**A/B 테스트를 통한 광고 효과 분석 결과 (실험데이터)**

|  |
| --- |
| **전체 내용 두 줄 요약** |
| **Propensity Score Matching**을 통해 데이터 불균형 문제를 해결하고 신뢰할 수 있는 분석 결과를 도출. 이를 기반으로 광고 캠페인의 효과성을 검증 |

**1. 프로젝트 개요**

**1-1 프로젝트의 목적**

* 광고 캠페인의 효과성을 검증하고, 광고가 전환율(Conversion Rate) 증가를 유발했는지를 평가
* 캠페인이 성공적이었다면, 그 성공의 어떤 부분이 광고에 기인했는지를 파악
* 광고의 순수한 효과를 측정함으로써 데이터 기반의 의사결정을 지원

**1-2 분석 질문**

1. 광고 캠페인은 성공했는가?
2. 성공이 광고의 효과라면, 얼마나 광고의 영향 때문인가?

**1-3 데이터 구성**

* **총 데이터 크기**: 588,101개 샘플
* **주요 변수**:
  + user id: 사용자 ID(고유값)
  + test\_group: 광고 여부 (Treatment: 광고 시청, Control: PSA 시청)
  + converted: 사용자가 제품을 구매했는지 여부 (1: 구매, 0: 구매 안 함)
  + total\_ads: 사용자가 본 총 광고 수
  + most\_ads\_day: 사용자가 가장 많이 광고를 본 요일
  + most\_ads\_hour: 사용자가 가장 많이 광고를 본 시간대

**2. 데이터 분석 프로세스 및 결과**

**2-1 데이터 불균형 문제**

* **문제 정의**:
  + 원본 데이터는 **심각하게 불균형**함
    - Control 그룹(psa): 23,524개
    - Treatment 그룹(ad): 564,577개
    - **비율**: psa:ad = 1:24
  + 광고 그룹(Treatment)과 공익광고 그룹(Control)의 데이터 크기뿐만 아니라 변수(most\_ads\_hour, most\_ads\_day)의 분포에서도 차이를 보임

**2-2 불균형 문제 해결 방법**

**왜 데이터 불균형을 해결해야 하는가?**

* 데이터 불균형은 다음과 같은 문제를 초래함:
  1. **편향된 결과**:  
     Treatment 그룹의 데이터가 과다하면, Control 그룹의 특성이 제대로 반영되지 않아 **광고 효과를 과대평가하거나 과소평가**할 위험이 있음
  2. **공정한 비교 불가**:  
     두 그룹 간 변수 분포가 다르면 \*\*광고 효과 외의 다른 요인들(시간대, 요일 등)\*\*이 결과에 영향을 미칠 가능성이 있음
  3. **통계적 신뢰성 저하**:  
     데이터 크기와 분포 차이가 클 경우, 통계적 검증 결과의 신뢰성이 떨어짐

**불균형 해결 방법: Propensity Score Matching (PSM)**

* **PSM의 개념**:
  + Propensity Score(광고를 볼 확률)를 계산해 두 그룹(Treatment, Control)의 변수를 유사하게 맞추는 기법
  + 매칭 후 두 그룹의 외부 요인이 동일해지므로, **광고 효과만을 독립적으로 평가** 가능
* **PSM을 사용하는 이유**:
  + 단순 샘플링이나 통계적 보정보다 **광고 효과에 대한 더 정확한 추정**을 제공
  + **데이터 손실 최소화**:
    - Treatment 그룹의 중요한 데이터를 제거하지 않고, Control 그룹의 부족한 데이터를 보완
  + **비교의 공정성 확보**:
    - 두 그룹 간 변수 차이를 제거하여 **외부 요인을 통제**하고 광고 효과를 평가

