

# Data Analysis Storytelling :

## 고객 이탈 예방을 위한 이탈 고객 분석 및 대안 전략 수립

---

Workers : 5조 [ 김수현, 나한울, 정혜원, 한대희, 황유진 ]

- Data Analysis : 팀원 전체
- Edit Analysis Content : 팀원 전체
- Edit Article : 김수현, 나한울, 한대희
- Preparation of presentation : 정혜원, 황유진
- Translate : 정혜원, 황유진

Period for A to Z : 2 days (2024-01-15 ~ 2024-01-16)

# Contents



**FIRST  
STEP**

**SECOND  
STEP**

**THIRD  
STEP**

# Contents

**FIRST STEP : 1차적 자료 탐색**

**SECOND STEP :**

1차 분석 결과 토대로 방향 제시

**THIRD STEP :**

50대 고객 타겟으로 심층 분석

# Contents

**FIRST STEP : 1차적 자료 탐색**

**SECOND STEP :**

1차 분석 결과 토대로 방향 제시

**THIRD STEP :**

50대 고객 타겟으로 심층 분석

# Contents

**FIRST STEP : 1차적 자료 탐색**

**SECOND STEP :**

1차 분석 결과 토대로 방향 제시

**THIRD STEP :**

50대 고객 타겟으로 심층 분석



# Contents

**FIRST STEP : 1차적 자료 탐색**

**SECOND STEP :**

1차 분석 결과 토대로 방향 제시

**THIRD STEP :**

50대 고객 타겟으로 심층 분석



# ZERO STEP

# ZERO STEP

## 데이터 분석 소개

Period for Zero : 10 minutes brain storming ( 2024-01-16 )



문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

# WHAT IS PROBLEM?

# WHAT IS PROBLEM?

**: LOSE A CUSTOMER**

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

A 은행



문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



문제  
도입

목표  
설정

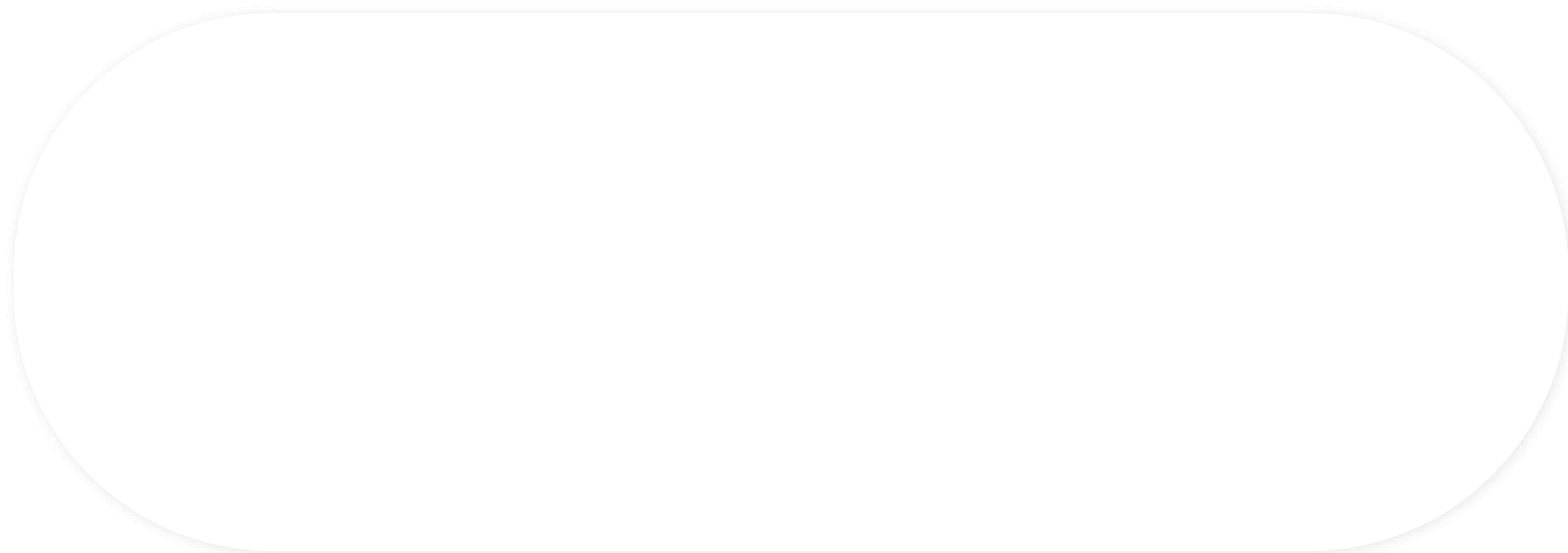
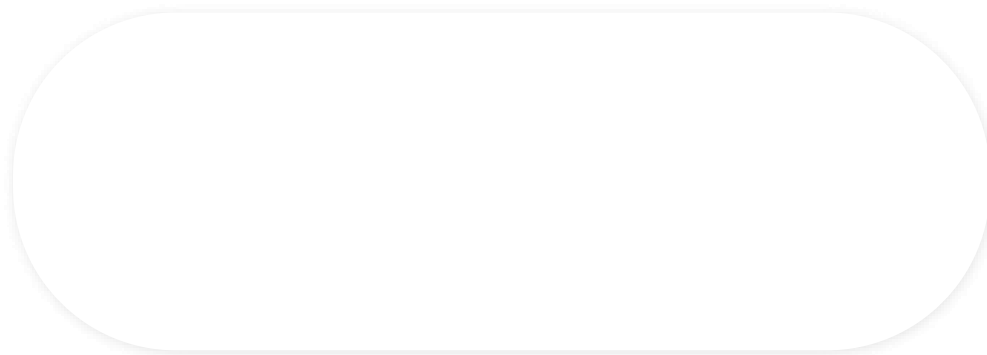
컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



고객 이탈률 증가의 원인은?

고객 이탈률 증가의 원인은?

이탈  
최소화

고객 이탈률 증가의 원인은?

이탈  
최소화

유지율  
향상



고객 이탈률 증가의 원인은?

이탈  
최소화

유지율  
향상

신규 회원  
유치

고객 이탈률 증가의 원인은?

이탈  
최소화

유지율  
향상

신규 회원  
유치

기업 가치  
상승

# Contents

**FIRST  
STEP**

SECOND  
STEP

THIRD  
STEP

# FIRST STEP

## 1차적 자료 탐색

Period for First : a quarter of a day ( 2024-01-15 )

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



사용 라이브러리



**numpy**  
**1.26.2**



**matplotlib**  
**3.80**



**seaborn**  
**0.12.2**



**scipy**  
**1.11.4**



**sklearn**  
**1.3.2**



**statsmodels**  
**0.14.0**

**TOTAL DATA**  
**165,034**

Credit Score: 고객의 신용점수

Geography: 고객이 거주하는 국가

Age: 고객의 나이

⋮

Tenure: 고객이 은행을 이용한 연수

Balance: 고객의 계좌 잔액

Exited: 이탈된 고객 여부

**TOTAL DATA**  
**165,034**  
  
**with**  
**13**  
**COLUMNS**

Credit Score: 고객의 신용점수  
Geography: 고객이 거주하는 국가  
Age: 고객의 나이  
.  
.  
.  
Tenure: 고객이 은행을 이용한 연수  
Balance: 고객의 계좌 잔액  
Exited: 이탈된 고객 여부

## 01 전체 데이터의 기술 통계량

- 신용카드 보유 여부

(1: 보유 중인 고객, 0: 보유하지 않은 고객)

- 활성화된 고객 여부

(1: 활성화된 고객, 0: 활성화되지 않은 고객)

- 이탈 여부

(1: 이탈 고객, 0: 이탈하지 않은 고객)

```
import pandas as pd
file_path = '/content/train.csv'
df = pd.read_csv(file_path)
df.describe()
```

	mean	std	min	25%	50%	75%	max
신용점수	656.4544	80.10334	350	597	659	710	850
나이	38.12589	8.867205	18	32	37	42	92
거래 기간	5.020353	2.806159	0	3	5	7	10
계좌 잔액	55478.09	62817.66	0	0	0	119939.5	250898.1
이용 제품 수	1.554455	0.547154	1	1	2	2	4
카드 보유 여부	0.753954	0.430707	0	1	1	1	1
활성 회원 여부	0.49777	0.499997	0	0	0	1	1
예상 연봉	112574.8	50292.87	11.58	74637.57	117948	155152.5	199992.5
이탈 여부	0.211599	0.408443	0	0	0	0	1

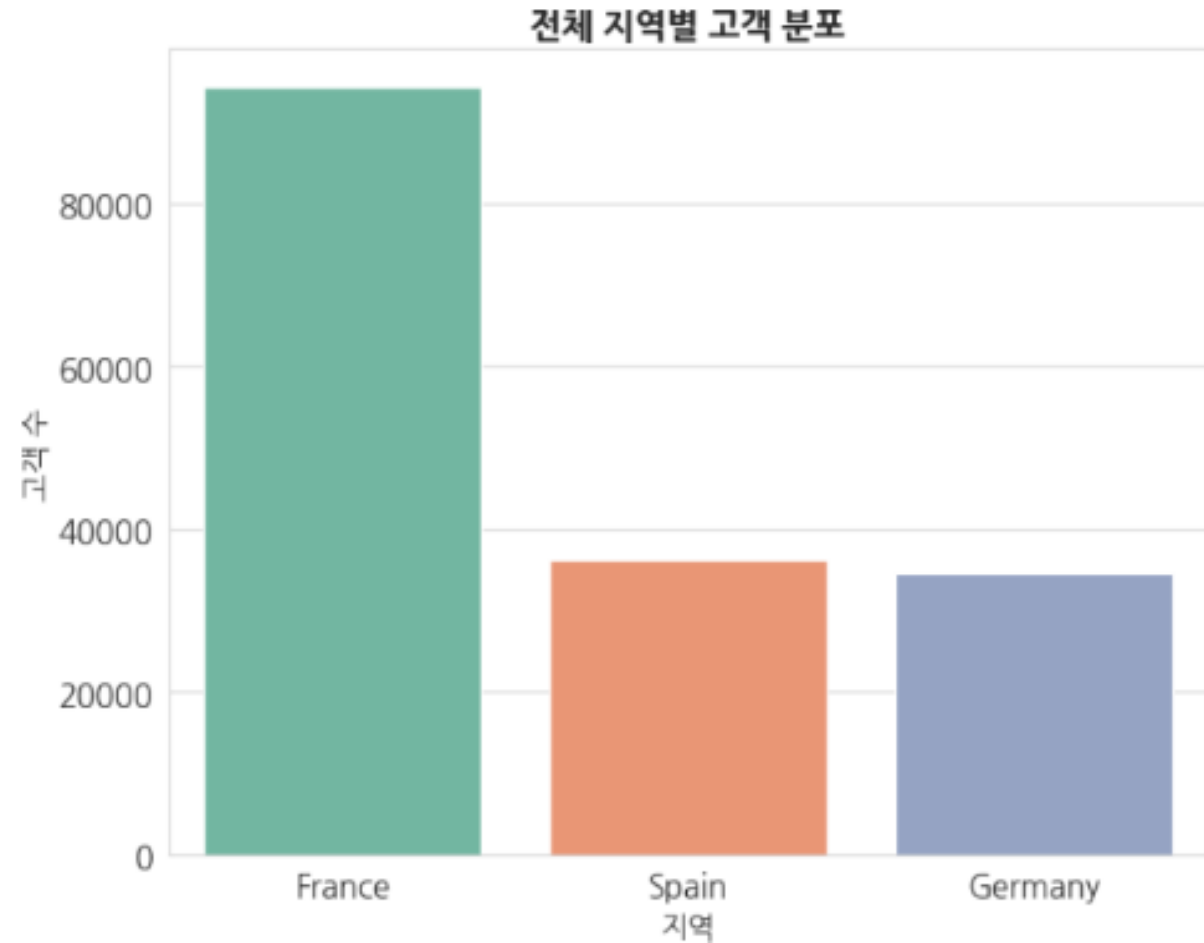


## 02 지역별 고객 분석

GERMANY < SPAIN <<< FRANCE

```
# 서브플롯 생성
fig, axs = plt.subplots(1, 2, figsize=(15, 6))

# 전체 지역별 고객 분포
sns.countplot(x='Geography', data=df, palette='Set2', ax=axs[0])
axs[0].set_title('전체 지역별 고객 분포', fontsize=15, fontweight='bold')
axs[0].set_xlabel('지역', fontsize=13)
axs[0].set_ylabel('고객 수', fontsize=13)
```

