

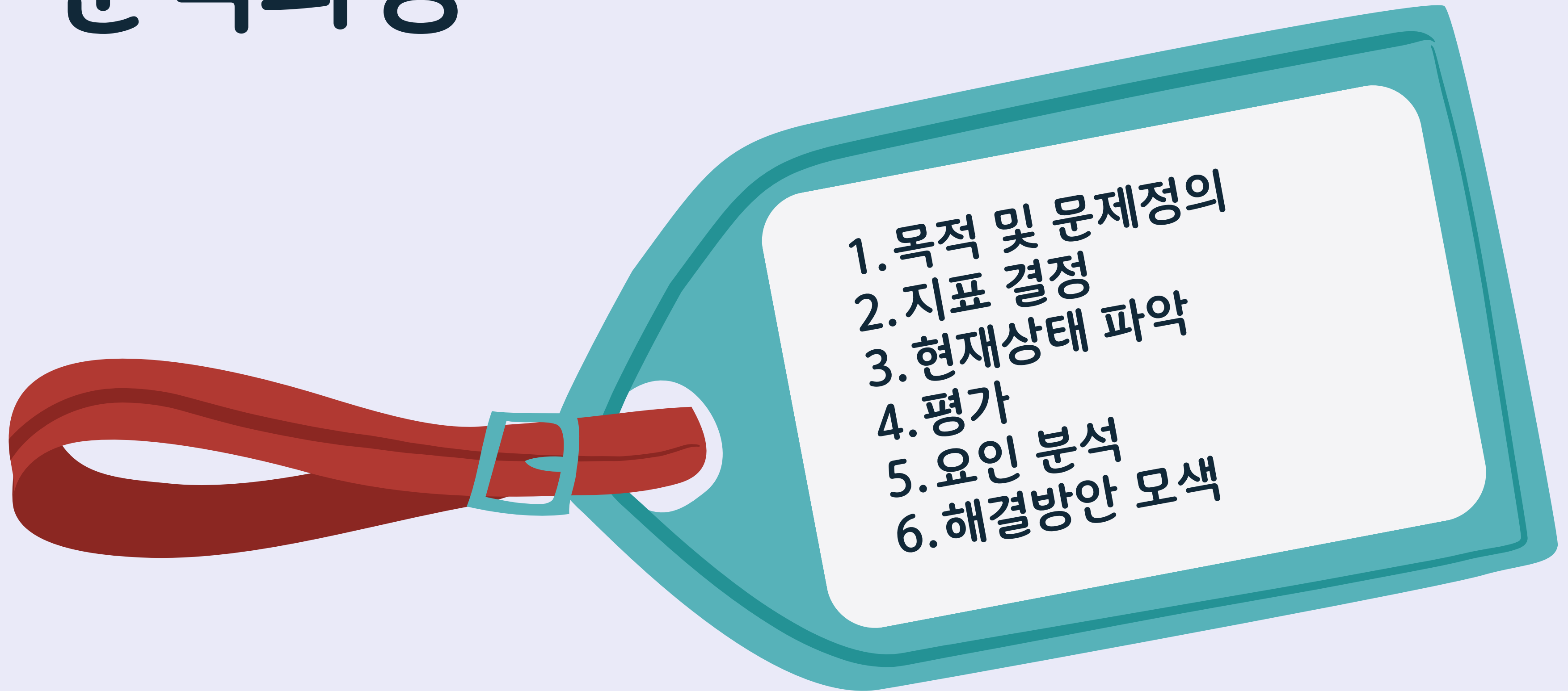


AIRLINE PASSENGER SATISFACTION

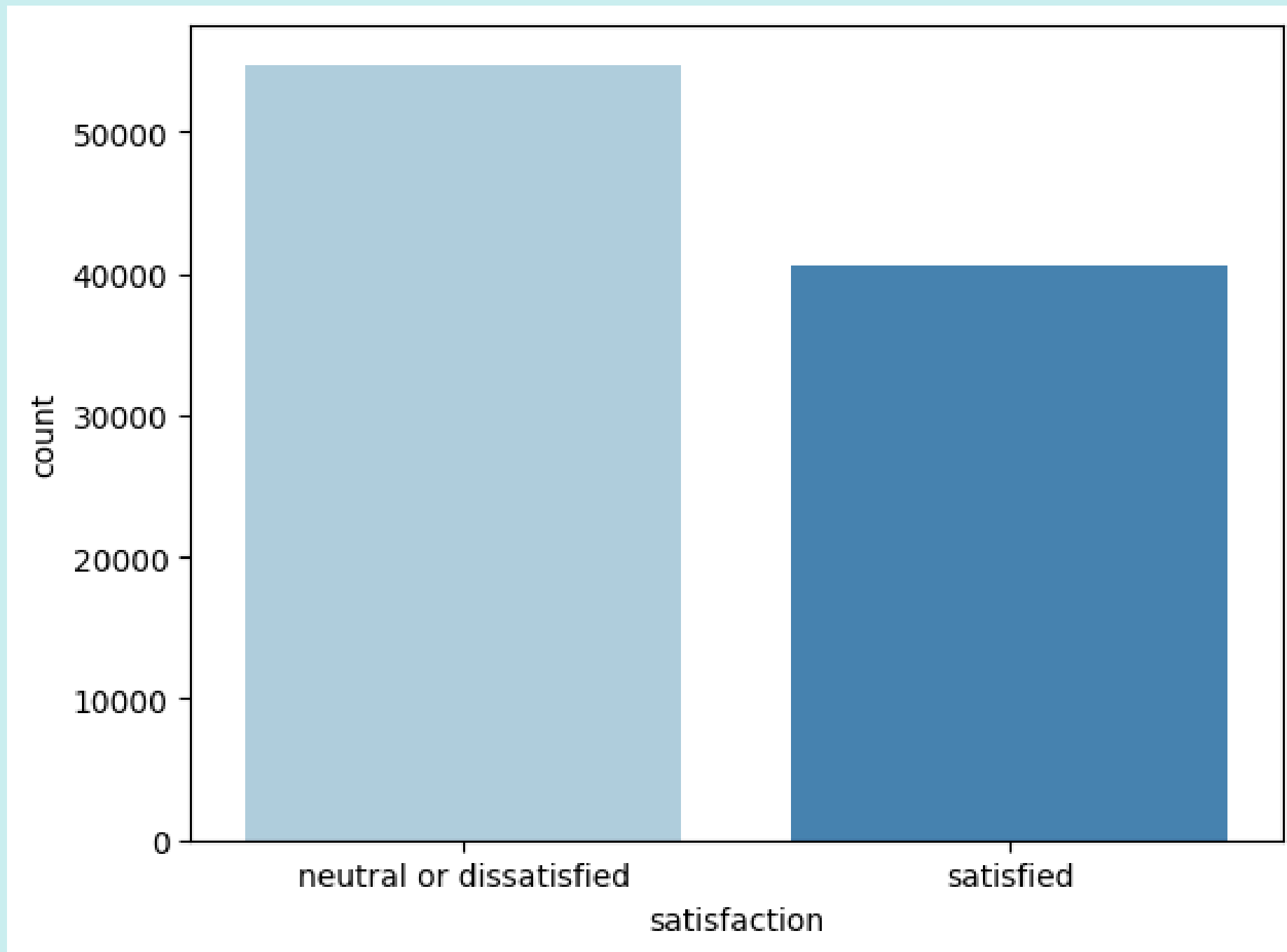
토크프로젝트 B-2
김근아, 신단하, 전주현



분석과정



1. 목적 및 문제정의



항공사에 대한 만족도가 satisfied보다
neutral or dissatisfied가 더 많다.

2. 지표설정

승객관련 정보

Gender: Gender of the passengers (Female, Male)

Customer Type: The customer type (Loyal customer, disloyal customer)

Age: The actual age of the passengers

Type of Travel: Purpose of the flight of the passengers (Personal Travel, Business Travel)

Class: Travel class in the plane of the passengers (Business, Eco, Eco Plus)

Flight distance: The flight distance of this journey

서비스 관련지표

Inflight wifi service: Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)

Departure/Arrival time convenient: Satisfaction level of Departure/Arrival time convenient

Ease of Online booking: Satisfaction level of online booking

Gate location: Satisfaction level of Gate location

Food and drink: Satisfaction level of Food and drink

Online boarding: Satisfaction level of online boarding

Seat comfort: Satisfaction level of Seat comfort

Inflight entertainment: Satisfaction level of inflight entertainment

On-board service: Satisfaction level of On-board service

Leg room service: Satisfaction level of Leg room service

Baggage handling: Satisfaction level of baggage handling

Check-in service: Satisfaction level of Check-in service

Inflight service: Satisfaction level of inflight service

Cleanliness: Satisfaction level of Cleanliness

Departure Delay in Minutes: Minutes delayed when departure

Arrival Delay in Minutes: Minutes delayed when Arrival

Satisfaction: Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

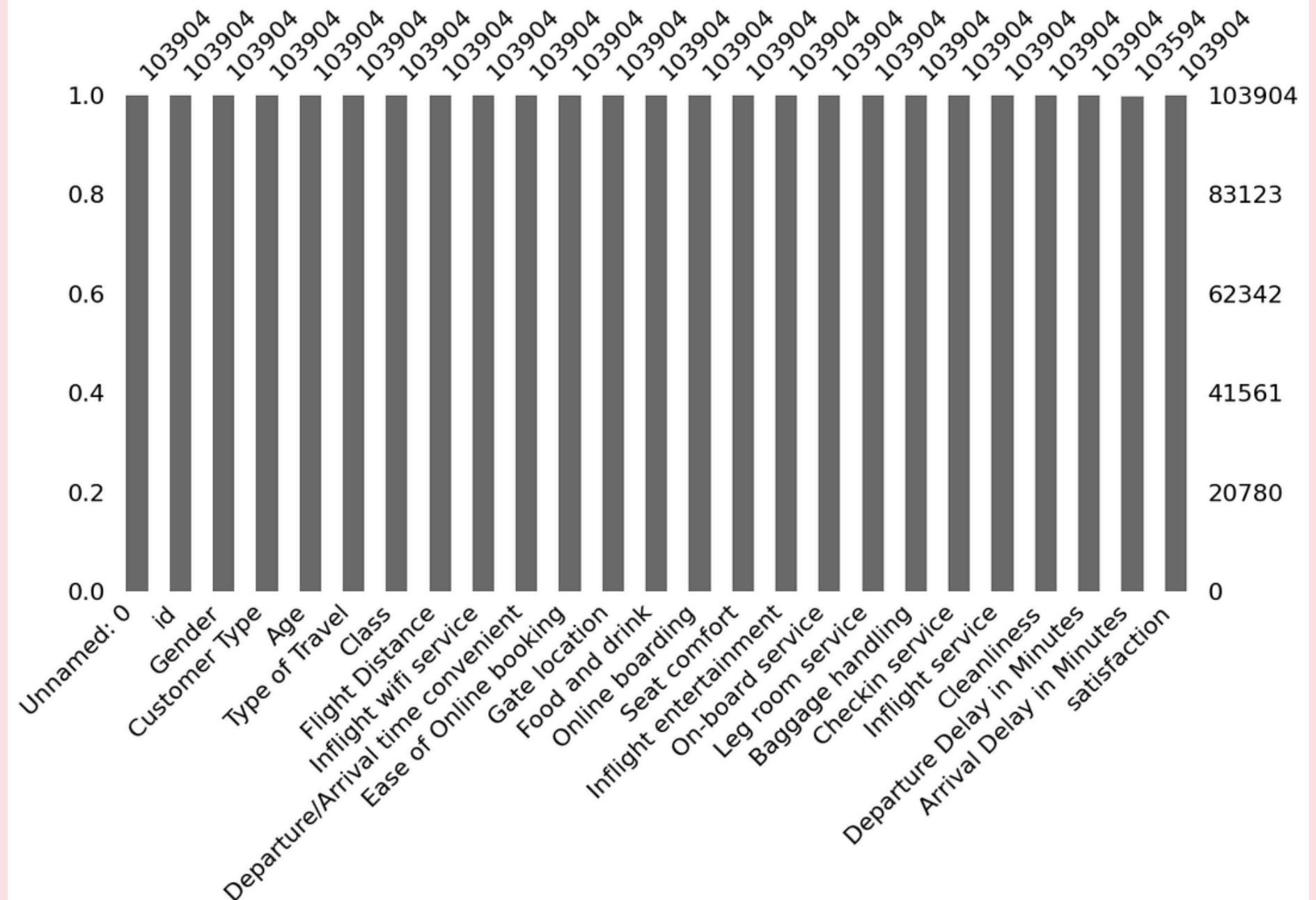
3. 현재 상태 파악

df.info()

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103904 entries, 0 to 103903
Data columns (total 25 columns):
#   Column                                     Non-Null Count  Dtype  
---  -
0   Unnamed: 0                               103904 non-null int64  
1   id                                         103904 non-null int64  
2   Gender                                    103904 non-null object 
3   Customer Type                             103904 non-null object 
4   Age                                        103904 non-null int64  
5   Type of Travel                            103904 non-null object 
6   Class                                     103904 non-null object 
7   Flight Distance                           103904 non-null int64  
8   Inflight wifi service                     103904 non-null int64  
9   Departure/Arrival time convenient         103904 non-null int64  
10  Ease of Online booking                    103904 non-null int64  
11  Gate location                             103904 non-null int64  
12  Food and drink                            103904 non-null int64  
13  Online boarding                           103904 non-null int64  
14  Seat comfort                              103904 non-null int64  
15  Inflight entertainment                    103904 non-null int64  
16  On-board service                          103904 non-null int64  
17  Leg room service                          103904 non-null int64  
18  Baggage handling                          103904 non-null int64  
19  Checkin service                           103904 non-null int64  
20  Inflight service                           103904 non-null int64  
21  Cleanliness                               103904 non-null int64  
22  Departure Delay in Minutes                103904 non-null int64  
23  Arrival Delay in Minutes                  103594 non-null float64 
24  satisfaction                               103904 non-null object 
dtypes: float64(1), int64(19), object(5)
```

결측치 확인



3. 현재 상태 파악

