

신규 사용자 획득을 위한 광고 캠페인 성과 분석

*데이터 요약

광고 캠페인 집계 데이터 : 광고ID, 캠페인ID, 페이스북 추적 ID, 나이, 성별, 관심사, 노출 횟수, 클릭 수, 지불 비용, 가입 인원수, 구매 인원수 (구분 캠페인 수 1143개)
데이터 개요 : 30-40대 남녀 대상 캠페인, Z캠페인은 수가 매우 적음, 그외 수집 기간, 비즈니스 모델 등에 대한 정보 전무

[데이터 사전 검토]

1. 데이터 전처리

[1] 오류 데이터 정리

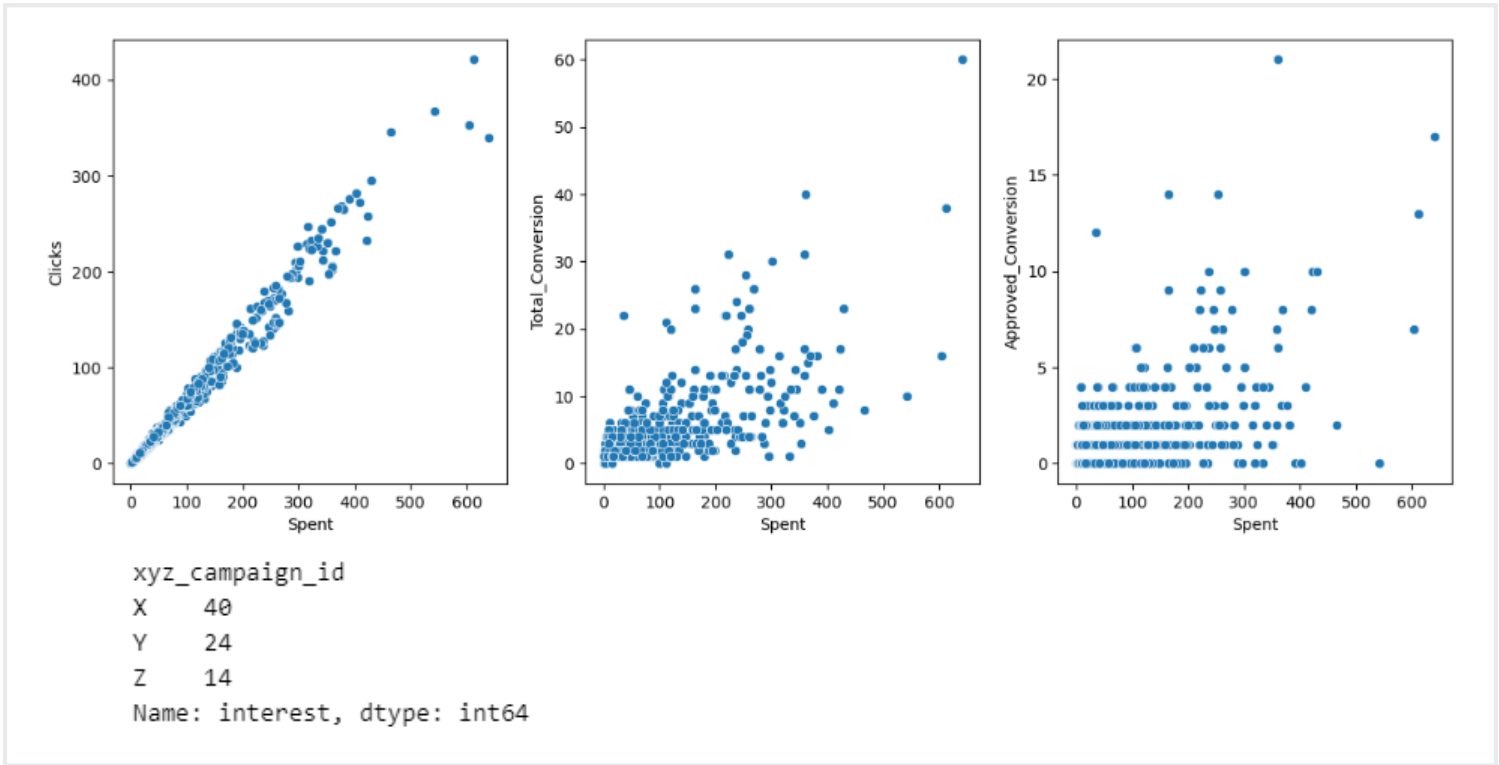
- 클릭 수 > 가입인원 수 > 구매인원 수 구조를 지키지 않는 데이터는 오류로 판단하여 제거

[2] 혼동을 주는 열 이름 및 데이터 변경

- Conversion0이 중복되는 열 이름을 사전 정의에 맞게 변경하고, 인코딩 된 캠페인 명을 변경

[3] 불필요한 데이터 정리

- 페이스북 추적 ID가 다른 열에 연동되는 것을 확인하여 제거, 관심사 데이터는 다각도로 검토한 끝에 유용성이 없어 제거



2. 비용 구조?

