

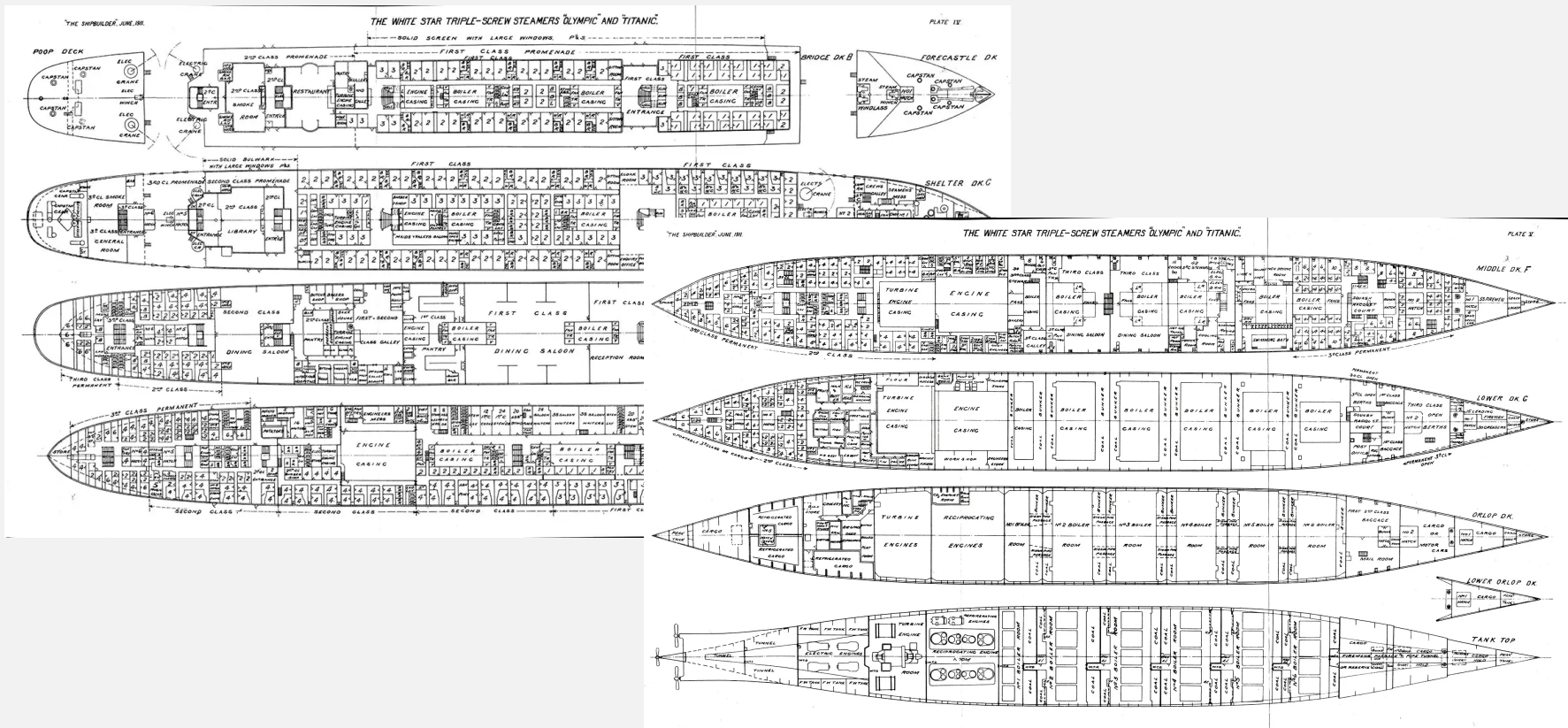
2주차 2조

팀원: 황동욱, 권도혁

```
lookup.KeyValue  
f.constant(['em  
=tf.constant([0  
ce = tf.lookup.StaticV  
init,  
num_oov_buckets=5)
```

```
lookup.StaticVocabular  
initializer,  
num_oov_buckets,  
lookup_key_dtype=None  
name=None,  
experimental_is_open
```

