

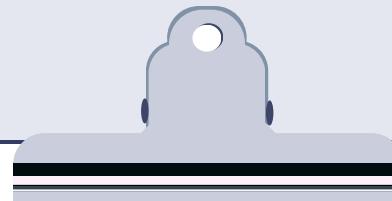


데이터 기반 유통 재고관리: 수요-공급 연계 예측 모델 보고서

DATA EATERS

김송현, 김수연, 유수연, 장수성





목차

01 분석 개요

02 기초 전처리

03 분석 모델링: 매출

04 분석 모델링: 매입

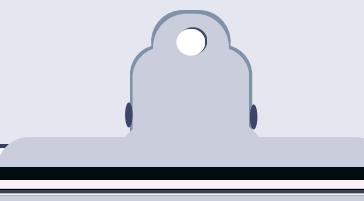
05 분석결과

06 결론 및 제안

07 부록

1. 분석 개요

01 분석 배경



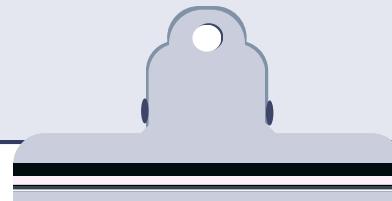
프로젝트 목적

- 정확한 매출 수요 예측을 통한 재고 유지 비용 최소화 및 결품률 감소
- 예측한 매출량 기반 매입 계획 수립으로 물류 효율성 극대화 도모



기대 효과

- **경제적 효과:** 정밀한 매입 계획 수립을 통해 재고 보유 비용을 절감하고, 재고 보관 공간 활용도를 향상시켜 운영 효율 향상
- **전략적 대응:** 계절적 변동이나 명절 등 특수 이벤트로 인한 수요 변화를 예측하여 선제적으로 재고를 확보, 품절 위험을 최소화하고 시장 상황에 신속하게 대응 가능



02 분석 개요

분석 목표

상품의 수요(매출량)와 공급(매입량)을 정밀하게 파악함으로써 재고 불일치 위험을 사전에 최소화하고, 이를 통해 효율적인 재고 운용 및 비용 절감을 실현

데이터 설명

도매 유통망의 역할을 수행하는 A, B 풀필먼트 센터의 데이터로, 250여곳 소매점을 대상으로 15,000여개의 상품의 매출/매입에 관한 정보를 포함

제공 데이터 기간

2021.01.01 ~ 2024.12.31 (4년)

예측 목표 (타겟 변수 Y)

2025년 1~3월의 매출량/매입량

예측 대상

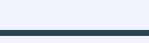
A 풀필먼트 센터 (중분류)



- 생수, 음료, 건강

- 신선식품

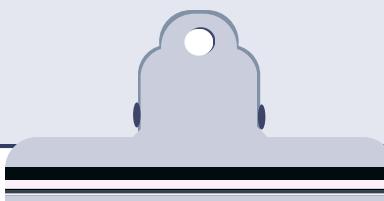
B 풀필먼트 센터 (소분류)



- 탄산음료

- 봉지라면

*편의성을 위해 이하 A 풀필먼트 센터를 'A 센터'로, B 풀필먼트 센터를 'B 센터'로 지칭함



03 분석 절차

데이터 확보

분석에 사용할 원천 데이터셋 및 분석 목적에 따른 외부 데이터를 확보한 후 데이터 품질 점검

데이터 전처리

예측 모델의 신뢰성 향상을 위해 결측치 및 이상치, 중복값 처리, 데이터 타입 통일 등 데이터 정제 작업 수행

변수 생성

예측 성능 향상에 기여할 수 있는 주요 요인을 가설적으로 도출하고, 이에 따라 변수를 설계 및 중요도 검증을 통해 최종 입력변수 선정

모델링 및 최종 모델 선정

데이터의 특성과 분석 목적을 고려하여 알고리즘을 비교·검토한 후 평가 지표를 사용하여 최적의 예측 모델을 선정하고, 카테고리별로 적합한 모델 설계

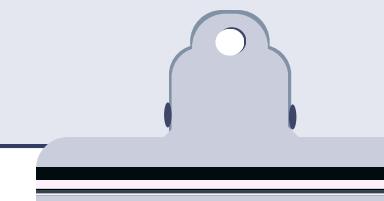
모델 성능 검증

모델의 일반화 성능 검증을 위해 홀드 아웃 백테스팅을 실시, 실제 값과 예측 값을 비교함으로써 모델의 현실적 예측 오차를 객관적으로 검증

결과 및 활용 방안 제시

최종 검증 모델로 카테고리별 예측을 실시하고, 결과를 바탕으로 실무적 활용 방안 및 모델의 한계점 제시

2. 기초 전처리



01 기초 전처리

데이터 통합 및 구조 정규화

- 데이터 컬럼 순서 정규화 후 연도별 데이터 통합
- '판매일' 컬럼 값 공백 제거 후 분석 목적으로 맞게 데이터 타입 변환(datetime)

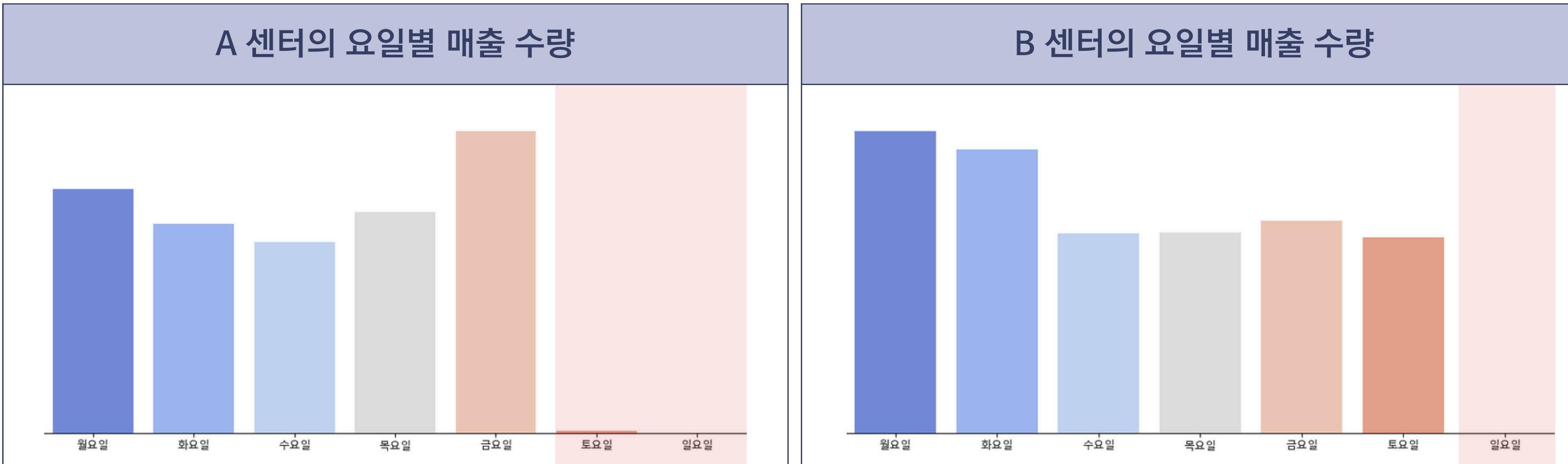
데이터 정제

- 모든 값이 동일한 중복 데이터 제거
- 문제 정의('매출/매입 수량' 예측)에 따라 '반출' 데이터는 제외하여, 순수 거래 이력 데이터 확보
- 공급금액이 0인 경우(전체 중 0.05% 미만) 이상치로 판단하여 제거
- 판매수량, 공급금액, 부가세가 음수인 데이터 제거