[1] 광고 계약 종류

- CPM, CPC, CPA 계약 방식 존재
- 클릭이 0일 때 비용이 발생하지 않고, 상관관계 고려 시, **CPC로 추정**

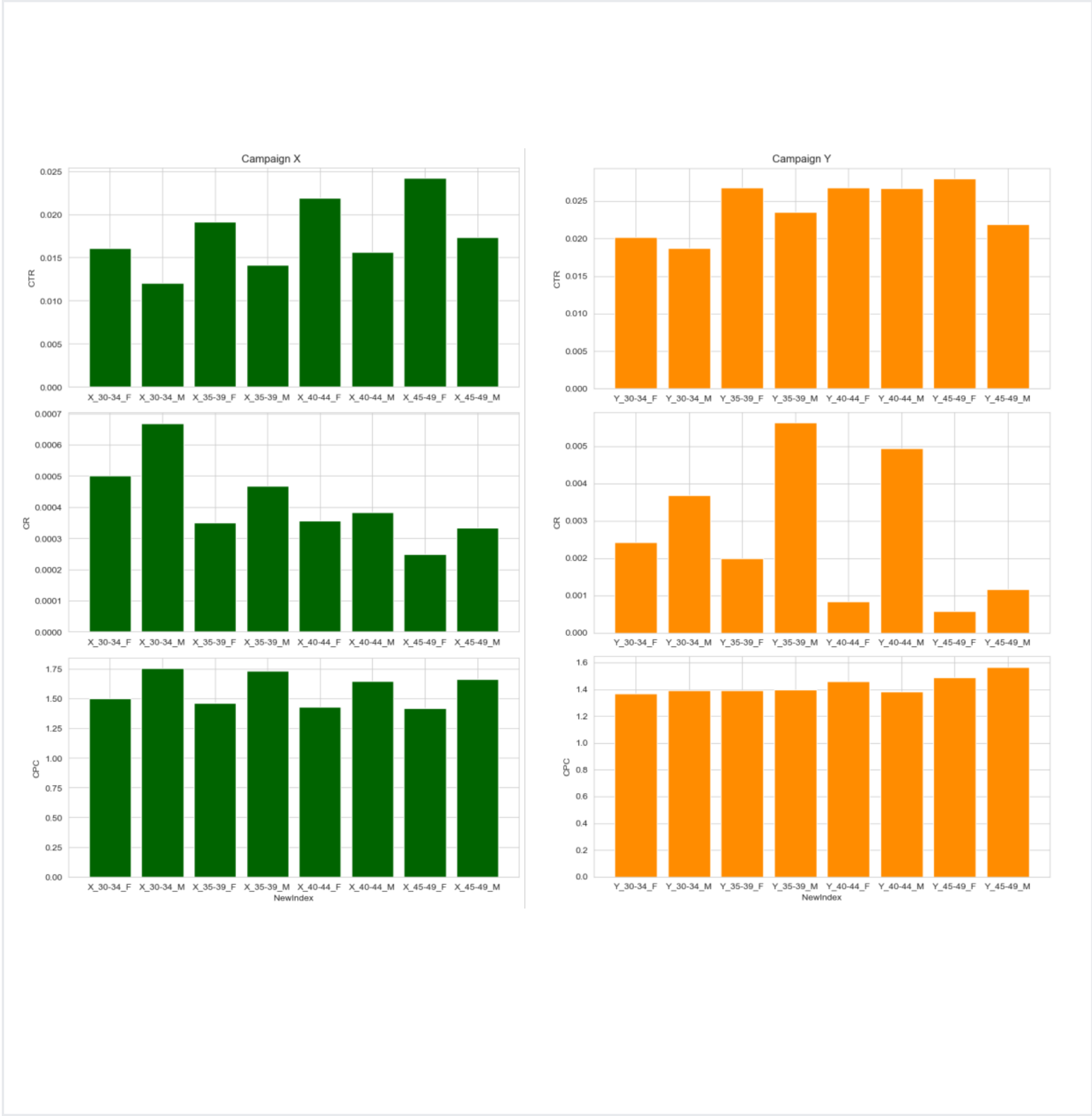
[2] 지불 비용

- 해당 열의 데이터는 **클릭당 비용**(CPC)가 아니라 총 클릭에 대한 비용
- 따라서 CPC를 따로 계산해주어야 함

3. 관심사에 대한 고려 여부

- 페이스북 내 계정의 관심사를 타겟으로 마케팅이 가능
- 하지만, 해당 데이터에서 최대 40개의 관심사가 발견되지만, 해당 인코딩이 어떤 방식인지, 어떤 내용인지 알 수 없음
- 사전에 광고주가 의도한 타겟팅 대상인지, 아니면 페이스북 알고리즘에 따라 노출된 관심사 그룹을 의미하는 것인지 판단하기 어려움
- 따라서 해당 데이터에선 **관심사 데이터는 고려 X**

[EDA]



[1] X 캠페인의 그룹별 CTR, CR, CPC 분석

CTR

- 여성이 남성보다 전반적으로 높고, 나이대가 높을 수록 상승하는 경향

CR

- CTR과 달리 남성이 여성보다 전반적으로 높고, 나이대가 높을 수록 하락

CPC

- 남성이 여성보다 모두 약 0.25정도 더 높고, 나이대에 따라서는 높아질수록 하락하지만 큰 차이는 없음

정리 :

- 남성 그룹은 CR이 높지만 CPC도 높다. 하지만 비용 대비 효과는 각 그룹별 **고객 가치**에 따라 달라질 수 있음
- 성별과 나이대를 타겟으로 마케팅 효과를 극대화하고자 한다면, 그룹별 가치 정보를 측정해서 활용해야 함

[2] Y 캠페인의 그룹별 CTR, CR, CPC 분석

CTR

- 두드러지는 경향은 관찰되지 않으나, 30대 초반의 CTR이 상대적으로 낮게 나타남

CR

- 남성이 여성보다 전 구간에 더 높은 경향
- 35~44세 두 구간에서 남녀의 차이가 더 확장됨

CPC

- 큰 경향성은 X, 40대 후반의 CPC가 가장 큼

정리 :

- 35~44세 구간의 남성들을 타겟으로 마케팅을 진행하면 다른 그룹에 비해 **비용대비 효과가 클 것**으로 예상
- 35~44세 구간의 여성들은 CR이 절대적으로 가장 낮게 나타나므로 이들 그룹의 고객가치가 특별히 높지 않다면, **비용대비 효과가 가장 작을 것**

[A/B 테스트]

1. 문제 정의	<ul style="list-style-type: none">기존에 하고 있던 X 마케팅 캠페인에 비해 신규로 진행할 Y 마케팅 캠페인이 더 효과적인가?통계적으로 유의미하다면 신규 마케팅으로 대체하자!
