스코틀랜드 로드는 후에 타이타닉 선체 내부에 물이 들어차기 시작했을 때 E갑판의 선수쪽 전체를 침수시키고 방수격실을 무력화시키는 원인 중 하나였다.)

F 갑판 - 객실은 3등실이 대부분을 차지했지만 2등실 객실과 선원 숙소도 있었다. 3등실 식당은 여기에 위치했다. 또 1등실 승객을 위한 수영장과 터키탕도 있었다.

G 갑판 - 수면 위에서 가장 낮은 층으로, 선원과 승객 객실이 있는 가장 낮은 갑판이었다. 스쿼시 코트도 있었으며 우편 저장소도 여기에 있었다. 대다수의 영역이 보일러실이 있는 최하 갑판과 겹쳐 있어 선미와 후미가 각각 따로 떨어져 막혀 있었다.

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

위와 같이 나이, 객실, 탑승지 이 3개에 결측치가 존재한다.

결측치를 대충 넣으면 나중에 정확도가 떨어지기 때문에, 정확도가 생명이다.

나이의 경우 이름에 있는 호칭을 통해 유추

객실번호의 경우 티켓의 가격이나 객실등급으로 유추해보고

탑승지의경우 조사를 해봐야겠다…

1. 모델 학습 및 평가
2. 모델 적용