데이터 전처리, 모델 설계, 모델 훈련



Feature 수정방법

1. Passenger_ID

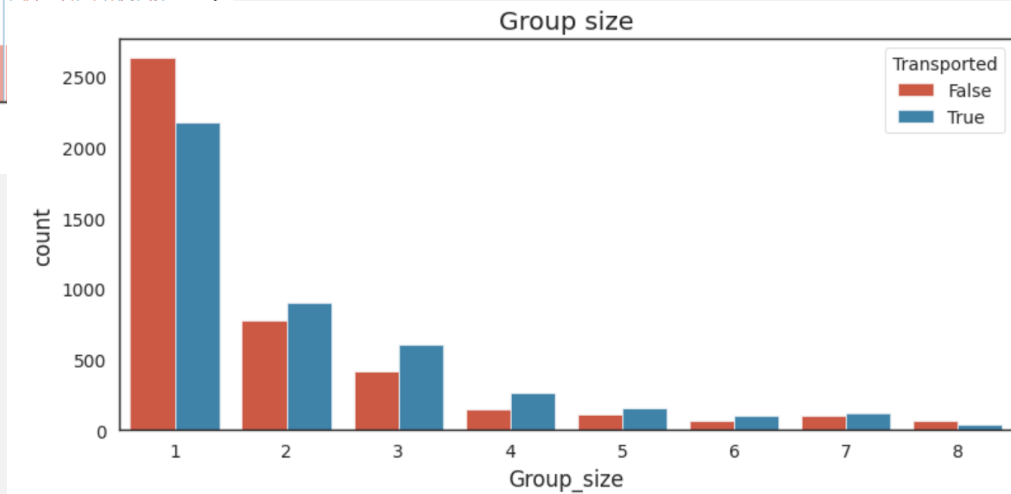
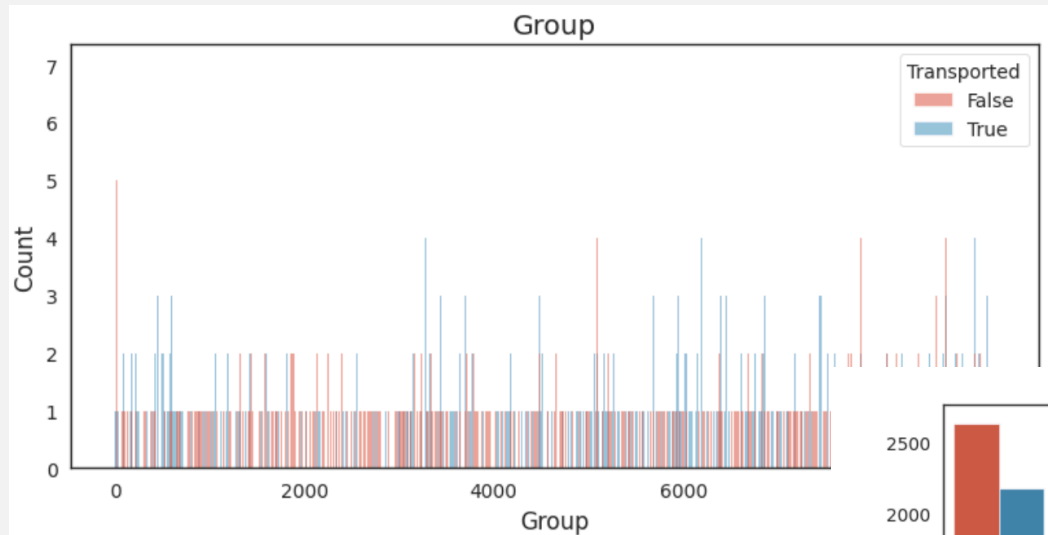
- 기존 Feature : 'group number_index' 형태로 해당 사람이 group으로 탑승을 하였는지, group이 몇 명인지, group 번호는 몇 번인지에 대한 다양한 정보를 담고 있음



