

CODE AIRLINE

코드에어 다시 한 번 No.1 브랜드로 날아보자



목차

CONTENTS

01 데이터 설명

02 가설 설정

03 데이터 모델링

04 해석 및 전략



01. 비행기 데이터

승객의 정보

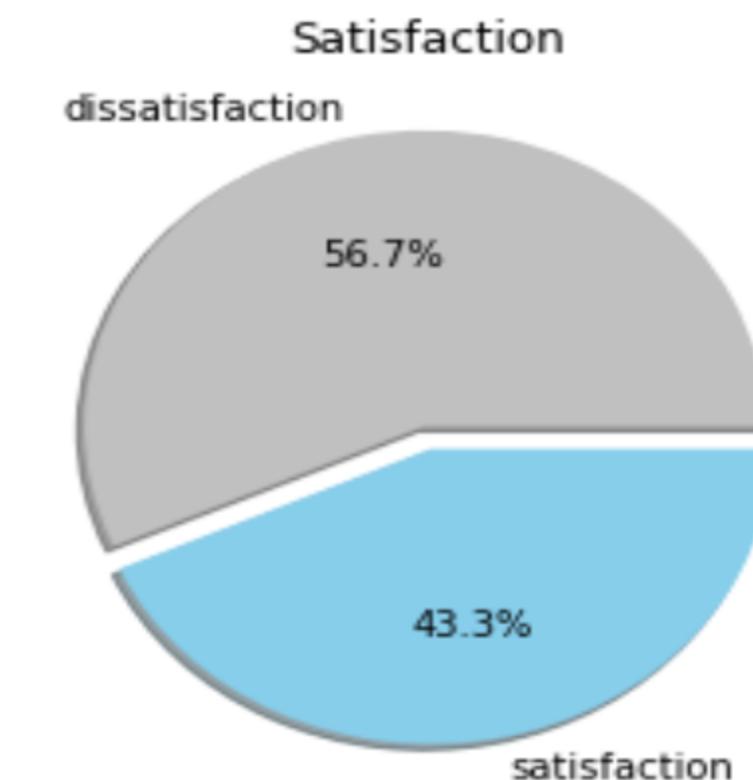
성별
고객 태입(Loyal, Disloyal)
비행 목적(Business, Personal)
클래스(Business, Eco Plus, Eco)
비행거리
출발/도착 몇분 지연됐는지

승객의 만족도

기내 와이파이, 출발/도착시간,
온라인예약, 탑승구, 식사,
온라인게시판, 자리 편안함,
기내 엔터테이먼트, 기내,
다리 공간,
수하물 처리, 체크인, 기내 서비스,
청결 만족도
(0: 응답하지 않음)
(1-5: 만족점수)

만족도

비행 만족도
(neutral or dissatisfaction,
satisfaction)



01. 비행기 데이터

	satisfaction	neutral or dissatisfied	satisfied
Age	37.566688	41.750583	
Flight Distance	928.919971	1530.140255	
Inflight wifi service	2.399633	3.161288	
Departure/Arrival time convenient	3.129112	2.970305	
Ease of Online booking	2.546850	3.031582	
Gate location	2.976121	2.977879	
Food and drink	2.958050	3.521310	
Online boarding	2.656125	4.027474	
Seat comfort	3.036295	3.966530	

Inflight entertainment	2.894156	3.964931
On-board service	3.019158	3.857324
Leg room service	2.990812	3.822143
Baggage handling	3.375991	3.966396
Checkin service	3.042952	3.646041
Inflight service	3.388814	3.969461
Cleanliness	2.936123	3.744342
Departure Delay in Minutes	16.503728	12.608084
Arrival Delay in Minutes	17.127536	12.630799
sum_satisfaction	41.350193	50.627007

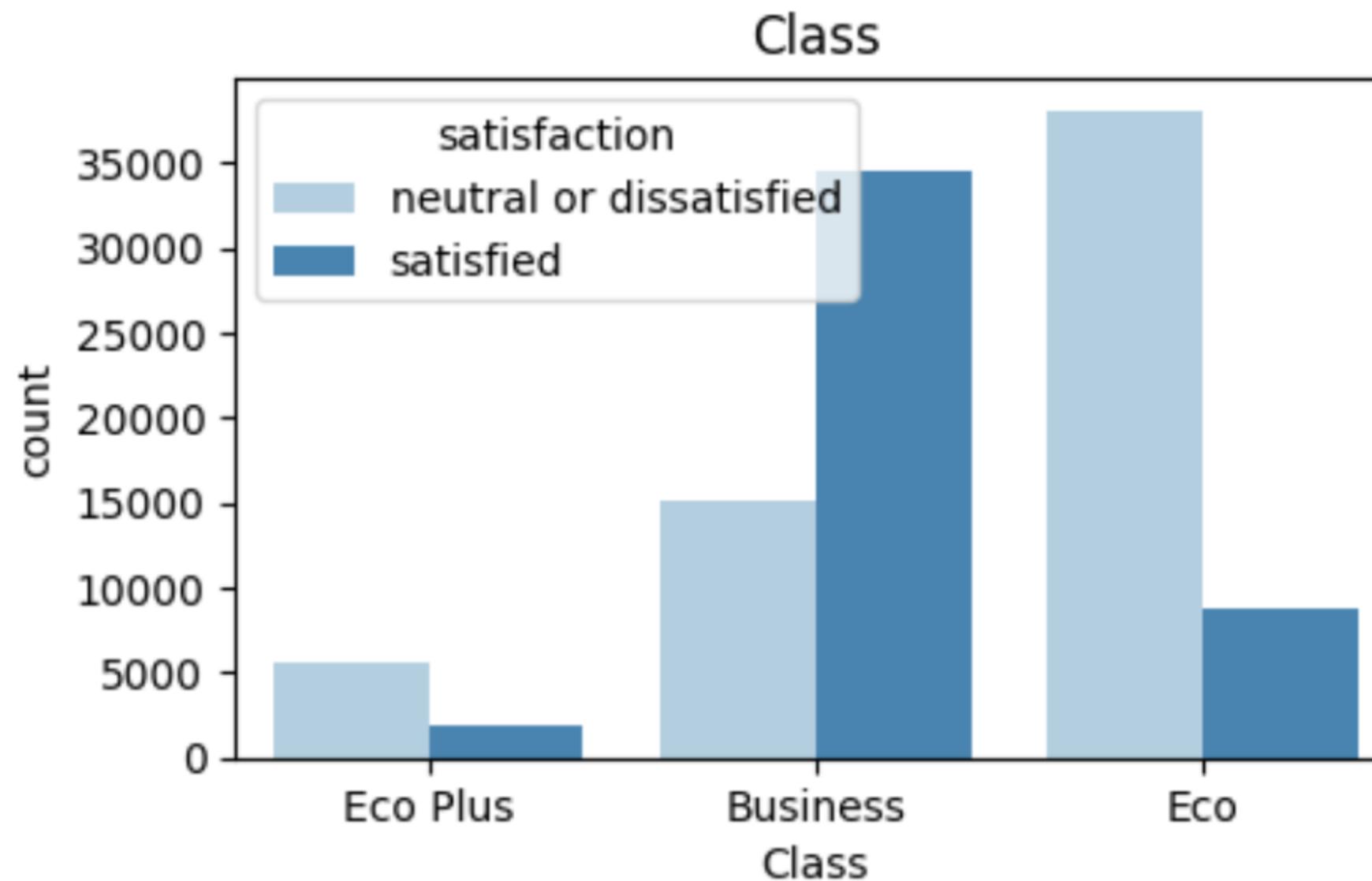
만족감에 따른 수치형 데이터들의 평균을 확인해봤다.

다른 데이터들은 예상대로 만족할 때 점수가 더 높게 나왔지만,

편리한 출발/도착 시간에 대한 점수는 반대로 나왔다. 이것을 봤을 때, 비행기 시간은 스스로 고를 수 있어 만족감에는 큰 영향을 주지 않는 것이 아닐까 생각이 든다.

