

Chapter. 13

직접 해봐야 내것이 된다

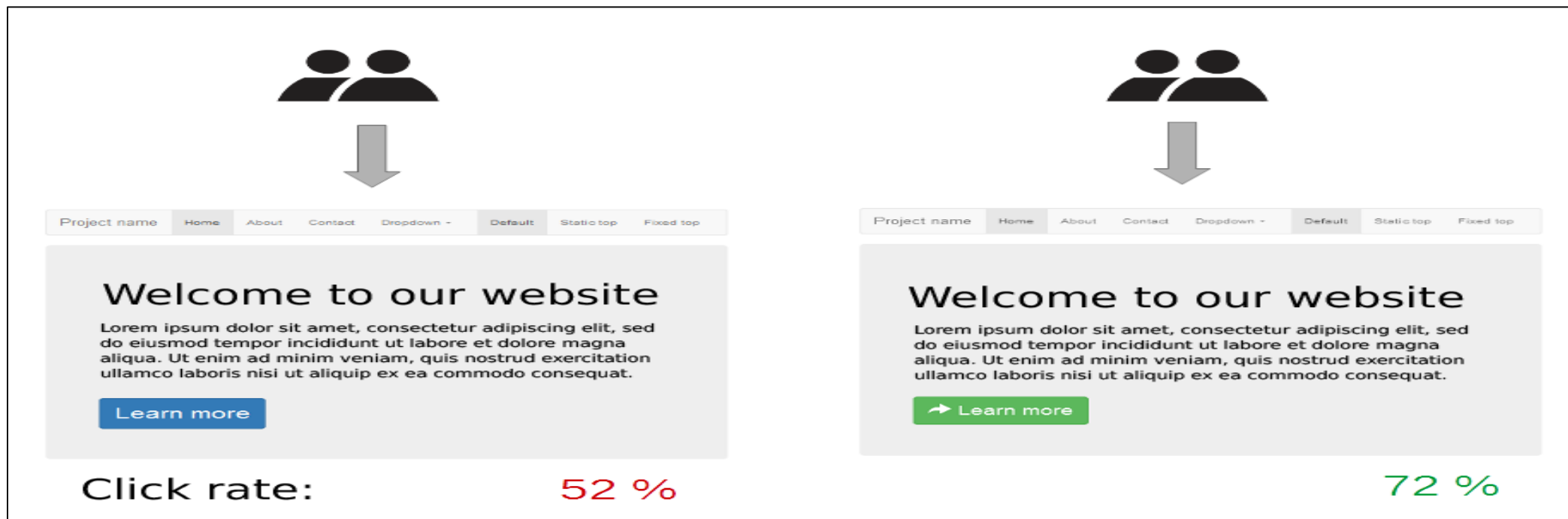
A/B 테스트: 홈페이지 화면을 어떻게 구성할 것인가?

FAST CAMPUS
ONLINE
데이터 탐색과 전처리 I

강사. 안길승

I A/B 테스트란?

- 임의로 나눈 둘 이상의 집단에 서로 다른 콘텐츠를 제시한 뒤, **통계적 가설 검정**을 이용하여 어느 콘텐츠에 대한 반응이 더 효과적인지를 파악하는 방법



https://ko.wikipedia.org/wiki/A/B_%ED%85%8C%EC%8A%A4%ED%8A%B8

I 문제 상황

- 온라인 쇼핑몰 페이지 구성에 따른 다양한 실험 결과를 바탕으로 **전환율이 최대**가 되는 구성을 하고 싶다!

현황 파악

상품 배치와
상품 구매 금액에
따른 관계 분석

사이트맵 구성에 따른
체류 시간 차이 분석

할인 쿠폰의 효과 분석

체류 시간과 구매 금액 간
관계 분석

구매버튼 배치에 따른
구매율 차이 분석

- 관련 데이터: AB테스트 폴더 내 모든 데이터

I Step 1. 현황 파악

- 관련 데이터: 일별현황데이터.csv
- 분석 내용
 - (1) 구매자수, 방문자수, 총 판매 금액에 대한 기술 통계
 - (2) 일자별 방문자수 추이 파악
 - (3) 일자별 구매자수 추이 파악
 - (4) 일자별 총 판매 금액 추이 파악

I Step 2. 상품 배치와 상품 구매 금액에 따른 관계 분석

- 관련 데이터

- 상품배치_A.csv
- 상품배치_B.csv
- 상품배치_C.csv

- 분석 내용

- (1) 일원분산분석을 이용한 상품 배치에 따른 상품 구매 금액 평균 차이 분석 (상품 구매 금액 0원 **미포함**)
- (2) 일원분산분석을 이용한 상품 배치에 따른 상품 구매 금액 평균 차이 분석 (상품 구매 금액 0원 **포함**)
- (3) 카이제곱 검정을 이용한 구매 여부와 상품 배치 간 독립성 파악

I Step 3. 사이트맵 구성에 따른 체류 시간 차이 분석

- 관련 데이터
 - 사이트맵_A.csv
 - 사이트맵_B.csv
 - 사이트맵_C.csv
- 분석 내용
 - (1) 사이트맵별 체류시간 평균 계산
 - (2) 일원분산분석을 이용한 사이트맵에 따른 체류 시간 평균 차이 분석 (박스플롯 포함)