```
Unnamed: 0      0
id              0
Gender          0
Customer Type   0
Age            0
Type of Travel  0
Class           0
Flight Distance 0
Inflight wifi service 0
Departure/Arrival time convenient 0
Ease of Online booking 0
Gate location   0
Food and drink  0
Online boarding 0
Seat comfort    0
Inflight entertainment 0
On-board service 0
Leg room service 0
Baggage handling 0
Checkin service 0
Inflight service 0
Cleanliness     0
Departure Delay in Minutes 0
Arrival Delay in Minutes 0
satisfaction    0
dtype: int64
```

결측치 처리

세부 서비스 만족도의 점수가
0(무응답)인 경우,
결측치 처리 후 제거

제거 후에도 약 10만개의 데이터가
남았기 때문에 과감하게 제거하였음.



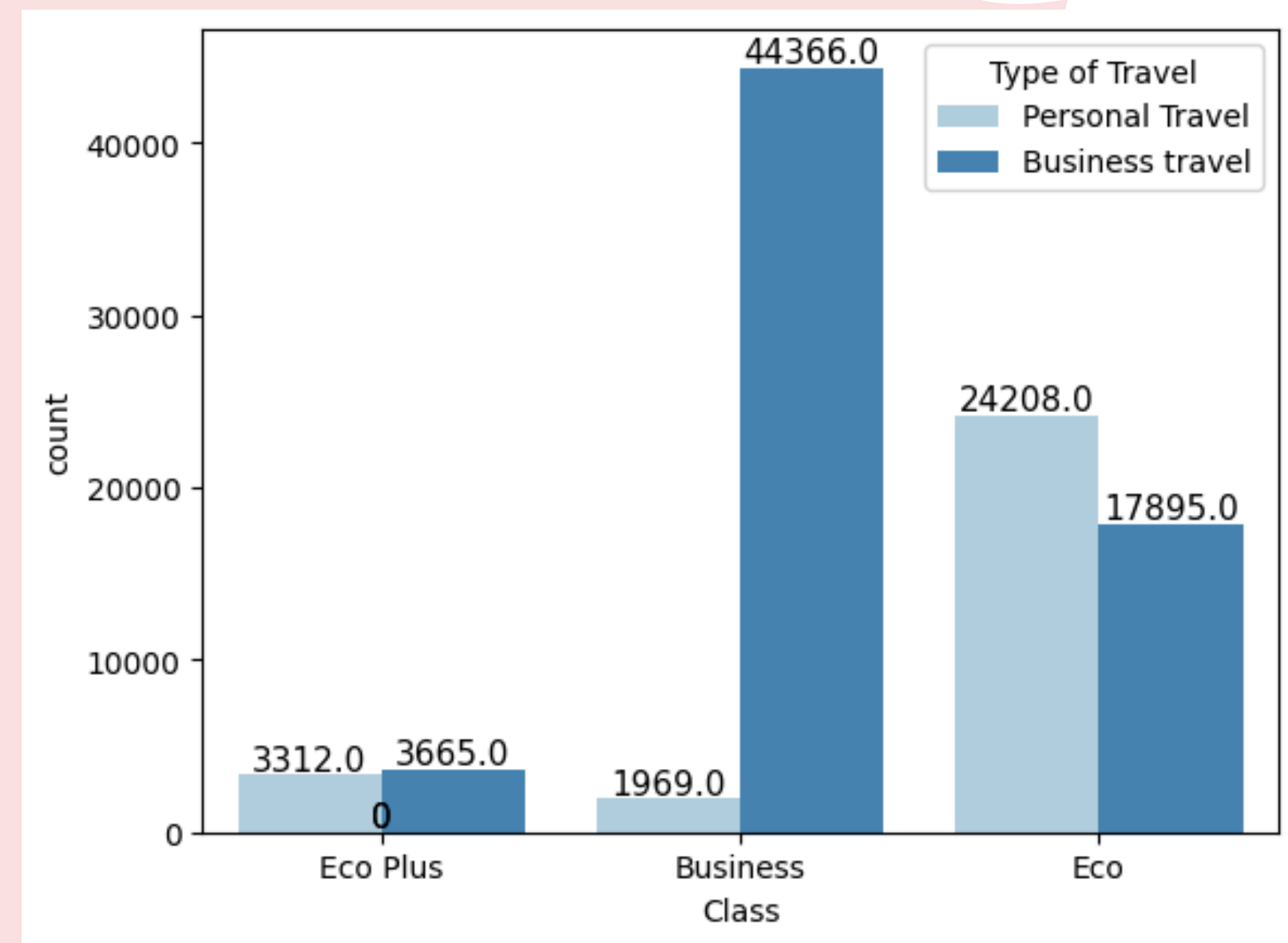
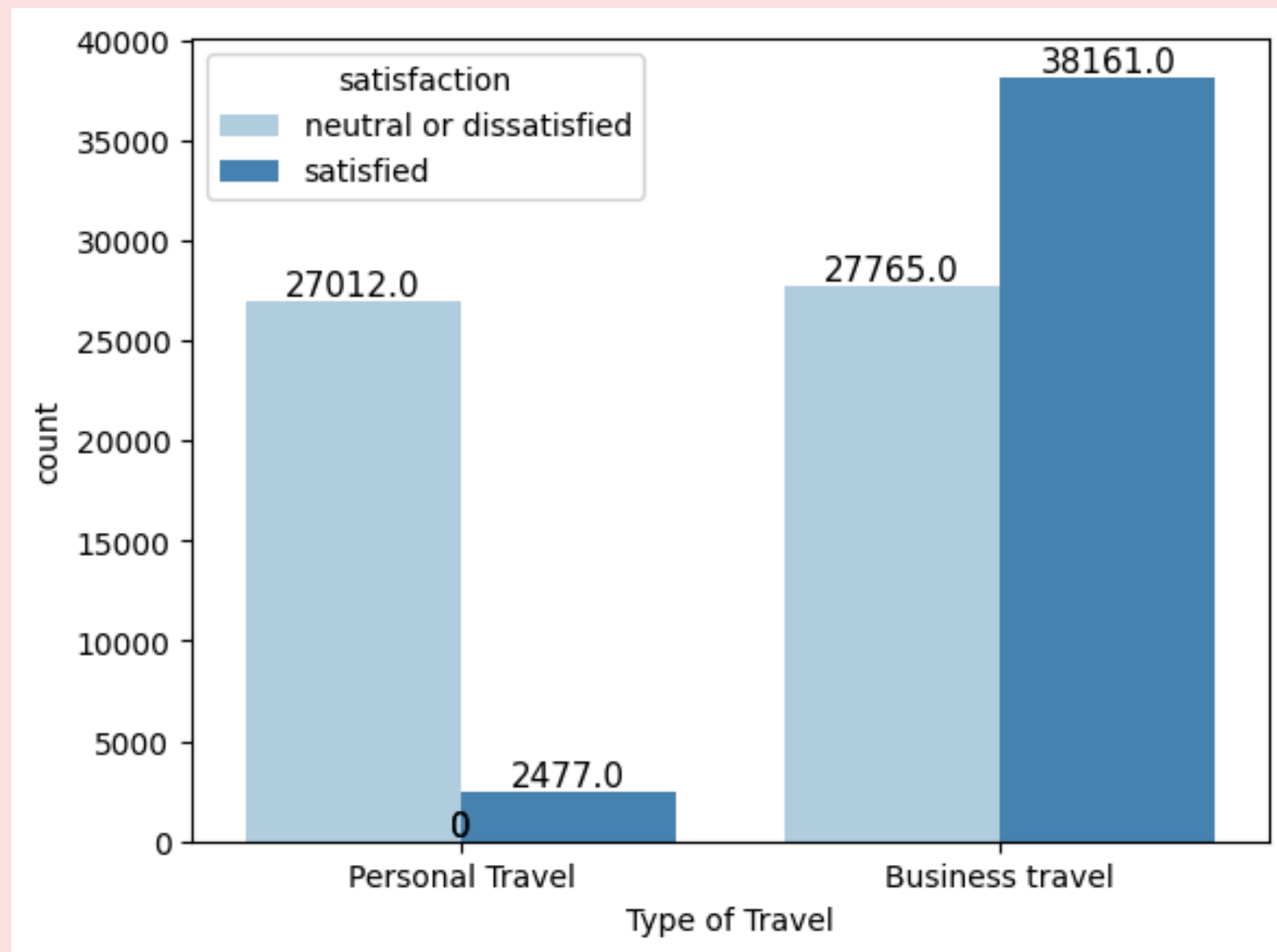
3. 현재 상태 파악

Inflight wifi service	Departure/Arrival time convenient	...	On- board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes	satisfaction	satisfaction_binary