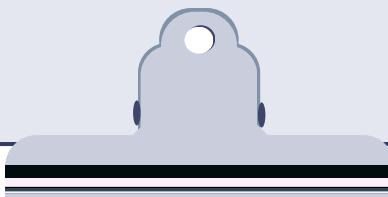
	판매일	바코드	상품명	판매수량	공급금액
0	2021-01-04	1.701005e+12	오츠카]데미소다<오렌지/250ml/캔>	1	11300
1	2021-01-04	1.701006e+12	네퓨어매실캔<175ml*30>XXX	1	6200
2	2021-01-04	1.701006e+12	코카]스프라이트<1.5>	1	17200
3	2021-01-04	1.701006e+12	코카]조지아<캔/오리지널/240ml>	1	11700
4	2021-01-04	1.701006e+12	코카]암바사<1.5L>	1	14700
5	2021-01-04	1.701006e+12	코카]환타<오렌지/1.5>	2	35600
6	2021-01-04	1.701006e+12	코카]조지아<캔/카페라테/240ml>	1	13600
7	2021-01-04	1.701006e+12	코카]환타<오렌지/600ml>	1	22800
8	2021-01-04	1.701006e+12	코카콜라<500ml>	7	209300
9	2021-01-04	1.701006e+12	코카]환타<파인/600ml>	1	23300
10	2021-01-04	1.701006e+12	코카콜라<250ml/캔>	14	277600
11	2021-01-04	1.701006e+12	코카콜라<1.5>	12	307700
12	2021-01-04	2.818830e+12	코카콜라병<355ml>	3	42900
13	2021-01-04	2.818830e+12	칠성사이다병<340ml>	5	67000

*데이터프레임 예시

02 거래 비발생일 데이터 제외

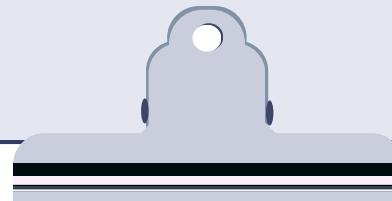


- A/B 센터 모두 본 프로젝트에서 정의한 데이터 활용 기간 내 일요일 매출량/매입량이 나타나지 않음
- A 센터의 토요일 매출량/매입량은 불규칙적으로 일부 발생함
- 이같은 데이터는 모델의 학습 시 노이즈가 될 가능성이 높아 학습 데이터에서 제거하며, 예측 시에도 해당 요일의 매출/매입 수량은 0(개)로 가정하고 제외함



03 카테고리 분류값 보완

문제	예시	해결																																																
<ul style="list-style-type: none"> 분석 대상인 카테고리(예: 중분류 - 생수, 음료, 건강) 분류 값이 결측치인 데이터 다수 	<table border="1"> <thead> <tr> <th>상품명</th> <th>대분류</th> <th>중분류</th> <th>소분류</th> <th>공급가액</th> <th>부가세</th> </tr> </thead> <tbody> <tr> <td>칠성]게토레이<600ml></td> <td>식품</td> <td>NaN</td> <td>NaN</td> <td>19800</td> <td>1980</td> </tr> </tbody> </table>	상품명	대분류	중분류	소분류	공급가액	부가세	칠성]게토레이<600ml>	식품	NaN	NaN	19800	1980	<ul style="list-style-type: none"> 거래량 예측 정확도 향상을 위해, 상품 목록 검토 후 내부 정의 기준에 따라 미분류 및 오분류 상품 재분류 																																				
상품명	대분류	중분류	소분류	공급가액	부가세																																													
칠성]게토레이<600ml>	식품	NaN	NaN	19800	1980																																													
<ul style="list-style-type: none"> ‘대분류’ 값 오류 	<table border="1"> <thead> <tr> <th>상품명</th> <th>대분류</th> <th>중분류</th> <th>소분류</th> </tr> </thead> <tbody> <tr> <td>건전지 벡셀 9V XXX</td> <td>None</td> <td>None</td> <td>None</td> </tr> <tr> <td>잡-주방]수세미<3M/철/35g></td> <td>None</td> <td>None</td> <td>None</td> </tr> <tr> <td>건전지 벡셀 9V XXX</td> <td>None</td> <td>None</td> <td>None</td> </tr> <tr> <td>사천 짜파게티멀티XXX</td> <td>휴지</td> <td>None</td> <td>None</td> </tr> </tbody> </table>	상품명	대분류	중분류	소분류	건전지 벡셀 9V XXX	None	None	None	잡-주방]수세미<3M/철/35g>	None	None	None	건전지 벡셀 9V XXX	None	None	None	사천 짜파게티멀티XXX	휴지	None	None	<ul style="list-style-type: none"> ‘대분류’가 결측치이거나 분류 기준이 부정확한 상품은 분석 대상에서 제외 																												
상품명	대분류	중분류	소분류																																															
건전지 벡셀 9V XXX	None	None	None																																															
잡-주방]수세미<3M/철/35g>	None	None	None																																															
건전지 벡셀 9V XXX	None	None	None																																															
사천 짜파게티멀티XXX	휴지	None	None																																															
<ul style="list-style-type: none"> 상품명 오탈자 및 분류 오류 	<p>'액티비아병복승아210ml' '잡곡]병아이콩<500g*2>면세'</p>	<ul style="list-style-type: none"> 상품명 확인 후, 내부 정의 기준에 따라 카테고리 분류값 입력 																																																
<ul style="list-style-type: none"> 동일 상품 수량 단위 미통일 → 별개 상품으로 취급 오류 → 판매수량 정확도 저해 우려 	<table border="1"> <thead> <tr> <th>판매일</th> <th>매출처</th> <th>우편번호</th> <th>매출처코드</th> <th>판매수량</th> <th>옵션 코드</th> <th>규격</th> <th>입수</th> </tr> </thead> <tbody> <tr> <td>809</td> <td>2021-01-04</td> <td>37793</td> <td>2144</td> <td>1</td> <td>EA</td> <td>None</td> <td>1</td> </tr> <tr> <td>873</td> <td>2021-01-04</td> <td>37823</td> <td>2156</td> <td>5</td> <td>EA</td> <td>1입</td> <td>1</td> </tr> <tr> <td>1488</td> <td>2021-01-04</td> <td>37632</td> <td>99</td> <td>1</td> <td>EA</td> <td>None</td> <td>1</td> </tr> <tr> <td>1556</td> <td>2021-01-04</td> <td>37886</td> <td>104</td> <td>1</td> <td>BX</td> <td>137*5*8</td> <td>40</td> </tr> <tr> <td>2370</td> <td>2021-01-04</td> <td>37836</td> <td>34</td> <td>11</td> <td>EA</td> <td>1입</td> <td>1</td> </tr> </tbody> </table>	판매일	매출처	우편번호	매출처코드	판매수량	옵션 코드	규격	입수	809	2021-01-04	37793	2144	1	EA	None	1	873	2021-01-04	37823	2156	5	EA	1입	1	1488	2021-01-04	37632	99	1	EA	None	1	1556	2021-01-04	37886	104	1	BX	137*5*8	40	2370	2021-01-04	37836	34	11	EA	1입	1	<ul style="list-style-type: none"> 예측 데이터에서도 동일한 오류 상품이 비슷한 비율로 존재할 것으로 가정, 추가 처리하지 않음
판매일	매출처	우편번호	매출처코드	판매수량	옵션 코드	규격	입수																																											
809	2021-01-04	37793	2144	1	EA	None	1																																											
873	2021-01-04	37823	2156	5	EA	1입	1																																											
1488	2021-01-04	37632	99	1	EA	None	1																																											
1556	2021-01-04	37886	104	1	BX	137*5*8	40																																											
2370	2021-01-04	37836	34	11	EA	1입	1																																											