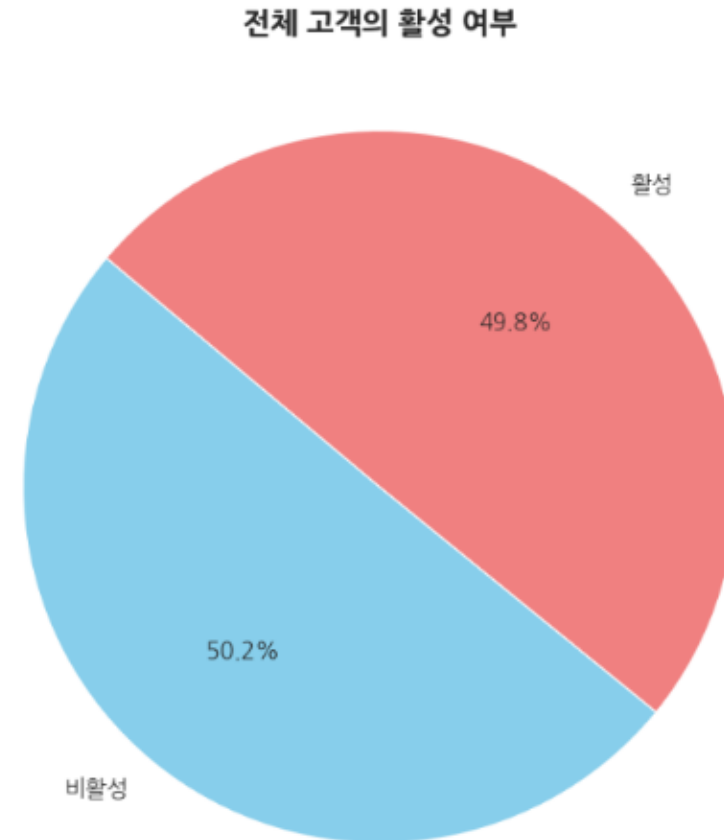


### 03 전체 고객 활성화 여부

#### 과반수의 회원 비활성화 상태

```
plt.figure(figsize=(16, 8))
plt.subplot(1, 2, 1)
plt.pie(active_counts, labels=['비활성', '활성'],
        autopct='%1.1f%%', colors=['skyblue', 'lightcoral'],
        startangle=140)
plt.title('전체 고객의 활성화 여부', fontsize=18, fontweight='bold')

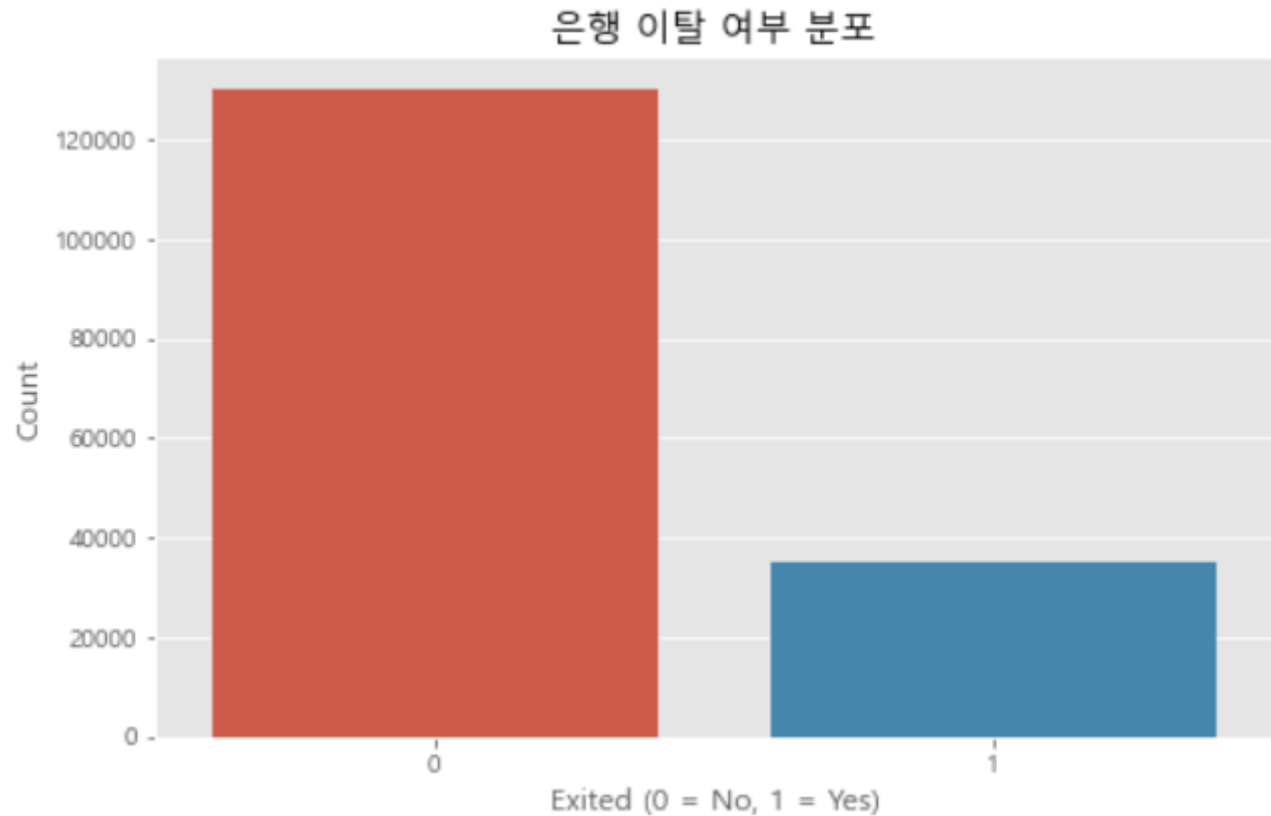
# 데이터 객체 정의 확인
active_counts = data['IsActiveMember'].value_counts()
```



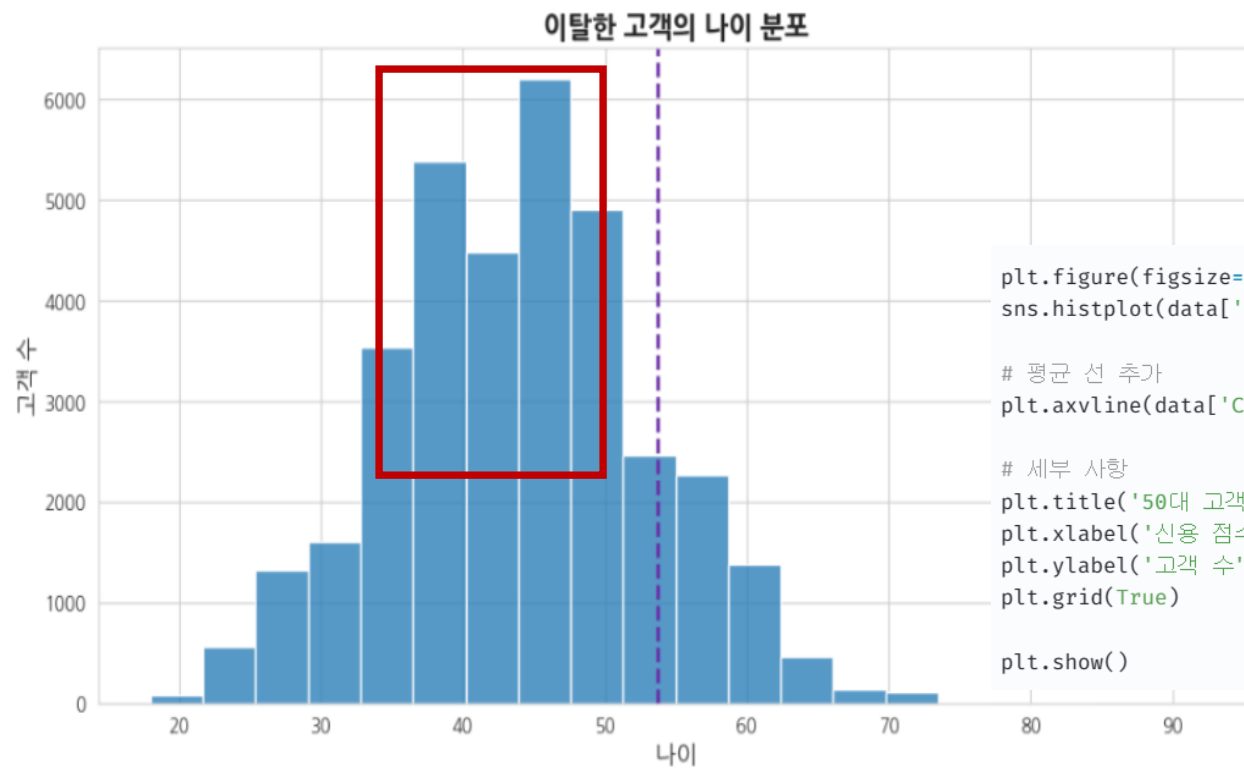
## 04 전체 고객 이탈 및 비이탈

이탈 수 < 비이탈 수 **4배** 많음

```
import pandas as pd
file_path = '/content/train.csv'
data = pd.read_csv(file_path)
exited_0 = data[data['Exited'] == 0]
exited_1 = data[data['Exited'] == 1]
statistics_exited_0 = exited_0.describe()
statistics_exited_1 = exited_1.describe()
print("Exited=0 그룹:")
print(statistics_exited_0)
print("\nExited=1 그룹:")
print(statistics_exited_1)
```



## 05 이탈 고객 연령층 분석



30대 중반에서 40대 중반 **최다**

```
plt.figure(figsize=(12, 6))
sns.histplot(data['CreditScore'], bins=30, color='#C8A4E2', edgecolor='#B477DD')

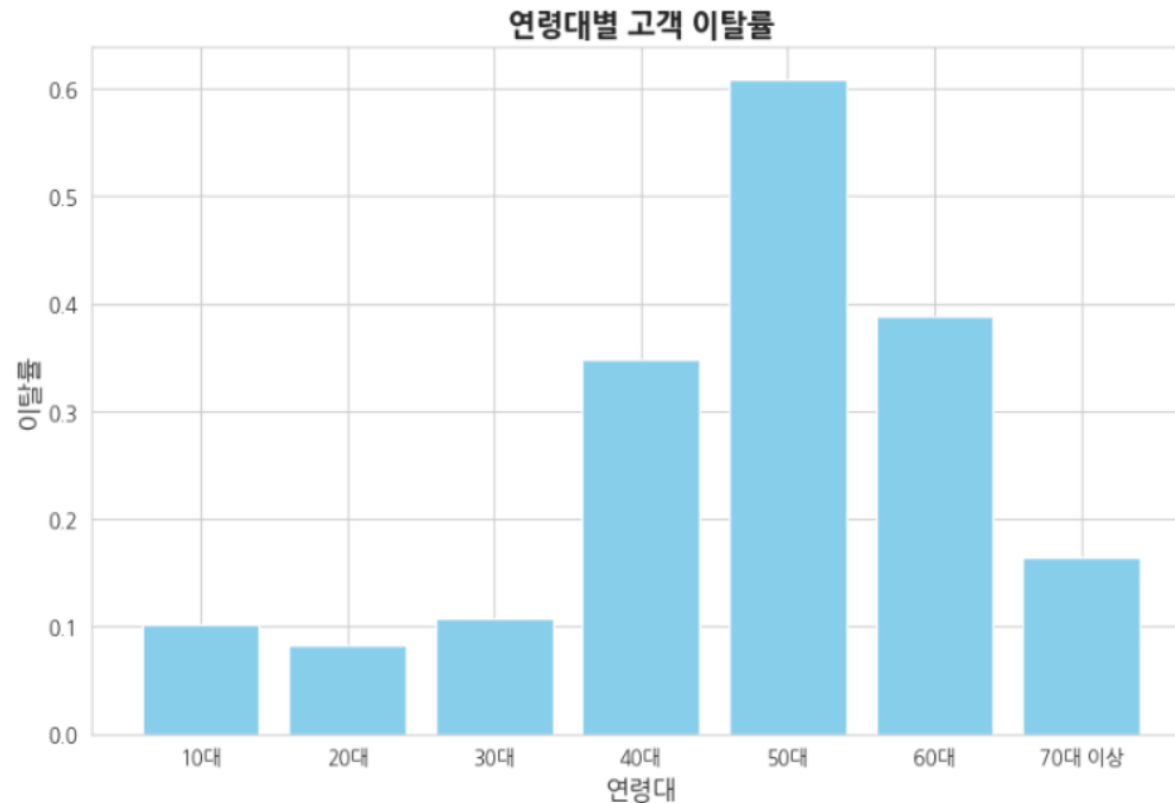
# 평균 선 추가
plt.axvline(data['CreditScore'].mean(), color='#2989DF', linestyle='dashed', linewidth=2)

# 세부 사항
plt.title('50대 고객의 신용 점수 분포', fontsize=15, fontweight='bold')
plt.xlabel('신용 점수', fontsize=13)
plt.ylabel('고객 수', fontsize=13)
plt.grid(True)

plt.show()
```

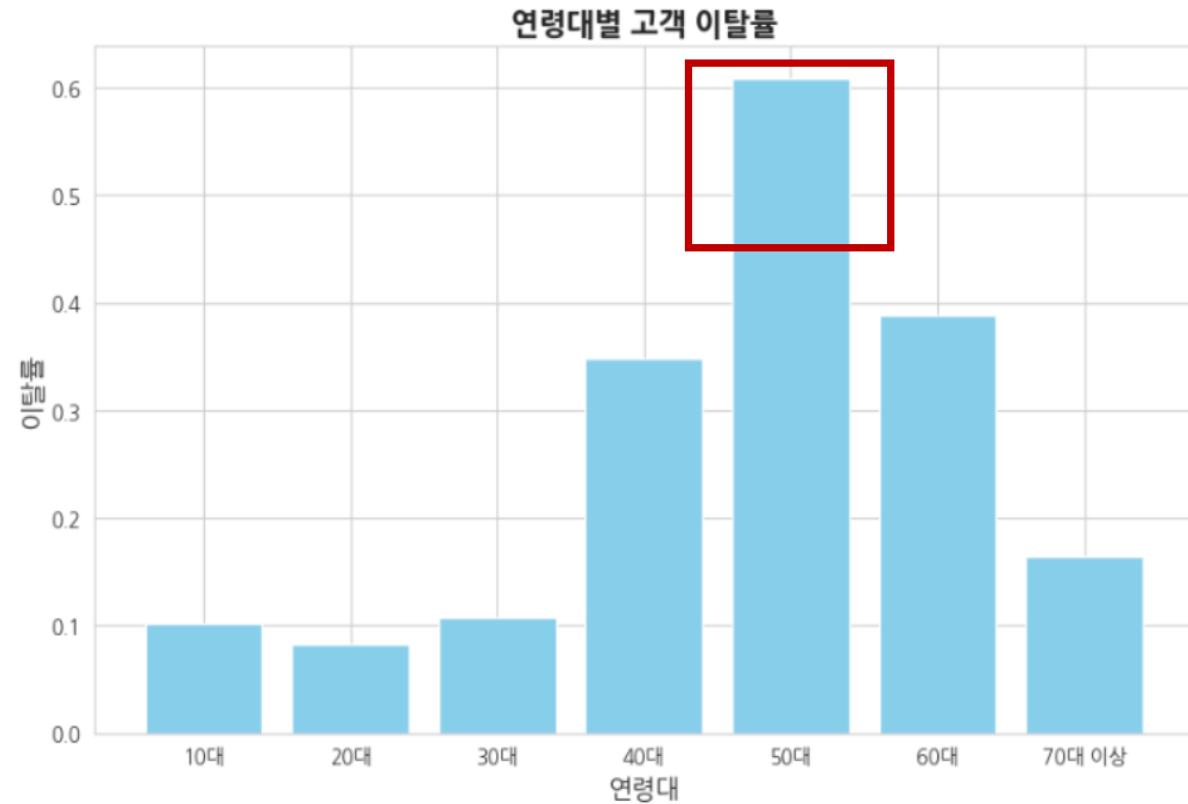
## 06 전체 고객 활성화 여부

```
def create_age_group(age):  
    if age < 20:  
        return '10대'  
    elif age < 30:  
        return '20대'  
    elif age < 40:  
        return '30대'  
    elif age < 50:  
        return '40대'  
    elif age < 60:  
        return '50대'  
    elif age < 70:  
        return '60대'  
    else:  
        return '70대 이상'  
  
df['AgeGroup'] = df['Age'].apply(create_age_group)  
# 연령대별 이탈률  
AgeGroup = ['10대', '20대', '30대', '40대', '50대', '60대', '70대 이상']  
AgeExitRates = []  
  
for i in AgeGroup:  
    AgeExited = df.loc[(df['AgeGroup'] == i) & (df['Exited'] == 1), :].shape[0]  
    AgeData = df.loc[df['AgeGroup'] == i].shape[0]  
    exit_rate = round(AgeExited / AgeData, 3) if AgeData > 0 else 0  
    AgeExitRates.append(exit_rate)
```