02. 가설 설정

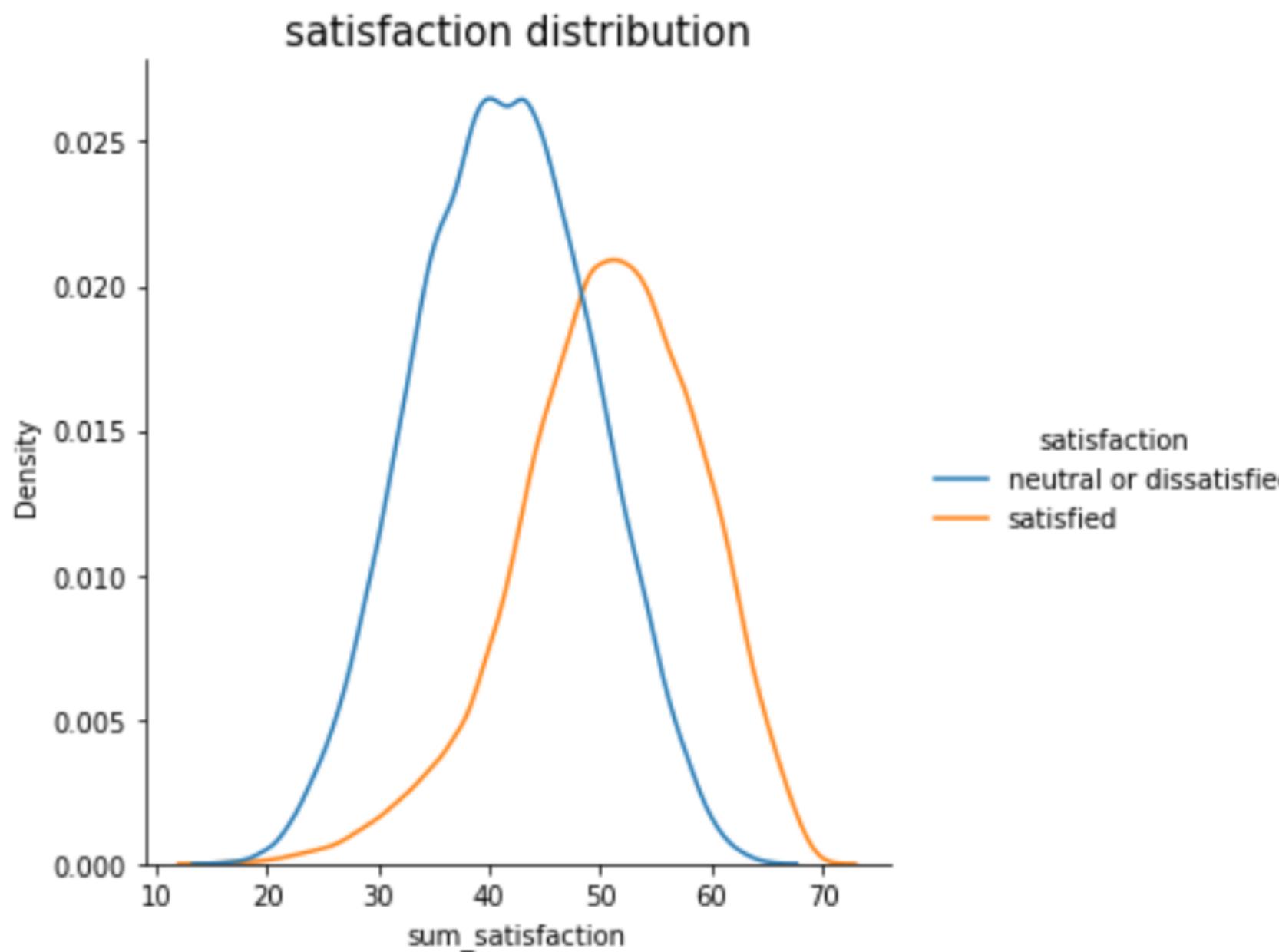
1. 승객의 항공 클래스가 좋을 수록 만족도가 높다



그래프를 보면 비즈니스 좌석은 만족도가 높게 나왔지만 상대적으로 싼 에코플러스와 에코 좌석은 불만족이라는 응답이 더 많이 나온 것을 확인할 수 있다.
비율을 확인해 본 결과, 비즈니스석은 0.47, 에코+에코플러스는 0.53으로 비슷하기 때문에 에코석에도 더 신경을 써줘야 승객 만족도가 올라갈 수 있다.

02. 가설 설정

2. 설문조사 만족도의 합이 높을 수록 비행 만족도가 높다

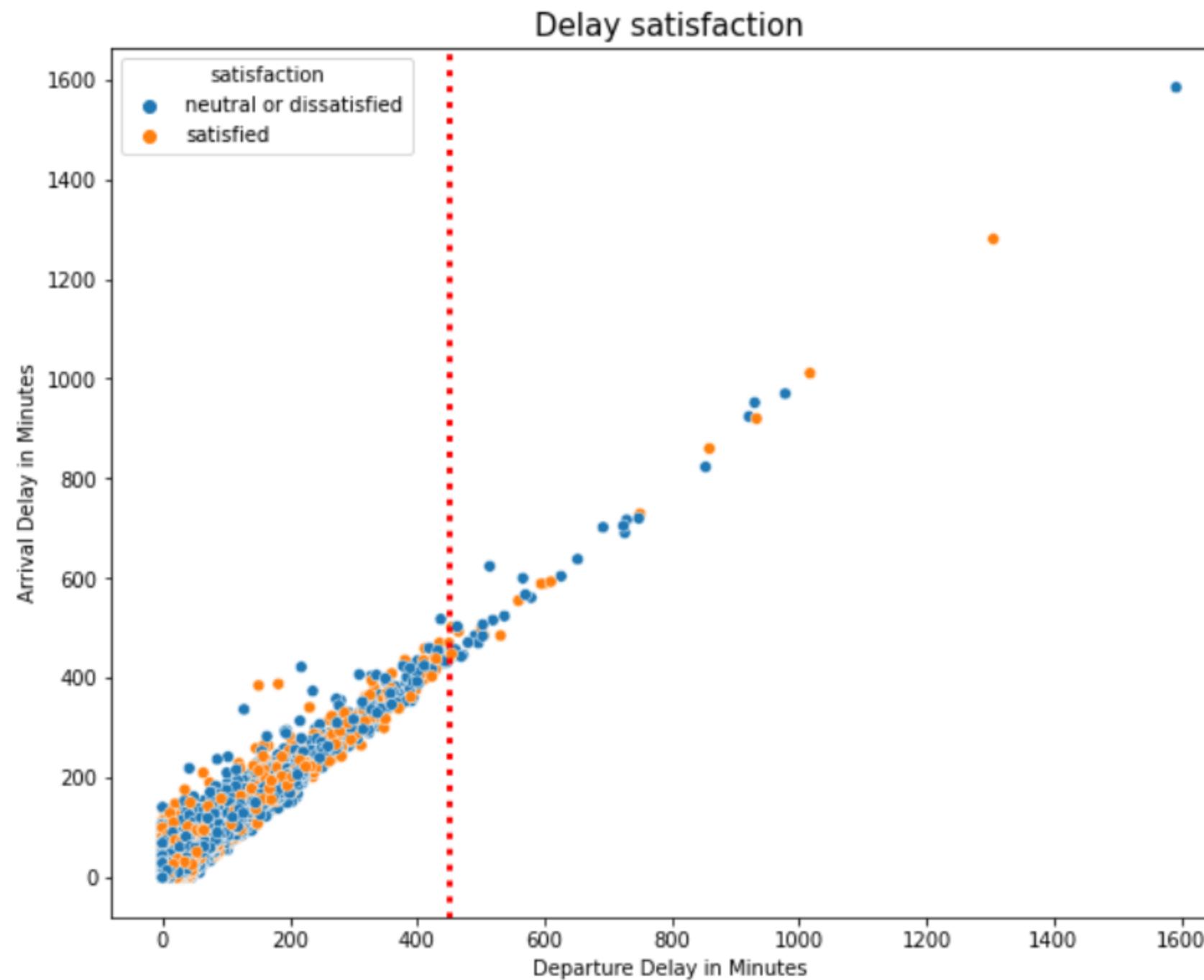


```
satisfaction  
neutral or dissatisfied    41.350193  
satisfied                  50.627007  
Name: sum_satisfaction, dtype: float64
```

각각의 만족도를 다 더해서 새로운 특징을 만들었고
그것을 분포로 그려봤을 때,
만족도의 합이 높을수록 비행 만족한다는
응답비율이 많은 것을 확인할 수 있다.

02. 가설 설정

3. 출발/도착 지연이 길어질수록 비행만족도가 떨어진다



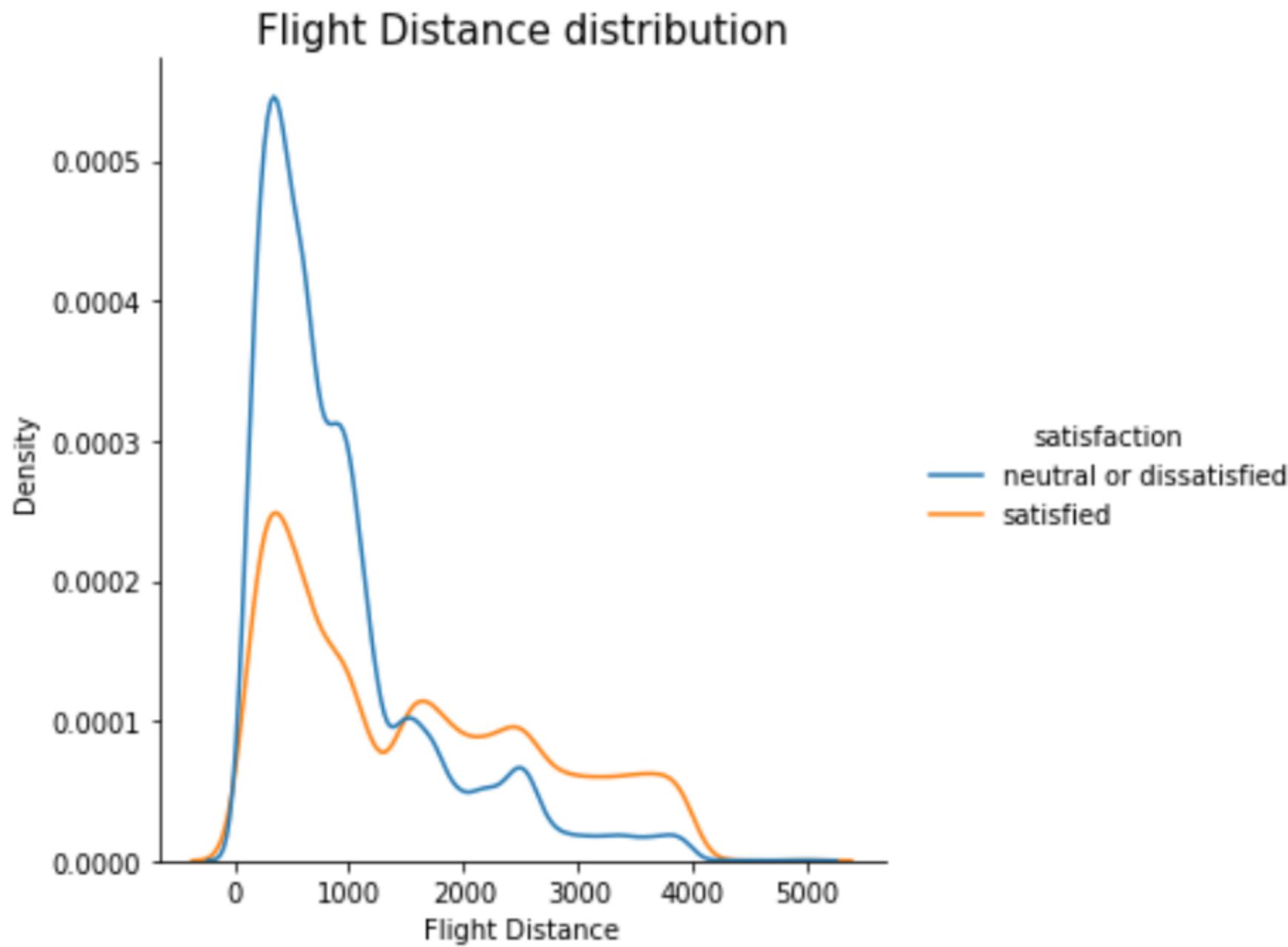
그래프를 보면 출발과 도착 간의 상관관계가 높은 것을 알 수 있고, 450분 이상부터는 만족하지 못한다는 답변이 더 많이 보인다.