04 거래 기록 전처리

1. 최근 1년 동안 판매되지 않거나 매입되지 않은 물품들은 각각 매출/매입 데이터에서 삭제

→ 장기간 거래되지 않은 상품을 학습/예측에서 제외함으로써 최근 거래량 예측 정확도 향상을 기대할 수 있음

2. 비정기적 특수 거래로 추정되는 거래 내역은 이상치로 간주하여 삭제

→ 아래 조건을 모두 만족하는 거래는 일반적인 흐름과 다른 특수 케이스로 판단되어 모델링 대상에서 제외하였음

[비정기적 특수 거래 조건]

거래 분류	조건	기대효과
즉시 출고 거래	입고 후 0~3일 이내 판매 발생	동일 수량이 입고 즉시 출고되는 특수 케이스 제거
이상치 거래	평균 판매량의 2배 이상 발생	급격한 이상치 제거로 예측 안정성 확보
대량 집중 거래	전체 판매량의 30% 이상 집중	특정 대량 납품 케이스 제거로 왜곡 방지

3. 분석 모델링: 매출

01 학습 데이터 준비: 변수 생성

- 카테고리 단위 일자별 매출량 집계 후 매출량 변동을 설명하는 데 기여할 가능성이 높은 요인들을 변수로 생성

시간 · 주기 변수

시계열 정보를 통해 장기적/주기적 매출 패턴 파악

- 연도
- 월
- 요일
- 월중분기 (1~10일 = 월초, 11~20일 = 월중순, 21~30일 = 월말)

외부 환경 변수

계절성 및 물가 등 외부 요인이 매출에 미치는 영향 반영

- 계절
- 월 평균 기온/강수량 (센터별 주요 매출 지역 기준)
- 월별 소비자물가상승률

이벤트 변수

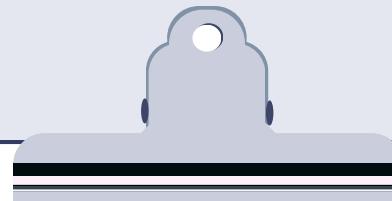
특수한 이벤트 전후의 수요 급변 현상 포착

- 공휴일, 명절, 롱위켄드, 대형마트 휴무 등 이벤트 해당 여부
- 명절연휴 전후 기간 해당 여부
- 대형마트 휴무 전 1~2일 해당 여부

과거 추세 변수

과거 판매 데이터 기반 단/장기적 추세 및 변동성 반영

- 직전 (1, 7, 28)일 판매량
- 직전 (7, 28)일간 이동평균 판매량
- 직전 (7, 28)일간 평균 대비 전일 판매량
- 전년도 동일 날짜 전후 3일 평균 판매량



02 학습 데이터 준비: 센터별 추가 전처리

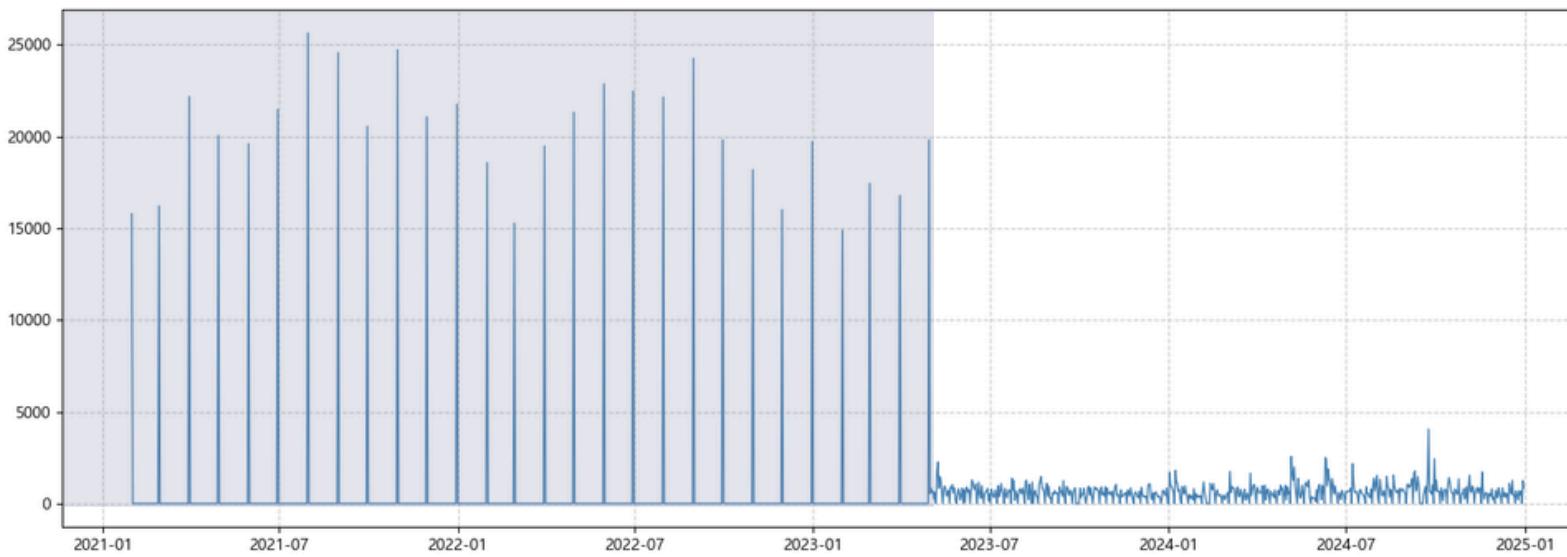
- 센터별 데이터 특성에 따라 모델 학습에 적합한 기간만을 선별하여 사용