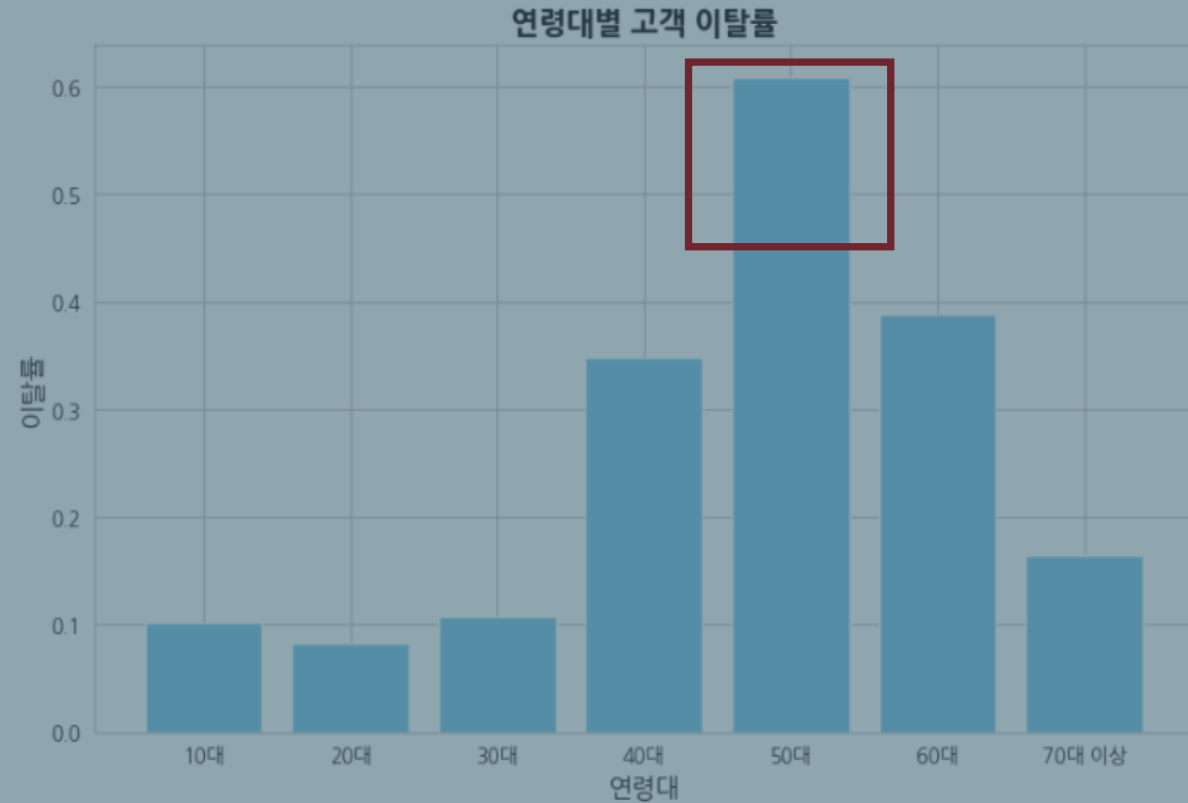
## 02 전체 고객 활성화 여부

BUT 50대 고객 이탈률 **최다**



## 02 전체 고객 활성화 여부

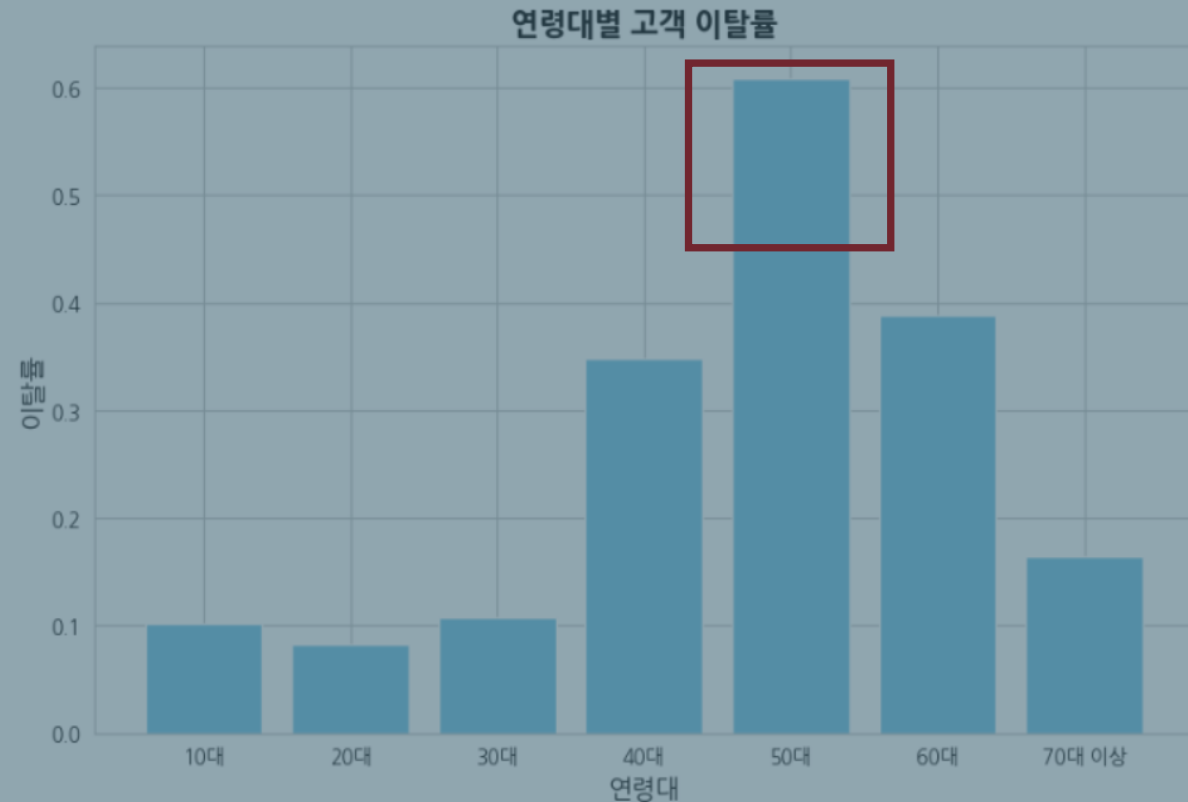
BUT 50대 고객 이탈률 **최다**



## 02 전체 고객 활성화 여부

# IT'S OURS!

BUT **50대** 고객 이탈률 **최다**





# Contents

FIRST  
STEP

**SECOND  
STEP**

THIRD  
STEP

# SECOND STEP

## 1차 분석 결과 토대로 방향 제시

Period for Zero : 4 hours ( 2024-01-16 )

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



IMPORTANT ISSUE

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

IMPORTANT ISSUE  
50 대 기존 고객 대거 이탈

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

2차 EDA 분석

IMPORTANT ISSUE  
50 대 기존 고객 대거 이탈

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

2차 EDA 분석

IMPORTANT ISSUE  
50 대 기존 고객 대거 이탈

타겟 심층 분석

## 01 50대 이탈자의 기술 통계량

- 총 데이터 : 4,387 개
- 결측치 없음
- 계좌 잔액이 0인 고객은 제외함

```
import pandas as pd
file_path = '/content/train.csv'
df = pd.read_csv(file_path)
data = df[(df['Balance'] != 0) & (df['Age'] >= 50) & (df['Age'] < 60)]
data.describe()
```

	Mean	Std	Min	25%	50%	75%	Max
신용점수	654.06	80.2	350	594	652	710	850
거래 기간	4.86	2.83	0	2	5	7	10
계좌 잔액	122428.91	22906.84	40685.92	108127.07	122314.5	136526.03	238387.56
이용 제품 수	1.31	0.68	1	1	1	1	4
카드 보유 여부	0.74	0.44	0	0	1	1	1
활성 회원 여부	0.31	0.46	0	0	0	1	1
예상 연봉	115362.19	50084.51	90.07	76569.55	121239.65	159793.16	199775.67

## 02 전체 고객과 50대 이탈자의 데이터 비교

50대 고객층은

계좌 잔액이 비교적 **많음**

	신용 점수	나이	거래 기간	계좌 잔액	이용 제품 수	카드 보유	활성 회원	예상 연봉	이탈 여부
평균 값	656.4 점	38세	5.02년	\$ 55,4 778.1	1.55개	75%	50%	\$112, 574.8	21.16 %
50대 이탈자 평균	652.4 점	54세	4.86년	\$72,3 26.4	1.31개	74%	31%	\$115, 075.9	



## 02 전체 고객과 50대 이탈자의 데이터 비교

50대 고객층은

활성회원 비율이 매우 낮은 편임

	신용 점수	나이	거래 기간	계좌 잔액	이용 제품 수	카드 보유	활성 회원	예상 연봉	이탈 여부
평균 값	656.4 점	38세	5.02년	\$ 55,4 778.1	1.55개	75%	50%	\$112, 574.8	21.16 %
50대 이탈자 평균	652.4 점	54세	4.86년	\$72,3 26.4	1.31개	74%	31%	\$115, 075.9	

## 03 지역별 50대 은행의 데이터 조사

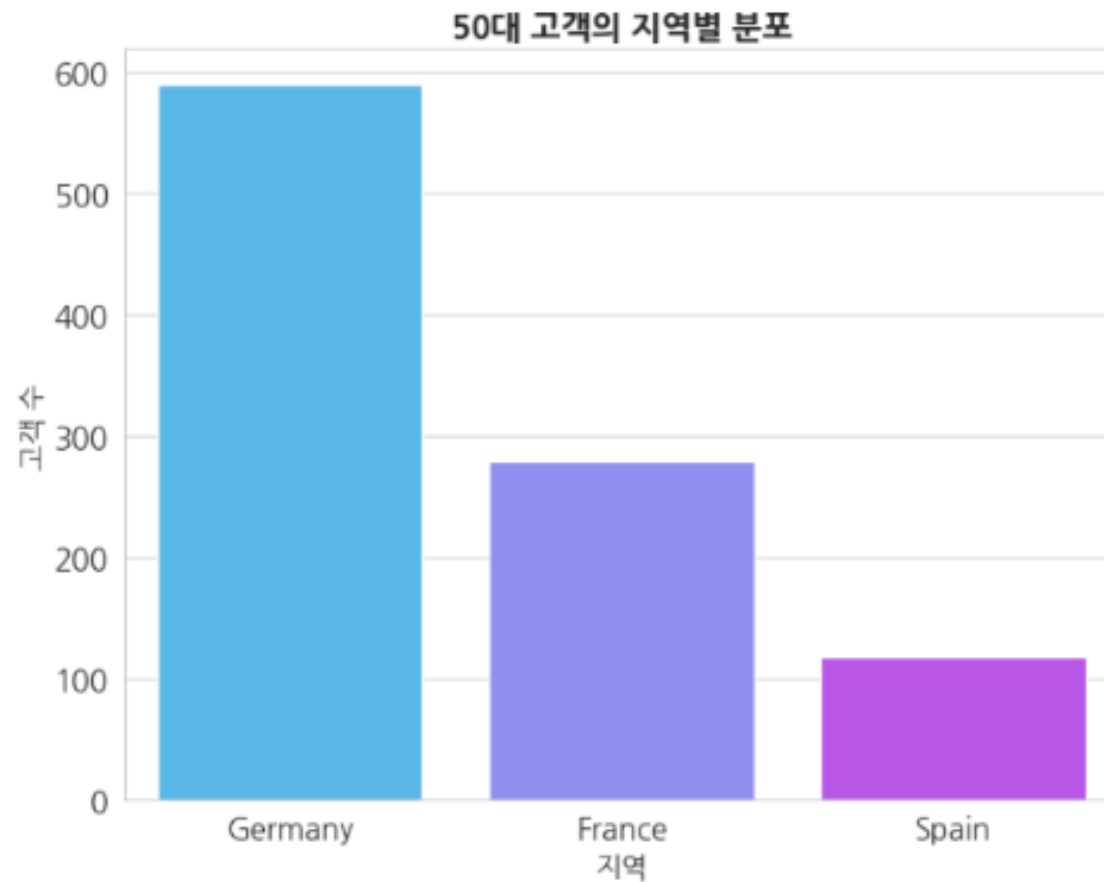
50대 고객층은

**독일** 거주 고객이 가장 많음

```
# 50대 고객의 지역별 분포
sns.countplot(x='Geography', data=data, palette='cool', ax=axes[1])
axes[1].set_title('50대 고객의 지역별 분포', fontsize=15, fontweight='bold')
axes[1].set_xlabel('지역', fontsize=13)
axes[1].set_ylabel('고객 수', fontsize=13)

# 그래프 간의 간격 조절
plt.tight_layout()

plt.show
```