2. 가설	<div>1. X 캠페인과 Y 캠페인의 전환율 차이는 통계적으로 유의미할 것이다. ➡ 이원(양쪽 꼬리) 가설 검정</div> <div>2. Y 캠페인의 전환율이 X 캠페인의 전환율보다 2%이상 높을 것이다. ➡ 일원(한쪽 꼬리) 가설 검정</div>
독립 변수 : 광고 캠페인	<ul style="list-style-type: none">귀무 가설 (H0) : 1. X 캠페인과 Y 캠페인의 전환율 차이는 유의미하지 않다. 2. Y 캠페인의 전환율이 X 캠페인의 전환율 보다 2% 이상 높지 않다.
종속 변수 : 전환율	<ul style="list-style-type: none">대립 가설 (H1) : 1. X 캠페인과 Y 캠페인의 전환율 차이는 유의미하다. 2. Y 캠페인의 전환율이 X 캠페인의 전환율 보다 2% 이상 높다.
임의로 2%로 가정 ➡	<div>목표 수준 - 전환율 2%에 대해</div> <ul style="list-style-type: none">목표 수준을 정하는 것은 비즈니스 모델에 따라 실질적인 효과의 크기 및 영향력을 고려해야 함ARPPU 등 수익 관련 데이터가 있다면, 효과 크기를 구할 수 있다.효과 크기와 테스트 비용을 고려해 목표 수준을 제시할 수 있다.
3. 시나리오 및 가정	<div>1. 본 데이터의 광고주는 기존에 X 마케팅 캠페인을 진행하고 있었고, 신규 Y 마케팅 캠페인으로 새롭게 변경할지 실험하고자 한다.</div> <div>2. X 마케팅 캠페인을 진행하는 와중에 Y 광고 캠페인을 새로이 투입하여 동시에 진행하였고, Y 마케팅 캠페인의 성과가 더 낮을 경우를 대비하여 Y 마케팅 캠페인의 크기를 3 : 1 정도로만 나누어 실시하였다. (X : Y = 3 : 1)</div> <div>3. 두 마케팅 캠페인을 동시에 진행하면서 고객은 두 캠페인 중 하나에만 노출되도록 통제하였다. 그 외 동시에 진행하는 다른 캠페인은 없으며, 이 외 통제 변수는 잘 관리되었다고 가정한다.</div> <div>4. Z 캠페인은 X, Y 캠페인과 진행 기간이 다르다고 가정하여 A/B 테스트 대상에서 제외한다.</div> <div>5. X 캠페인이 기존에 진행하고 있어서 생기는 샘플 크기의 차이는 독립 표본 t-검정을 이용해 샘플 크기 차이 문제를 회피한다.</div>