A 센터

- ‘전년도 동일 날짜 전후 3일간 평균 매출량’ 변수를 산출할 수 없는 2021년 데이터는 분석 대상에서 제외함
- 최종 학습 데이터 구간: 2022-01-04 ~ 2024-12-31

B 센터

- 2023년 4월 이전 기간은 매월 말일에만 매출 데이터가 존재하여 일자별 추세와 변동성을 반영할 수 없으므로 제외함
- 최종 학습 데이터 구간: 2023-05-01 ~ 2024-12-31



03 머신러닝 모델링: 최종 학습 모델 도출

5-Fold Cross Validation

단일 split의 우연성에 의존하지 않고 모델의 일반화 성능을 더욱 안정적으로 추정하는 CV 결과를 통해 가장 우수한 모델 선택

Random Forest	Gradient Boosting	XGBoost	Light GBM
---------------	-------------------	---------	-----------

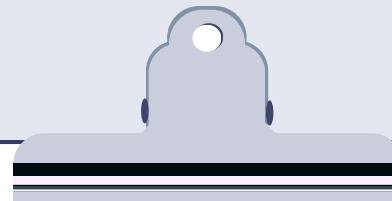


Grid Search CV

하이퍼 파라미터 조합 실험을 거쳐
모델의 성능을 최적화

월간 평가 지표가 가장 우수한 최종 모델 도출

	최적 모델	SMAPE(%)	R ²
생수·음료·건강	Random Forest	16.43	0.86
신선식품	Gradient Boosting	23.77	0.85
탄산음료	Random Forest	20.86	0.79
봉지라면	Random Forest	28.92	0.45



04 머신러닝 모델링: 미래 매출량 예측

- 상황 정의: 2025년 1~3월의 매출량을 2024년 말일 시점에 예측
→ 미래(2025년)에 대해서는 변수별로 전망 데이터 또는 예측 데이터를 사용해야 함

2025년 일자별 데이터를 생성하고 학습시와 동일한 변수 생성

과거 추세 이외 변수

- 연도, 월, 요일, 공휴일/명절 여부 등
- 날짜 정보만 있으면 미래에 대해서도 값 산출 가능

과거 추세 변수

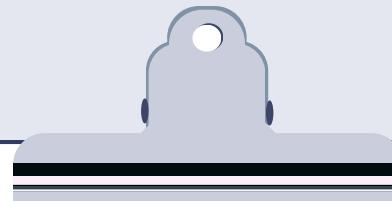
- 직전 7일 판매량 등
- *날마다 이전 일자들(=과거)의 매출량(예측값)이 필요

	판매일	직전 1일 판매량	판매수량
524	2025-02-03	519.866028	563.991175
525	2025-02-04	563.991175	594.653312
526	2025-02-05	594.653312	571.668979
527	2025-02-06	571.668979	515.849310
528	2025-02-07	515.849310	486.374452

*앞서 학습된 모델을 사용, 재귀적 예측 방식으로 일자별 매출량 예측값을 순차적으로 생성

미래 매출량 예측

4. 분석 모델링: 매입



01 학습 데이터 준비: 변수 구성

예상 매출량

미래 예상 수요를 반영하여
현재 매입량 결정의 핵심 근거로 활용
*3/31까지 매입량을 예측하기 위해
3/31 이후 일자들까지 매출량 예측 후 활용

과거 매입량

과거의 매입 주기 및 변동성을 파악하여
적정 매입량을 산출

주기성 계절성

장기적/주기적 매입 패턴과 더불어
특정 이벤트나 외부 환경 변화에 따른 변동성 포착

변수 구성

*직후 7/14/28일 간 판매량 누적합

최근 7/14/28일 간 매입량 누적합

최근 28일 매입량 표준편차

직후 14일 내 이벤트(명절/롱위켄드) 유무

이벤트까지 남은 일수(최대 14일)

날짜(연도/월/일/요일/월중분기)

계절

월 평균 기온

월 평균 강수량

02 학습 데이터 준비: 변수 생성

- 데이터 특성에 따라 데이터 사용 구간을 재정의하고, 절차에 따라 학습 데이터를 구축하였음

초기 구간 컷오프	
28일	학습 구간
과거 데이터 불충분 구간	
데이터 시작일 +28일	
구분	최종 사용 구간
A 센터	2021-02-10 ~ 2024-12-31
B 센터	2023-04-29 ~ 2024-12-31

