

Instacart

데이터 분석



AI - 16 기
김정겸



Table of Contents

01

프로젝트 개요

02

프로젝트 구성
및 절차



03

결과

04

보완점



01

프로젝트 개요

Project Overview

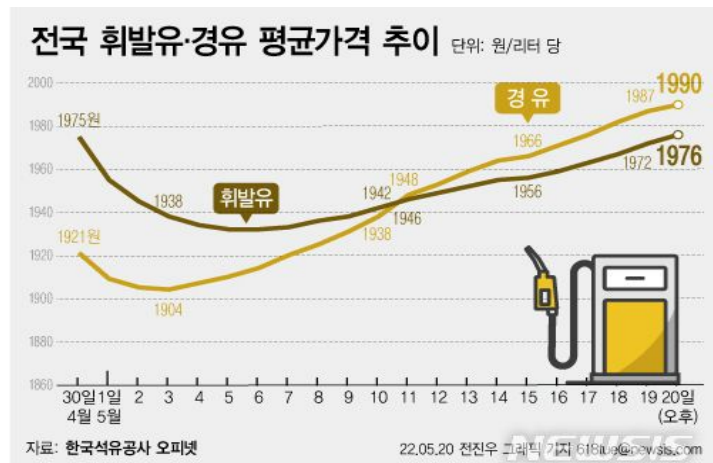


1. 프로젝트 개요

● 러우 전쟁 장기화

전쟁 장기화로 인한 연료비 상승과

그로 인한 냉방기구 사용의 거부감 확대



1. 프로젝트 개요

● 2018년도 폭염

무더위와 열대야로 인한 과한 냉방기구 사용 부담

➔ 2018년과 비슷한 소비 패턴 예측



1. 프로젝트 개요

- **Instacart** 데이터

과거 2018년도 데이터를 사용하여 매출 패턴 분석





02

프로젝트 구성 및 절차

Project Sequence

2. 프로젝트 구성

● EDA

- ▶ 데이터 병합 및 데이터 특성 분석
- ▶ 데이터 줄이기 (절반의 주문 기록 사용)

● 가설 설정 / 시각화

- ▶ 관련 뉴스에 따라 적절한 가설설정
- ▶ 가설에 맞는 결과의 시각화 그래프 제작



2. 프로젝트 절차

● EDA(데이터 구성)

- ▶ aisle (소 카테고리 이름)
- ▶ department (카테고리 이름)
- ▶ prior (주 데이터)
 - > 주문 크기 순 홀수 주문만 사용
- ▶ order (주문 데이터)
- ▶ product (상품 이름)



총 15개 데이터 특성

order_id
product_id
user_id
Order_hour_of_day
size
...



2. 프로젝트 절차

● **EDA**(기본적 데이터 특징 파악)

- ▶ 주문 크기 통계 (size)
- ▶ 요일별 주문량 추이 (size/day)
- ▶ 주문 시간대 통계 (size / hour)
- ▶ 주문 시간대별 상품 구성 (product / hour)
- ▶ 상품별 재주문 비율 (product / reordered)



2. 프로젝트 절차

● 가설 설정

- ▶ 폭염시 수박/닭고기 판매량 증가

h1 : 관심 상품군과 다른 상품군간의 연관 관계가 있다

- ▶ 더위에 배달주문 증가

h2 : 주문 횟수와 주문 크기에 상관관계가 있다

- ▶ 야간 시간의 현장 방문객 증가

h3 : 야간 주문 비율이 높은 상품군이 있다



2. 프로젝트 절차

● 가설 검증

h1: 해당 상품군과 다른 상품군간의 연관 관계가 있다

- 주문별로 각 카테고리의 상품 개수 count
- 판매량이 비슷하면 연관성이 높다고 가정하여 기준지표 설정

$$\text{기준지표 식} = \frac{\text{기준 상품 수}}{|\text{각 상품 수} - \text{기준 상품 수}| + \text{기준 상품 수}}$$