4. A/B 테스트 결과

- 두 가설 모두 귀무 가설을 기각하고 대립가설을 받아들인다.
- X 캠페인과 Y 캠페인의 전환율 차이는 통계적으로 유의미하다.
- Y 캠페인의 전환율이 X 캠페인 보다 2% 이상 높을 것이다.
- X 캠페인에서 Y 캠페인으로 변경하는 조치는 타당하다고 결론 내릴 수 있다.

import scipy.stats as stats

캠페인 x와 y에 대한 A/B테스트 가설1 검정

가설1 : x 캠페인과 y 캠페인의 전환율 차이는 통계적으로 유의미할 것이다.

campaign_X_data = ad_campaign[ad_campaign['xyz_campaign_id'] == 'X']

campaign_Y_data = ad_campaign[ad_campaign['xyz_campaign_id'] == 'Y']

conversion_rate_X = campaign_X_data['Conversion'].sum() / campaign_X_data['Clicks'].sum()

conversion_rate_Y = campaign_Y_data['Conversion'].sum() / campaign_Y_data['Clicks'].sum()

양쪽 꼬리 가설 검정

t_stat, p_value = stats.ttest_ind(campaign_X_data['Conversion'], campaign_Y_data['Conversion'], equal_var=False)

print(f"x 캠페인의 전환율: {conversion_rate_X}")

print(f"y 캠페인의 전환율: {conversion_rate_Y}")

print(f"p-value: {p_value}")

if p_value < 0.05:

print("두 캠페인의 전환율 차이는 통계적으로 유의미합니다.")

else:

print("두 캠페인의 전환율 차이는 통계적으로 유의미하지 않습니다.")

x 캠페인의 전환율: 0.024014863703169628

y 캠페인의 전환율: 0.058168942842690946

p-value: 7.652257563602614e-25

두 캠페인의 전환율 차이는 통계적으로 유의미합니다.

< 가설 1 >

캠페인 x와 y에 대한 A/B테스트 가설2 검정

가설2 : y 캠페인의 전환율이 x 캠페인의 전환율보다 2%이상 높을 것이다.

conversion_rate_difference = conversion_rate_Y - conversion_rate_X

일원 가설 검정

t_stat, two_tailed_p_value = stats.ttest_ind(campaign_Y_data['Conversion'], campaign_X_data['Conversion'], equal_var=True)

one_tailed_p_value = two_tailed_p_value / 2

print(f"x 캠페인의 전환율: {conversion_rate_X}")

print(f"y 캠페인의 전환율: {conversion_rate_Y}")

print(f"전환율 차이: {conversion_rate_difference}")

print(f"일원 가설 검정 p-value: {one_tailed_p_value}")

if one_tailed_p_value < 0.05 and conversion_rate_difference > 0.02:

print("y 캠페인의 전환율이 x 캠페인보다 2% 이상 높으며, 이 차이는 통계적으로 유의미합니다.")

else:

print("y 캠페인의 전환율이 x 캠페인보다 2% 이상 높지 않거나, 이 차이가 통계적으로 유의미하지 않습니다.")