일별 매입량 집계		
매입일	상품명	매입수량
2023. 4. 29.	A	10
2023. 4. 29.	B	40
2023. 4. 29.	C	30

매입일	매입수량
2023. 4. 29.	80
2023. 5. 1.	100
2023. 5. 2.	20

▶

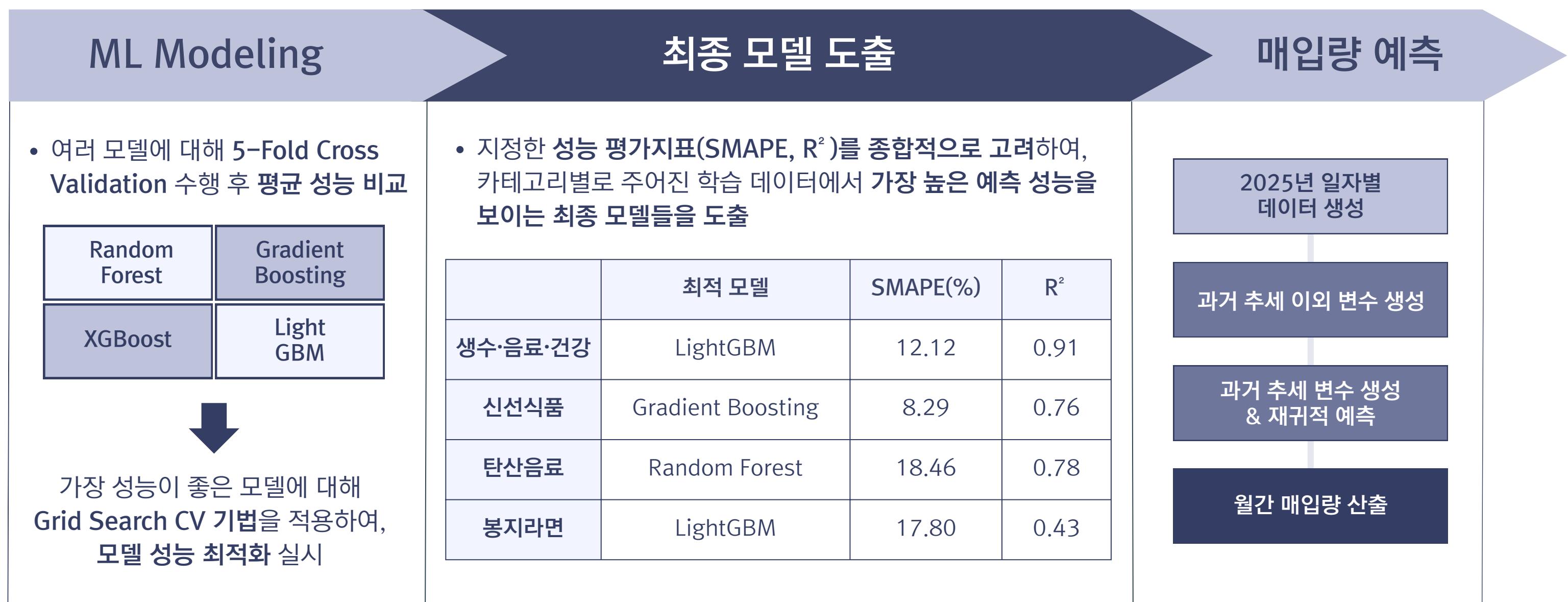
변수 생성				
매입일	직후7일_누적판매량	최근7일_누적매입량	계절	
2024-02-24	523	3702.0	겨울	
2024-02-26	504	5006.0	겨울	
2024-02-27	553	5382.0	겨울	
2024-02-28	740	5856.0	겨울	
2024-02-29	743	6256.0	겨울	
2024-03-01	816	2932.0	봄	
2024-03-02	839	2932.0	봄	
2024-03-04	885	2908.0	봄	
2024-03-05	946	2052.0	봄	
2024-03-06	986	1718.0	봄	

*데이터프레임 예시

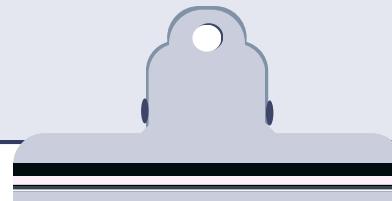
- A – 생수·음료·건강: 1,015행
- A – 신선식품: 1,015행
- B – 탄산음료: 525행
- B – 봉지라면: 525행

03 최종 학습 모델 도출 및 미래 매입량 예측

- 교차검증 및 성능 평가를 통해 최적 모델을 도출하고, 재귀적 예측 방식으로 미래 매입량을 산출하도록 함



5. 분석 결과



01 모델 평가 지표 및 변수 중요도 도출 기법 선정 이유

1. 예측 정확도 평가: SMAPE (Symmetric Mean Absolute Percentage Error)

분석 시 실제 거래량이 발생하지 않는(값이 0인) 날짜의 데이터가 포함되어, 단순 MAPE(Mean Absolute Percentage Error) 사용 시에는 0분모로 인한 왜곡 문제가 발생할 수 있음 SMAPE는 과대예측과 과소예측에 대칭적인 패널티를 부여하여, 이같은 문제를 보완하기 때문에 예측값 변동성이 높은 데이터 환경에서도 안정적인 오차율을 제공함

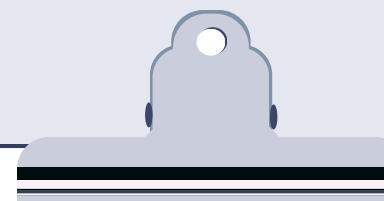
2. 모델 적합도 평가: R-squared (R^2)

R^2 는 모델이 타겟 변수(Y)의 전체 분산을 얼마나 설명 가능한지 0~1 사이 값으로 직관적으로 파악할 수 있는 지표로서, 회귀모델의 기본적인 설명력을 평가하기 위한 보조지표로 사용하였음

3. 피처 중요도 해석: Permutation Importance

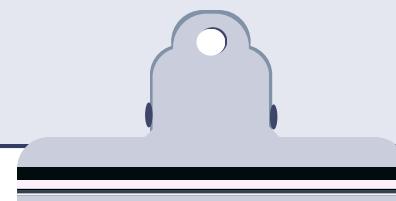
모델 해석의 투명성을 확보하고 핵심 변수를 도출하기 위해 Permutation Importance를 사용하였음. 이 분석 기법은 모델 구조에 관계없이(Model-Agnostic) 특정 변수의 값을 무작위로 섞었을 때 모델 성능이 감소하는 정도를 측정하여 중요도를 산출함. 단순 Impurity 기반 변수 중요도 도출 기법에 비해, 예측에 대한 각 변수의 실질적 기여도에 대한 직관적이고 신뢰성 있는 해석이 가능함

02 최종 모델 성능 평가



센터 구분	분류	매출/매입	최적 모델	SMAPE (주요 지표)	R ² (보조 지표)	해석
A 센터	생수, 음료, 건강	매출	Random Forest	16.43	0.86	<ul style="list-style-type: none"> 매입 예측 오차율이 매출보다 약 4%p 낮아 매입 통제력이 더 높음 모델이 데이터 변동성의 86% 이상을 설명(R^2)하여 적합도가 가장 높음
		매입	LightGBM	12.12	0.91	
	신선식품	매출	Gradient Boosting	23.77	0.85	<ul style="list-style-type: none"> 매출 예측에 대한 변동성은 크나, 매입은 비교적 계획적으로 잘 예측됨 매입 예측 오차율이 가장 낮아 매우 안정적인 예측을 보여줌 상대적으로 매입에 비해 매출의 변동성을 더 잘 설명함
		매입		8.29	0.76	
B 센터	탄산음료	매출	Random Forest	20.86	0.79	<ul style="list-style-type: none"> 매출은 변동성을 잘 설명하나, 매입의 경우는 비교적 낮음 변동성 포착은 약간 미흡할 수 있으나 예측 오차는 작음
		매입		18.46	0.78	
	봉지라면	매출	Random Forest	28.92	0.45	<ul style="list-style-type: none"> 상대적으로 수요 변동성이 크거나 외부학습이 조금 어려울 수 있음 변동성의 설명력이 상대적으로 낮은 편 하지만 SMAPE, R² 모두 실무 상 신뢰가능한 수준
		매입	LightGBM	17.80	0.43	