출발/도착 지연이 1분 이상인 경우 평균은 약 34분으로 지연이 없는 것이 가장 좋은 방법이겠지만

비행기 지연이 길어질수록 만족감이 떨어지기 때문에 평균 34분 이상을 넘기지 않는 것이 좋은 방법이다.

02. 가설 설정

4. 비행 거리가 길어질수록 비행 만족도가 떨어진다



예상과는 다르게,
비행 거리가 길어질수록 만족을 응답한 사람이 많고
비행 거리가 짧을 수록 불만족에 많이 응답한 것을 볼
수 있다.
이것을 통해, [비행 거리보다는 서비스에 대해](#)
[만족/불만족을 선택한다고 볼 수 있다.](#)

03. 데이터 모델링

기준모델 : 0.567

03. 데이터 모델링

1. DecisionTreeClassifier

훈련정확도 0.9380556524668263

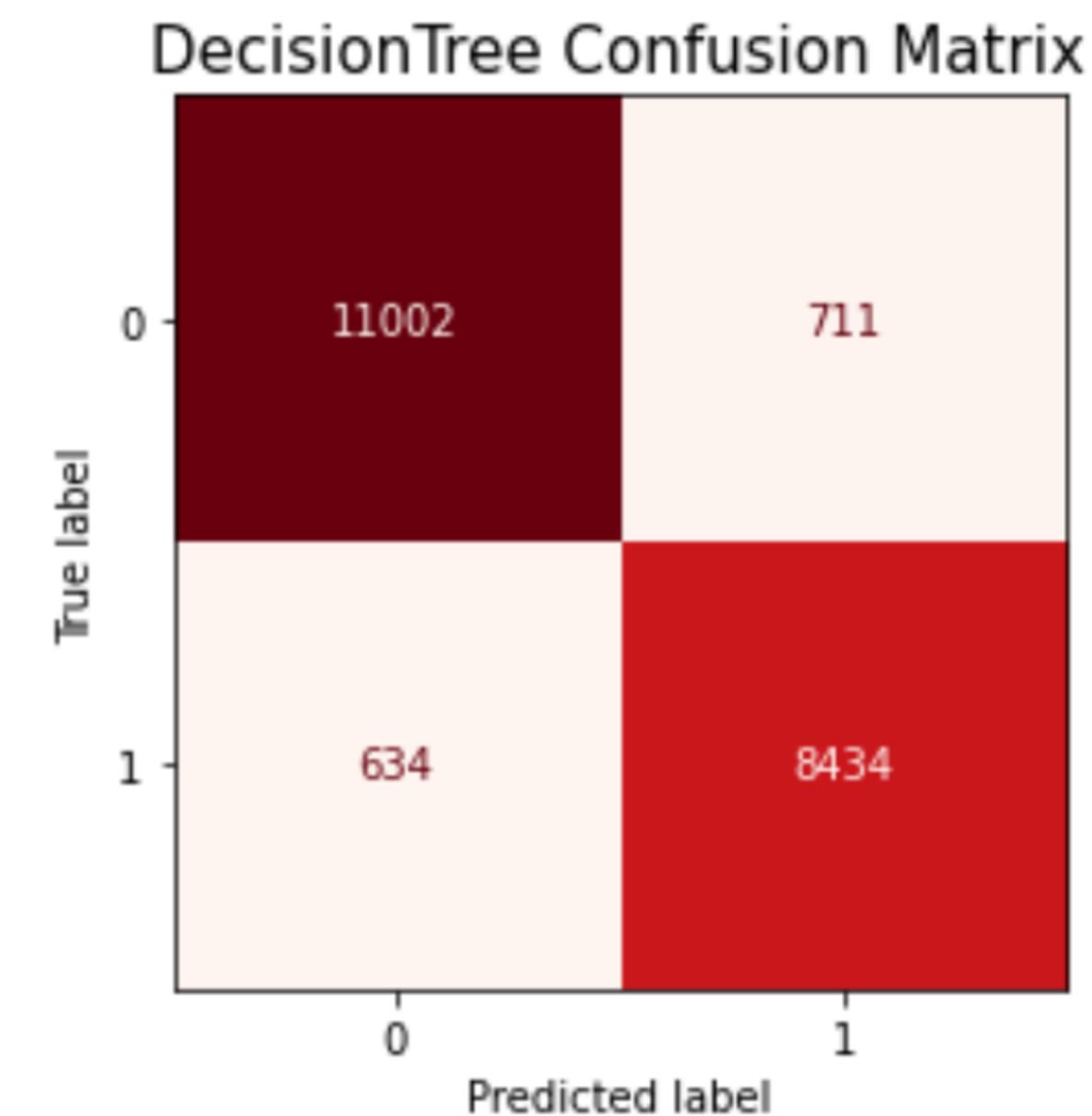
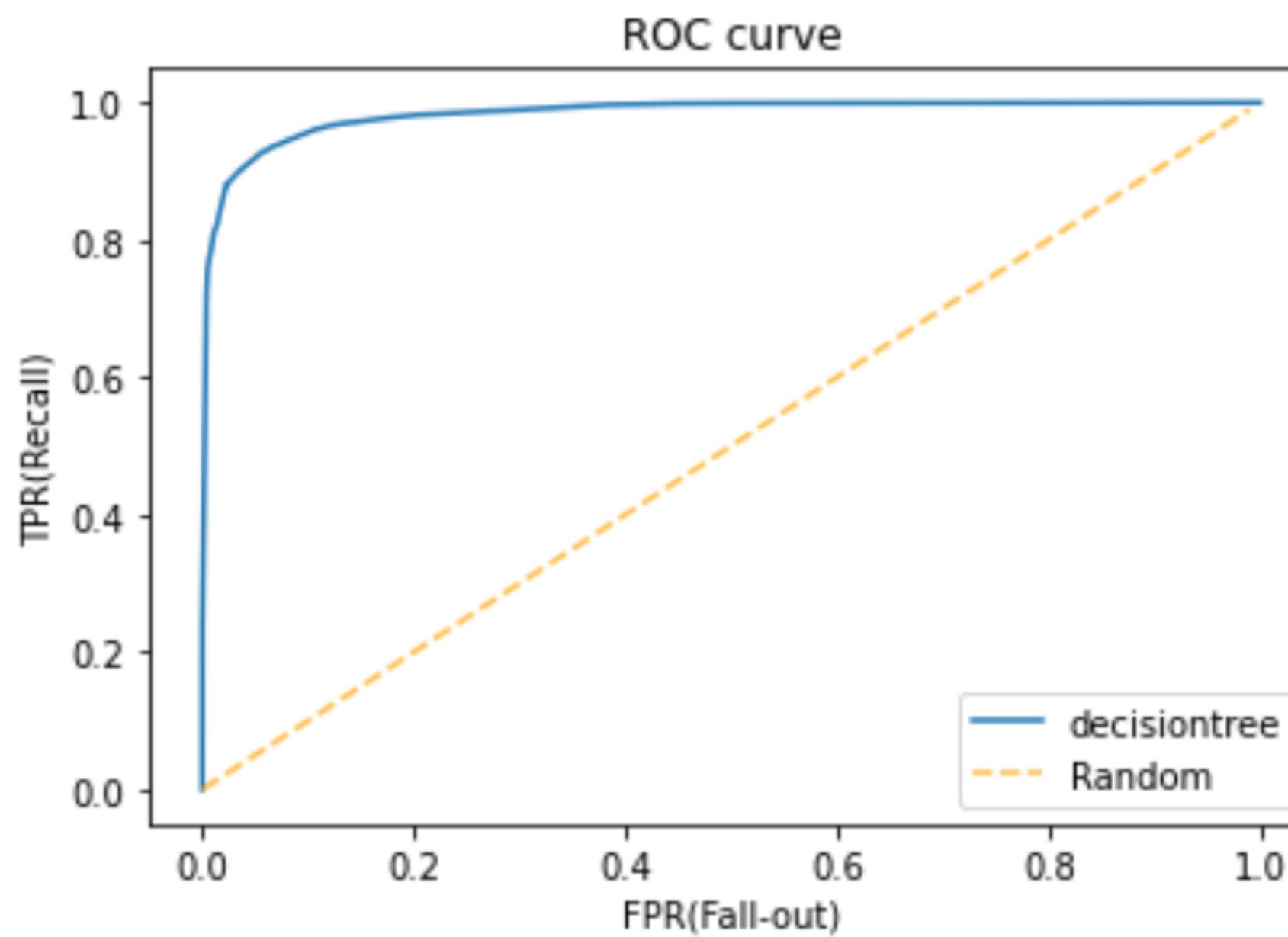
검증정확도 0.9352774168711804

roc_auc 0.983591297127585

	precision	recall	f1-score	support
0	0.95	0.94	0.94	11713
1	0.92	0.93	0.93	9068
accuracy			0.94	20781
macro avg	0.93	0.93	0.93	20781
weighted avg	0.94	0.94	0.94	20781