- 연관점수 평균과 총 주문량을 병합하여 그래프 제작



2. 프로젝트 절차

● 가설 검증

h2 : 주문 횟수와 주문 크기에 상관관계가 있다

- 사용자별로 주문 크기와 주문 횟수 count
- 사용자별 **주문 크기**(size) 평균과 **주문 횟수**(cnt) 데이터 제작
- 주문 크기 & 주문 횟수 데이터 시각화



2. 프로젝트 절차

● 가설 검증

h3 : 야간 주문 비율이 높은 상품군이 있다

- 주문 시간 기준으로 **오전(6~12), 오후(12~22) 야간(22~6)** 데이터 구분
- 소 카테고리별로 **야간 시간대 주문 비율** 확인
- 소 카테고리 와 야간 주문 비율 그래프 시각화



03

가설 검증 결과

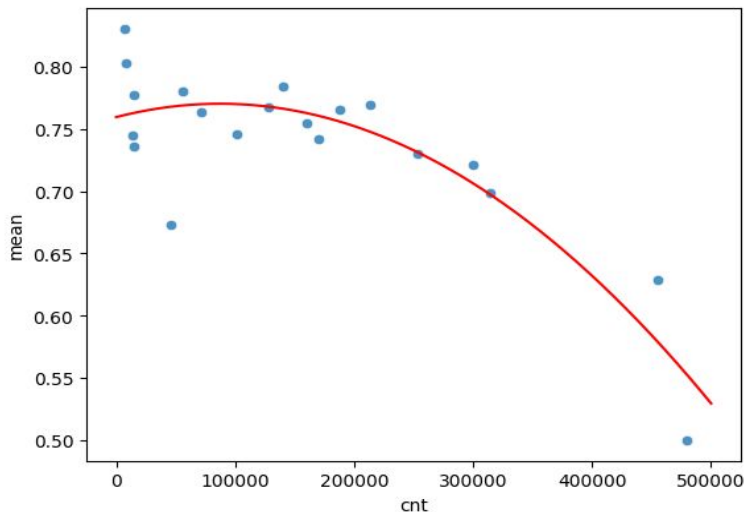
Results



3. 결과

● 시각화 결과

h1 : 관심 상품군과 다른 상품군간의 연관 관계가 있다



냉동 식품군과 다른 상품군과의 관계
(frozen)

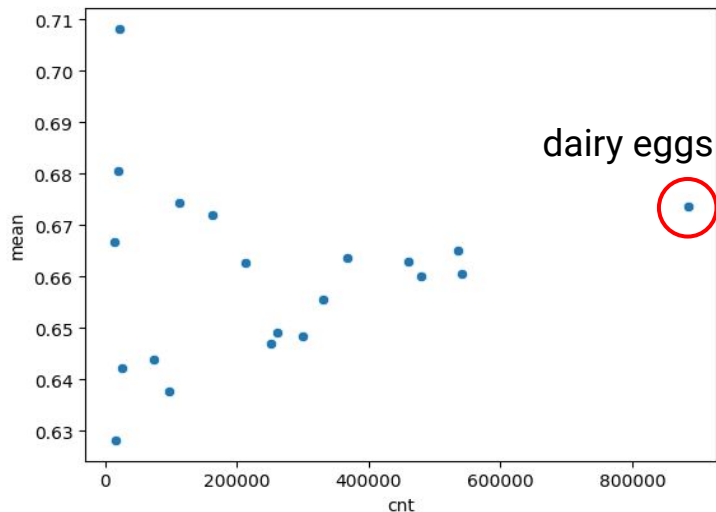


주문 크기가 커질수록 평균에 회귀할 뿐
뚜렷한 상관관계 X

3. 결과

● 시각화 결과

h1 : 관심 상품군과 다른 상품군간의 연관 관계가 있다



포장 식품군과 다른 상품군과의 관계
(produce)