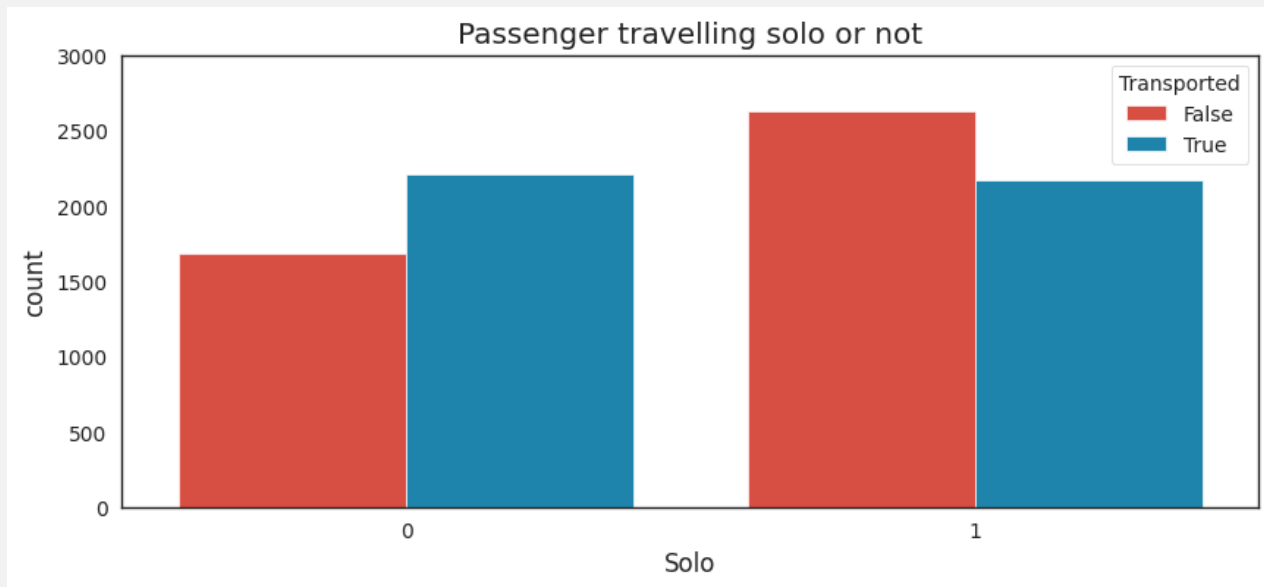
- Feature 추가 :
 1. 'Group_num' : Group 번호
 2. 'Group_Size' :
 3. 'is_Group' : Group으로 탑승하였는지, 혼자서 탑승하였는지 True / False