3.0	4.0	...	4.0	3.0	4	4.0	5.0	5.0	25	18.0	neutral or dissatisfied	1
3.0	2.0	...	1.0	5.0	3	1.0	4.0	1.0	1	6.0	neutral or dissatisfied	1
2.0	2.0	...	4.0	3.0	4	4.0	4.0	5.0	0	0.0	satisfied	5
2.0	5.0	...	2.0	5.0	3	1.0	4.0	2.0	11	9.0	neutral or dissatisfied	1
3.0	3.0	...	3.0	4.0	4	3.0	3.0	3.0	0	0.0	satisfied	5
...
2.0	1.0	...	3.0	1.0	4	2.0	3.0	2.0	3	0.0	neutral or dissatisfied	1
4.0	4.0	...	5.0	5.0	5	5.0	5.0	4.0	0	0.0	satisfied	5
1.0	1.0	...	3.0	2.0	4	5.0	5.0	4.0	7	14.0	neutral or dissatisfied	1
1.0	1.0	...	4.0	5.0	1	5.0	4.0	1.0	0	0.0	neutral or dissatisfied	1
1.0	3.0	...	1.0	1.0	4	4.0	3.0	1.0	0	0.0	neutral or dissatisfied	1

히트맵을 위해 binary 생성

3. 현재 상태 파악

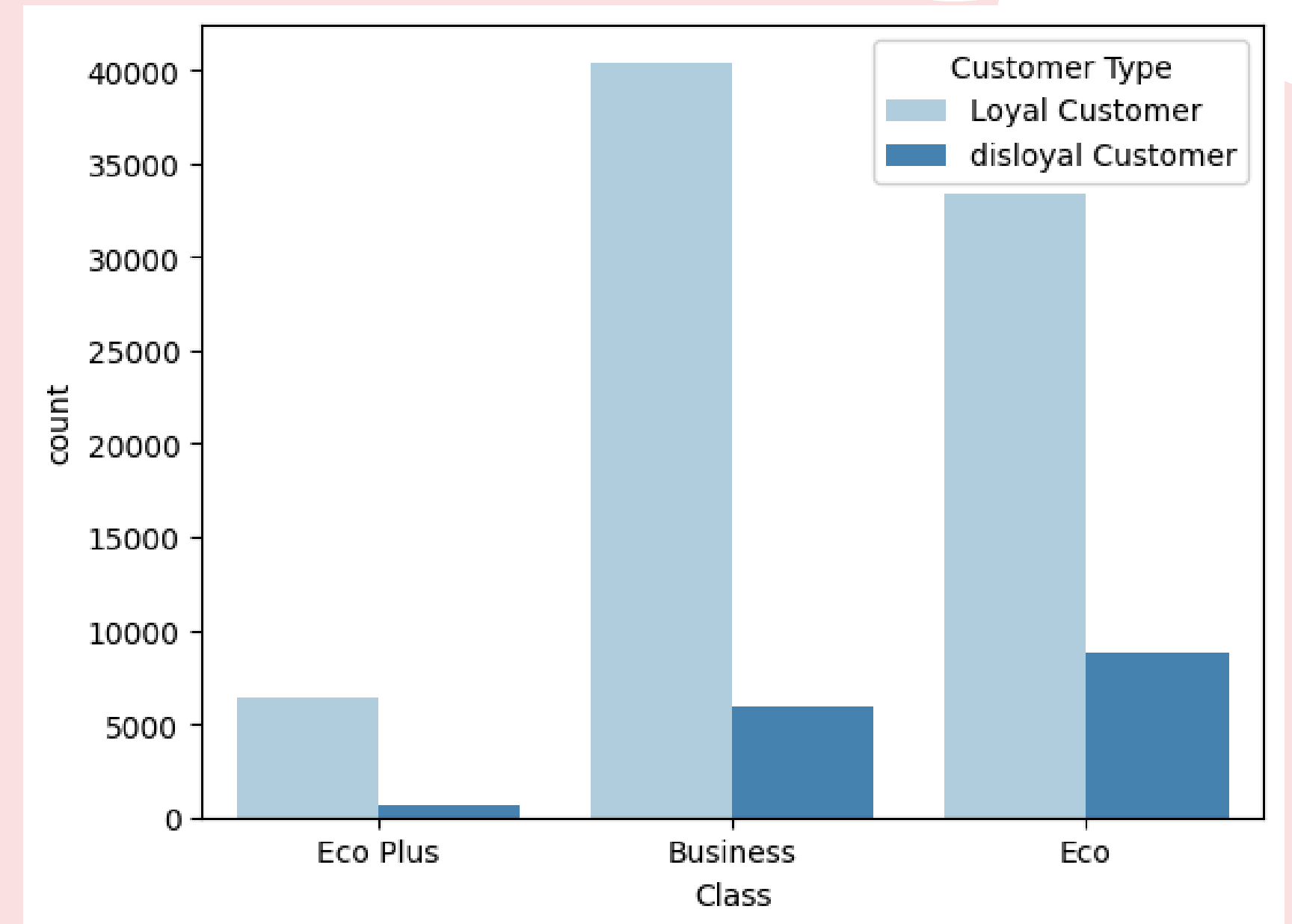
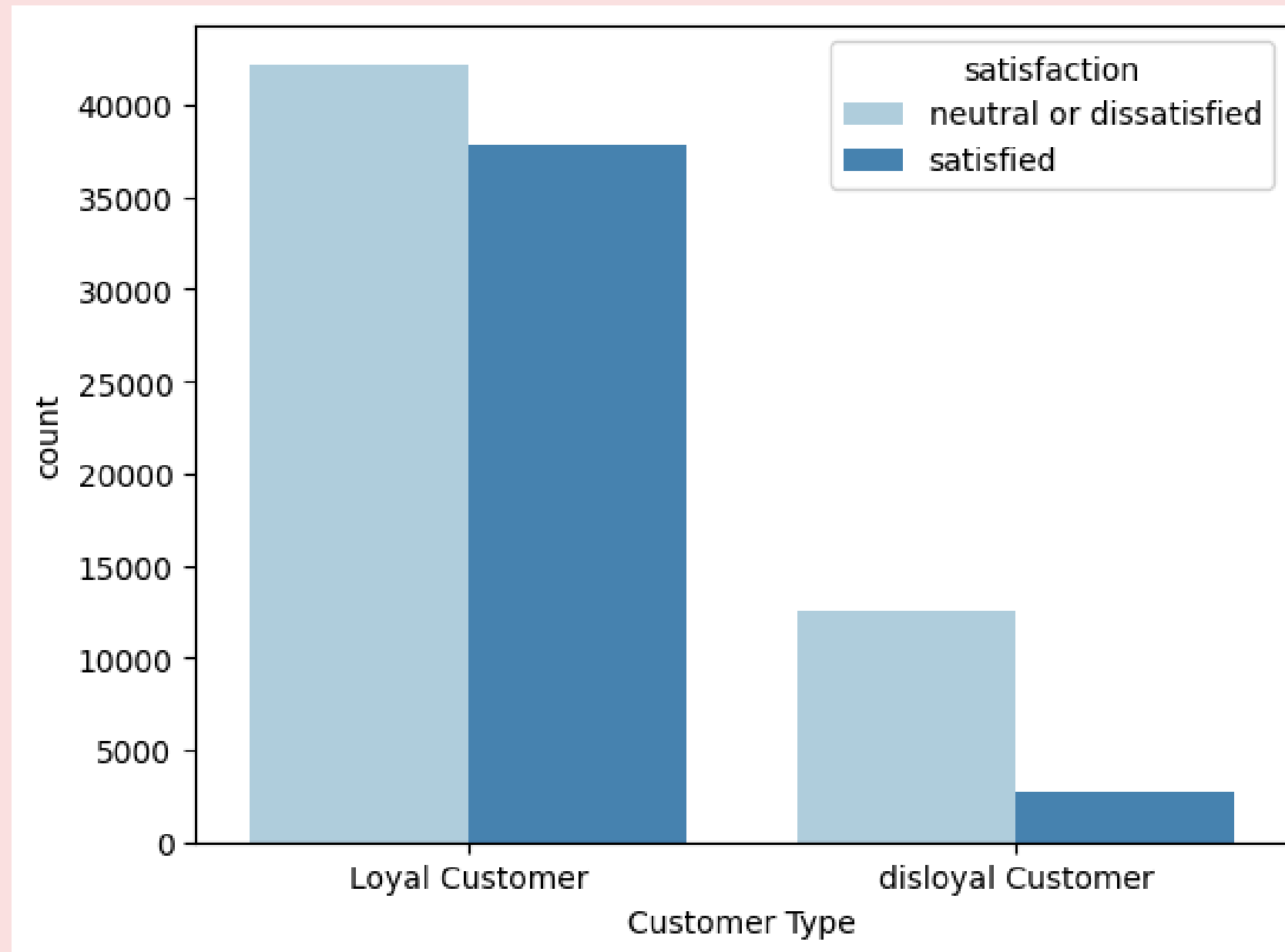