03. 데이터 모델링

1. DecisionTreeClassifier



03. 데이터 모델링

2. RandomForestClassifier

- 훈련정확도 0.8312380448251386
- 검증정확도 0.8322506135412155
- roc_auc 0.9178286487617713

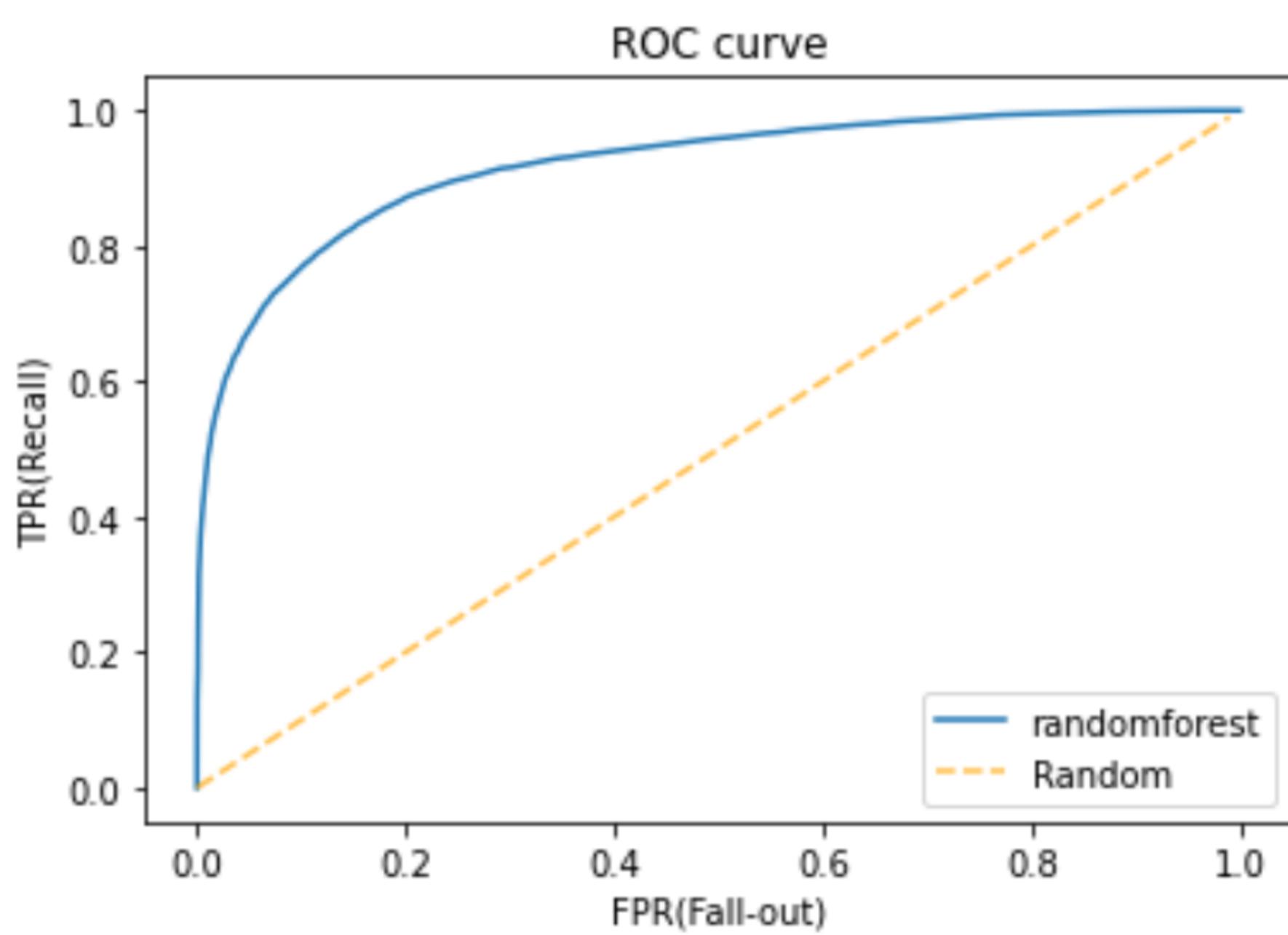
	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.80	0.95	0.86	11713
1	0.91	0.69	0.78	9068

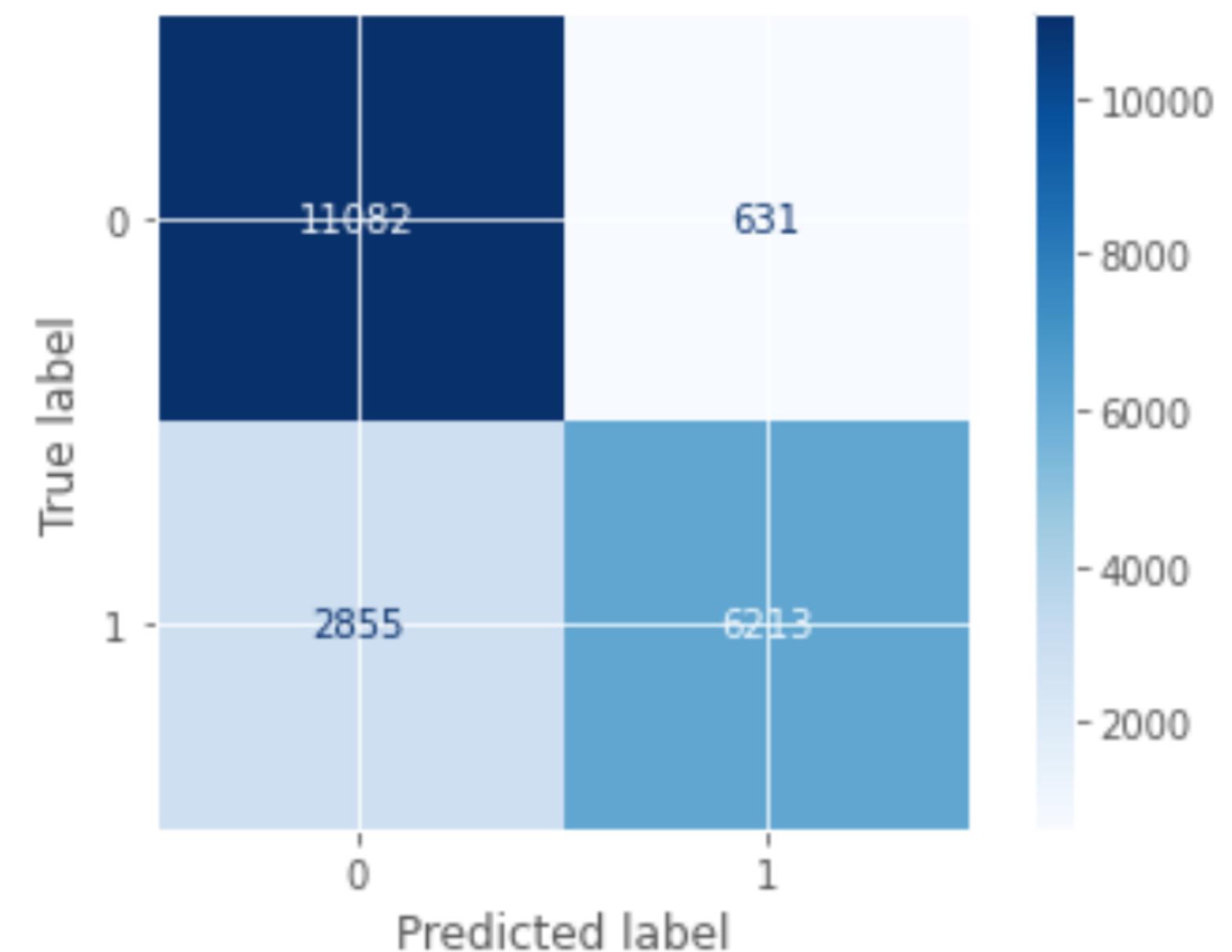
accuracy			0.83	20781
macro avg	0.85	0.82	0.82	20781
weighted avg	0.84	0.83	0.83	20781