특히나 group은

Feature 수정방법



Feature 수정방법



Feature 수정방법

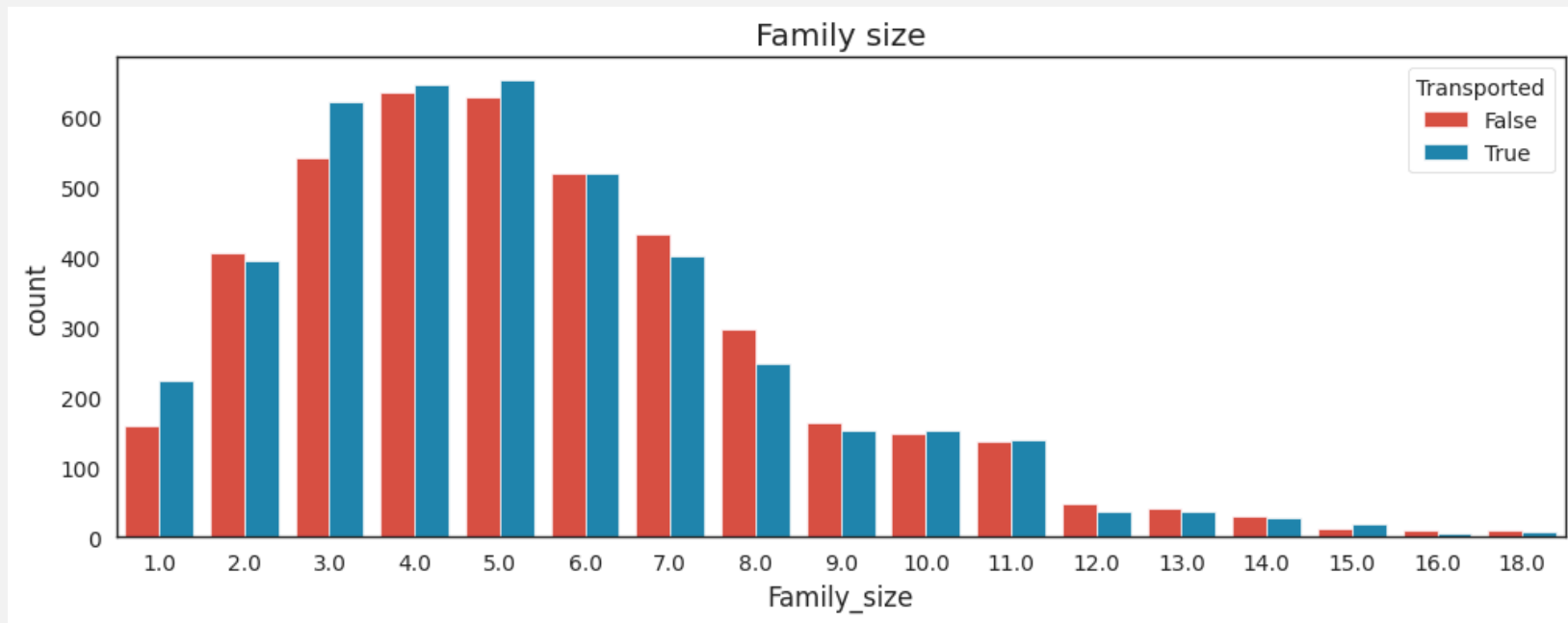
2. Name

- 기존 Feature : 사람 이름으로 기존에는 일종의 index로만 작용할 것으로 판단하고 아예 학습에서 해당 feature를 제거하여 학습을 시키기도 하였음.



- Feature 추가 : 그러나 group에서도 가족인 group이 있고 아닌 경우가 있으며, 해당 정보도 동일 group내에서 유의미한 결과를 줄 수 있을 것으로 판단함.
 1. 'Last name' : 각 사람의 Last name을 새로운 feature로 만들었음.
 2. 'is_Family' : 본인 group 내에 동일한 Last name을 가진 사람이 있을 경우 family로 판단하였음