Insight 1 : 비즈니스 여행보다 개인 여행의 만족도가 더 낮음



Class와 연관성 파악

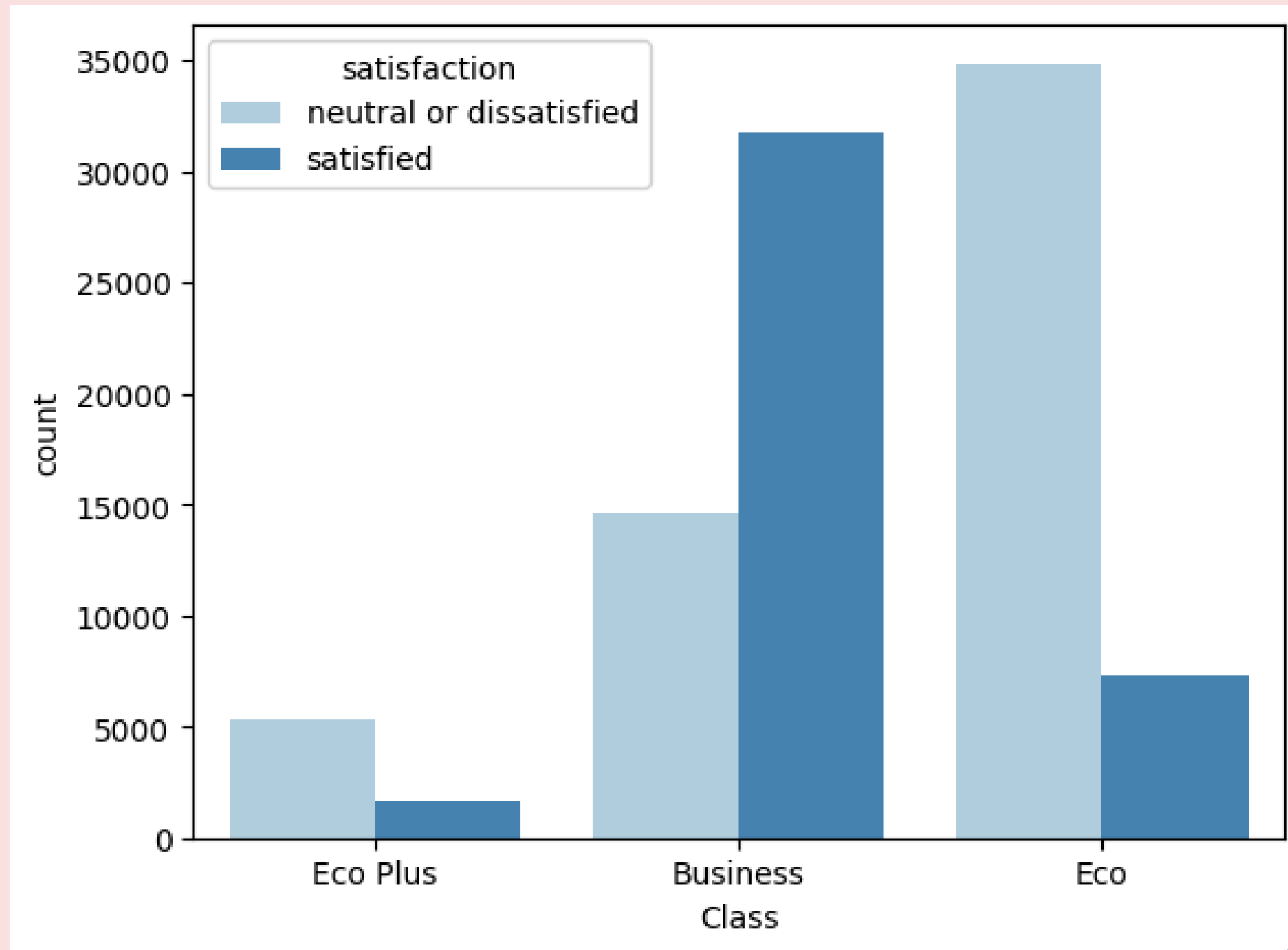
3. 현재 상태 파악

Insight 2 : 충성고객보다 비충성고객의 만족도가 더 낮음



Class와 연관성 파악

3. 현재 상태 파악

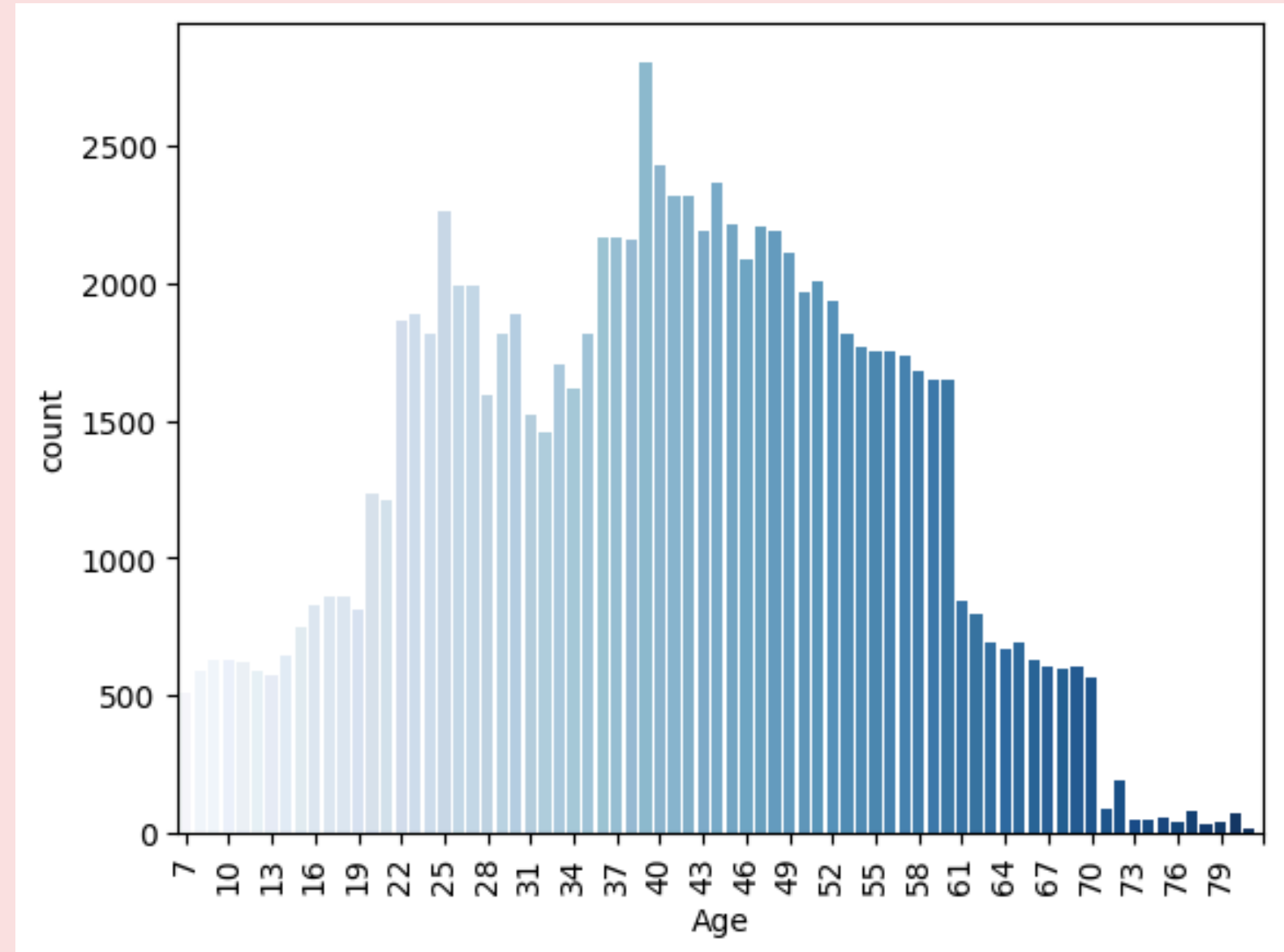
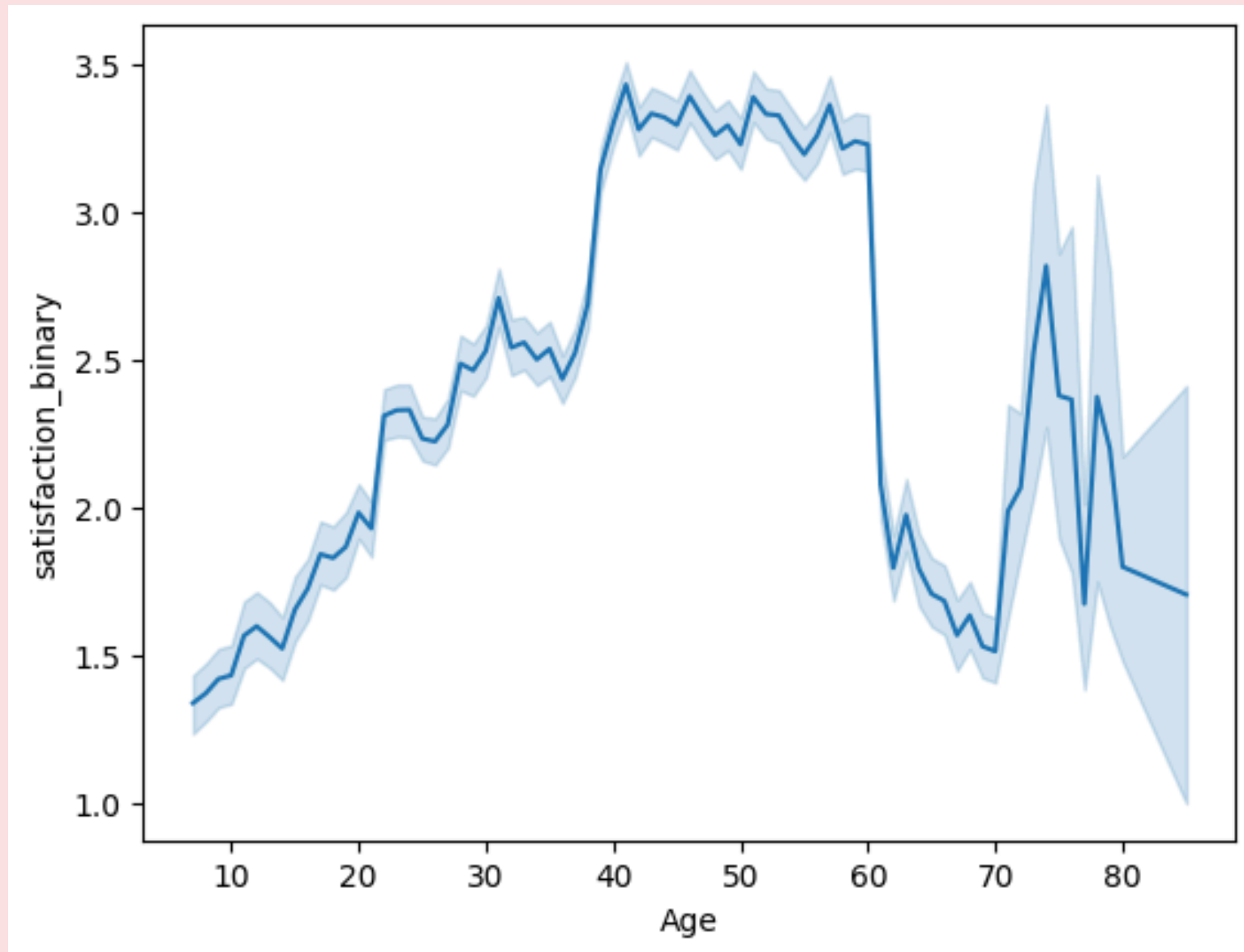


Insight 3:
Eco와 Eco Plus Class에서
유독 만족도가 낮음



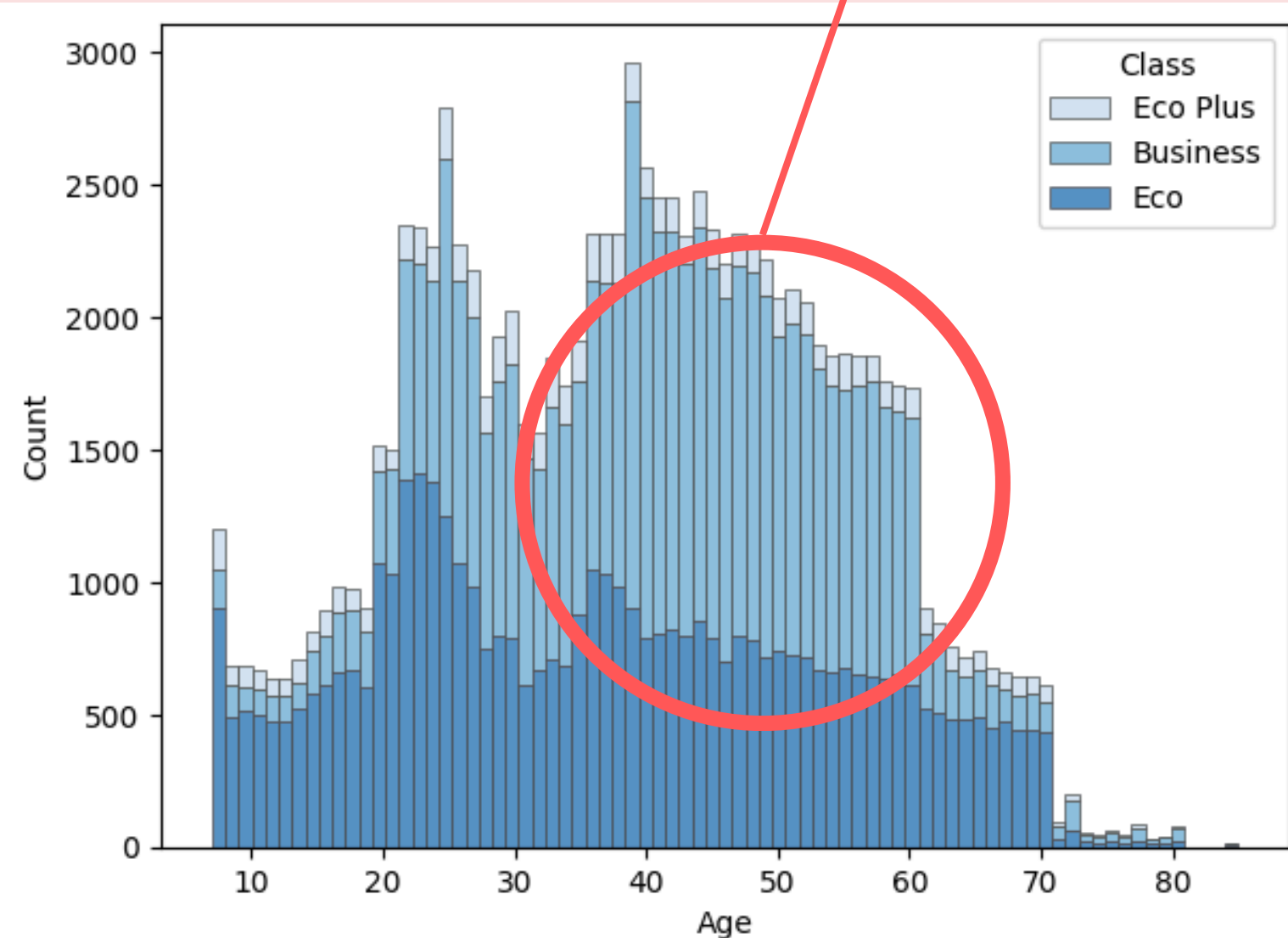
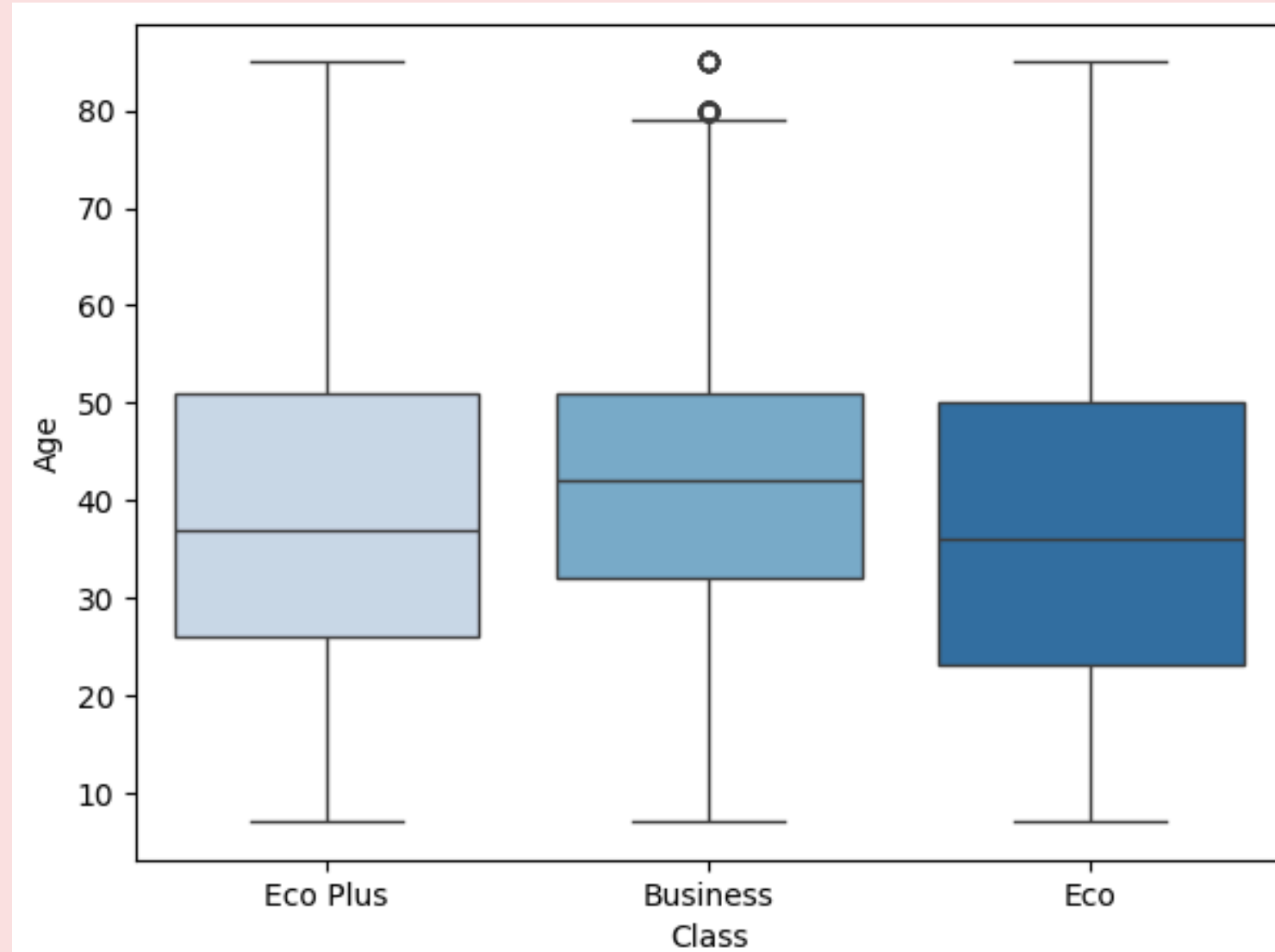
3. 현재 상태 파악

Insight 4: 40-60대 사이의 만족도가 높은 경향이 있음



70대 이상의 데이터 개수가 매우 작아, 이를 제외하면 40-60대 사이의 만족도가 유독 높음.
수입이 높고, 출장이 많은 40-60대의 경우, 비즈니스 클래스를 주로 이용할 것으로 예상함.

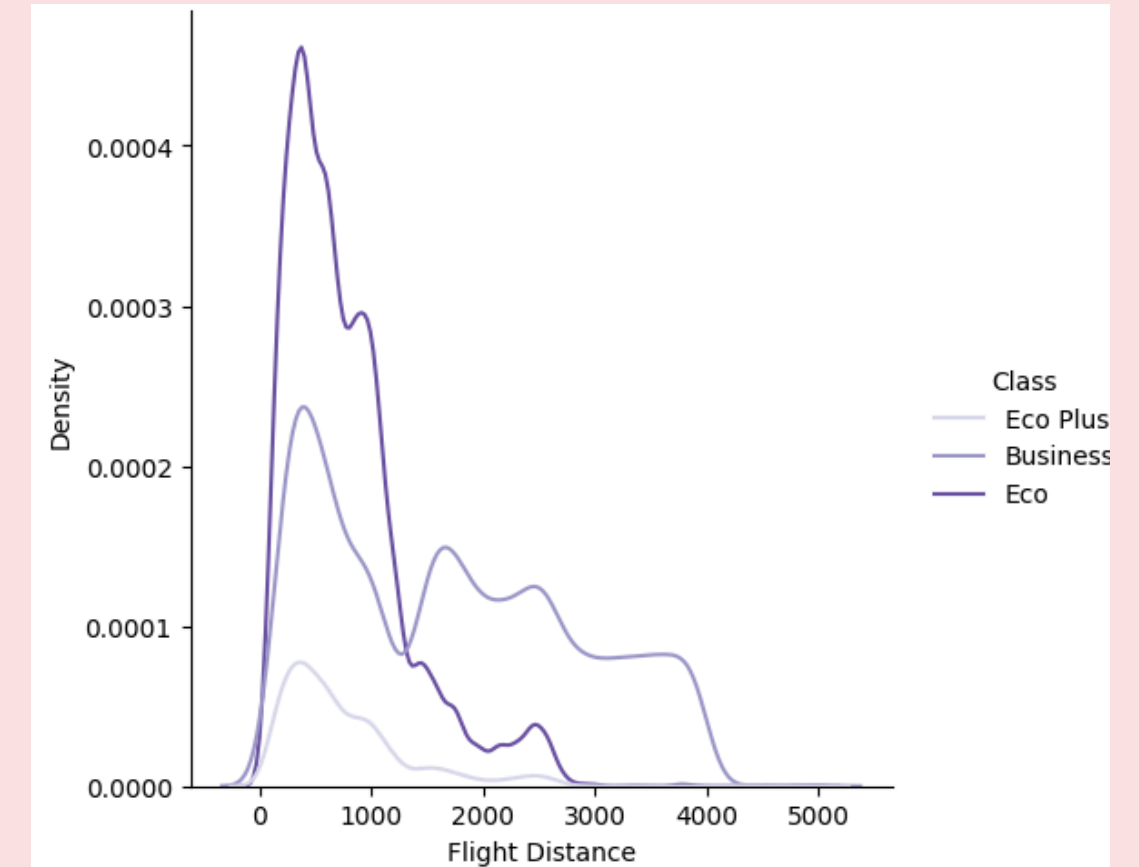
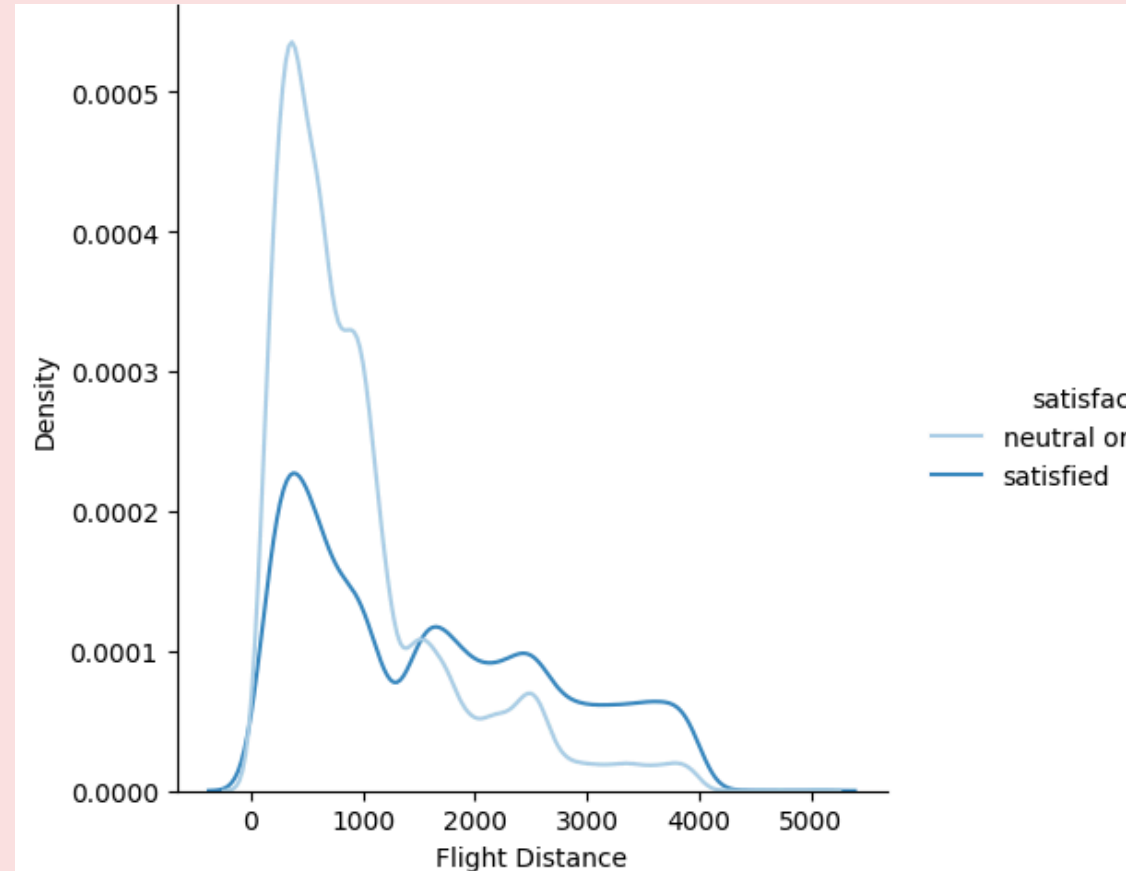
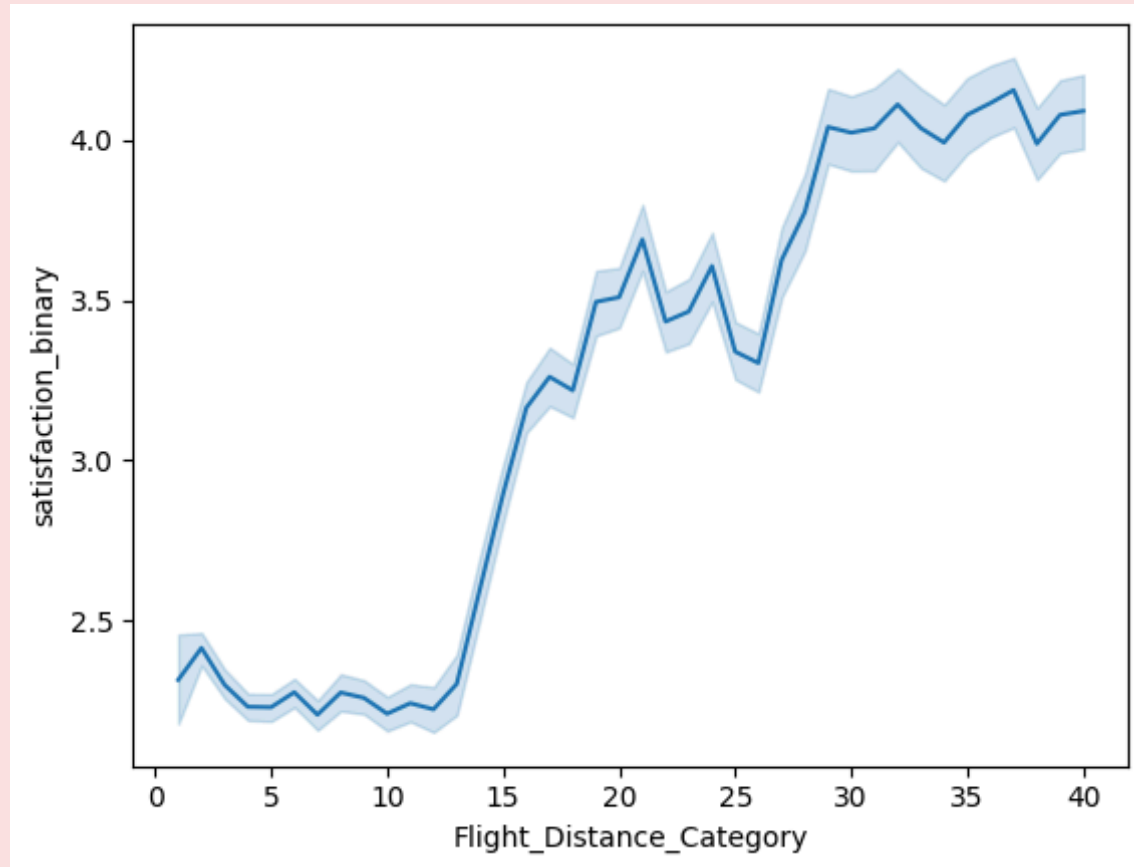
3. 현재 상태 파악



-> 연령대와 Class의 연관성 존재.

3. 현재 상태 파악

Insight 5: 비행 거리가 늘어날수록 만족도가 증가

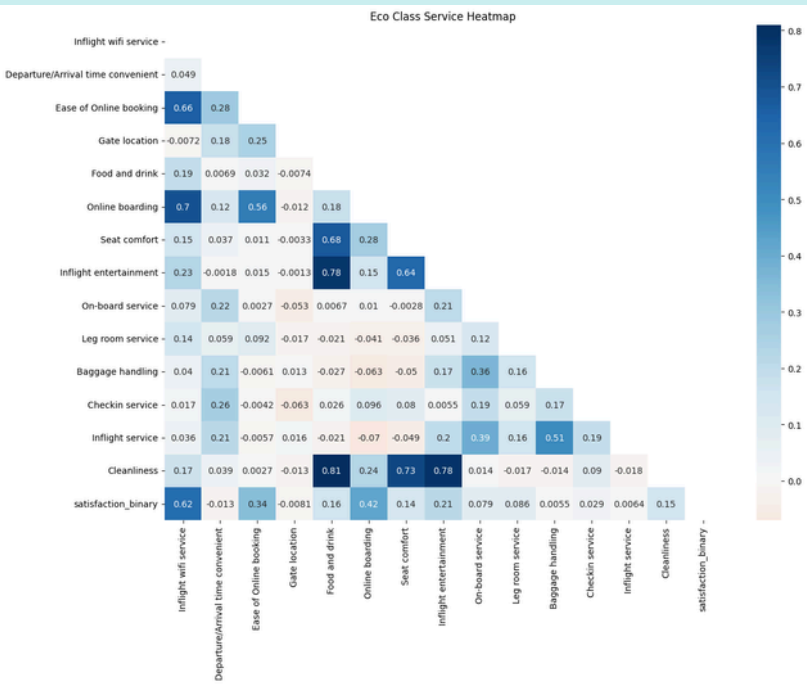