03 예측 결과: 생수·음료·건강



매출

예측 매출량

- 2025년 1월: 31,965개
- 2025년 2월: 31,008개
- 2025년 3월: 32,761개

주요 결과 해석

- 월 평균 기온의 변수 중요도가 압도적으로 높은 것으로 보아, 생수·음료·건강 종류에 해당되는 제품들은 날씨 또는 기온이 가장 강력한 핵심요인임을 시사함 (대체로 기온이 높은 4~9월에 판매량 증가)
- 공휴일이나 수요일에 특별한 소비 패턴이 발생하는 것일 수 있음
- 장기 추세(28일 이동 평균 판매량)뿐만 아니라 최근 일주일간의 수요 변화 또한 민감하게 반영하고 있음

순열 변수 중요도 (Permutation Importance)	
변수명	중요도 (SMAPE 기준)
경북_월평균기온(°C)	12.82
요일	7.46
공휴일	5.22
28일 이동평균 판매량	4.76
7일 이동평균 판매량	4.04

매입

예측 매입량

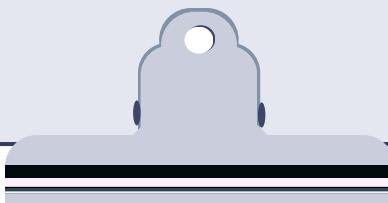
- 2025년 1월: 45,104개
- 2025년 2월: 32,648개
- 2025년 3월: 35,617개

주요 결과 해석

- 매입 계획이 해마다의 장기적인 성장 추세에 가장 크게 의존하고 있음
- 매출과 마찬가지로 단기, 장기적인 매입 패턴 뿐만 아니라 2주 단위 패턴 또한 존재하며, 매입량에 영향을 많이 미치는 요인이 되는 것으로 보임
- 상대적으로 중요도가 낮으나, 월 평균 기온 또한 예상되는 외부 수요인 월 평균 기온을 선제적으로 반영하고 있음
- 즉, 예측된 수요를 바탕으로 매입이 이루어지는 선행 학습 특성을 보임

순열 변수 중요도 (Permutation Importance)	
변수명	중요도 (SMAPE 기준)
연도	5.68
최근14일_누적매입량	4.34
최근28일_누적매입량	3.16
경북_월평균기온(°C)	2.24
최근7일_누적매입량	2.03

* 모든 예측은 2024년 11월까지의 데이터를 학습하고 2024년 12월 매출량을 예측하는 백테스팅을 통해 최종 모델의 성능을 재검증하였으며, 이후 2025년도 예측 모델링을 진행하였음



04 예측 결과: 신선식품

매출

예측 매출량

- 2025년 1월: 3,570개
- 2025년 2월: 3,461개
- 2025년 3월: 3,564개

주요 결과 해석

- 주간 스케줄(요일)과 장기적인 트렌드(연도)가 예측의 성패를 결정하는 가장 큰 요인
- 특히 화, 수, 목요일은 소매점 주문이나 물류 배송 주기가 포함되어 있는 등 주간 수요 패턴이 중요한 관계성이 내포되어 있는 것으로 추측
- 월 평균기온도 중요하지만 주간 패턴에 비하면 영향력이 조금 약한 편

순열 변수 중요도 (Permutation Importance)	
변수명	중요도 (SMAPE 기준)
요일	18.04
연도	11.48
28일 이동평균 판매량	8.04
직전 7일 판매량	4.13
경북월평균기온(°C)	3.42

매입

예측 매입량

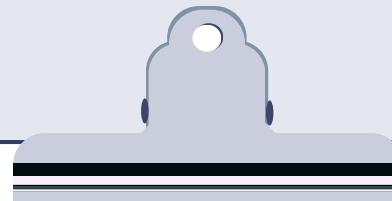
- 2025년 1월: 36,622개
- 2025년 2월: 18,794개
- 2025년 3월: 22,419개

주요 결과 해석

- 매입량 예측은 특히 과거 장기 매입 수준에 가장 크게 의존
- 미래의 예측 수요(직후7일 판매량)도 중요하게 반영하고 있음
- 과거 데이터를 통한 안정성과 미래 예측을 위한 정확성을 모두 고려하는 균형 잡힌 매입 계획 모델임을 보여줌
- '최근28일 누적 매입 표준편차'의 높은 중요도는 모델이 매입량의 변동성을 예측 과정에 반영하고 있어 안정성 향상에 기여

순열 변수 중요도 (Permutation Importance)	
변수명	중요도 (SMAPE 기준)
최근28일_누적매입량	3.33
최근7일_누적매입량	2.99
최근28일누적매입표준편차	2.43
요일	2.24
직후7일_누적판매량	1.87

* 모든 예측은 2024년 11월까지의 데이터를 학습하고 2024년 12월 매출량을 예측하는 백테스팅을 통해 최종 모델의 성능을 재검증하였으며, 이후 2025년도 예측 모델링을 진행하였음



05 예측 결과: 탄산음료

매출

예측 매출량

- 2025년 1월: 11,300개
- 2025년 2월: 12,155개
- 2025년 3월: 12,915개

주요 결과 해석

- 어제 판매량과 오늘 요일의 정보를 얼마나 정확하고 빠르게 반영하는지에 따라 예측의 정확도를 향상시킬 수 있음
- 직전 1일 판매량과 요일 등 초단기적/구조적 요인에 가장 크게 의존하여 재고 관리나 발주 계획이 데일리 또는 주간 단위로 매우 빠르게 이루어져야 함을 시사
- 중요도가 7일, 14일, 28일 순으로 점차 낮아지므로, 장기 추세보다는 최근 몇 주 간의 판매 흐름에 더 큰 영향을 받음을 의미
- 빠른 반응 속도가 중요한 상품군

순열 변수 중요도 (Permutation Importance)	
변수명	중요도 (SMAPE 기준)
직전 1일 판매량	8.46
요일	8.41
7일 이동평균 판매량	4.95
직전 14일 판매량	3.82
28일 이동평균 판매량	3.70

매입

예측 매입량

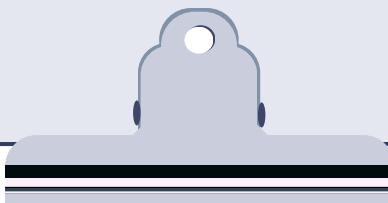
- 2025년 1월: 15,153개
- 2025년 2월: 13,568개
- 2025년 3월: 15,679개

주요 결과 해석

- '요일' 변수를 중심으로 단기적이고 구조화된 주간 수요 패턴을 가장 중요하게 학습하고 예측 안정성을 확보하고 있음
- '연도'나 명절 관련 변수의 중요도가 음수이므로, 해당 상품군이 장기적 추세가 없거나 주요 이벤트에 큰 영향을 받지 않을 수 있음
- 시기에 따라 매입 패턴이 확연히 바뀐 B센터의 특성상, A센터에 비해 학습 기간이 짧게 슬라이싱되어 장기적 패턴을 포착하기 어려운 문제 또한 존재

