

항공화물 수요예측 프로젝트

데이터 분석 수련생

박찬성 이재은

목차

서론

- 프로젝트 배경
- 분석 프로세스

본론

- 데이터 수집
- 데이터 전처리
- 분석 모델링

결론

- 분석 결과
- 향후 과제

서론

- 프로젝트 배경
- 분석 프로세스

1

항공산업의 발전에 따른 **항공화물 수요 증가**

2

여객 중심, 장기적 관점 등 **기존 연구의 한계**

3

화물, 공항 중심의 수요 예측 필요

1

항공산업의 발전에 따른 항공화물 수요 증가

월 단위 항공화물 수요예측

2

여객 중심, 장기적 관점 등 기존 연구의 한계

→ 항공산업의 이윤과 효율성 극대화

3

화물, 공항 중심의 수요예측 필요

STEP 1 데이터 수집

- 항공화물 실적
- 공휴일 정보

STEP 2 데이터 전처리

- 파생 변수 생성
- 주 단위 데이터 생성
- 일 단위 데이터 생성

STEP 3 분석 모델링

- 주 단위 ARIMA
- 주 단위 Auto ML
- 일 단위 Auto ML

STEP 4 결과 도출

- 분석 결과 가시화
- 향후 과제 제시

본론

- 데이터 수집 및 전처리

- 분석 모델링

✓ 항공화물 실적

한국공항공사 내부 데이터

김포공항 내 **일일 항공 화물량** (2000.01 - 2022.09)

→ '운항 일자', '운항횟수', '화물 출발/도착' 추출

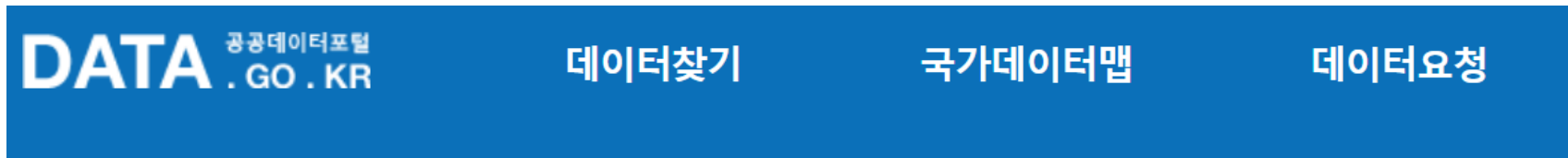
→ '노선 수' 파생 변수 생성

	화물	운항횟수	노선수
일자			
20000101	125649.0	390.0	27
20000102	316415.0	395.0	28
20000103	645382.0	379.0	28
20000104	748145.0	371.0	25
20000105	797219.0	352.0	26
...
20220926	206704.0	389.0	28
20220927	197240.0	375.0	31
20220928	202744.0	376.0	31
20220929	196823.0	376.0	30
20220930	205855.0	431.0	34

[일자 별 화물 데이터 집계]