Feature 수정방법



Feature 수정방법

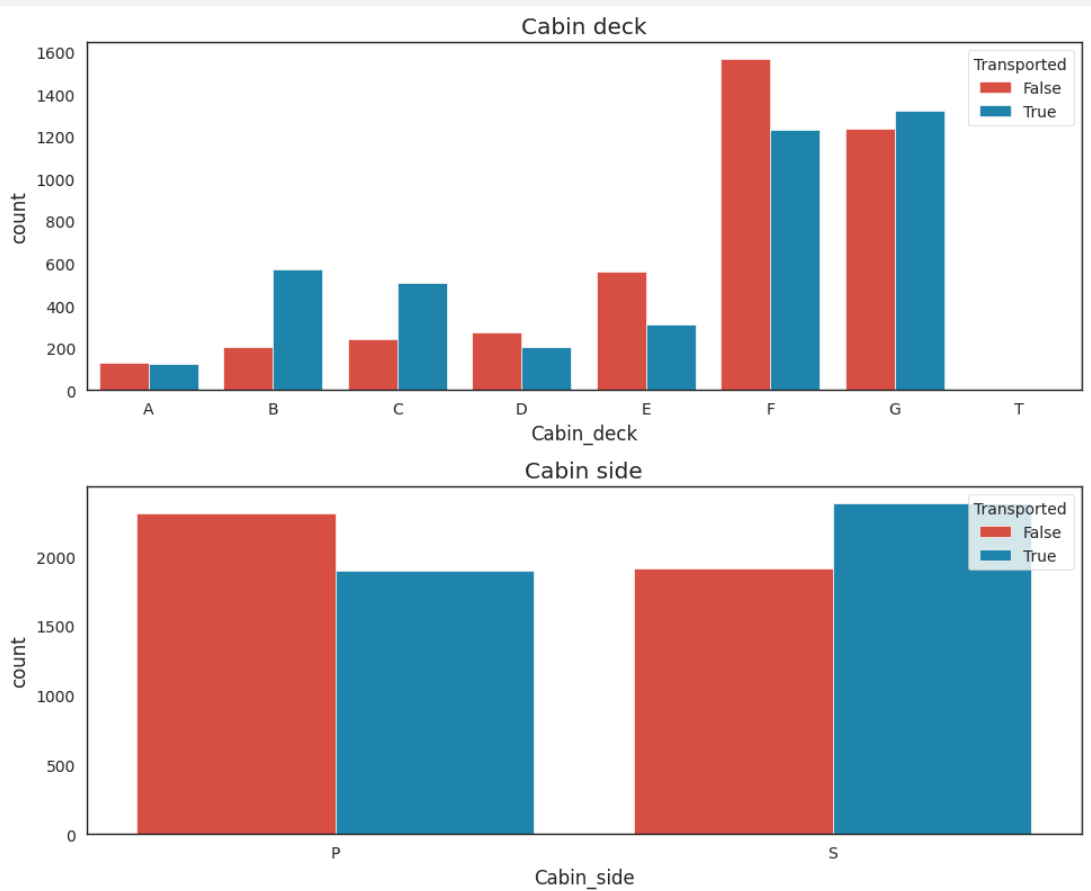
3. Cabin

- 기존 Feature : 'Deck / Cabin number / Port'의 형태로 해당 사람이 몇 층에 위치하였는지, 방 번호가 몇 번인지, 왼쪽과 오른쪽 중 어떤 사이트에 있는지의 정보를 포함