순열 변수 중요도 (Permutation Importance)	
변수명	중요도 (SMAPE 기준)
요일	2.33
일	0.65
명절_14일내_있음	-0.02
연도	-0.04
명절_남은일수_최대14일	-0.06

* 모든 예측은 2024년 11월까지의 데이터를 학습하고 2024년 12월 매출량을 예측하는 백테스팅을 통해 최종 모델의 성능을 재검증하였으며, 이후 2025년도 예측 모델링을 진행하였음



06 예측 결과: 봄지라면

매출

예측 매출량

- 2025년 1월: 2,637개
- 2025년 2월: 2,734개
- 2025년 3월: 2,680개

주요 결과 해석

- '요일' 정보를 중심으로 예측의 안정성을 확보하고, 특히 토요일 판매 패턴에 가장 크게 의존하고 있음.
- 이는 주말 판매량이 높거나, 소매점의 주말 주문 패턴이 주중과 완전히 달라 예측 시 토요일을 가장 중요한 특성으로 취급해야 함을 나타냄
- '평균 대비 전일 판매량' 변수들이 상위권에 위치하여 평소 대비 최근 판매량의 변동성을 민감하게 포착하고 있음
- 예측이 장기적인 추세(14일, 28일 변수)보다는 주간 구조(요일) 및 단기 변동성에 더 크게 의존하고 있음

순열 변수 중요도 (Permutation Importance)	
변수명	중요도 (SMAPE 기준)
요일	7.44
28일 평균 대비 전일 판매량	1.62
직전 14일 판매량	1.57
공휴일	1.31
7일 평균 대비 전일 판매량	1.23

매입

예측 매입량

- 2025년 1월: 24,553개
- 2025년 2월: 21,510개
- 2025년 3월: 22,578개

주요 결과 해석

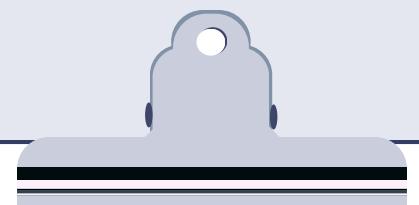
- 과거의 공급 이력('최근 14일 매입량')과 미래의 예측 수요('직후 7일/28일 판매량')를 모두 높은 중요도로 활용 → 안정성과 수요 민감도 동시 반영
- '요일'과 '일' 변수가 상위권에 위치하므로, 매입 결정이 주간 및 월간의 고정된 물류 및 발주 주기에 크게 의존하고 있는 것으로 해석 가능
- 장기적인 예측보다는 단기(1~2주) 수요 변화에 더 민감하게 반응하며, 매입 예측의 유연성을 높이고 있음을 시사함

순열 변수 중요도 (Permutation Importance)	
변수명	중요도 (SMAPE 기준)
요일	3.72
직후 7일 누적 판매량	3.02
최근 14일 누적 매입량	3.01
일	2.97
직후 28일 누적 판매량	2.63

* 모든 예측은 2024년 11월까지의 데이터를 학습하고 2024년 12월 매출량을 예측하는 백테스팅을 통해 최종 모델의 성능을 재검증하였으며, 이후 2025년도 예측 모델링을 진행하였음

6. 결론 및 제안

01 인사이트 및 액션 제시



생수, 음료, 건강

매출

- 기온이 높은 4~9월 판매량 증가
- 특정 요일/공휴일 소비 패턴 존재
- 장기·단기 수요 모두 모델에 반영됨

매입

- 장기 성장 추세가 가장 큰 영향
- 최근 수요인 단기·중기 패턴(7·14·28일 누적 매입량)의 영향을 받음
- 기온도 일부 영향을 받는 것은 예측 수요가 선제적으로 반영된 것으로 보임

액션 제시

- 여름철 기온 상승에 맞춘 재고 선제 확보
- 단/장기 수요 변동에 맞춘 발주주기 조정
- 요일·공휴일 소비 패턴 및 원인 파악

신선식품

매출

- 주간 스케줄(요일)과 장기 트렌드(연도)가 가장 큰 영향 → 특히 화·수·목요일 매출량 집중
- 월평균 기온도 일부 영향이 있으나 주간 패턴이 더 중요
- 단기·장기 추세가 고루 모델에 반영됨

매입

- 과거 장기 매입 수준에 가장 크게 의존 → 안정적 매입 계획 수립
- 미래 예측 수요 또한 반영되어 선제적 발주 가능
- 누적 매입 표준편차를 통해 변동성 고려

액션 제시

- 주간 수요 패턴 기반 선제적 발주
- 장기·단기 추세 반영하여 재고 수준 조정
- 매입 변동성 관리로 과잉/부족 재고 방지

탄산음료

매출

- 단기·주간 수요 패턴이 가장 중요
- 연도·명절 변수 중요도 낮음 → 장기적 트렌드나 이벤트 영향 적음
- 빠른 반응 속도가 중요한 카테고리므로 매출 변화가 즉각적

매입

- 매입 또한 단기 수요 패턴 중심
- 장기적 패턴 포착 어려움 → 학습 기간이 짧아 반영된 B센터 특성

액션 제시

- 장기 추세보다 최근 판매 흐름 반영
- 재고 관리 및 발주 계획을 데일리/주간 단위로 신속히 수립
- 빠른 회전율 상품군에 맞춰 유연한 운영

봉지라면

매출

- 요일 중 특히 토요일 판매 패턴이 가장 큰 영향 → 주말 수요 패턴 존재 시사
- 단기 변동성(평균 대비 전일 판매량) 중요 → 최근 판매 흐름 민감하게 반영
- 장기 추세보다 단기·주간 구조에 의존

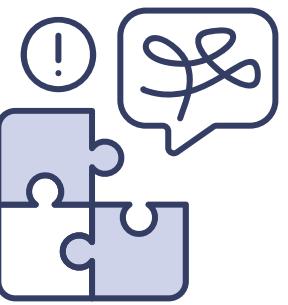
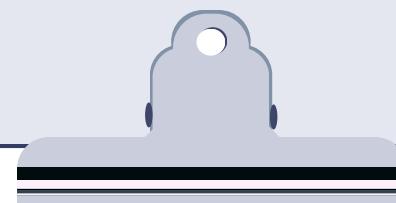
매입

- 과거 매입과 미래 예측 수요 모두 반영하여 안정성과 수요 민감도 동시에 고려
- 매입 결정이 주간·월간 고정 발주 주기에 크게 의존
- 단기(1~2주) 수요 변화에 민감

액션 제시

- 토요일 등 요일별 판매 패턴 파악 및 반영
- 최근 매입 이력과 미래 수요 예측 활용
- 주간·월간 물류 및 발주 주기 최적화

02 한계점

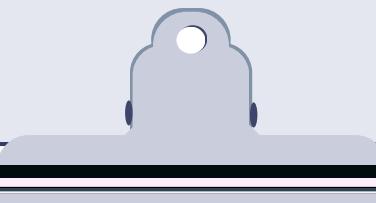


한계

- 카테고리 기준 학습으로 상품별 특성을 정밀하게 반영한 예측은 불가
- 현재 데이터에 포함된 기간의 특수성을 반영한 모델이므로, 또다른 미래 특정 기간에 적용할 경우 해당 기간의 특수성에 대한 고려가 다시 필요함