✓ 공휴일 정보

한국천문연구원 API(공공데이터포털)*



오픈API 상세

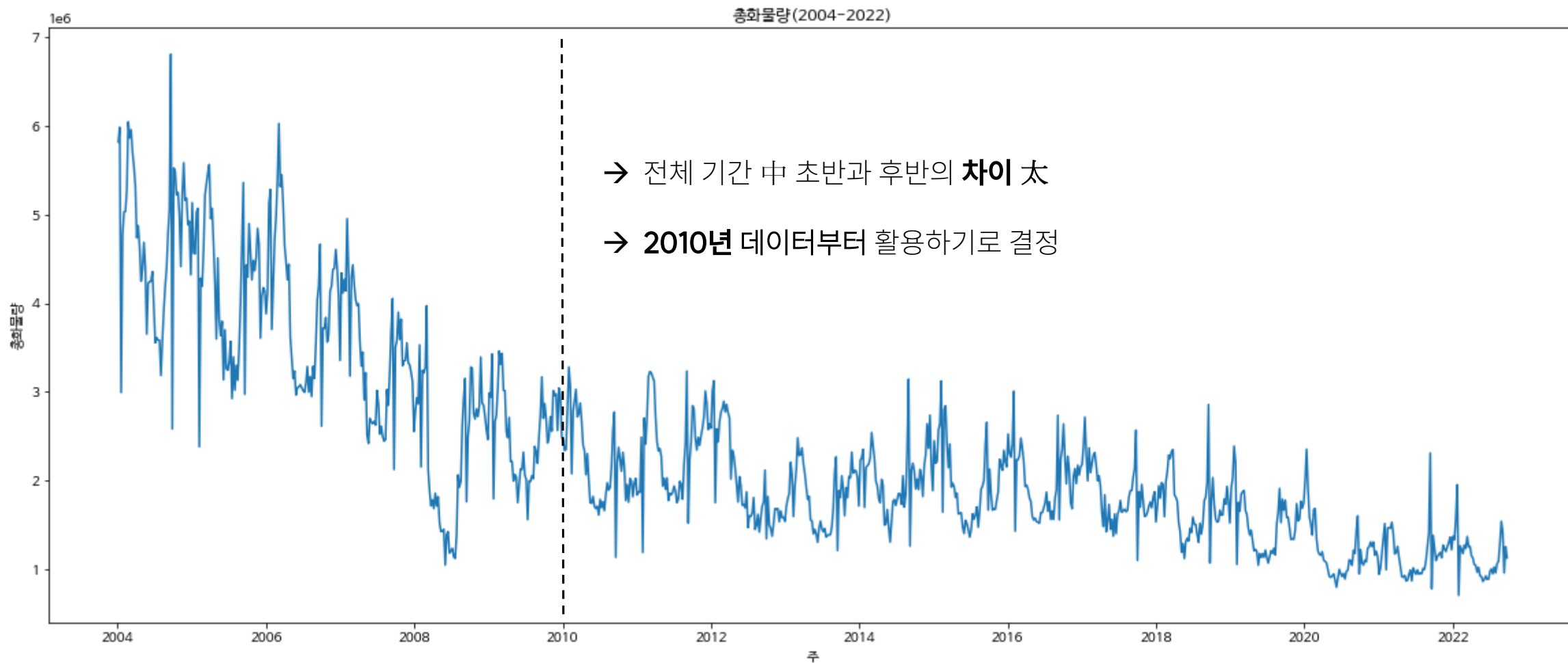
XML 한국천문연구원_특일 정보

(천문우주정보)국경일정보, 공휴일정보, 기념일정보, 24절기정보, 잡절정보를 조회하는 서비스 입니다.
활용시 날짜, 순번, 특일정보의 분류, 공공기관 휴일 여부, 명칭을 확인할 수 있습니다.

	locdate	dateName
0	20040101	신정
1	20040121	설날
2	20040122	설날
3	20040123	설날
4	20040301	삼일절
...
297	20220815	광복절
298	20220909	추석
299	20220910	추석
300	20220911	추석
301	20220912	대체공휴일

[공휴일 정보 데이터]

✓ 총화물량 그래프 (2004.01 - 2022.09)



✓ 공휴일 명칭 통일

동일한 공휴일이나, 이름이 다른 경우 존재

ex) 신정, 1월 1일 → '신정'으로 통일

설날	40
추석	38
삼일절	13
현충일	13
광복절	13
어린이날	12
개천절	12
기독탄신일	12
대체공휴일	10
신정	8
석가탄신일	8
한글날	8
부처님오신날	5
1월1일	5
대통령선거일	3
임시공휴일	2
대체휴무일	2
국회의원선거일	2
전국동시지방선거	2
동시지방선거일	1
어린이날	1
제21대 국회의원선거	1



설날	40
추석	38
신정	13
삼일절	13
어린이날	13
석가탄신일	13
현충일	13
광복절	13
기독탄신일	12
개천절	12
대체공휴일	10
한글날	8
국회의원선거	3
대통령선거	3
전국동시지방선거	3
임시공휴일	2
대체휴무일	2

[수정 전]

[수정 후]

✓ 일자별 요일 추출

주말 / 주말제외휴일 / 주말동시휴일 파생 변수 추가 생성

일자	요일	월요일	화요일	수요일	목요일	금요일	토요일	일요일	주말	주말제외휴일	주말동시휴일
0 20100101	4	0	0	0	0	1	0	0	0	1	0
1 20100102	5	0	0	0	0	0	1	0	1	0	0
2 20100103	6	0	0	0	0	0	0	1	1	0	0
3 20100104	0	1	0	0	0	0	0	0	0	0	0
4 20100105	1	0	1	0	0	0	0	0	0	0	0

[일자 별 요일 및 주말 관련 컬럼 생성]

✓ 공휴일 관련 변수

공휴일 별 당일 / 1주 전후 여부 파생 변수 생성

	일자	총화물량	운항횟수	노선수	공휴일여부	주말제외휴일	주말동시휴일	개천절	광복절	국회의원선거	...	주말동시휴일1주전	주말동시휴일1주후	개천절1주전	개천절1주후	광복절1주전	광복절1주후	국회의원선거1주전	국회의원선거1주후	기독탄신일1주전	기독탄신일1주후
51	2010-12-26	2023616.0	1890.0	135	1.0	0	1	0	0	0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
104	2012-01-01	2590271.0	1964.0	142	1.0	0	1	0	0	0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
156	2012-12-30	1538186.0	2020.0	146	1.0	1	0	0	0	0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
208	2013-12-29	1712237.0	1999.0	150	1.0	1	0	0	0	0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
260	2014-12-28	1885667.0	2273.0	124	1.0	1	0	0	0	0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

[공휴일 별 컬럼 생성 후 '기독탄신일 1주 후'인 주 출력]

✓ 일 단위 → 주 단위 변환

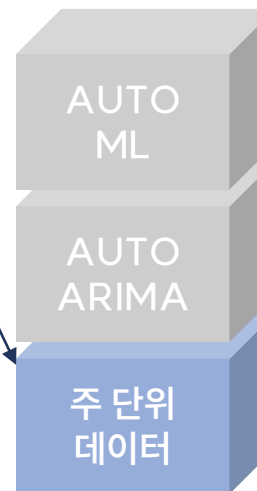
일자	총화물 량	운항 횟수	노선 수
2010-01-03	347016.0	298.0	20
2010-01-04	123718.0	78.0	21
2010-01-05	401365.0	278.0	18
2010-01-06	481274.0	280.0	19
2010-01-07	414219.0	224.0	19
...