Eco Class 이용객은 짧은 거리(0-1000)에 밀집되어 있음.

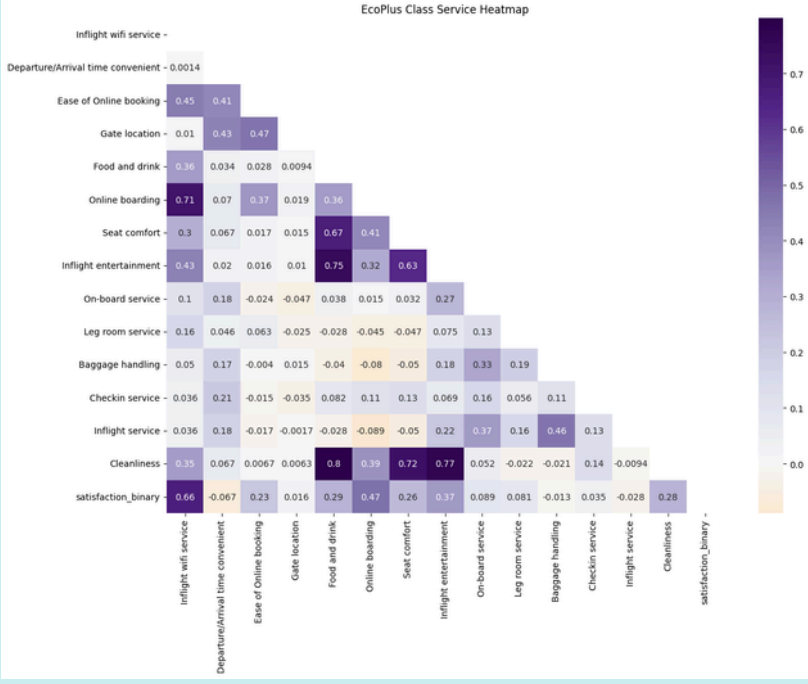
-> 비행거리와 Class의 연관성 존재.

4. 평가 - 상관관계 파악

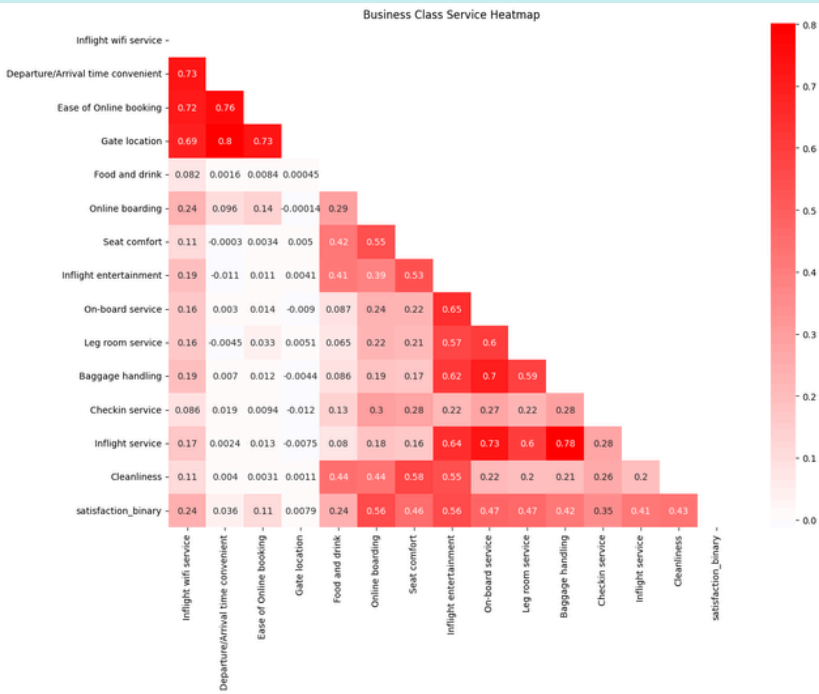
히트맵으로 Class에 따라 서비스 지표 간 상관관계 파악



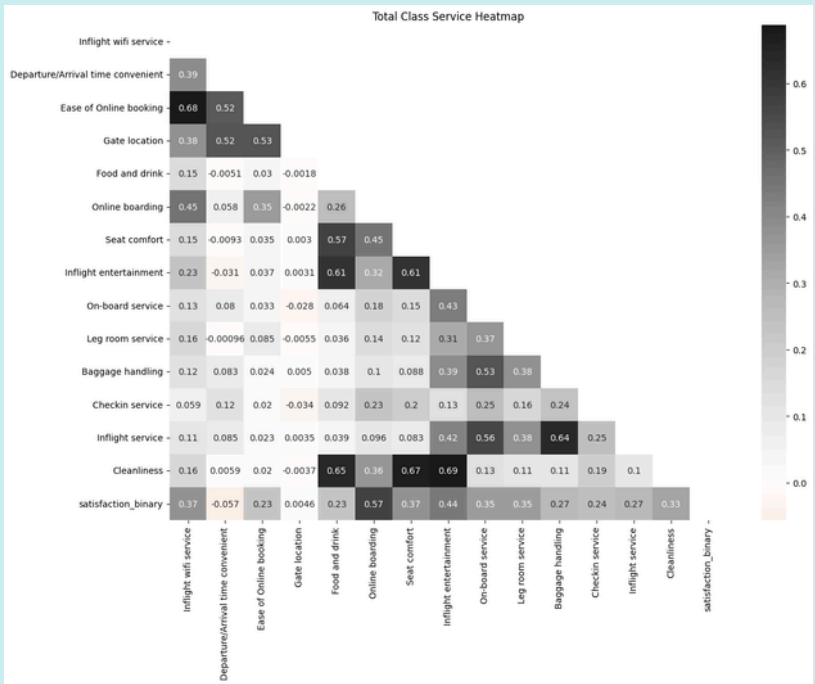
Economy



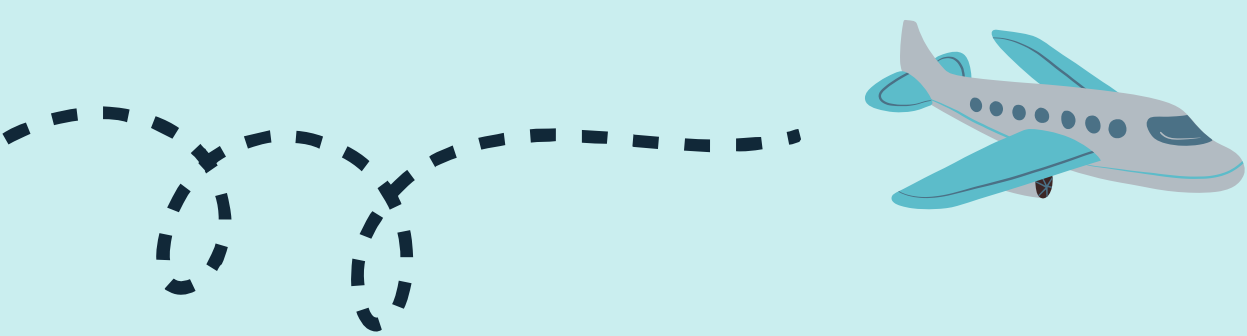
Economy Plus



Business

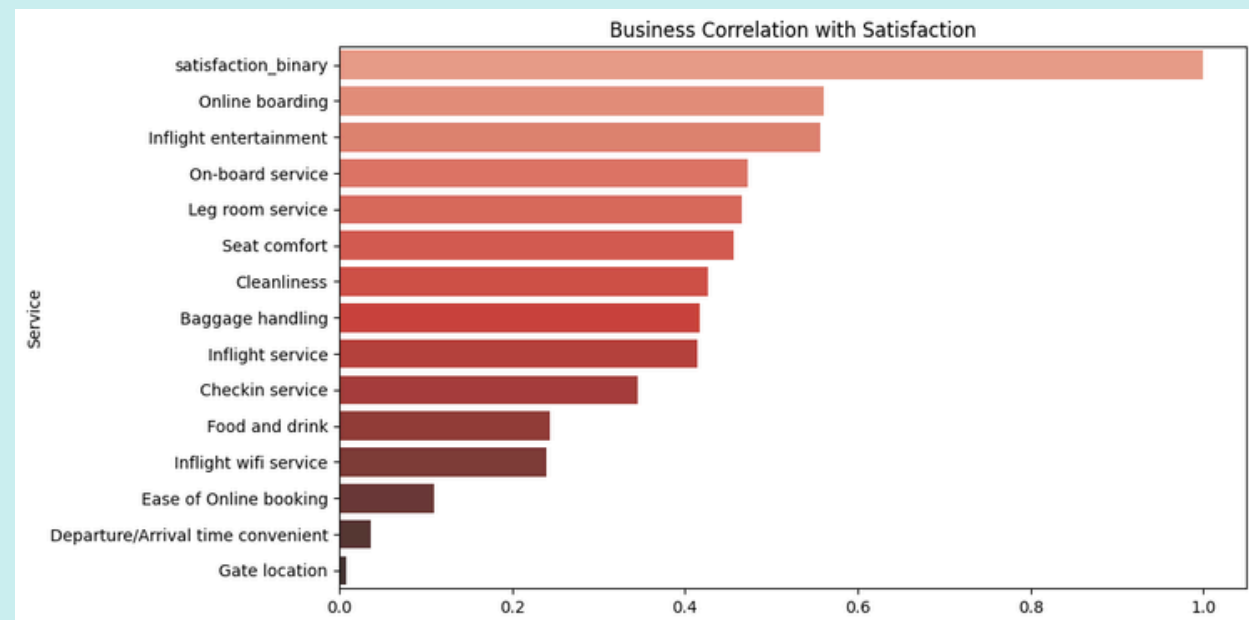
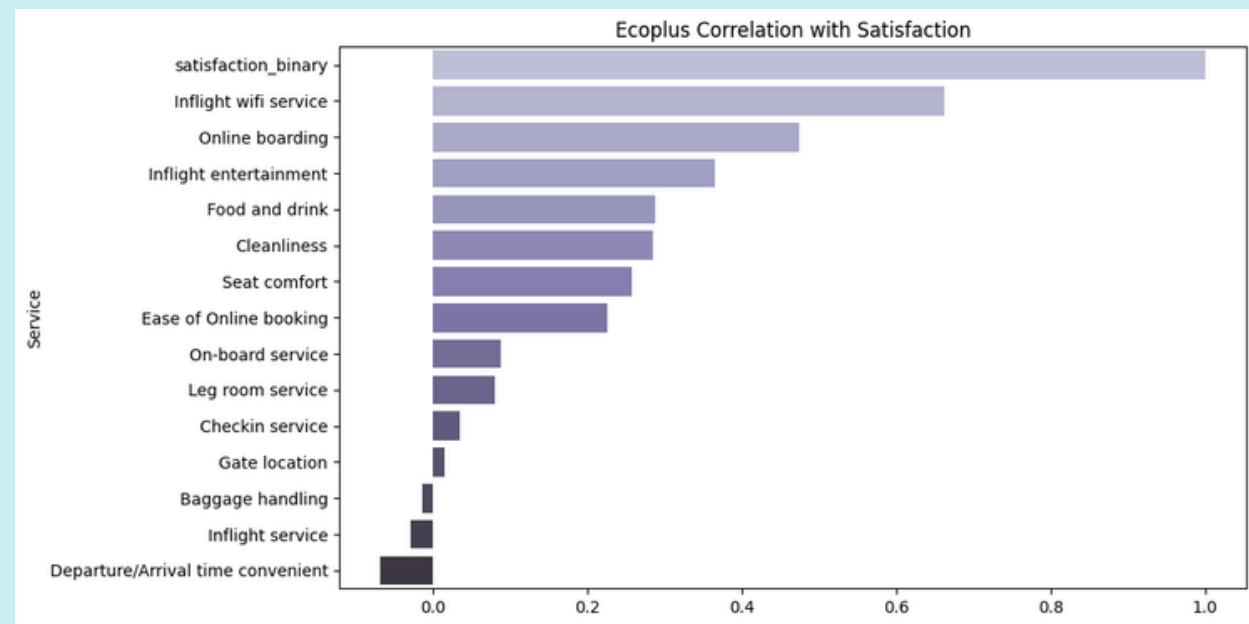
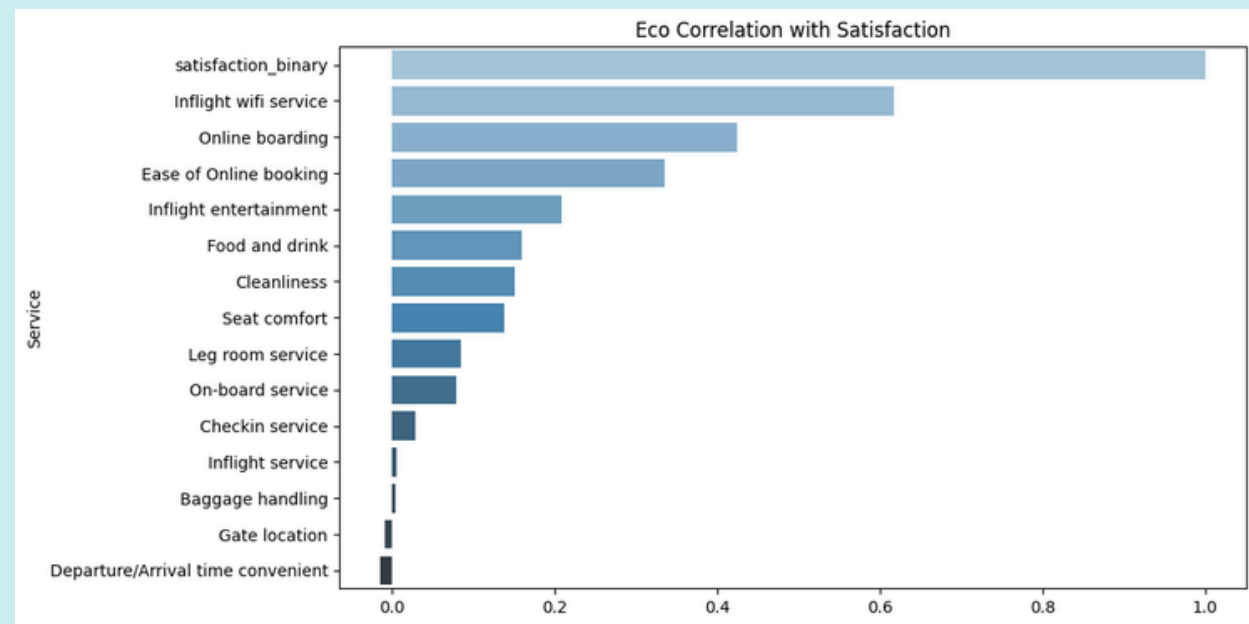


total correlation



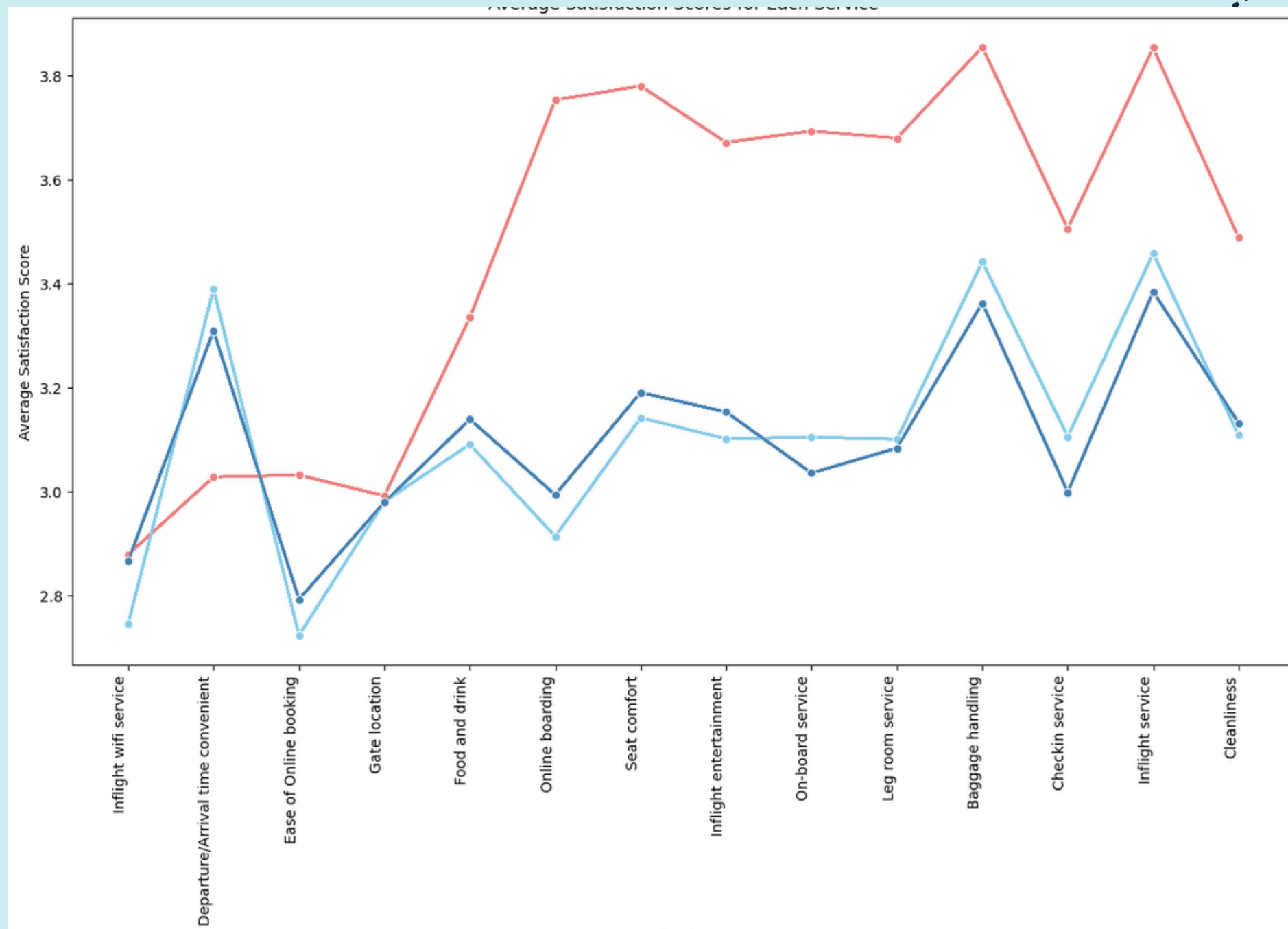
4. 평가 - 상관관계 파악

좌석 Class에 따라 만족도와 상관관계가 가장 높은 서비스 지표 상위 3개 추출



	Economy Class Correlation Rank	Economy Plus Class Correlation Rank	Business Class Correlation Rank
1	Inflight wifi service	Inflight wifi service	Online boarding
2	Online boarding	Online boarding	Inflight Entertainment