I Step 4. 할인 쿠폰의 효과 분석

- 관련 데이터: 할인쿠폰발행효과.csv
- 분석 내용
 - (1) 발행후와 전의 구매 횟수 차이에 대한 기술 통계
 - (2) 발행전, 발행후의 구매 횟수에 대한 박스플롯 시각화
 - (3) 쌍체 표본 t-검정을 이용한 차이 유의성 검정

I Step 5. 체류 시간과 구매 금액 간 관계 분석

- 관련 데이터: 체류시간_구매금액.csv
- 분석 내용
 - (1) 구매금액과 체류 시간의 산점도 시각화
 - (2) 구매금액과 체류 시간 간 상관관계 분석

I Step 6. 구매버튼 배치에 따른 구매율 차이 분석

- 관련 데이터: 구매버튼_버튼타입_통계.xlsx
- 분석 내용
 - (1) 결측 대체
 - (2) pivot table을 이용한 교차 테이블 생성
 - (3) 카이제곱검정을 이용한 독립성 검정

Chapter. 13

직접 해봐야 내것이 된다

| 고객 세분화를 통한 마케팅 전략 수립

FAST CAMPUS
ONLINE
데이터 탐색과 전처리 I

강사. 안길승

I 문제 상황

- 주문 내역 데이터를 바탕으로 고객들을 주문 특성과 주문 상품에 따라 그룹화하고 싶다!

주문 특성 기준 군집화

- 환불 데이터와 주문 데이터로 분할
- 특징 부착
- 코사인 유사도 기반의 계층 군집화 수행

주문 상품 기준 군집화

- 판매 상위 100개 상품 출현 여부를 나타내는 데이터로 변환
- 자카드 유사도 기반의 계층 군집화 수행

I 주문 특성 기준 군집화 - 데이터 분할

- 해당 데이터에는 주문 데이터와 환불 데이터가 동시에 포함되어 있음
- 환불 데이터는 **주문 수량이 음수**라는 특징이 있어, 데이터를 분할하는데 활용함

I 주문 특성 기준 군집화 - 특징 추출

1. 군집화 데이터를 유니크한 **고객 ID 컬럼만 있는 데이터프레임으로 초기화**
2. 주문/반품 횟수 계산 및 부착: 고객 ID와 주문 ID를 기준으로 **중복을 제거**하는 방식으로 유니크한 (고객 ID, 주문 ID)를 추출한 뒤, 추출한 데이터에서 **고객 ID의 수를 카운트**하여 군집화 데이터에 부착
3. 주문량 계산 및 부착: **고객 ID에 따른 주문량의 합계**를 계산하여 군집화 데이터에 부착
4. 주문 금액 계산 및 부착: 주문량과 단가를 곱하여 **주문 금액을 계산**한 뒤, 고객 ID에 따른 주문 금액의 합계를 계산하여 군집화 데이터에 부착
5. 최근성 계산 및 부착: **현재 날짜에서 주문 날짜의 차이**를 뺀 뒤, 고객별 해당 값의 최소 값을 군집화 데이터에 부착

I 주문 특성 기준 군집화 – 코사인 유사도 기반의 계층 군집화 수행

- 군집화 데이터에 (군집 개수 = 5, 군집 간 거리 = 평균, 거리 척도 = 코사인 유사도)를 갖는 군집화 모델을 학습하여, 군집별 주요 특성을 파악

	주문횟수	반품횟수	주문량	주문금액합계	최근성
주문특성_군집					
0	5.482738	1.075466	1502.736633	3073.161575	59.893064
1	3.370717	0.485981	1823.274143	1395.034455	69.040498
2	1.345733	0.188184	148.695842	272.678600	231.516411
3	1.086331	0.140288	59.593525	127.332122	272.694245
4	1.066667	0.266667	8.466667	27.213333	293.200000

I 주문 상품 기준 군집화

- 주문 횟수가 상위 100등안에 드는 상품들을 기준으로 **고객을 상품 구매 여부로 구성된 벡터로 표현**
- 이 데이터에 (군집 개수 = 5, 군집 간 거리 = 평균, 거리 척도 = 자카드 유사도)를 갖는 군집화 모델을 학습하여, 군집별 주요 특성을 파악함

Chapter. 13

직접 해봐야 내것이 된다

| 이탈 고객의 고객 여정 탐색하기

FAST CAMPUS
ONLINE
데이터 탐색과 전처리 I

강사. 안길승

I 문제 상황

- 고객 로그 데이터를 바탕으로 이탈한 고객과 이탈하지 않은 고객이 보이는 주요 행동 패턴을 탐색하자!
- (1) 이탈 고객과 비이탈 고객 분리
- (2) 이탈 고객과 비이탈 고객 데이터 내 주요 행동 패턴 추출
- (3) 주요 행동 패턴의 등장 비율 비교

Chapter. 13

직접 해봐야 내것이 된다

| 감사합니다

FAST CAMPUS
ONLINE
데이터 탐색과 전처리 I

강사. 안길승