x 캠페인의 전환율: 0.024014863703169628

y 캠페인의 전환율: 0.058168942842690946

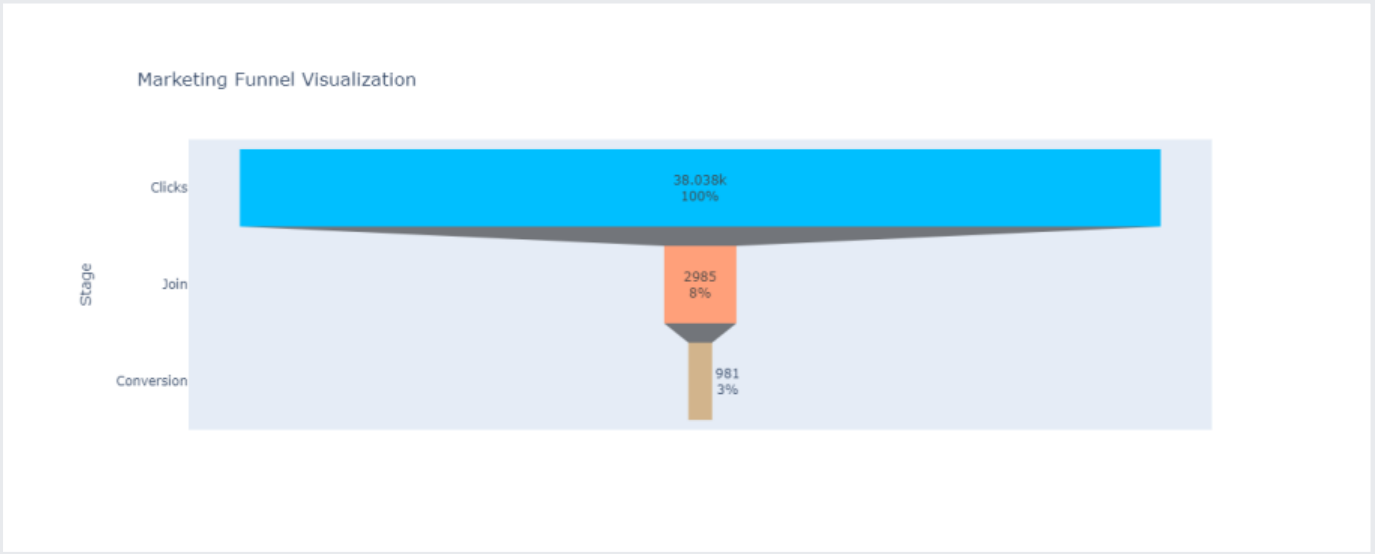
전환율 차이: 0.034154079139521315

일원 가설 검정 p-value: 3.826128781801307e-25

y 캠페인의 전환율이 x 캠페인보다 2% 이상 높으며, 이 차이는 통계적으로 유의미합니다.

< 가설 2 >

[퍼널 시각화]



방문 부터 실제 구매까지의 퍼널

- 약 38,000건의 방문자 중 가입까지 이어진 것은 2,985명 (방문 이후 전환율 7.8%)
- 2,985명의 가입자 중 실제 구매자 수는 981명 (가입 이후 전환율 32.9%)

[프로젝트 한계점 및 개선 아이디어]

[1] 고객 수익 지표와 비즈니스 모델 부재의 아쉬움

- 개선 후 행동 방안 : 특정 기간에 대한 ARPPU 혹은 고객 생애 매출(LTR) 등의 데이터를 보충한다면, 고객 획득 비용(CAC)을 이용해 수익화 분석을 추가할 수 있을 것.

또한, A/B 테스트 수행 설계 시점에서 효과 크기와 테스트 비용을 고려할 수 있어 목표 수준을 명확히 하여 테스트를 수행할 수 있을 것.

[2] 유저 방문 부터 구매까지의 퍼널 단계 세부 데이터의 부족

- 개선 후 행동 방안 : 방문 이후 '회원 가입 시작' 및 '전화번호 인증' 단계, 가입 이후 '상품 페이지 조회' 및 '장바구니' 등 최종 구매 까지 유저 행동에 대한 로그를 통해 퍼널 단계를 세밀하게 추적할 수 있다면, 고객 이탈 시점을 분석하고 유저 경험 개선을 통해 서비스를 개선할 수 있을 것