3	Ease of Online booking	Inflight Entertainment	On-board service

4. 평가 - 비교



1

Economy Class
Correlation Rank

Inflight wifi service

2

Online boarding

3

Ease of Online booking

Economy Plus Class
Correlation Rank

Inflight wifi service

Online boarding

상관관계가 높은 변수 중에서,
점수를 개선해야할 지표 선정

class에 따라 서비스 지표들 간 평균적인 만족도 비교

5. 요인분석 - 가설검정

가설1

Economy Class 탑승객 중 다음 세가지에서 낮은 점수(1-3)를 준 응답자는 dissatisfied 했을 것이다.

- 1) Inflight wifi service
- 2) Online boarding
- 3) Ease of Online booking

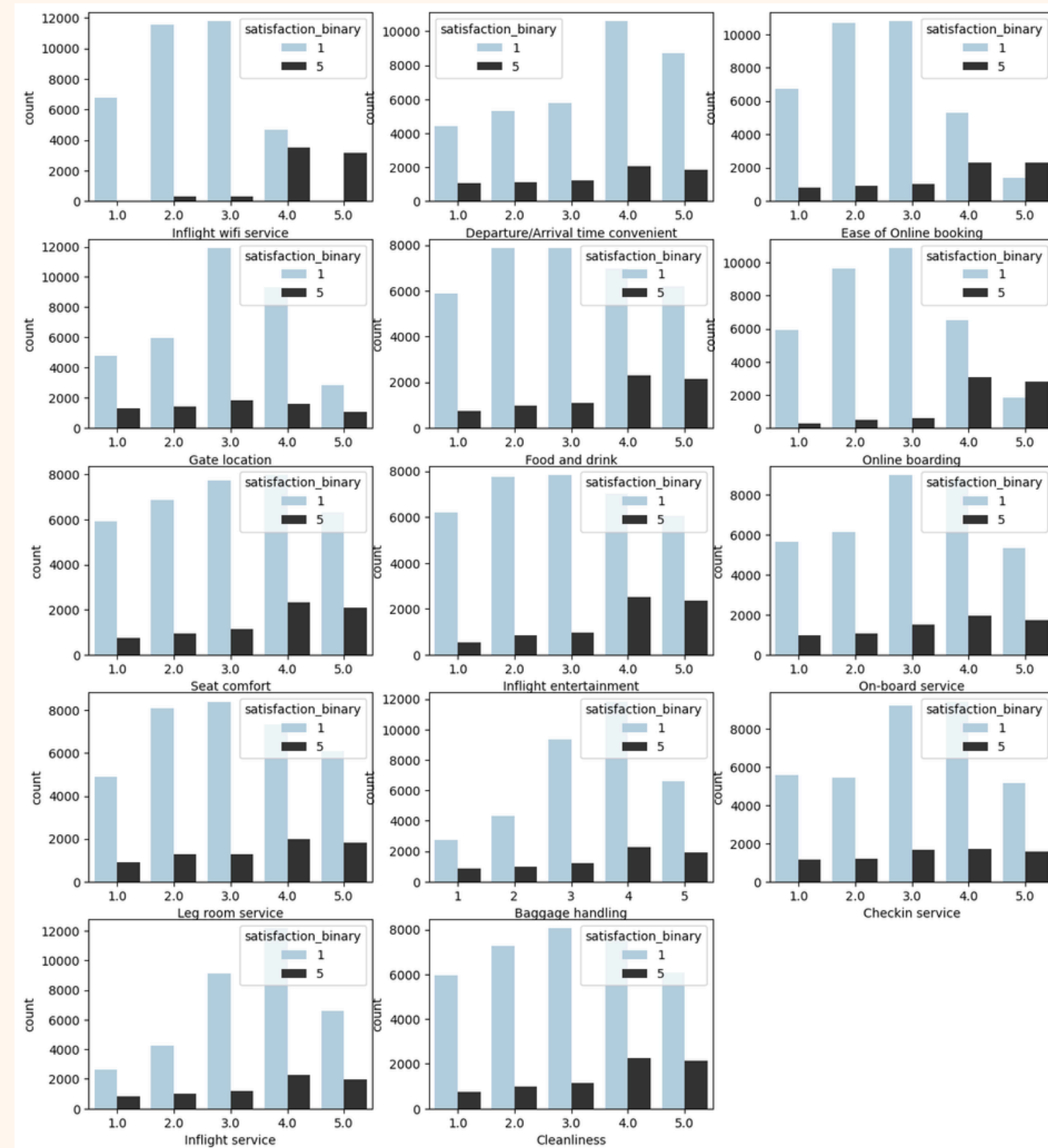
가설 설정 및 검정 과정

H0: Inflight wifi/online boarding/Ease of Online booking service 만족도에서 낮은 점수를 준 응답자가 Dissatisfied했을 것이라 할 수 없다.

H1: Inflight wifi/online boarding/Ease of Online booking service 만족도에서 낮은 점수를 준 응답자가 Dissatisfied했을 것이다.



ECONOMY CLASS



1) Inflight wifi service