03. 데이터 모델링

2. RandomForestClassifier



RandomForest Confusion Matrix



03. 데이터 모델링

3. XGBClassifier

훈련정확도 0.9780686452606379

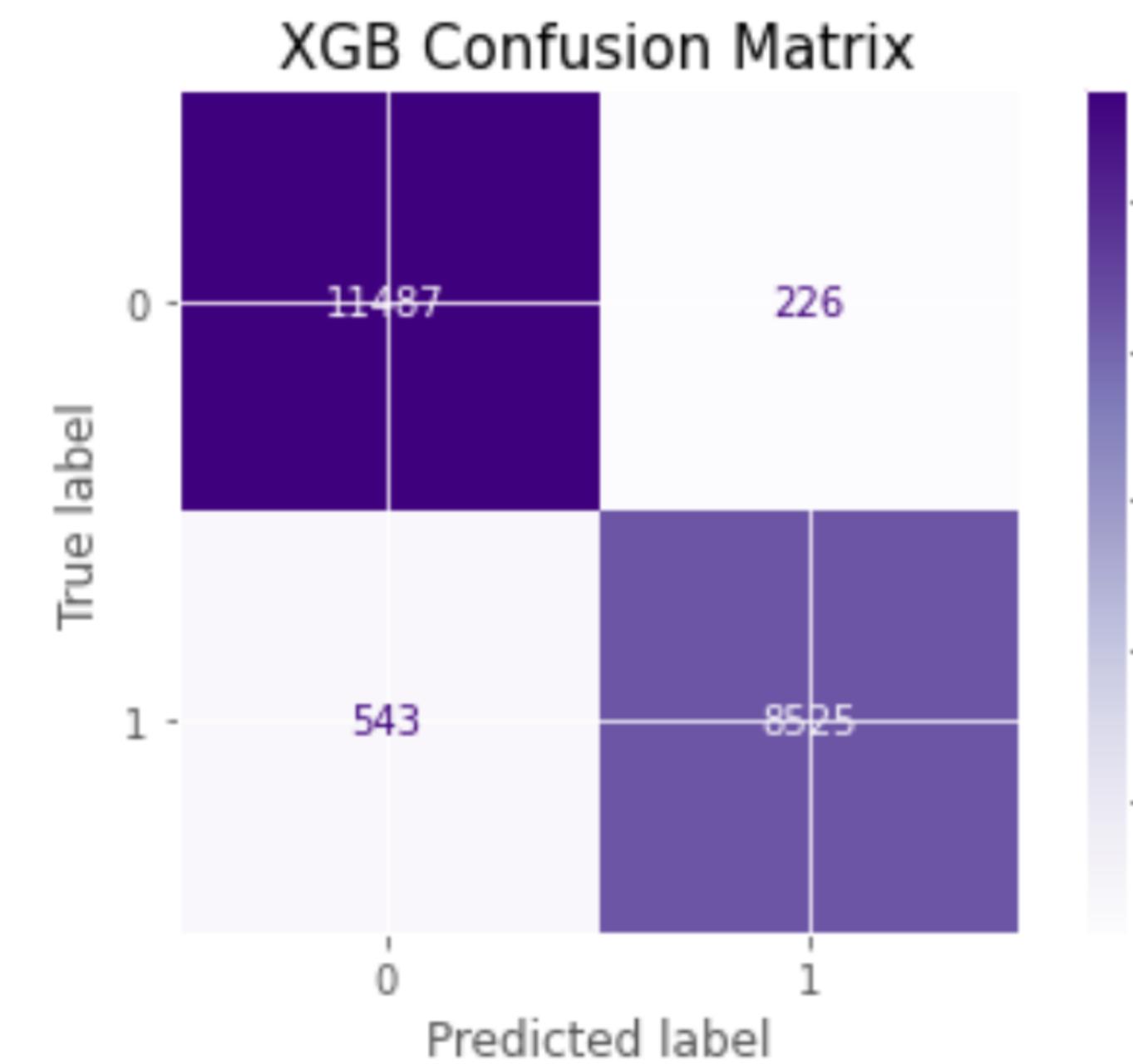
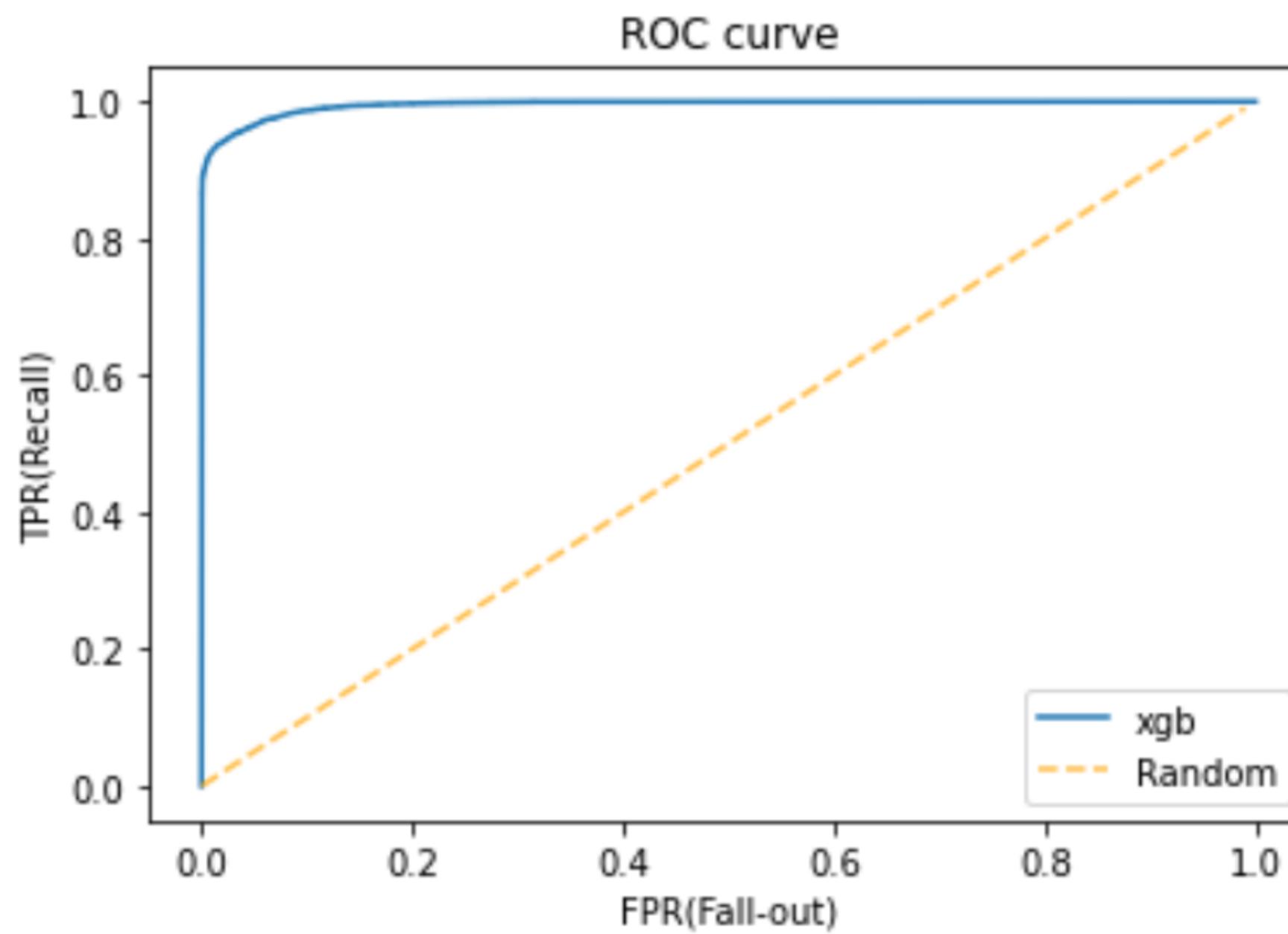
검증정확도 0.9629950435493961

roc_auc 0.9951569708418566

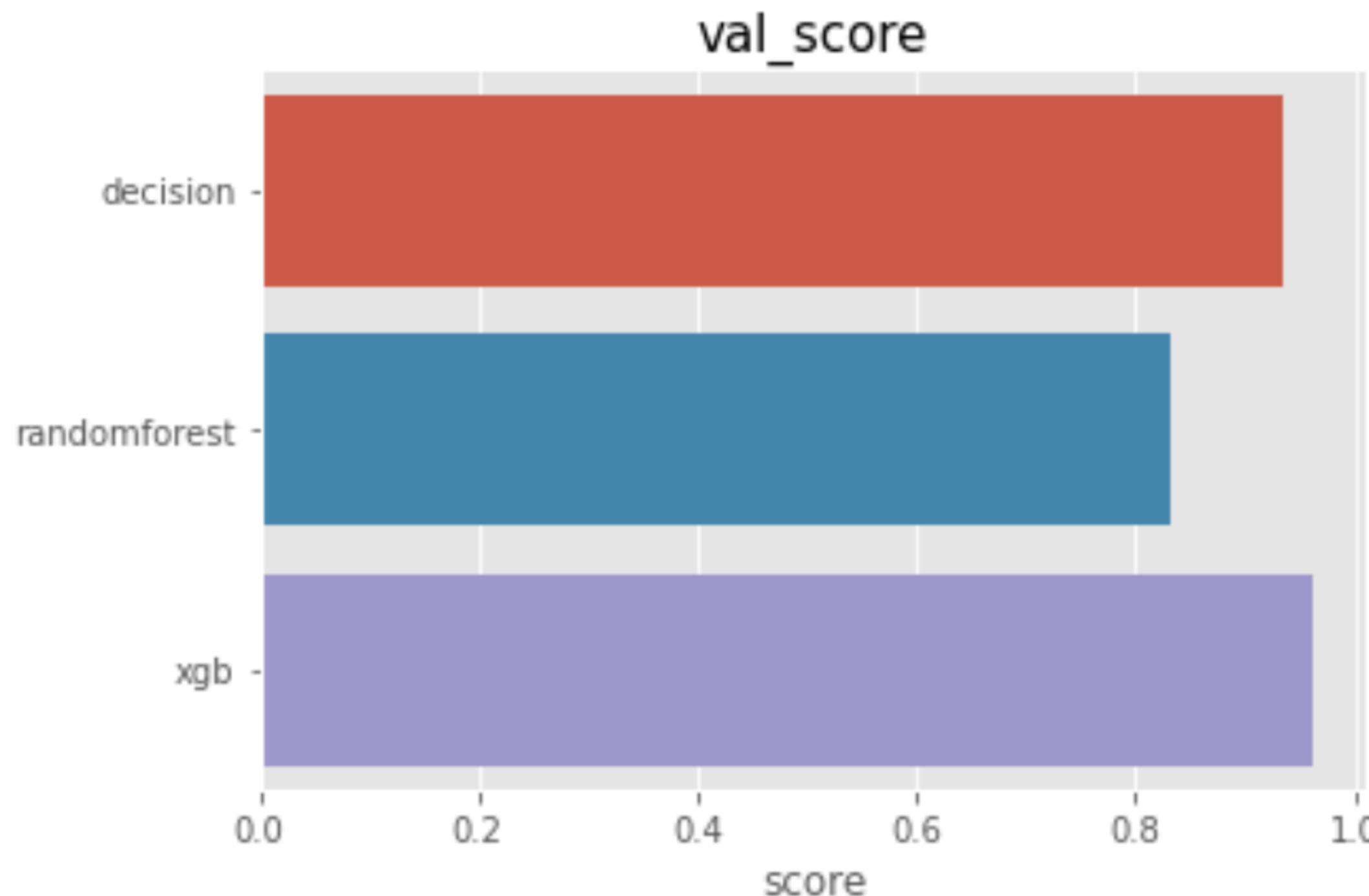
	precision	recall	f1-score	support
0	0.95	0.98	0.97	11713
1	0.97	0.94	0.96	9068
accuracy			0.96	20781
macro avg	0.96	0.96	0.96	20781
weighted avg	0.96	0.96	0.96	20781