## 04 50대 이탈리아의 거주 지역 데이터 분석

```
# 데이터 설정
labels = ['Exited', 'Stayed']
france_sizes = [48.6, 51.4]
germany_sizes = [75.9, 24.1]
spain_sizes = [51.5, 48.5]
colors = ['#ff9999', '#66b3ff']

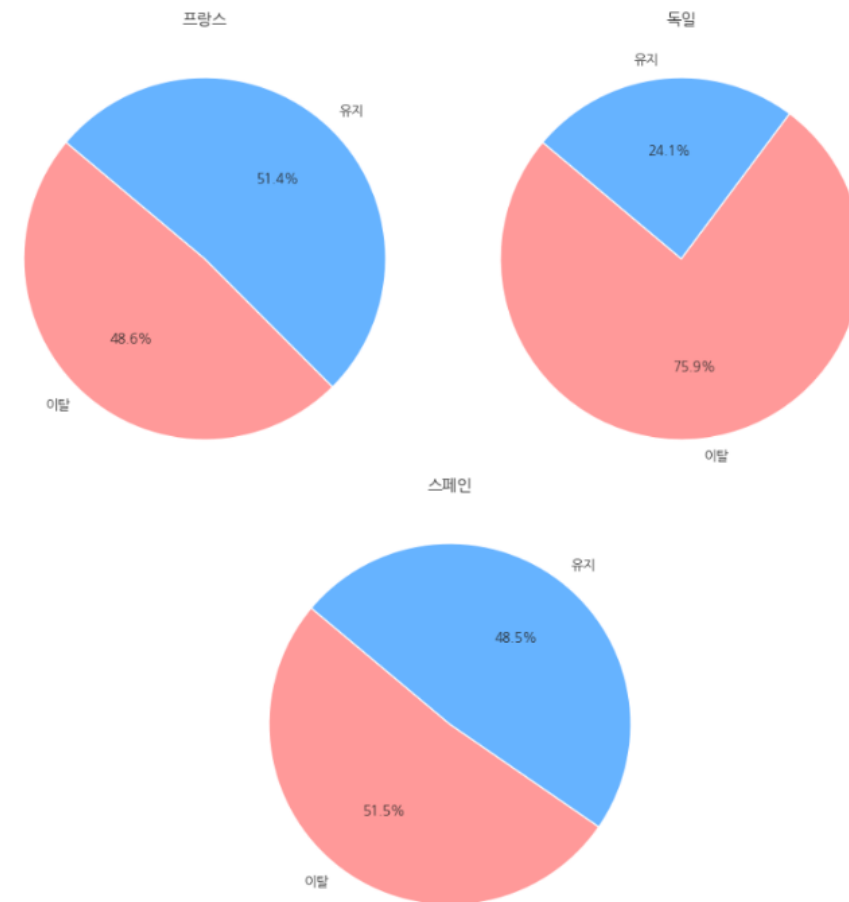
# 파이 차트 생성
fig, axs = plt.subplots(1, 3, figsize=(15, 5))

# 프랑스 파이
axs[0].pie(france_sizes, labels=['이탈', '유지'], colors=colors, autopct='%1.1f%%',
axs[0].set_title('프랑스')

# 독일 파이
axs[1].pie(germany_sizes, labels=['이탈', '유지'], colors=colors, autopct='%1.1f%%',
axs[1].set_title('독일')

# 스페인 파이
axs[2].pie(spain_sizes, labels=['이탈', '유지'], colors=colors, autopct='%1.1f%%',
axs[2].set_title('스페인')

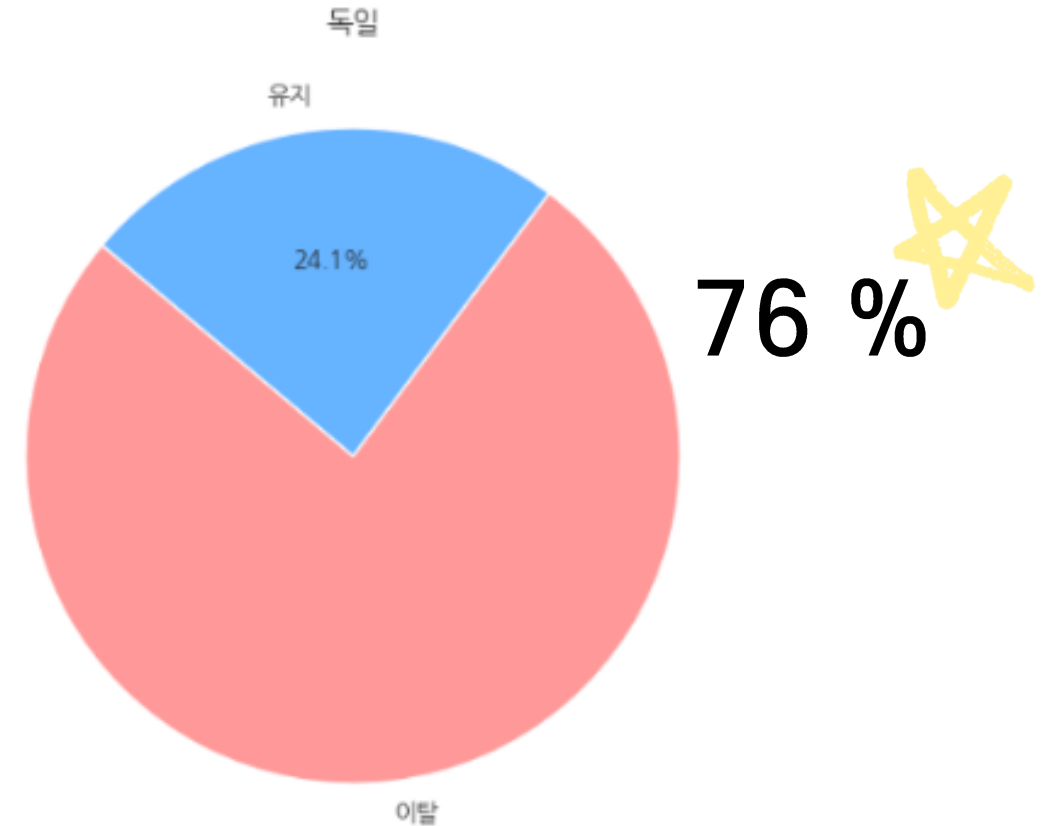
plt.tight_layout()
plt.show()
```



## 04 50대 이탈자의 거주 지역 데이터 분석

**독일**에 거주하고 있는

50대 은행 고객의 **76%** 가 이탈

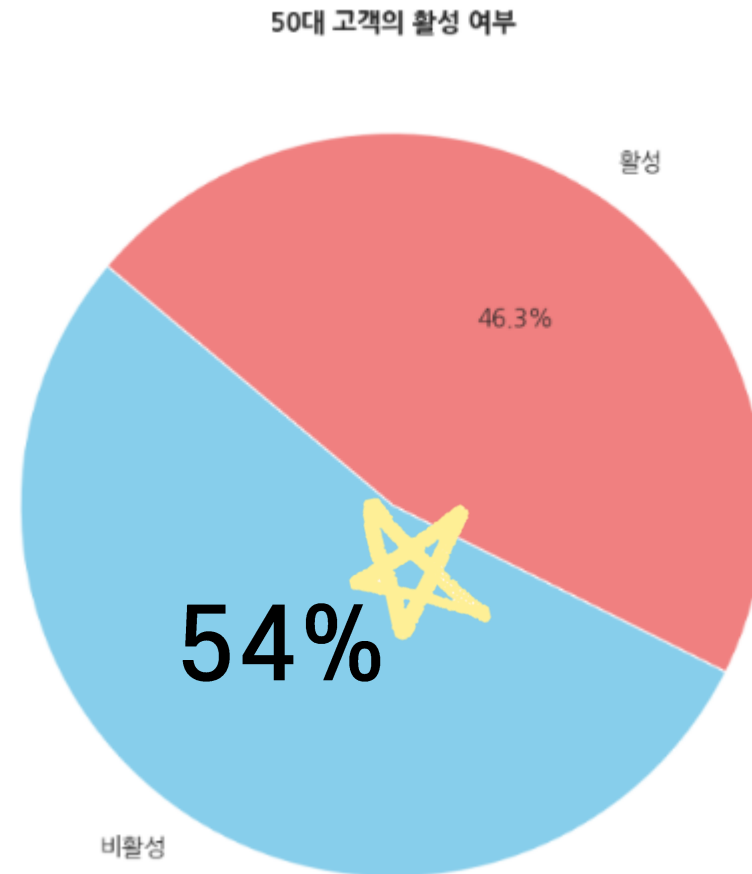


## 04 50대 전체 고객의 활성 여부 분석

50대 고객 중,

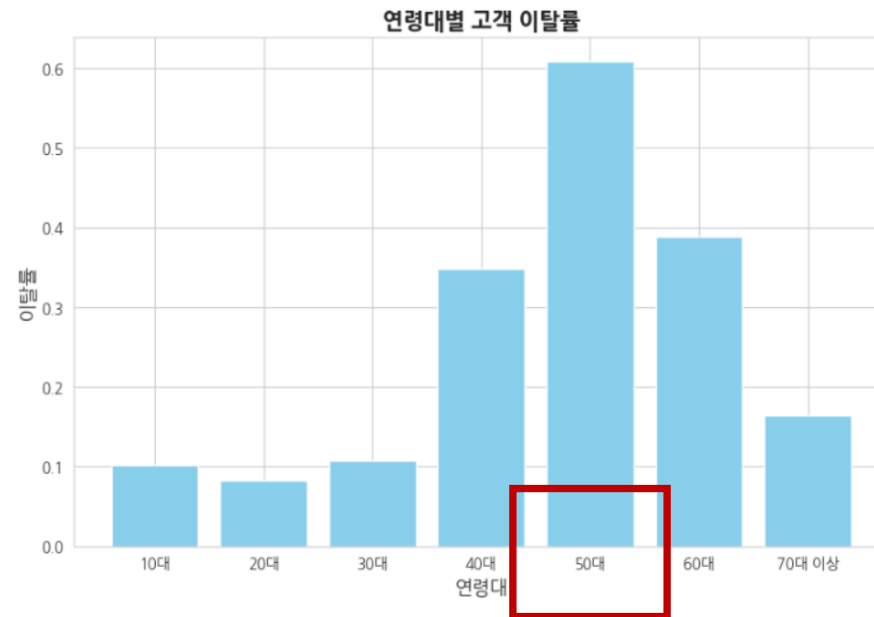
**비활성화 고객이 과반수를 넘음**

```
# 50대 고객의 활성 파이
plt.subplot(1, 2, 2)
plt.pie(active_counts, labels=['비활성', '활성'], autopct='%1.1f%%', colors=['skyblue', 'red'])
plt.title('50대 고객의 활성 여부', fontsize=15, fontweight='bold')
plt.tight_layout()
plt.show()
```