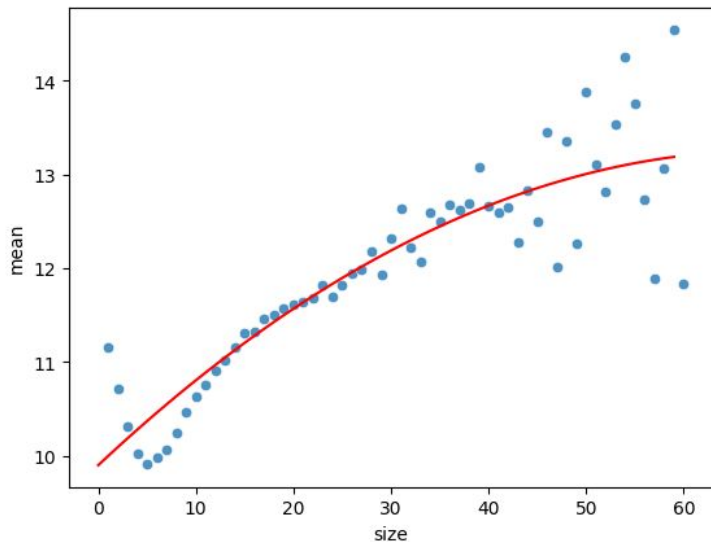
→ 다른 상품보다 상대적으로 **dairy eggs**와의
연관성이 높다

→ **produce**와 **dairy eggs**의 연계판매 추천

3. 결과

● 시각화 결과

h2 : 주문 횟수와 주문 크기에 상관관계가 있다



주문크기와 평균 방문 주기는
대체적으로 선형 관계를 가짐



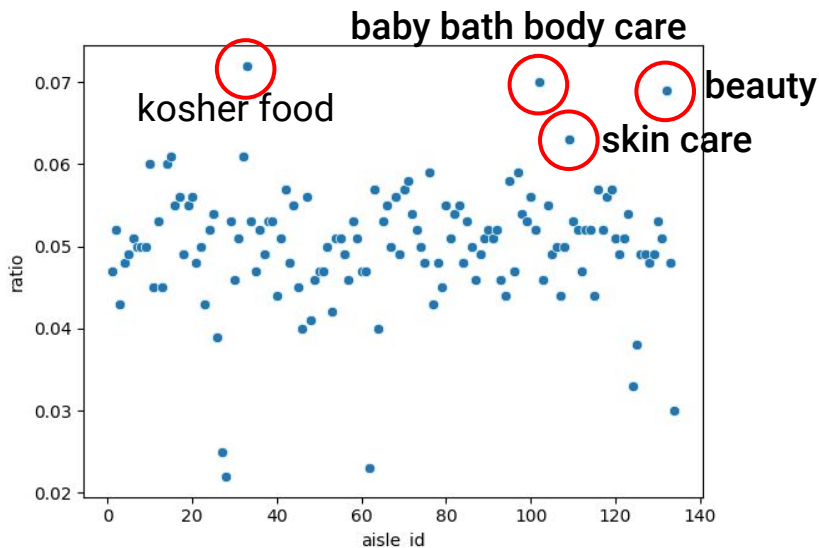
소규모 주문에 따른 상품구성을
분석하여 추천 상품 구성 필요



3. 결과

● 시각화 결과

h3 : 야간 주문 비율이 높은 상품군이 있다



→ **Baby bath body care와 Beauty, skin care 등 여성 제품의 비율 높음**

→ **해당 상품군의 현장 판매 추진 / 온라인 판매 비중 줄이기**



04

보완점

Plans to do

4. 보완점

- 연도 데이터 **X**

다른 연도의 같은 기간 데이터를 확보하여 **연도의 흐름에 따른** 판매 트렌드 분석 필요

- 온도 데이터 **X**

당시의 계절 및 온도 데이터를 확보하여 **날씨의 흐름에 따른** 판매 분석 필요

- 사용자 데이터 **X**

개인정보를 침해하지 않는 범위의 **사용자 데이터(ex. 성별)**를 추가하여 패턴 분석 필요



The image features a central teal-colored rectangular area. In the four corners of the image, outside this teal area, are fresh green leafy vegetables, likely lettuce or parsley, arranged in a symmetrical pattern. The word "The" is positioned above a horizontal white line, and the word "End" is positioned below it, both in a large, bold, white sans-serif font.

The

End