03. 데이터 모델링

3. XGBClassifier



03. 데이터 모델 선택

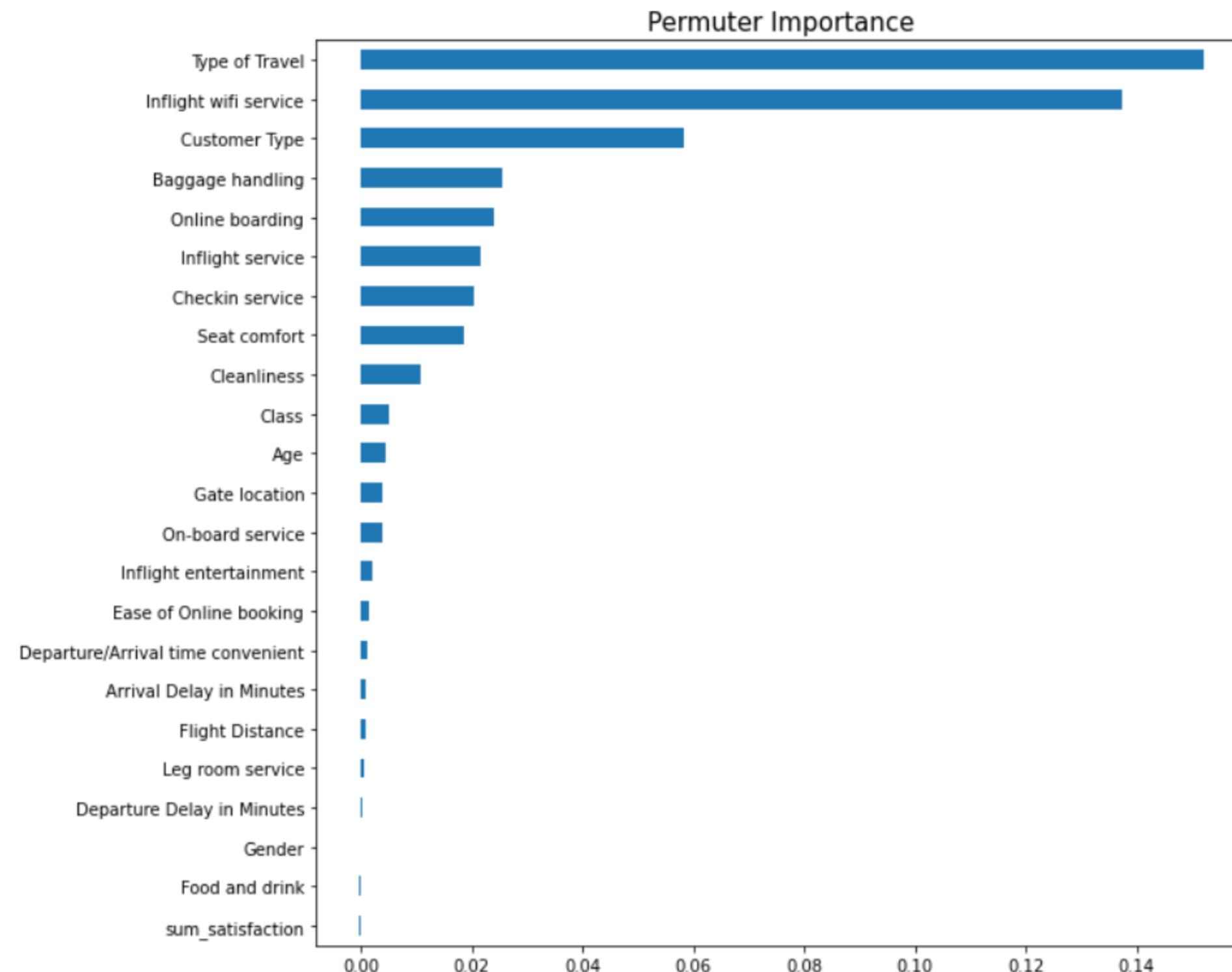


모델 성능이 가장 좋은 것은
XGBClassifier

03. 데이터 모델 특성중요도

XGBClassifier Permuter Importance

Weight	Feature
0.1520 ± 0.0023	Type of Travel
0.1372 ± 0.0034	Inflight wifi service
0.0583 ± 0.0021	Customer Type
0.0256 ± 0.0012	Baggage handling
0.0241 ± 0.0030	Online boarding
0.0217 ± 0.0011	Inflight service
0.0205 ± 0.0020	Checkin service
0.0187 ± 0.0020	Seat comfort
0.0108 ± 0.0010	Cleanliness
0.0052 ± 0.0010	Class
0.0046 ± 0.0014	Age
0.0039 ± 0.0007	Gate location
0.0038 ± 0.0009	On-board service
0.0021 ± 0.0008	Inflight entertainment
0.0014 ± 0.0006	Ease of Online booking
0.0011 ± 0.0009	Departure/Arrival time convenient
0.0009 ± 0.0007	Arrival Delay in Minutes
0.0008 ± 0.0009	Flight Distance
0.0005 ± 0.0007	Leg room service
0.0002 ± 0.0006	Departure Delay in Minutes
0.0001 ± 0.0003	Gender
-0.0002 ± 0.0004	Food and drink
-0.0004 ± 0.0004	sum_satisfaction

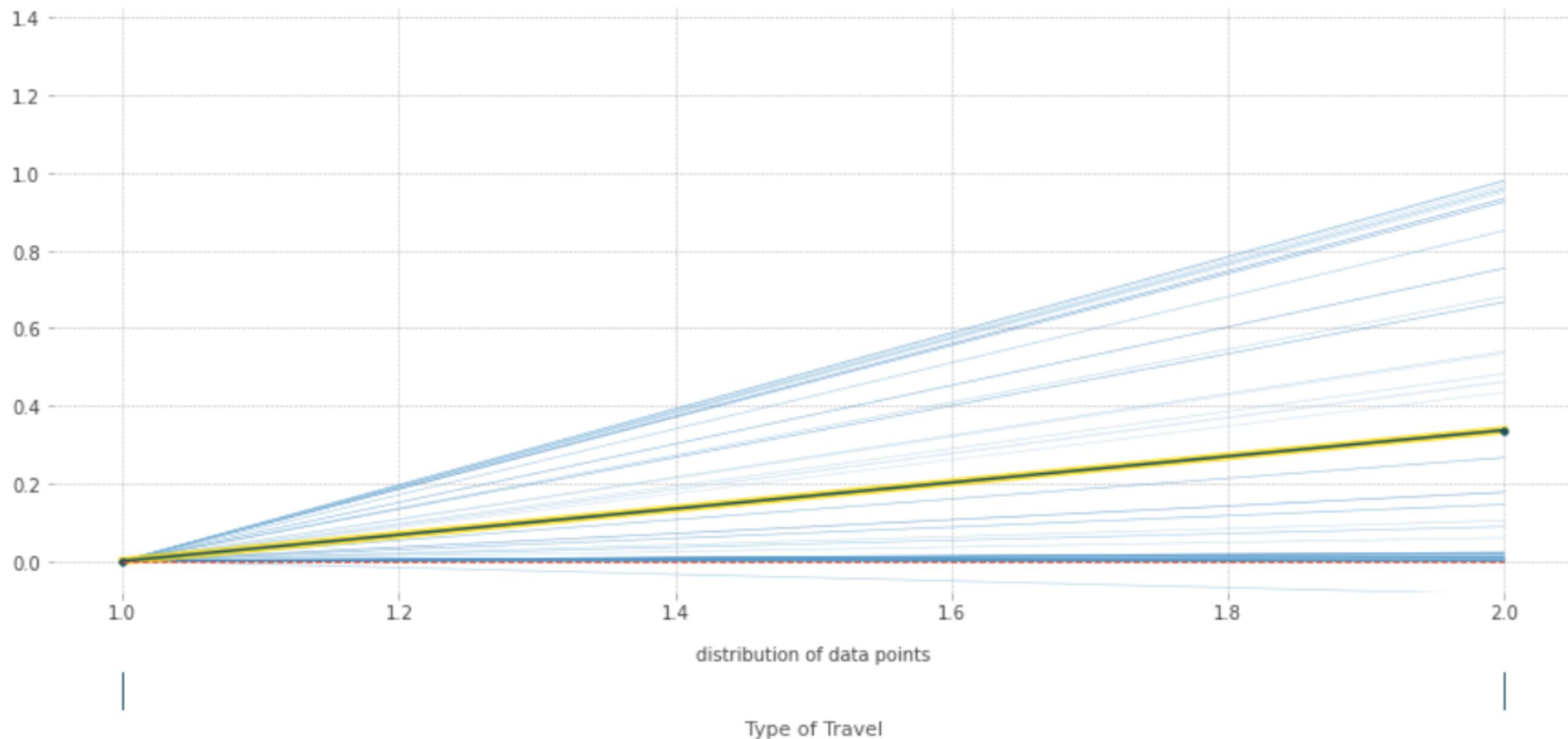


03. 데이터 모델 PDP Plot

XGBClassifier PDP Plot - Type of Travel

PDP for feature "Type of Travel"