## Acquired Insight

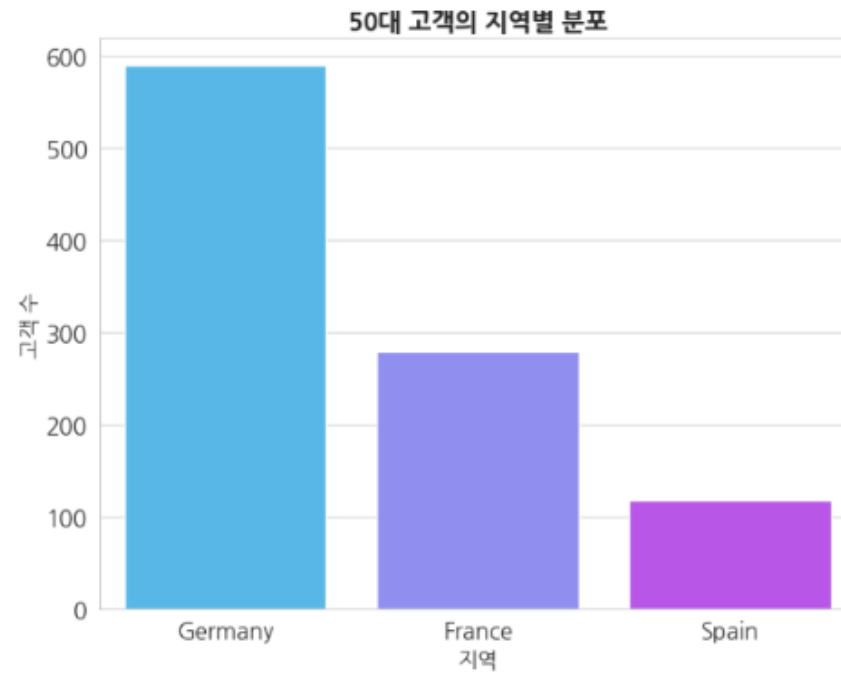
이탈률이 **가장 높은** 50대 고객



## Acquired Insight

이탈률이 가장 높은 50대 고객

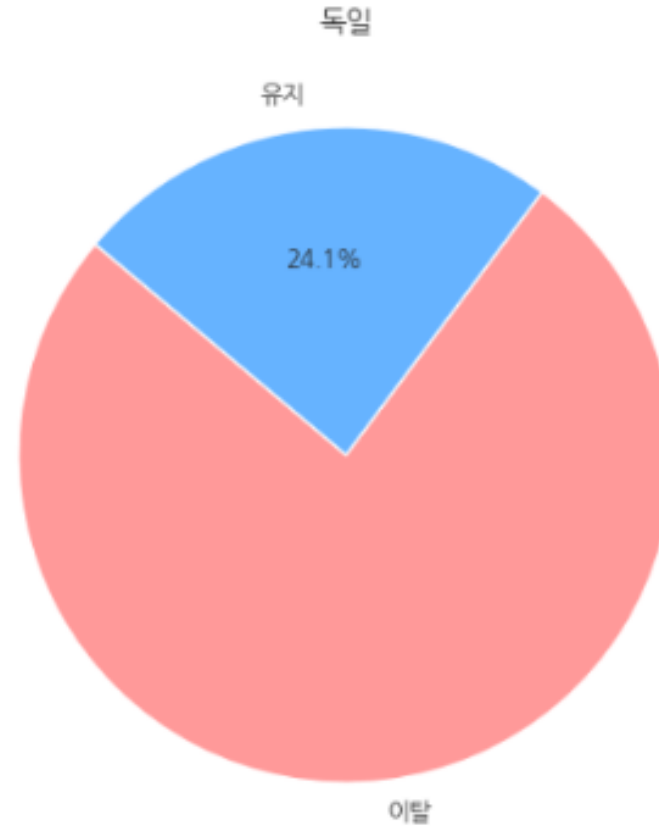
50대 고객이 가장 많은 독일 지역




## Acquired Insight

이탈률이 가장 높은 50대 고객

독일 지역 50대 **이탈 고객 : 80 %**



76 % 



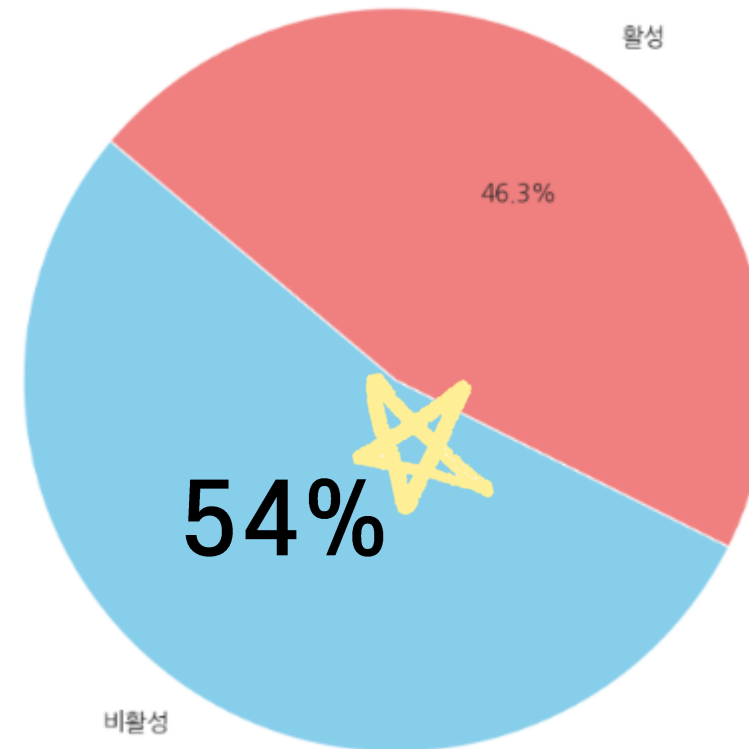
## Acquired Insight

이탈률이 가장 높은 50대 고객

50대 고객이 가장 많은 독일 지역

**과반수**가 넘는 비활성화 회원

50대 고객의 활성 여부



# Contents

FIRST  
STEP

SECOND  
STEP

**THIRD  
STEP**

# THIRD STEP

50대 고객 타겟으로 심층 분석

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

2차 EDA 분석

IMPORTANT ISSUE  
50 대 기존 고객 대거 이탈

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

2차 EDA 분석

IMPORTANT ISSUE  
50 대 기존 고객 대거 이탈

타겟 심층 분석



문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

교차분석  
& 회귀분석

문제  
도입

목표  
설정

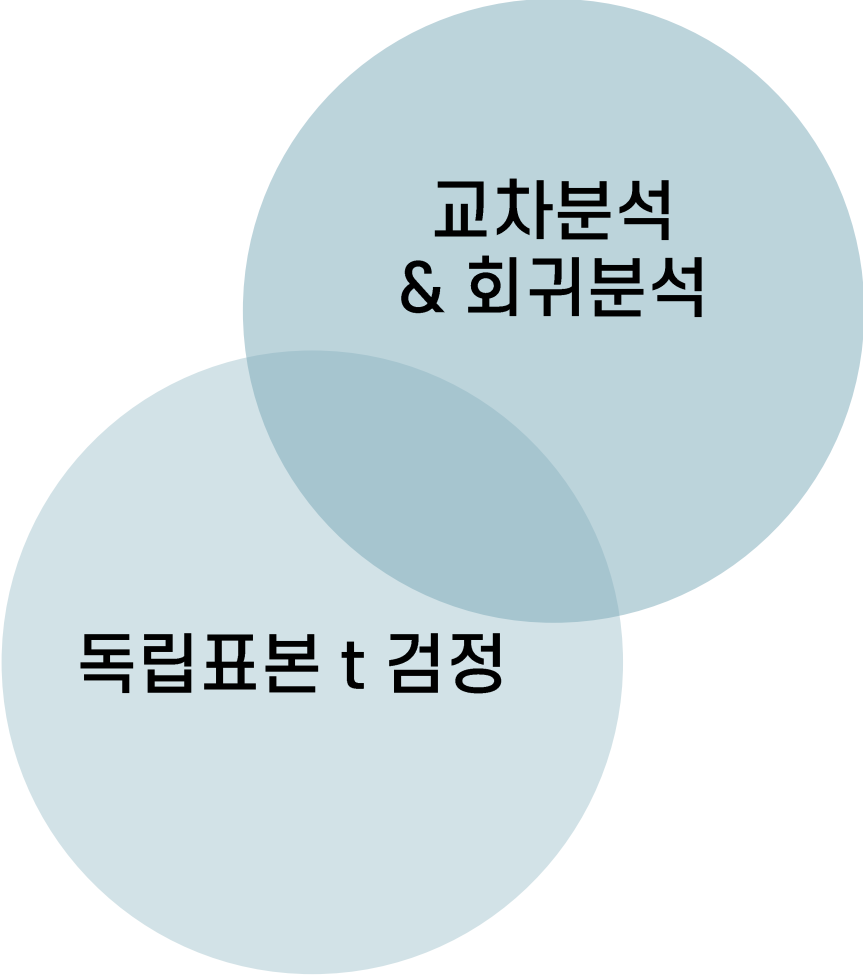
컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



교차분석  
& 회귀분석

독립표본 t 검정

문제  
도입

목표  
설정

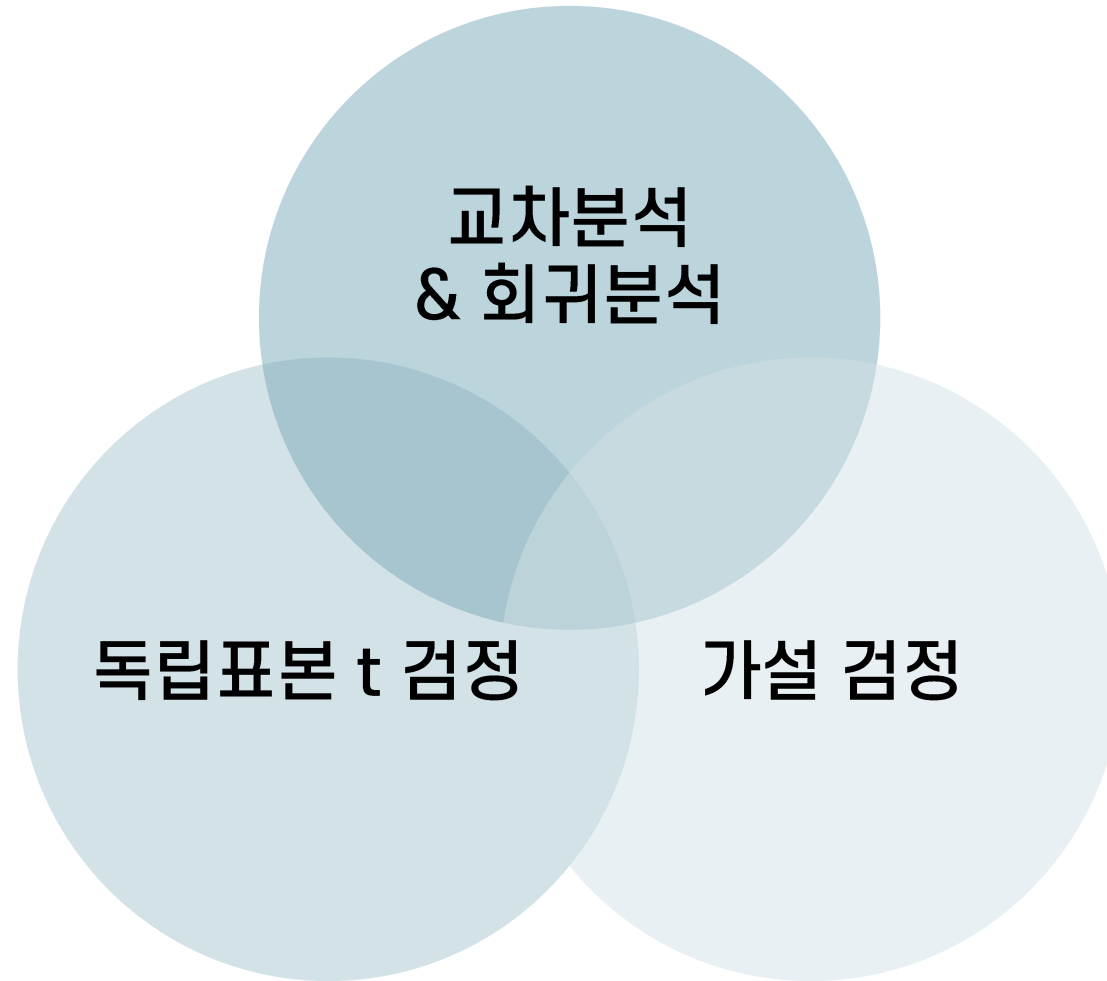
컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망





문제  
도입

목표  
설정

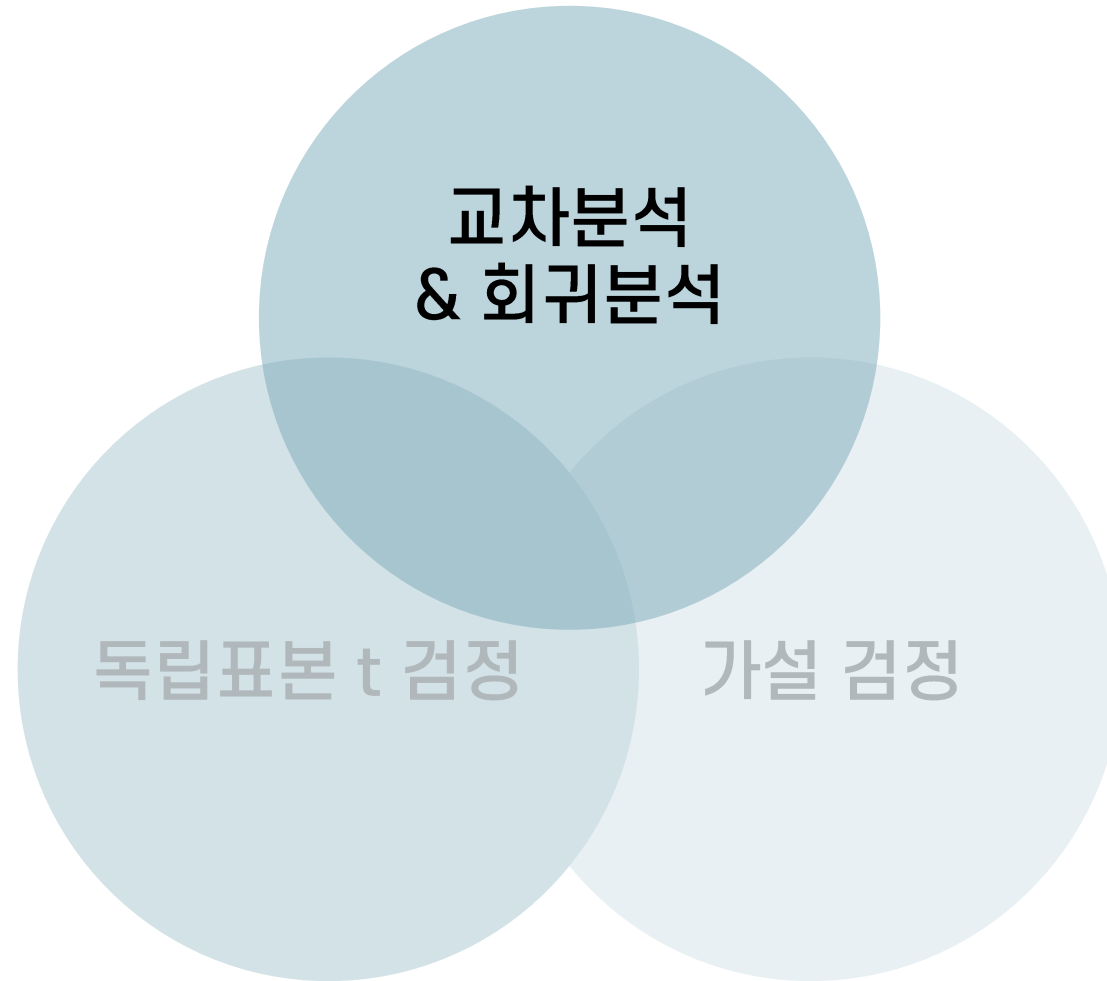
컨텍스트  
제시

교차분석  
& 회귀분석

인사이트  
제공

해결책  
제시

미래  
전망



## 교차분석 진행

종속변수 : Exited

독립변수

HasCrCard

Gender

BalLevel

IsActiveMember

## 교차분석 진행

종속변수 : Exited

가설 설정

H0: Exited 와 독립변수는 서로 독립이다.  
H1: Exited 와 독립변수는 서로 독립이 아니다

## 교차분석 진행

```
col = ['HasCrCard', 'IsActiveMember', 'Gender_Male', 'BalLevel_중간소득', 'BalLevel_고소득']
for i in col:
    cross_data = pd.crosstab(index = dy5_dm_sc['Exited'], columns = dy5_dm_sc[i],
                              margins=False)

    from scipy.stats import chi2_contingency
    chi2, p, dof, expected = chi2_contingency(cross_data)
    msg = 'Test Statistic: {}\nnp-value: {}\nDegree of Freedom: {}'.format(chi2, p, dof)
    print(f'<Exited and {i}>')
    print(msg)
    if p < 0.05:
        print('dependent')
    else:
        print('independent')
    print(expected)
    print('-----')
```