보완 방안

- 적절한 데이터셋이 확보 되고 관리/운영 비용이 충족된 경우, 상품별 모델을 개발한다면 상품별 특성까지 더욱 잘 반영할 수 있을 것임
- 제품별 재고현황 및 리드타임 정보가 있다면 모델 성능 향상에 도움



감사합니다

THANK YOU

7. 부록

01 부록: 외부 데이터 및 분석 도구

1. 활용 외부 데이터 및 출처

활용 데이터	출처
기온 및 강수량	기상청 기상자료개방포털 https://data.kma.go.kr/stcs/grnd/grndRnList.do?pgmNo=69
소비자 물가 상승률	국가통계포털 https://kosis.kr/statHtml/statHtml.do?sso=ok&returnurl=https%3A%2F%2Fkosis.kr%3A443%2FstatHtml%2FstatHtml.do%3Fconn_path%3Di3%26tblId%3DDT_1J22003%26orgId%3D101%26

* 국가정보자원관리원 화재로 인해 미래 매출량/매입량 예측 과정에서 사용할 날씨 예보, 전망 데이터를 구할 수 없게 되어 부득이 실측값으로 대체하여 사용함

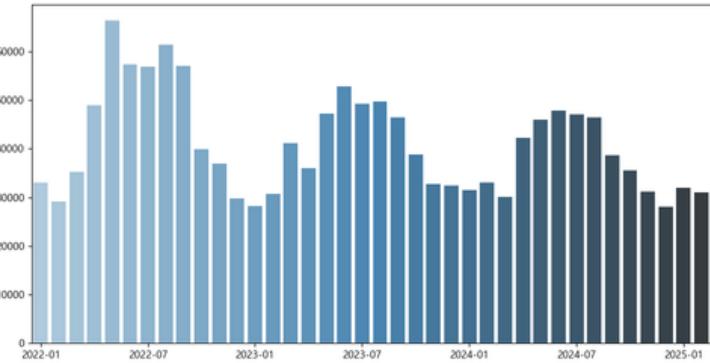
2. 분석 도구



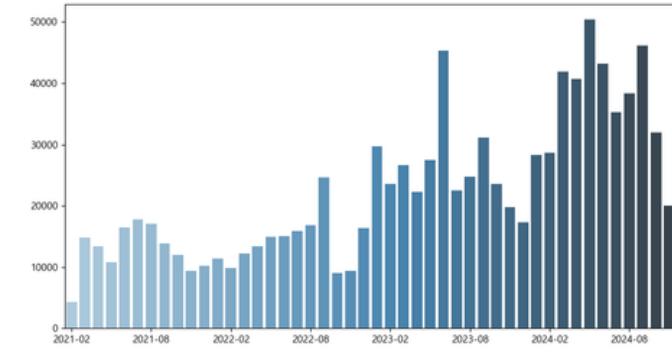
02 부록: 매출량/매입량 총 추이(~25년 3월) 그래프

1. 생수, 음료, 건강

매출

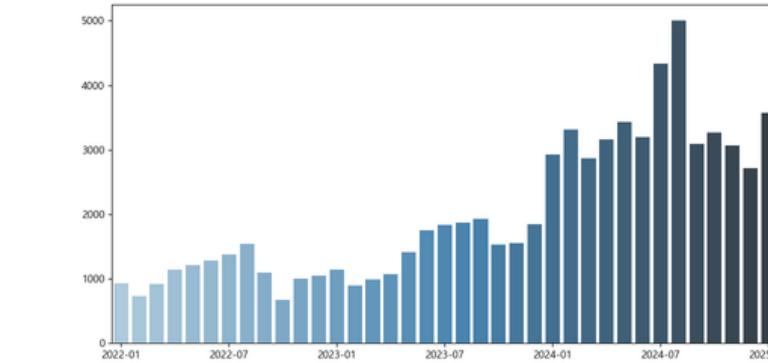


매입

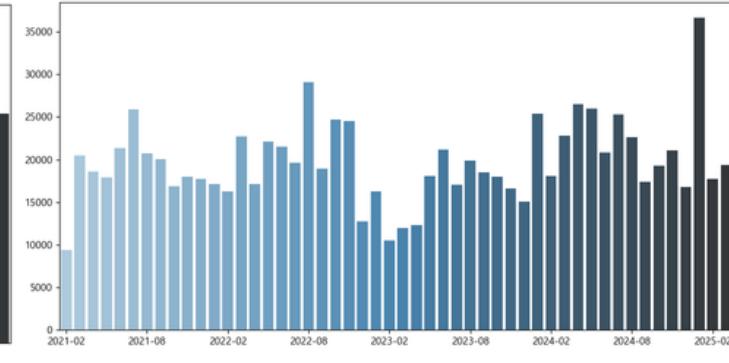


2. 신선식품

매출

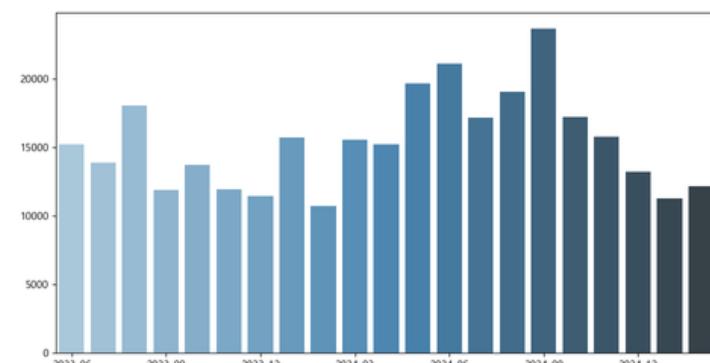


매입

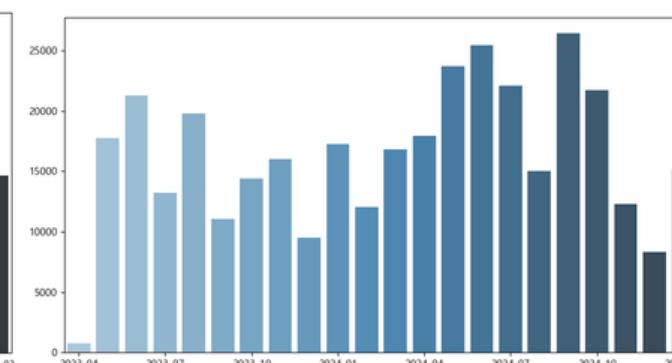


3. 탄산음료

매출



매입



4. 봉지라면

매출