```
from scipy.stats import chi2_contingency

# Inflight wifi service와 satisfaction_binary 사이의 교차 테이블 생성
cross_tab = pd.crosstab(echo_df["Inflight wifi service"] <= 3, echo_df["satisfaction_binary"])

# 카이제곱검정 시행
chi2, p_val, _, _ = chi2_contingency(cross_tab)

# 결과 출력
print("카이제곱검정 결과 p-value:", p_val)

# 유의수준 설정 (예: 0.05)
alpha = 0.05

# 유의수준과 비교하여 결과 해석
if p_val < alpha:
    print("귀무가설을 기각합니다. Dissatisfied 응답자는 Inflight wifi service 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.")
else:
    print("귀무가설을 기각하지 않습니다. Dissatisfied 응답자는 Inflight wifi service 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단할 수 없습니다.")

카이제곱검정 결과 p-value: 0.0
귀무가설을 기각합니다. Dissatisfied 응답자는 Inflight wifi service 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.
```

✓ 귀무가설
기각

2) Online boarding

```
[ ] from scipy.stats import chi2_contingency

# Online boarding와 satisfaction_binary 사이의 교차 테이블 생성
cross_tab = pd.crosstab(echo_df["Online boarding"] <= 3, echo_df["satisfaction_binary"])

# 카이제곱검정 시행
chi2, p_val, _, _ = chi2_contingency(cross_tab)

# 결과 출력
print("카이제곱검정 결과 p-value:", p_val)

# 유의수준 설정 (예: 0.05)
alpha = 0.05

# 유의수준과 비교하여 결과 해석
if p_val < alpha:
    print("귀무가설을 기각합니다. Dissatisfied 응답자는 Online boarding 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.")
else:
    print("귀무가설을 기각하지 않습니다. Dissatisfied 응답자는 Online boarding 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단할 수 없습니다.")

카이제곱검정 결과 p-value: 0.0
귀무가설을 기각합니다. Dissatisfied 응답자는 Online boarding 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.
```

✓ 귀무가설
기각

3) Ease of Online booking

```
[ ] from scipy.stats import chi2_contingency

# Ease of Online booking와 satisfaction_binary 사이의 교차 테이블 생성
cross_tab = pd.crosstab(echo_df["Ease of Online booking"] <= 3, echo_df["satisfaction_binary"])

# 카이제곱검정 시행
chi2, p_val, _, _ = chi2_contingency(cross_tab)

# 결과 출력
print("카이제곱검정 결과 p-value:", p_val)

# 유의수준 설정 (예: 0.05)
alpha = 0.05

# 유의수준과 비교하여 결과 해석
if p_val < alpha:
    print("귀무가설을 기각합니다. Dissatisfied 응답자는Ease of Online booking 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.")
else:
    print("귀무가설을 기각하지 않습니다. Dissatisfied 응답자는 Ease of Online booking만족도에서 낮은 점수(1-3)를 주었을 것으로 판단할 수 없습니다.")

카이제곱검정 결과 p-value: 0.0
귀무가설을 기각합니다. Dissatisfied 응답자는Ease of Online booking 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.
```

✓ 귀무가설
기각

5. 요인분석 - 가설검정

가설2

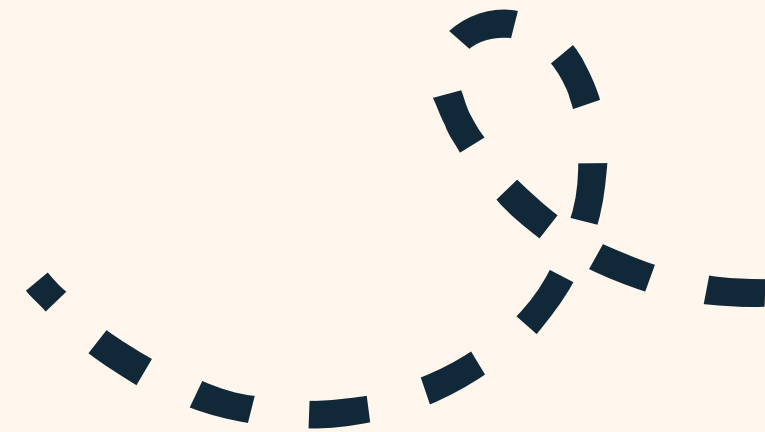
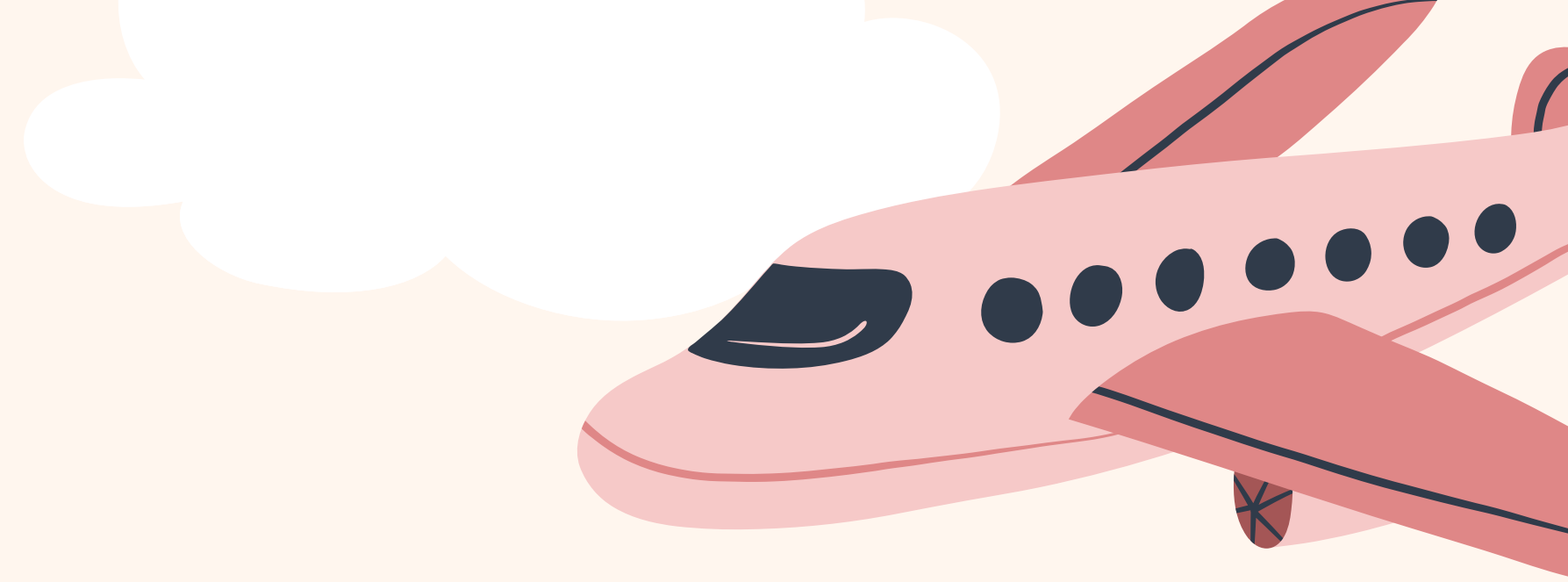
Economy Plus 탑승객 중 다음 두가지에서 낮은 점수(1-3)를 준 응답자는 ‘dissatisfied’ 했을 것이다.

- 1) Inflight wifi service
- 2) Online boarding

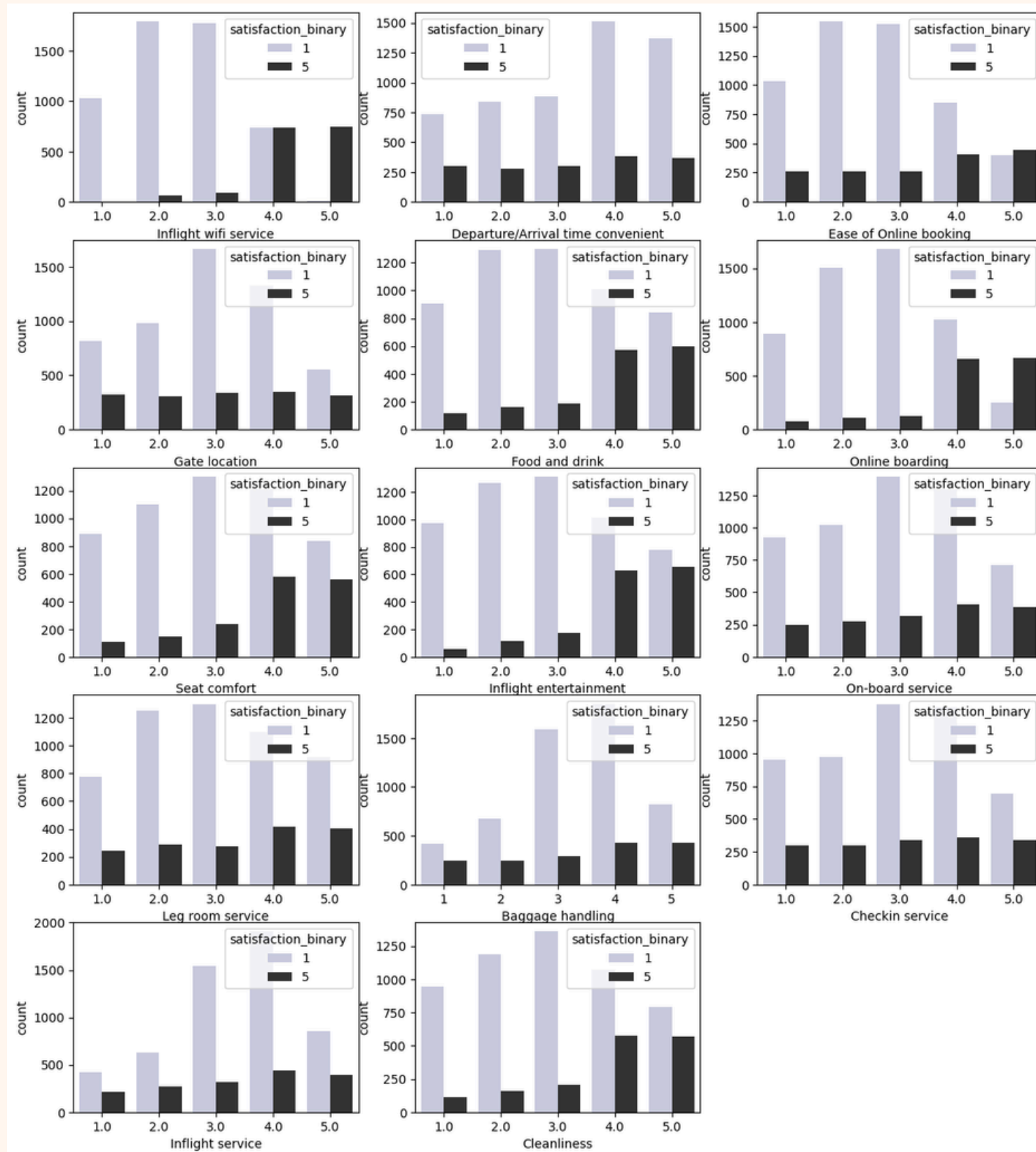
가설 설정 및 검정 과정

H0: Inflight wifi service/Online boarding 만족도에서 낮은 점수 준 응답자가 Dissatisfied 했을 것이라 할 수 없다.

H1: Inflight wifi service/Online boarding 만족도에서 낮은 점수 준 응답자가 Dissatisfied 했을 것이다.



ECONOMY PLUS CLASS



1) Inflight wifi service

카이제곱 (inflight wifi service)

```
[ ] from scipy.stats import chi2_contingency

# Ease of Online booking와 satisfaction_binary 사이의 교차 테이블 생성
cross_tab = pd.crosstab(ecoplus_df["Inflight wifi service"] <= 3, ecoplus_df["satisfaction_binary"])

# 카이제곱검정 시행
chi2, p_val, _, _ = chi2_contingency(cross_tab)

# 결과 출력
print("카이제곱검정 결과 p-value:", p_val)

# 유의수준 설정 (예: 0.05)
alpha = 0.05

# 유의수준과 비교하여 결과 해석
if p_val < alpha:
    print("귀무가설을 기각합니다. Dissatisfied 응답자는Inflight wifi service 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.")
else:
    print("귀무가설을 기각하지 않습니다. Dissatisfied 응답자는Inflight wifi service 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단할 수 없습니다.")