## 교차분석 진행

```
col = ['HasCrCard', 'IsActiveMember', 'Gender_Male', 'BalLevel_중간소득', 'BalLevel_고소득']
for i in col:
    cross_data = pd.crosstab(index = dy5_dm_sc['Exited'], columns = dy5_dm_sc[i],
                              margins=False)

    from scipy.stats import chi2_contingency
    chi2, p, dof, expected = chi2_contingency(cross_data)
    msg = 'Test Statistic: {}\nnp-value: {}\nDegree of Freedom: {}'.format(chi2, p, dof)
    print(f'Exited and {i}>')
    print(msg)
    if p < 0.05:
        print('dependent')
    else:
        print('independent')
    print(expected)
    print('-----')
```

결과

HasCrCard

$p \approx 0.061$   
H0 채택

Exited와 HasCrCard는 서로 독립이다.

결과

Test Statistic: 3.512031760205843

p-value: 0.06092469740333798

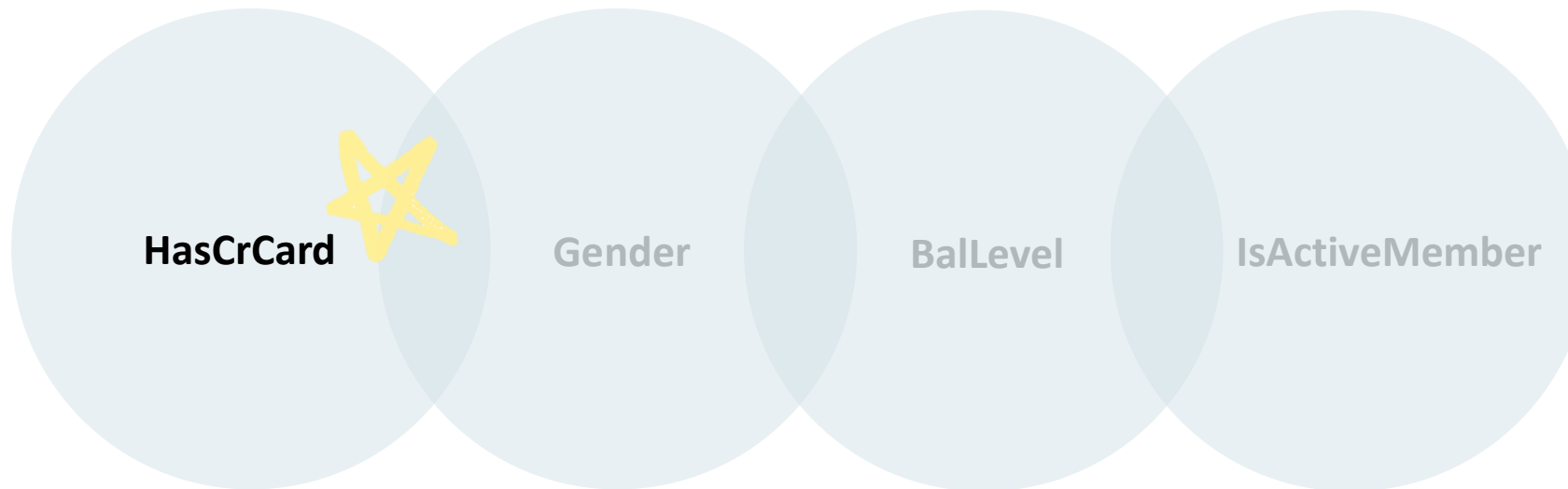
Degree of Freedom: 1

independent

[[ 648.71940431 1746.28059569][1188.28059569 3198.71940431]]

결과 : Exited와 HasCrCard만 서로 독립 관계를 가진다.

독립변수



문제  
도입

목표  
설정

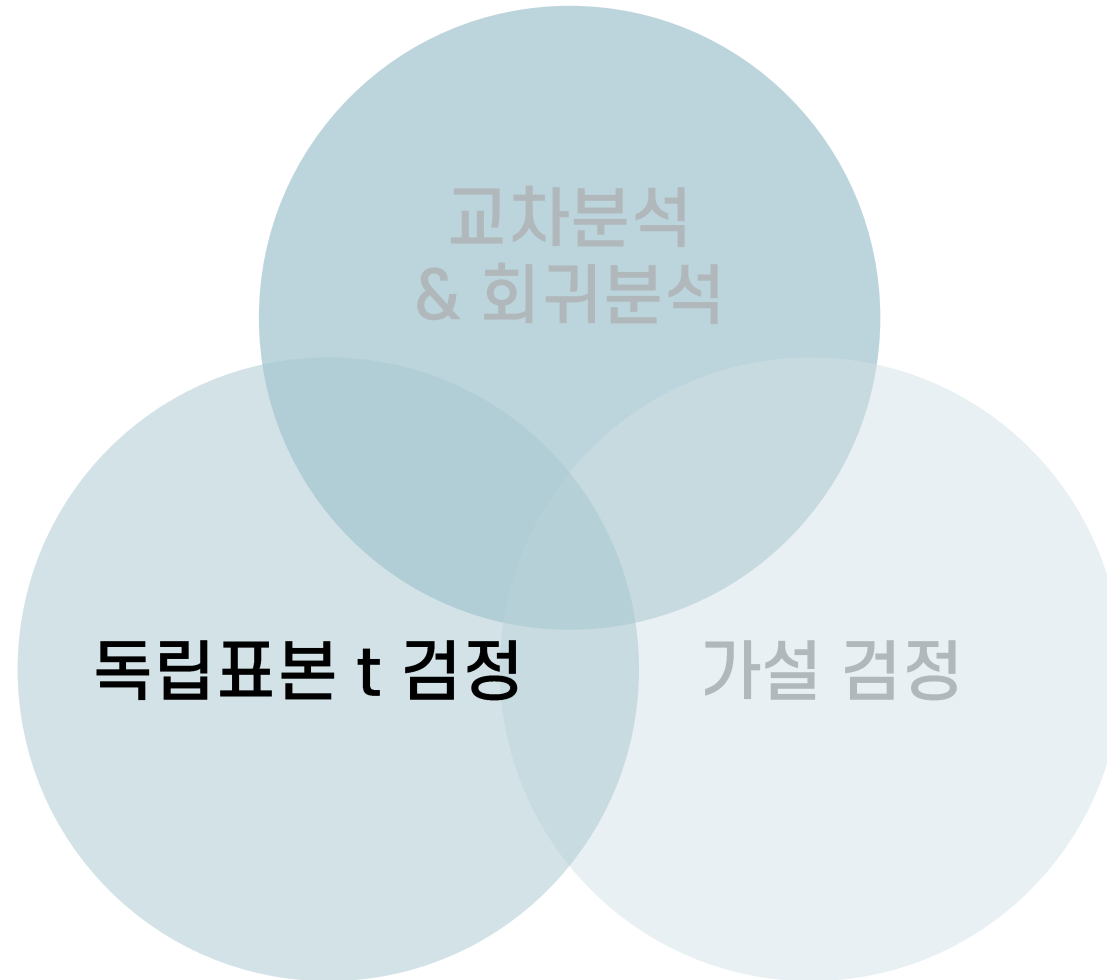
컨텍스트  
제시

독립표본  
t 검정

인사이트  
제공

해결책  
제시

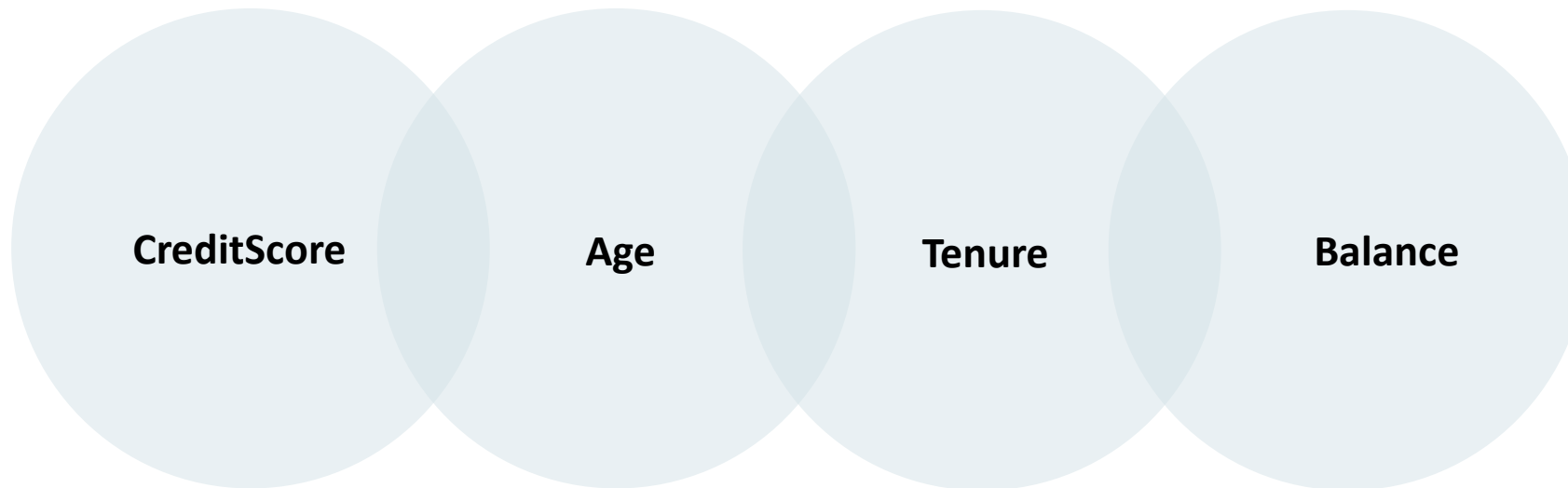
미래  
전망



## 독립표본 T 검정 ( 전체 대상 )

종속변수 : Exited

독립변수





## 독립표본 t 검정 ( 전체 대상 )

종속변수 : Exited

가설 설정

H0: Exited 그룹간의 종속변수에 대한 평균은 같다.  
H1: Exited 그룹간의 종속변수에 대한 평균은 다르다.

## 독립표본 t 검정 ( 전체 대상 )

```
nonzero=train_data[train_data['Balance']!=0].reset_index(drop=True)
df_0=nonzero.loc[nonzero['Exited']==0,['CreditScore','Age','Tenure','Balance']].reset_i
df_1=nonzero.loc[nonzero['Exited']==1,['CreditScore','Age','Tenure','Balance']].reset_i
```

### CreditScore

$p \approx 0$   
H0기각  
*Exited 그룹간의 CreditScore는 통계적으로 유의미한 차이가 존재함.*

#### 코드

```
t, p=stats.ttest_ind(df_0.CreditScore,df_1.CreditScore)
t, p
if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p,5))
```

### Age

$p \approx 0$   
H0기각  
*Exited 그룹간의 Age는 통계적으로 유의미한 차이가 존재함.*

#### 코드

```
t, p=stats.ttest_ind(df_0.Age,df_1.Age)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p,5))
```

### Tenure

$p \approx 0$   
H0기각  
*Exited 그룹간의 Tenure는 통계적으로 유의미한 차이가 존재함.*

#### 코드

```
t, p=stats.ttest_ind(df_0.Tenure,df_1.Tenure)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p,5))
```

## 독립표본 T 검정 ( 전체 대상 )

```
nonzero=train_data[train_data['Balance']!=0].reset_index(drop=True)
df_0=nonzero.loc[nonzero['Exited']==0,['CreditScore','Age','Tenure','Balance']].reset_i
df_1=nonzero.loc[nonzero['Exited']==1,['CreditScore','Age','Tenure','Balance']].reset_i
```

CreditScore

$p \approx 0$

H0기각

Exited 그룹간의 CreditScore는 통계적으로 유의미한 차이가 존재함.

Age

H0기각

Exited 그룹간의 Age는 통계적으로 유의미한 차이가 존재함.

Tenure

H0기각

Exited 그룹간의 Tenure는 통계적으로 유의미한 차이가 존재함.