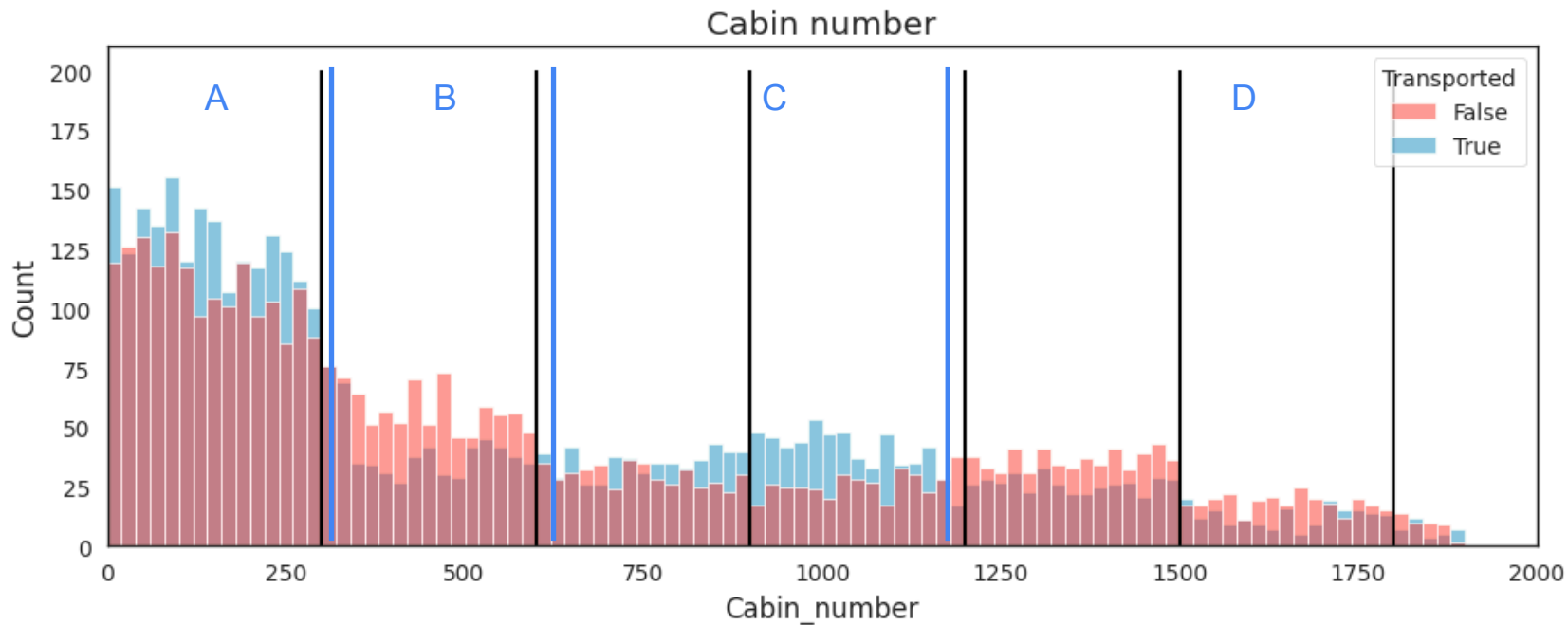


- 결측치 수정 :
 1. Raw data를 통해 보통 동일한 group끼리는 동일한 cabin을 사용하는 경우가 많아 본인 group에서 random으로 한 값을 가져와 결측치를 채움.
- Feature 추가 :
 1. 'Deck'
 2. 'Cabim_num'
 3. 'Port' : 이렇게 세 feature를 분리하였음
 4. 'Cabin_part' : 특정 Cabin 번호들에 따라 A,B,C,D로 나누었음