Number of unique grid points: 2



- 1: 여행
- 2: 출장

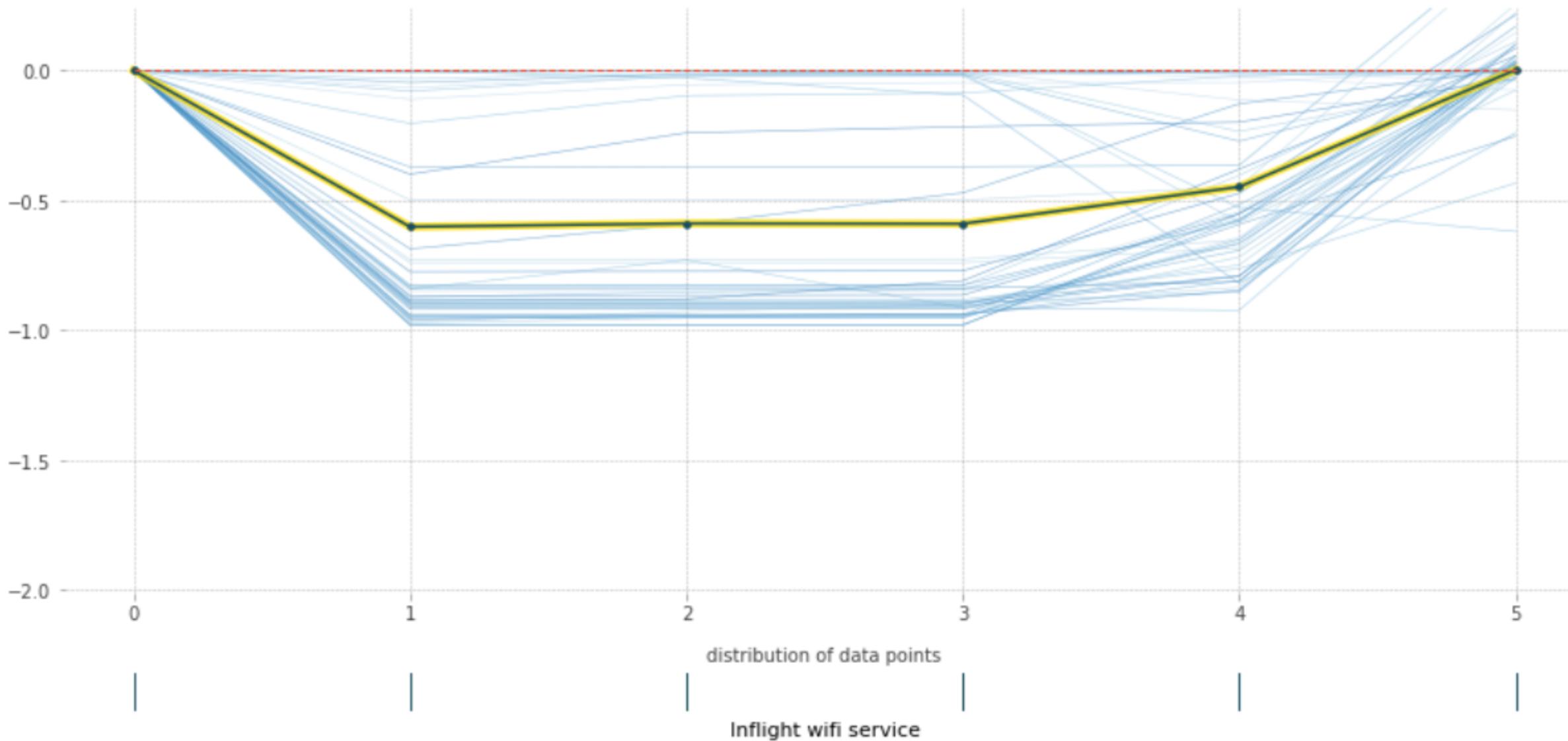
개인여행보다 출장이 더
만족도가 높다고
예측한다.

03. 데이터 모델 PDP Plot

XGBClassifier PDP Plot - Inflight wifi service

PDP for feature "Inflight wifi service"

Number of unique grid points: 6



0점은 응답하지 않음
1-5점은 와이파이 서비스만족도

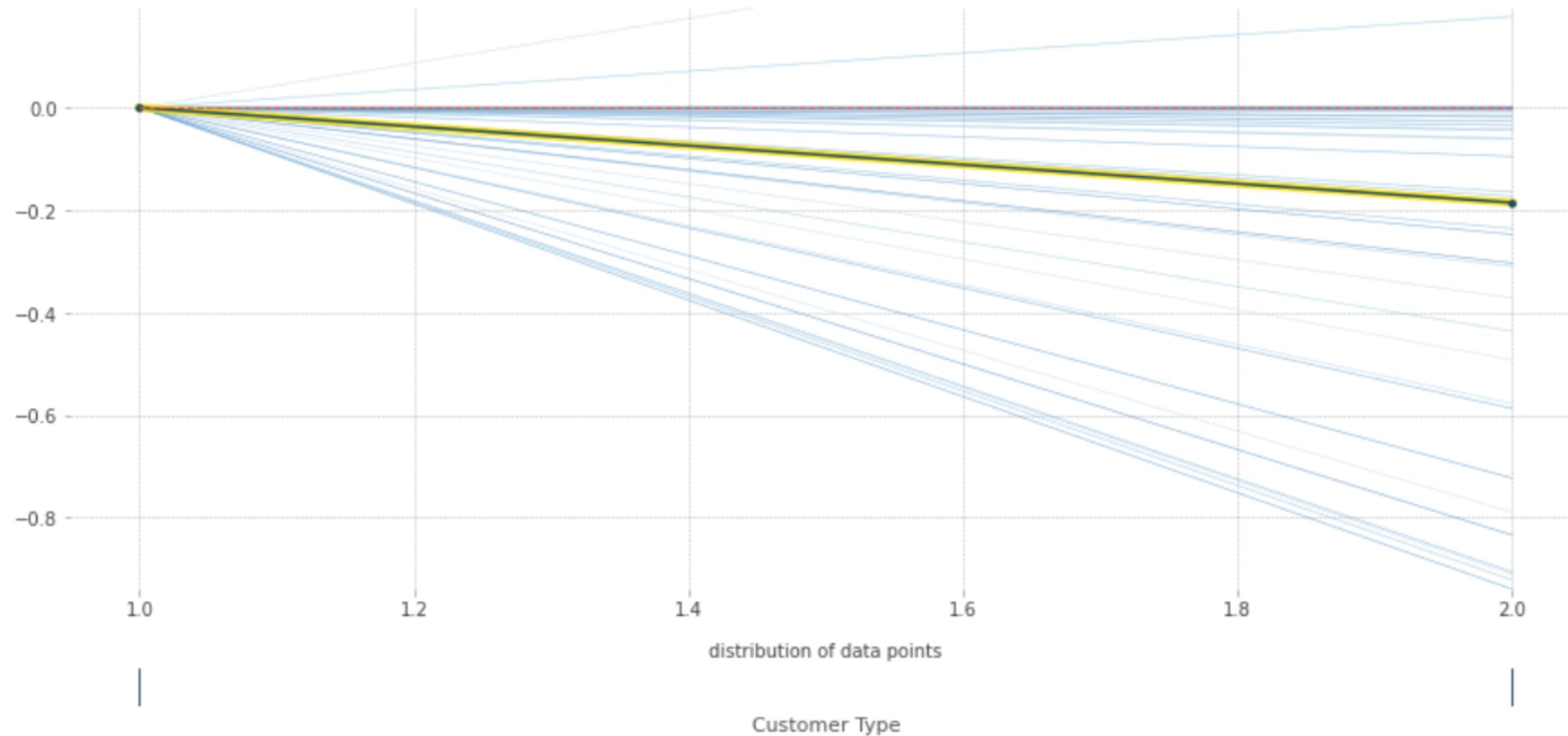
점수가 올라갈수록
만족한다고 응답하는 승객이
많다고 예측한다.

03. 데이터 모델 PDP Plot

XGBClassifier PDP Plot – Customer Type

PDP for feature "Customer Type"