**2-3 Propensity Score Matching 적용**

* **1단계: Propensity Score 계산**
  + 로지스틱 회귀(Logistic Regression)를 사용해 Propensity Score(광고를 볼 확률)를 계산
  + 사용된 변수:
    - most\_ads\_day (사용자가 광고를 가장 많이 본 요일)
    - most\_ads\_hour (사용자가 광고를 가장 많이 본 시간대)
* **2단계: 1:2 매칭 수행**
  + Propensity Score가 유사한 Control 그룹과 Treatment 그룹의 샘플을 1:2로 매칭
  + Caliper(허용 오차)를 설정하여 높은 품질의 매칭 수행
  + 처리 후 데이터 크기:
    - Control 그룹(psa): 23,524개
    - Treatment 그룹(ad): 47,048개
    - 비율: psa:ad = 1:2
* **3단계: 매칭 품질 평가**
  + Love plot을 사용해 매칭 전후 변수 균형 평가
    - 매칭 전: most\_ads\_hour 변수에서 두 그룹 간 분포 차이가 큼
    - 매칭 후: most\_ads\_hour 변수의 분포가 유사해짐
  + 표준화된 평균 차이(SMD)가 대부분의 변수에서 01 이하로 줄어듦

**2-4 로지스틱 회귀(Logit) 분석**

* **분석 설계**:
  + 종속 변수: converted (전환 여부)
  + 독립 변수:
    - test\_group\_binary: 광고 여부 (0: PSA, 1: 광고)
    - total\_ads: 총 광고 노출 횟수
    - most\_ads\_hour: 광고가 가장 많이 노출된 시간대
    - ads\_treatment\_interaction: 광고 여부와 광고 노출 횟수의 상호작용 변수
* **다중공선성(VIF) 평가**:
  + 대부분의 변수들이 낮은 VIF 값을 보여 다중공선성 문제가 없음
  + 상호작용항에서 약간 높은 VIF가 관찰되었으나 이는 예상된 결과
* **분석 결과**:
  + **광고 여부(test\_group\_binary)**:
    - 계수: -0.426
    - 전환율에 유의미한 영향을 미침 (p-value < 0.01)
  + **광고 노출 수(total\_ads)**:
    - 계수: -0.005
    - 노출 횟수가 많을수록 전환율 감소와 관련됨
  + **광고 노출 시간대(most\_ads\_hour)**:
    - 계수: +0.011
    - 시간대별 광고 효과의 차이를 시사
  + **상호작용 효과(ads\_treatment\_interaction)**:
    - 계수: +0.016
    - 광고 여부와 노출 횟수의 결합이 전환율에 강한 양의 영향을 미침

**2-5 전환율 분석**

* **원본 데이터 전환율**:
  + Control 그룹: 1.79%
  + Treatment 그룹: 2.55%
* **매칭 후 데이터 전환율**:
  + Control 그룹: 1.85%
  + Treatment 그룹: 2.66%
* **평균 처치 효과(ATE)**:
  + 광고 그룹(Treatment)의 전환율이 공익광고 그룹(Control)보다 **0.82%p 더 높음**
* **결론**:
  + 매칭 후 전환율 차이가 원본 데이터보다 신뢰할 수 있는 추정치를 제공
  + ATE는 광고가 전환율에 미친 순수한 효과를 보여줌

**3. 결론 및 비즈니스 인사이트**

**3-1 결론**

* 광고는 전환율 증가에 유의미한 긍정적 영향을 미침
* Propensity Score Matching을 통해 데이터 불균형 문제를 해결하고, 외부 요인 통제 후 광고 효과를 정확히 평가

**3-2 비즈니스 인사이트**

1. **광고 효과성 검증**:
   * 광고 캠페인이 유의미한 전환율 증가를 유발했음을 확인
2. **광고 노출 최적화 필요**:
   * 과도한 광고 노출은 전환율 감소와 관련될 수 있으므로 노출 전략 재구성 필요
3. **타겟팅 강화**:
   * 특정 시간대나 사용자 그룹에서 광고 효과가 높아질 가능성에 따라 타겟팅 전략 수립