## CreditScore, Age, Tenure에 대해서 통계적으로 유의미한 차이가 존재

코드

```
t, p=stats.ttest_ind(df_0.CreditScore,df_1.CreditScore)
t, p
if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p,5))
```

코드

```
t, p=stats.ttest_ind(df_0.Age,df_1.Age)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p,5))
```

코드

```
t, p=stats.ttest_ind(df_0.Tenure,df_1.Tenure)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p,5))
```

## 독립표본 T 검정 ( 50대 고객 대상 )

종속변수 : Exited

독립변수

CreditScore

Balance

Tenure

## 독립표본 t 검정 ( 50대 고객 대상 )

종속변수 : Exited

가설 설정

H0: Exited 그룹간의 독립변수에 대한 평균은 같다.  
H1: Exited 그룹간의 독립변수에 대한 평균은 다르다.

## 독립표본 T 검정 ( 50대 고객 대상 )

```
def categorize_age(age):  
    if age<20:  
        return '10s'  
    elif age<30:  
        return '20s'  
    elif age<40:  
        return '30s'  
    elif age<50:  
        return '40s'  
    elif age<60:  
        return '50s'  
    elif age<70:  
        return '60s'  
    else:  
        return '70+'  
  
train_data['Age_range']=train_data['Age'].apply(categorize_age)  
nonzero=train_data[train_data['Balance']!=0].reset_index(drop=True)  
df_0_50s=nonzero.loc[(nonzero['Exited']==0)&(nonzero['Age_range']=='50s'),['CreditScor  
df_1_50s=nonzero.loc[(nonzero['Exited']==1)&(nonzero['Age_range']=='50s'),['CreditScor
```

## 독립표본 t 검정 ( 50대 고객 대상 )

### CreditScore

$p \approx 0.406$

H0채택

Exited 그룹간의 CreditScore는 통계적으로 유의미한 차이가 존재하지않음.

### Tenure

$p \approx 0.066$

H0채택

Exited 그룹간의 Tenure는 통계적으로 유의미한 차이가 존재하지않음.

### Balance

$p \approx 0.305$

H0채택

Exited 그룹간의 Balance는 통계적으로 유의미한 차이가 존재하지않음.

#### 코드

```
t, p = stats.ttest_ind(df_0_50s.CreditScore, df_1_50s.CreditScore)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p, 5))
```

#### 코드

```
t, p = stats.ttest_ind(df_0_50s.Tenure, df_1_50s.Tenure)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p, 5))
```

#### 코드

```
t, p = stats.ttest_ind(df_0_50s.Balance, df_1_50s.Balance)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p, 5))
```

## 독립표본 t 검정 ( 50대 고객 대상 )

### CreditScore

p≈0.406

H0채택

Exited 그룹간의 CreditScore는 통계적으로 유의미한 차이가 존재하지않음.

### Tenure

p≈0.066

H0채택

Exited 그룹간의 Tenure는 통계적으로 유의미한 차이가 존재하지않음.

### Balance

p≈0.305

H0채택

Exited 그룹간의 Balance는 통계적으로 유의미한 차이가 존재하지않음.

50대에 한해서는 CreditScore, Tenure, Balance는 유의미한 차이가 존재 x

코드

```
t, p=stats.ttest_ind(df_0_50s.CreditScore,df_1_50s.CreditScore)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p,5))
```

코드

```
t, p=stats.ttest_ind(df_0_50s.Tenure,df_1_50s.Tenure)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p,5))
```

코드

```
t, p=stats.ttest_ind(df_0_50s.Balance,df_1_50s.Balance)
t, p

if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")

print(round(p,5))
```



## 독립표본 t 검정 ( 50대 고객 대상 )

### CreditScore

p≈0.406

H0채택

Exited 그룹간의 CreditScore는 통계적으로 유의미한 차이가 존재하지않음.

### Tenure

p≈0.066

H0채택

Exited 그룹간의 Tenure는 통계적으로 유의미한 차이가 존재하지않음.

### Balance

p≈0.305

H0채택

Exited 그룹간의 Balance는 통계적으로 유의미한 차이가 존재하지않음.

50대에 한해서는 CreditScore, Tenure, Balance는 유의미한 차이가 존재 x

코드

```
t, p=stats.ttest_ind(df_0_50s.CreditScore,df_1_50s.CreditScore)
t, p
```

```
if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")
print(round(p,5))
```

코드

```
t, p=stats.ttest_ind(df_0_50s.Tenure,df_1_50s.Tenure)
```

```
t, p
if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")
print(round(p,5))
```

코드

```
t, p=stats.ttest_ind(df_0_50s.Balance,df_1_50s.Balance)
```

```
t, p
if p < 0.05:
    print("H0 기각: 두 샘플 간에는 통계적으로 유의한 차이가 있습니다.")
else:
    print("H0 채택: 두 샘플 간에는 통계적으로 유의한 차이가 없습니다.")
print(round(p,5))
```

-> 50대의 이탈 여부를 결정하는 것은 다른 변수 또는 외부환경에 있을 확률이 크다

문제  
도입

목표  
설정

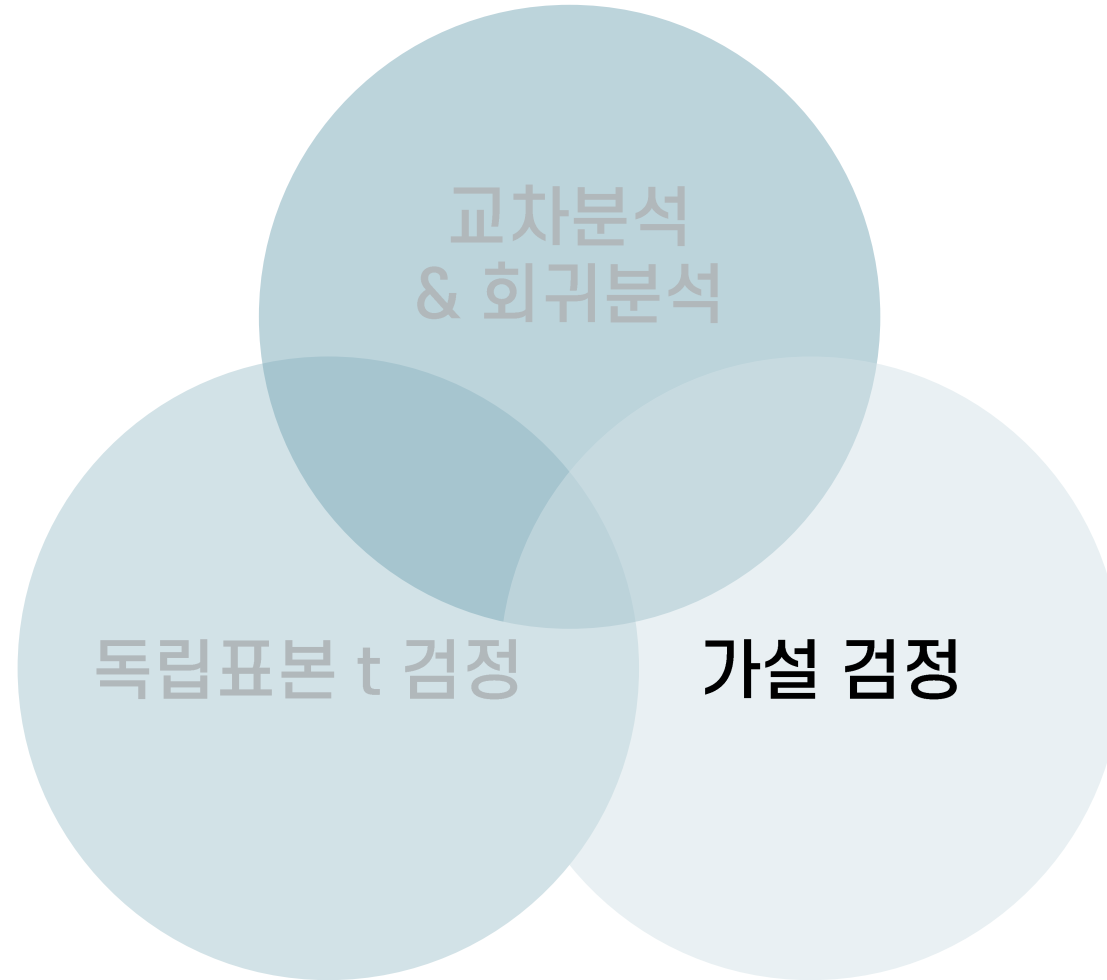
컨텍스트  
제시

가설 검정

인사이트  
제공

해결책  
제시

미래  
전망



## 가설 검증 ( 50대 고객 대상 )

### 가설 1

50대 고객(Age) 中  
거래기간(Tenure)  
이 길수록,  
이탈률이 낮다.

: Violin Plot을 활용

### 가설 2

50대 고객(Age) 中  
신용점수가  
낮을수록,  
이탈률이 높다.

: 히스토그램과 KDE  
활용

### 가설 3

50대 고객 中  
계좌잔액 (Balance)  
가 낮을수록,  
이탈률이 높다.

: 히스토그램과 KDE  
활용

## 가설 검증 ( 50대 고객 대상 )

### 가설 1

50대 고객(Age) 中  
거래기간(Tenure)  
이 길수록,  
이탈률이 낮다.

: Violin Plot을 활용

```
train_data = pd.read_csv('/kaggle/input/playground-series-s4e1/train.csv') # 데이터 파일 경

# 가설에 따른 분석
plt.figure(figsize=(12, 6))

# 50대 고객에 대한 데이터 필터링
age_50s_data = train_data[(train_data['Age'] >= 50) & (train_data['Age'] < 60)]

# 거래 기간(Tenure)에 따른 이탈률 시각화 (Violin Plot)
sns.violinplot(x=age_50s_data['Exited'], y=age_50s_data['Tenure'])

# 전체 표에 점선 추가
plt.axhline(y=age_50s_data['Tenure'].mean(), color='gray', linestyle='--', label='Mean Ten')
plt.legend() # 범례 표시

plt.title('Tenure vs. Exited for Customers in their 50s', fontsize=15)
plt.xlabel('Exited', fontsize=12)
plt.ylabel('Tenure', fontsize=12)
plt.xticks(ticks=[0, 1], labels=['Not Exited', 'Exited']) # x축 눈금 설정

# x축에 이탈률 표시
ax2 = plt.twinx()
ax2.set_xlim(0, 1)
ax2.set_xticks([0.25, 0.75])
ax2.set_xticklabels(['Low Churn Rate', 'High Churn Rate'])
ax2.set_xlabel('Churn Rate')

plt.yticks(ticks=range(int(age_50s_data['Tenure'].min()), int(age_50s_data['Tenure'].max())

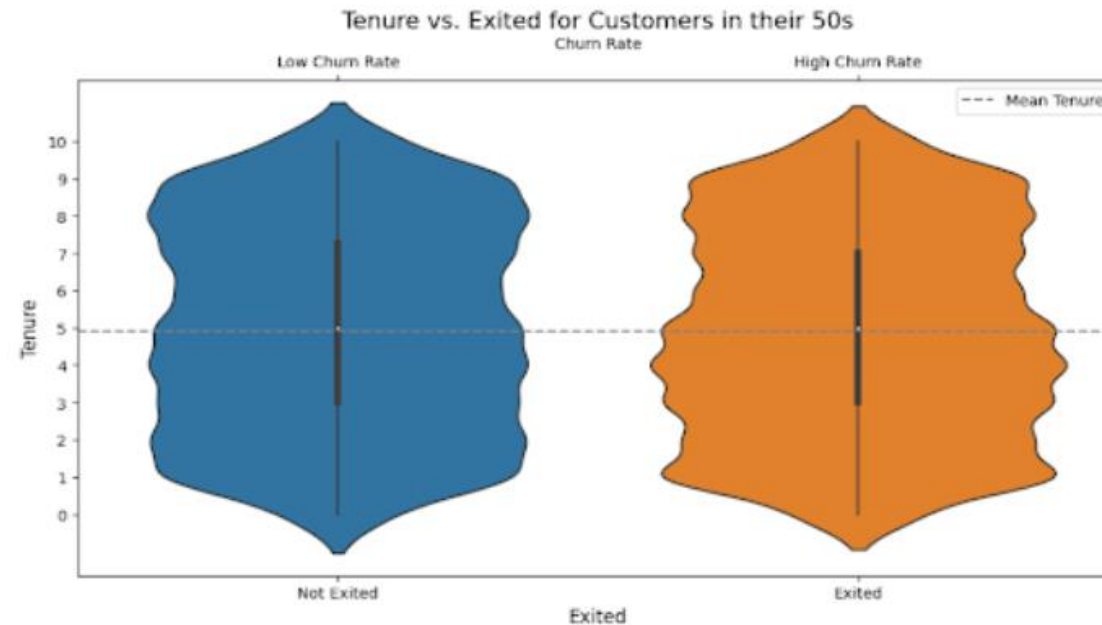
plt.show()
```

## 가설 검증 ( 50대 고객 대상 )

### 가설 1

50대 고객(Age) 中  
거래기간(Tenure)  
이 길수록,  
이탈률이 낮다.