Feature 수정방법



Feature 수정방법



Feature 수정방법

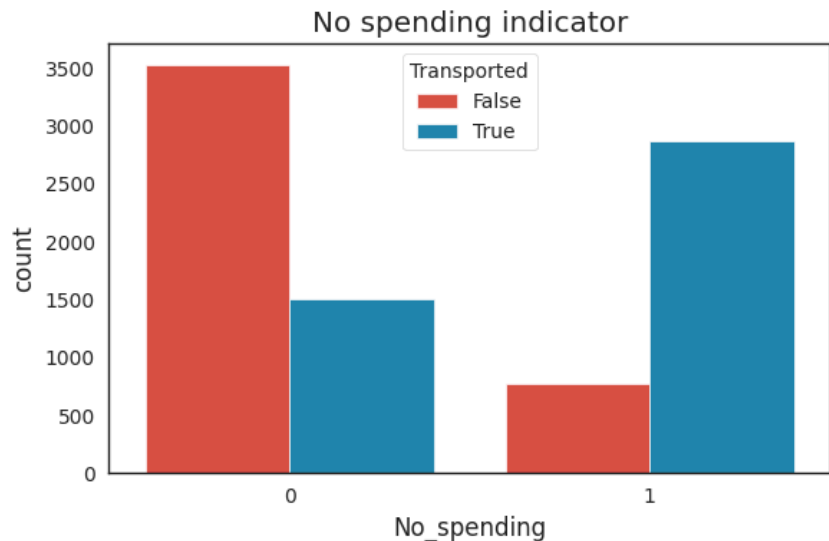
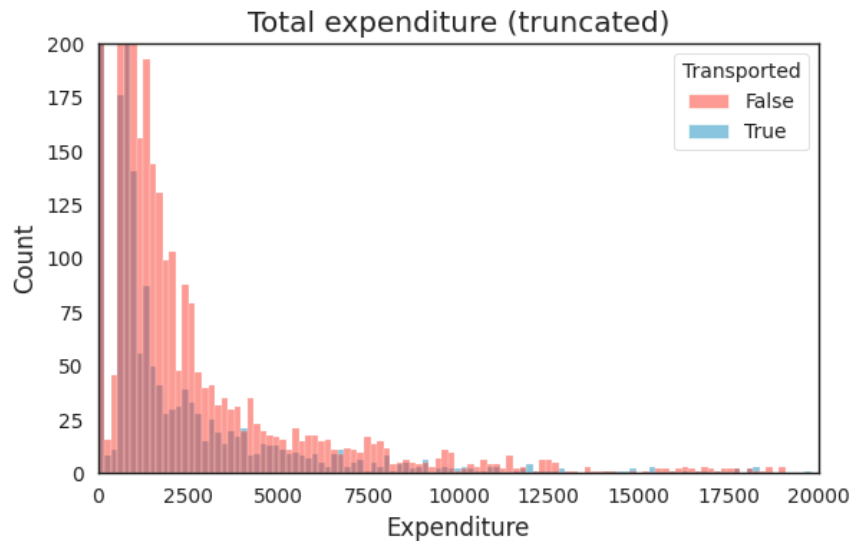
4. Services ('RoomService', 'FoodCourt', 'ShoppingMall', 'Spa', 'VRDeck')

- 기존 Feature : 각 사람별로 해당 서비스에 사용한 금액을 포함



- Feature 추가 : 1. 'Use_service' : 서비스를 이용한 금액의 합이 0일 경우 False / 그 외에는 True로 표시하는 새로운 feature를 추가.

Feature 수정방법



추가 전략

1. CryoSleep

- 분석 :
1. 지난주에 실시하였던 피어슨 상관계수분석을 통해 'CryoSleep'이 'Transported' 유무에 가장 직관적인 상관관계가 있음을 알 수 있었음. 이에 아래와 같이 data의 개수를 파악해보니 CryoSleep이 True일 때 Transported 확률이 매우 높음을 알 수 있음.
 2. CryoSleep이 해당 객실에서 이뤄지고 있다는 정보가 Kaggle 사이트에 적혀있는 것으로 보아, 특정 위치에 있던 사람들이 transported 되었을 것이라고 판단한 초기 직관이 옳은 직관일 수도 있다는 판단을 내리게 됨.

2. Deck 별 CryoSleep 관련

- CryoSleep == True 일 때 Transported

	A	B	C	D	E	F	G
True	50	332	237	75	88	417	706
False	3	2	2	1	47	6	366

추가 전략

1. CryoSleep

- 전략 :
1. 따라서 A,B,C,D,F deck에 있는 사람들 중 CryoSleep 중인 사람들은 무조건 Transported 되었다고 판단하기로 함.
 2. 추가로 CryoSleep의 결측치의 경우 여러 Service들을 이용한 기록이 전부 없었으므로 Use_Service가 False이면서 결측치인 경우에는 True를, 그 외에는 False로 채웠음.

2. Deck 별 CryoSleep 관련

- CryoSleep == True 일 때 Transported

	A	B	C	D	E	F	G
True	50	332	237	75	88	417	706
False	3	2	2	1	47	6	366

kaggle 제출

Kaggle 제출 결과



submission_10.csv

Complete · 1h ago

0.80196



submission_10.csv

Complete · 1h ago

0.80196



submission.csv

Complete · 1h ago

0.80056