카이제곱검정 결과 p-value: 0.0
귀무가설을 기각합니다. Dissatisfied 응답자는Inflight wifi service 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.
```

✓ 귀무가설
기각

2) Online boarding

```
[ ] from scipy.stats import chi2_contingency

# Ease of Online booking와 satisfaction_binary 사이의 교차 테이블 생성
cross_tab = pd.crosstab(ecoplus_df["Online boarding"] <= 3, ecoplus_df["satisfaction_binary"])

# 카이제곱검정 시행
chi2, p_val, _, _ = chi2_contingency(cross_tab)

# 결과 출력
print("카이제곱검정 결과 p-value:", p_val)

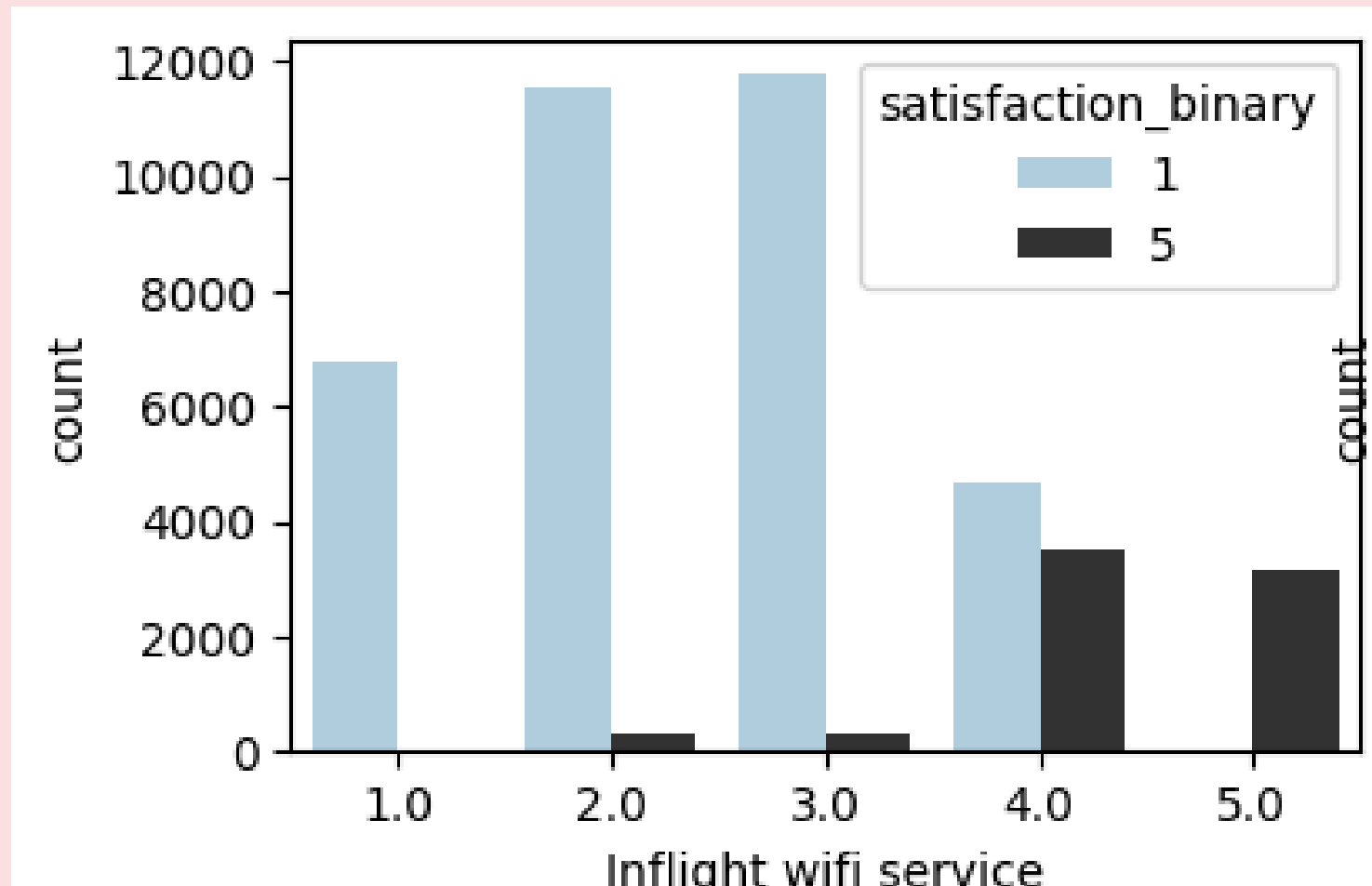
# 유의수준 설정 (예: 0.05)
alpha = 0.05

# 유의수준과 비교하여 결과 해석
if p_val < alpha:
    print("귀무가설을 기각합니다. Dissatisfied 응답자는Online boarding 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.")
else:
    print("귀무가설을 기각하지 않습니다. Dissatisfied 응답자는Online boarding 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단할 수 없습니다.")

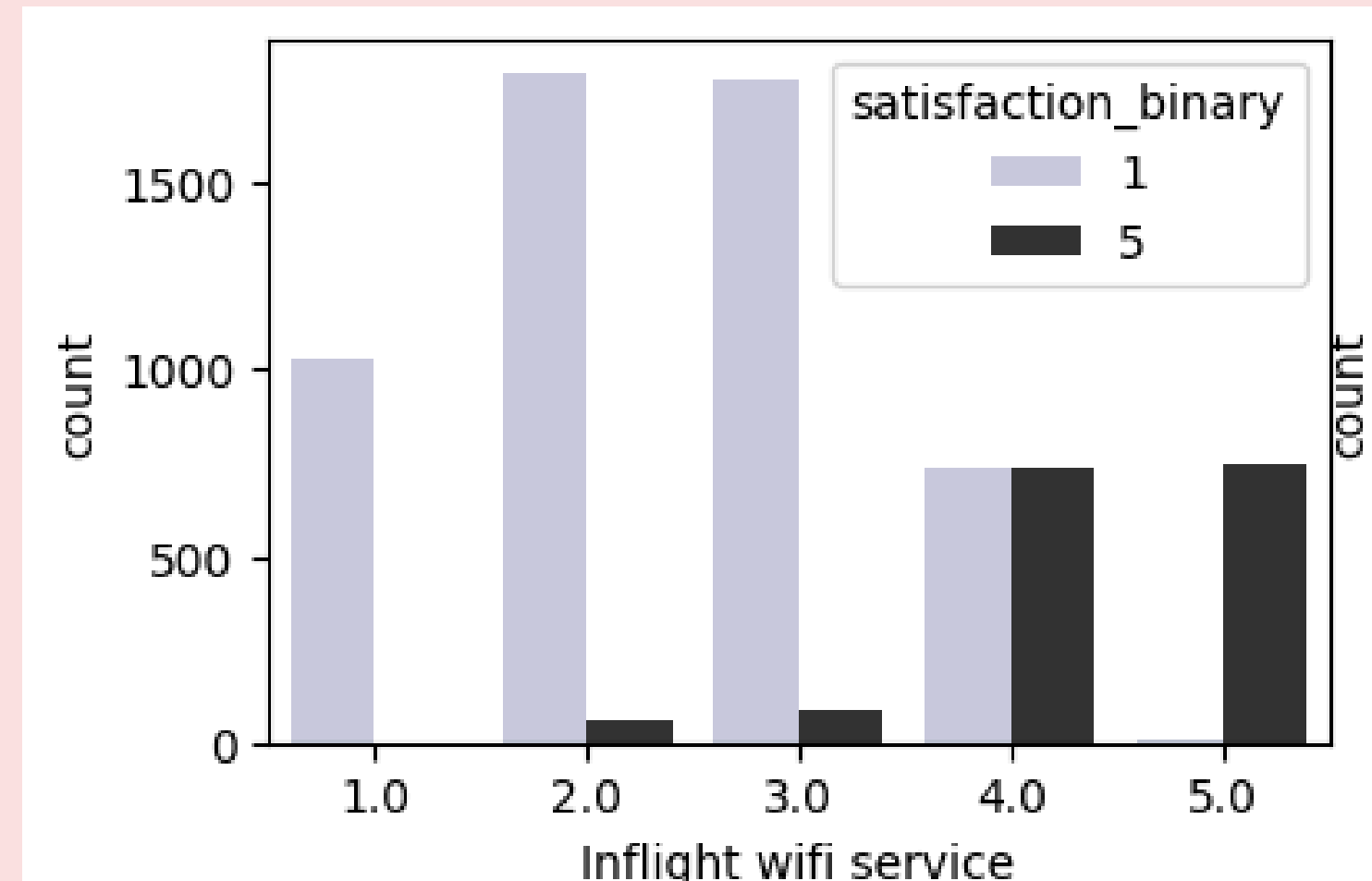
카이제곱검정 결과 p-value: 0.0
귀무가설을 기각합니다. Dissatisfied 응답자는Online boarding 만족도에서 낮은 점수(1-3)를 주었을 것으로 판단됩니다.
```

✓ 귀무가설
기각

6. 해결방안 제시



Economy

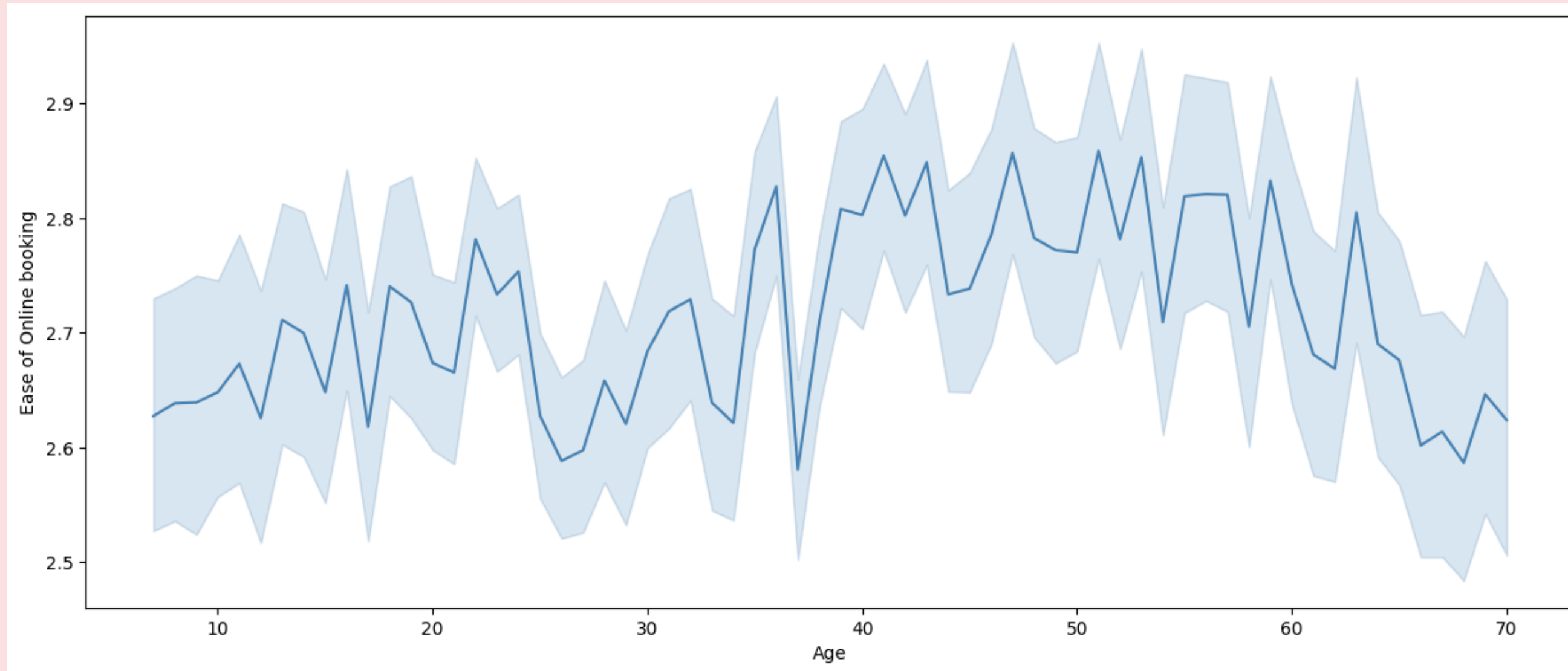


Economy Plus



Eco, Eco Plus Class에서 보다 더 나은
WIFI 서비스를 제공해야 이들의 만족도를 높일 수 있음

6. 해결방안 제시



50대부터 Ease of Online booking의 점수는 낮아지므로,
고령자들에게 Online booking에 대한 자세한 안내를 제공한다면
항공사에 대한 만족도를 높일 수 있을 것이라 기대됨

추가 : 모델링 결과

만족도를 예측하는 모델 생성

```
[ ] from sklearn.metrics import f1_score

for n,model in enumerate(models_predictions) :

    score=f1_score(y_test,model)

    print(f'F1 score for {label[n]} is {score:.3f}')
```

```
F1 score for XGB Classifier is 0.958
F1 score for logistic regression is 0.849
F1 score for Gradient Boosting Classifier is 0.933
F1 score for svc is 0.939
F1 score for CatBoostClassifier is 0.959
F1 score for KNN is 0.920
```

CatBoostClassifier의
F1 Score가 가장 높음.

(F1 Score는 0.0 ~ 1.0 사이의
값을 가지며 높을수록 좋다)



Q/A