: Violin Plot을 활용

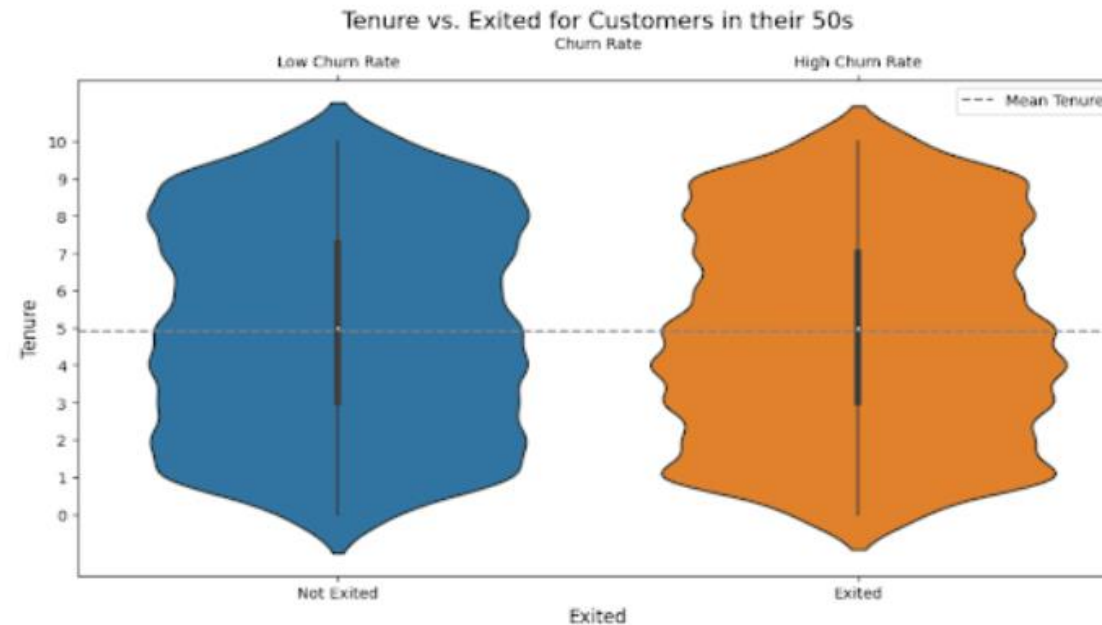


## 가설 검증 ( 50대 고객 대상 )

### 가설 1

50대 고객(Age) 中  
거래기간(Tenure)  
이 길수록,  
이탈률이 낮다.

: Violin Plot을 활용



Tenure는 Exited에  
영향을 주는 것으로 보임

## 가설 검증 ( 50대 고객 대상 )

### 가설 2

50대 고객(Age) 中  
신용점수가  
낮을수록,  
이탈률이 높다.

: 히스토그램과 KDE  
활용

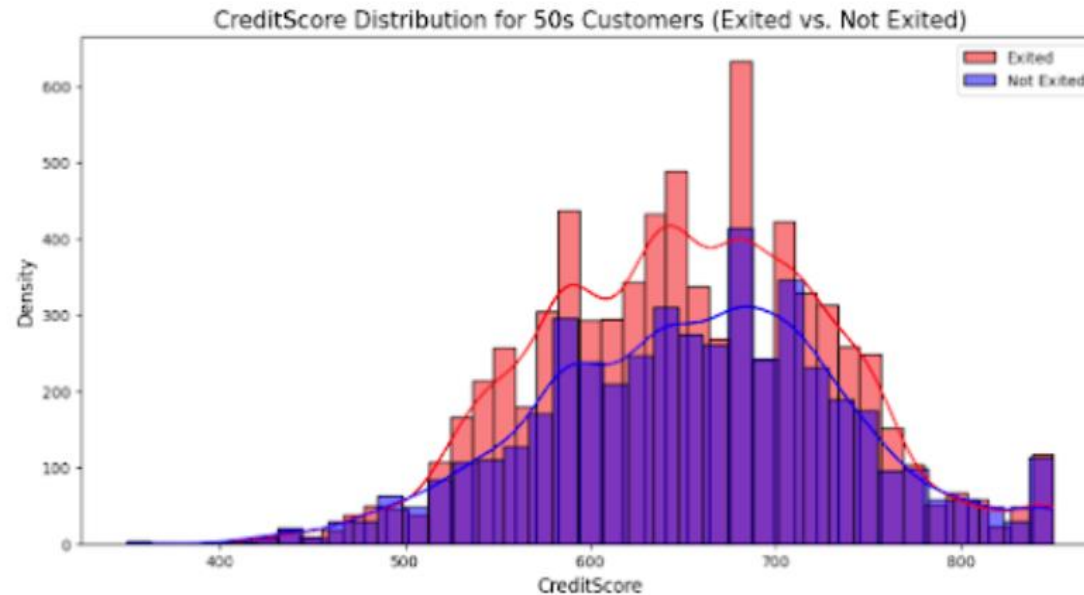
```
train_data = pd.read_csv('/kaggle/input/playground-series-s4e1/train.csv') # 데이터 파일 경.  
  
# 50대 고객에 대한 데이터 필터링  
age_50s_data = train_data[(train_data['Age'] >= 50) & (train_data['Age'] < 60)]  
  
# 이탈자와 비이탈자를 구분하여 데이터 분리  
exited_data = age_50s_data[age_50s_data['Exited'] == 1]  
not_exited_data = age_50s_data[age_50s_data['Exited'] == 0]  
  
# CreditScore 분포 시각화  
plt.figure(figsize=(12, 6))  
  
sns.histplot(exited_data['CreditScore'], label='Exited', color='red', alpha=0.5, kde=True)  
sns.histplot(not_exited_data['CreditScore'], label='Not Exited', color='blue', alpha=0.5,  
  
plt.title('CreditScore Distribution for 50s Customers (Exited vs. Not Exited)', fontsize=  
plt.xlabel('CreditScore', fontsize=12)  
plt.ylabel('Density', fontsize=12)  
plt.legend()  
plt.show()
```

## 가설 검증 ( 50대 고객 대상 )

### 가설 2

50대 고객(Age) 中  
신용점수가  
낮을수록,  
이탈률이 높다.

: 히스토그램과 KDE  
활용



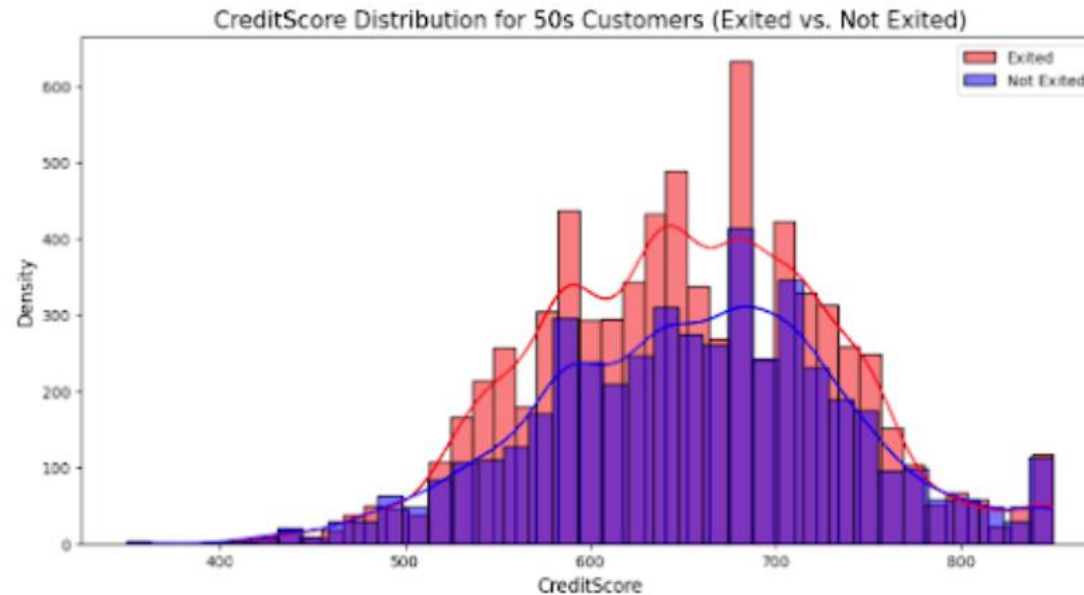


## 가설 검정 ( 50대 고객 대상 )

### 가설 2

50대 고객(Age) 中  
신용점수가  
낮을수록,  
이탈률이 높다.

: 히스토그램과 KDE  
활용



600점대 후반에서  
700점대 후반까지  
이탈률이 높게 측정됨

## 가설 검증 ( 50대 고객 대상 )

### 가설 3

50대 고객 中  
계좌잔액 (Balance)  
가 낮을수록,  
이탈률이 높다.

: 히스토그램과 KDE  
활용

```
train_data = pd.read_csv('/kaggle/input/playground-series-s4e1/train.csv')

# 50대 고객에 대한 데이터 필터링
age_50s_data = train_data[(train_data['Age'] >= 50) & (train_data['Age'] < 60)]

# 계좌잔액(Balance)에 따른 이탈률 시각화 (Violin Plot)
plt.figure(figsize=(12, 6))
sns.violinplot(x='Exited', y='Balance', data=age_50s_data, palette="muted", split=True)
plt.axhline(y=age_50s_data['Balance'].median(), color='black', linestyle='--', label='Median')
plt.title('Balance vs. Exited for Customers in their 50s', fontsize=15)
plt.xlabel('Exited', fontsize=12)
plt.ylabel('Balance', fontsize=12)
plt.xticks([0, 1], ['Not Exited', 'Exited'])
plt.legend()
plt.show()
```

## 가설 검증 ( 50대 고객 대상 )

### 가설 3

50대 고객 中  
계좌잔액 (Balance)  
가 낮을수록,  
이탈률이 높다.

: 히스토그램과 KDE  
활용



계좌 잔액이 낮은  
(잔액이 0원인)  
50대 고객이  
이탈하는 경향

## 가설 검증 ( 50대 고객 대상 )

### 가설 3

50대 고객 中  
계좌잔액 (Balance)  
가 낮을수록,  
이탈률이 높다.

: 히스토그램과 KDE  
활용



가설 채택



## 가설 검증 ( 50대 고객 대상 )

### 가설 3

50대 고객 中  
계좌잔액 (Balance)  
가 낮을수록,  
이탈률이 높다.

: 히스토그램과 KDE  
활용



계좌 잔액이  
낮은 고객에 대한  
서비스 개선이나  
타겟 마케팅이 필요함

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

## Acquired Insight

파생변수(BalLevel)와 연관성 보임

In 교차분석

계좌잔액과 이탈률의 상관관계

In 가설 검정

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망

## Acquired Insight

파생변수(BalLevel)와 연관성 보임

In 교차분석

계좌잔액과 이탈률의 상관관계

In 가설 검정

## Acquired Insight

파생변수(BalLevel)와 연관성 보임

In 교차분석

계좌잔액과 이탈률의 상관관계

In 가설 검정





문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

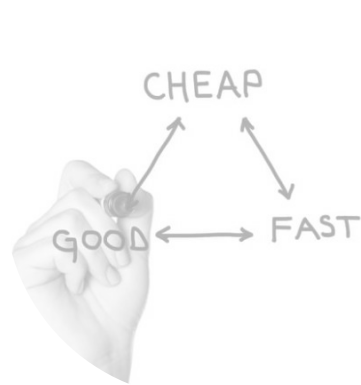
미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공



문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공



문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

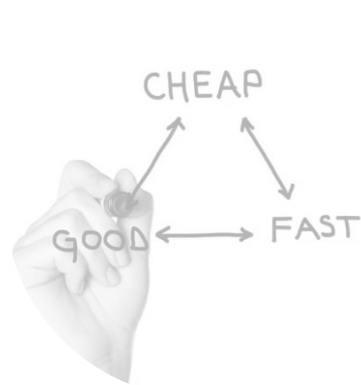
미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

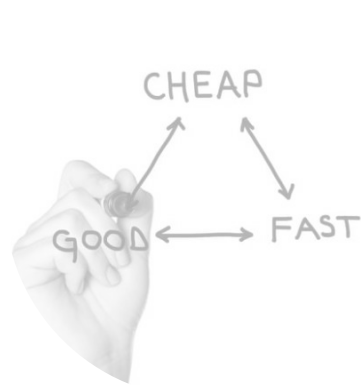
미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

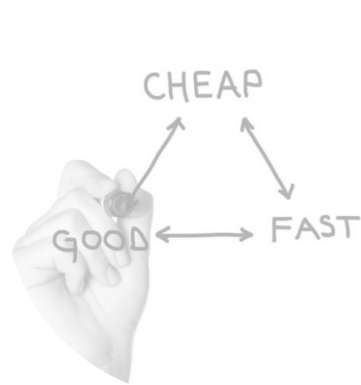
미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

고객 참여 유발

브랜드 충성도 증가

기업 경쟁력 확보

기업 글로벌화

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

**특정 국가 내 고객의 은행 이탈 예방**



문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

고객 참여 유발

브랜드 이미지 강화

고객 충성도 향상  
장기적 관계 구축

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

## 기업의 지역 사회 참여 ( CSR )

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

금전적 부담 ↓

은행 서비스 접근성 ↑

기존 고객 이탈 방지  
신규 고객 유치



문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

은퇴를 향한 50대의 부담감 해소

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

고객 호감도 ↑

기업 경쟁력 강화

기존 고객 이탈 방지  
신규 고객 유치

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

**고객 손실도 완화 & 고객 규모 확대**

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

고객의 신뢰 획득

고객 충성도 ↑

고객 불안감 감소

고객 이탈 방지

문제  
도입

목표  
설정

컨텍스트  
제시

주요 이벤트  
강조

인사이트  
제공

해결책  
제시

미래  
전망



국가별 특색 있는 마케팅



지역사회 협력



저렴하고 간편한  
은행 서비스



높은 이자 상품



50대를 위한  
컨설팅 제공

50대를 겨냥한 공격적 마케팅

# Q & A

---

Workers : 5조 [ 김수현, 나한울, 정혜원, 한대희, 황유진 ]

- Data Analysis : 팀원 전체
- Edit Analysis Content : 팀원 전체
- Edit Article : 김수현, 나한울, 한대희
- Preparation of presentation : 정혜원, 황유진
- Translate : 정혜원, 황유진

Period for A to Z : 2 days (2024-01-15 ~ 2024-01-16)

# Thank You for Listening !

---

Workers : 5조 [ 김수현, 나한울, 정혜원, 한대희, 황유진 ]

- Data Analysis : 팀원 전체
- Edit Analysis Content : 팀원 전체
- Edit Article : 김수현, 나한울, 한대희
- Preparation of presentation : 정혜원, 황유진
- Translate : 정혜원, 황유진

Period for A to Z : 2 days (2024-01-15 ~ 2024-01-16)