Number of unique grid points: 2

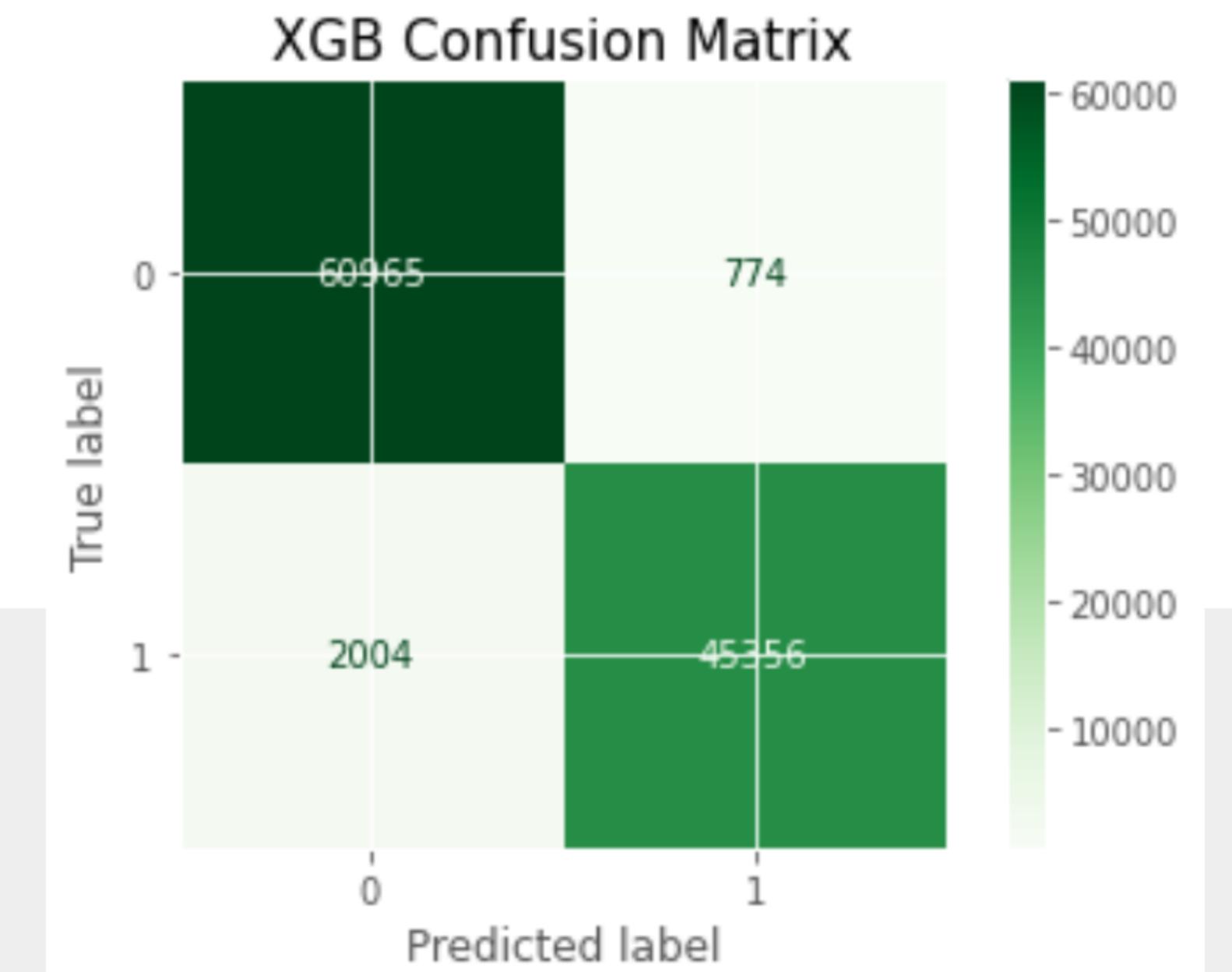
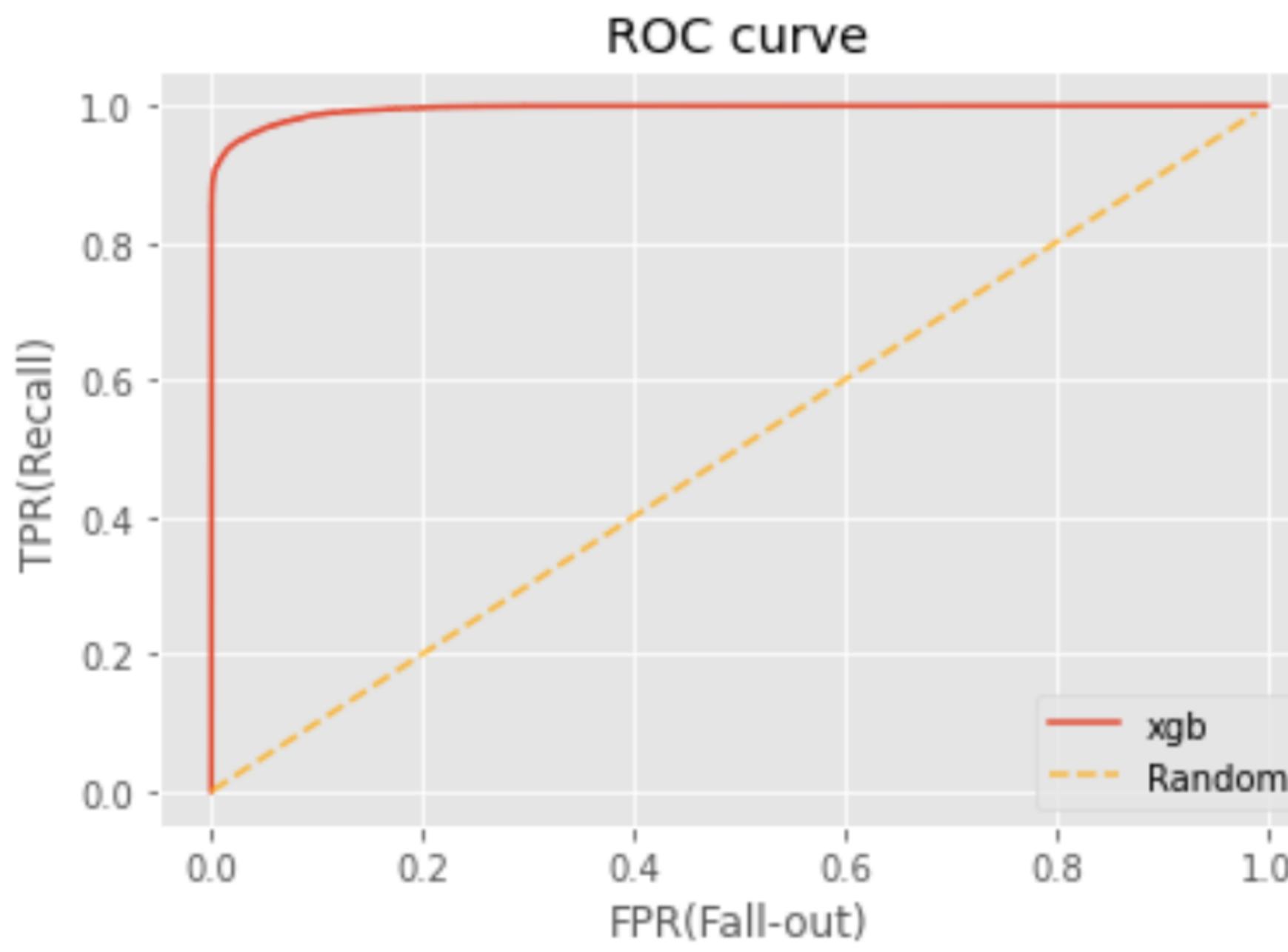


- 1: 로얄 승객
- 2: 일반 승객

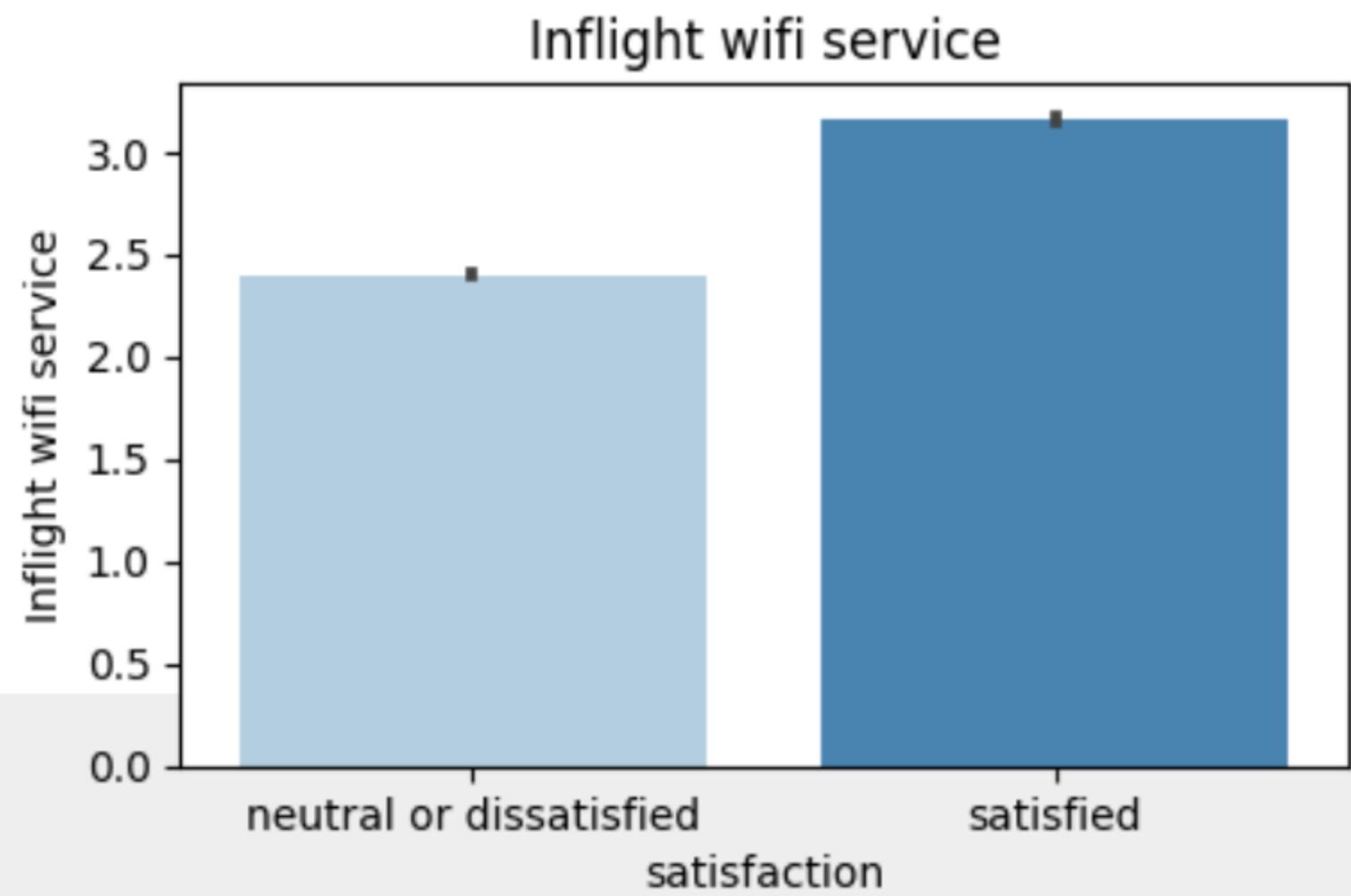
로얄 승객보다 일반 승객일 때 만족하지 않는다고 예측한다.

04. 최종모델(train + test)

roc_auc: 0.997
정확성: 0.97



04. 전략



06. 전략일정



THANK YOU!