[일 단위]



일자	총화물량	운항 횟수	노선 수
2010-01-03	2453515.0	1748.0	133
2010-01-10	2338951.0	1822.0	132
2010-01-17	2350311.0	1830.0	133
2010-01-24	2820431.0	1996.0	127
2010-01-31	3280586.0	1969.0	128
...

[주 단위]



✓ 주 단위 ARIMA 예측 값 추가

	총화물량	운항 횟수	
일자			
2010-01-03	2453515.0	1748.0	133
2010-01-10	2338951.0	1822.0	132
2010-01-17	2350311.0	1830.0	133
2010-01-24	2820431.0	1996.0	127
2010-01-31	3280586.0	1969.0	128
...

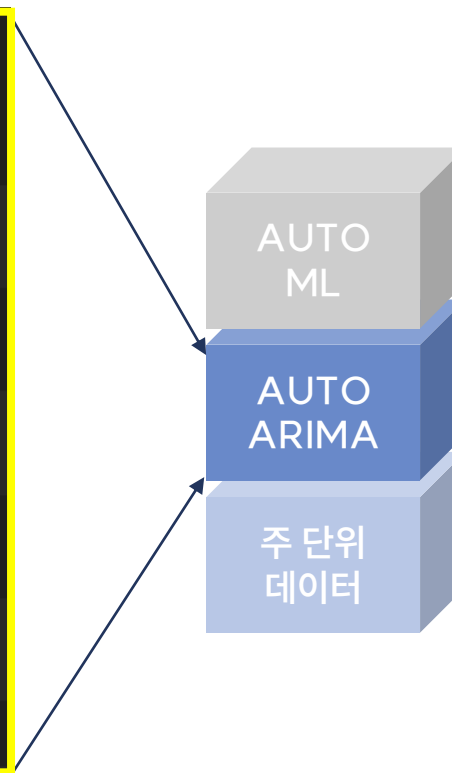
[추가 전]



	총화물량	운항 횟수	
일자			
2010-01-03	2453515.0	1748.0	133
2010-01-10	2338951.0	1822.0	132
2010-01-17	2350311.0	1830.0	133
2010-01-24	2820431.0	1996.0	127
2010-01-31	3280586.0	1969.0	128
...

[ARIMA 예측 값 추가]

arima_pred
1.174028e+06
1.767946e+06
2.043543e+06
2.192896e+06
2.500306e+06
...



✓ 주 단위 Auto ML 예측 값 추가

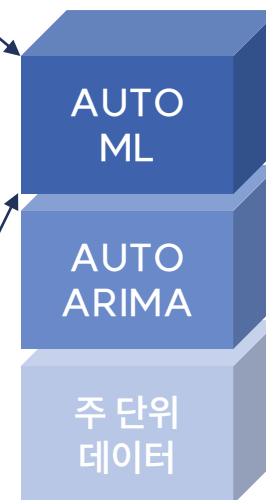
일자	총화물량	운항 횟수
2010-01-03	2453515.0	1748.0 133
2010-01-10	2338951.0	1822.0 132
2010-01-17	2350311.0	1830.0 133
2010-01-24	2820431.0	1996.0 127
2010-01-31	3280586.0	1969.0 128
...

...

arima_pred
1.174028e+06
1.767946e+06
2.043543e+06
2.192896e+06
2.500306e+06
...



일자	주단위예측값
2010-01-03	1.816007e+06
2010-01-10	2.078754e+06
2010-01-17	2.315227e+06
2010-01-24	2.480183e+06
2010-01-31	3.152892e+06
...	...
2022-08-28	1.673086e+06
2022-09-04	1.712906e+06
2022-09-11	1.216558e+06
2022-09-18	1.404903e+06
2022-09-25	1.356833e+06



[추가 전]

[Auto ML 예측 값 도출]

✓ 주 단위 → 일 단위 변환

마지막 주의 경우 **누락된 일자** 존재 가능

→ 부족한 일수만큼 **추가 생성**

ex) 일 단위로 변환 시

마지막 주의 경우 25일까지만 변환됨

→ 09.26 ~ 09.30 추가 보충 필요

주단위예측값	
일자	
2010-01-03	1.816007e+06
2010-01-10	2.078754e+06
2010-01-17	2.315227e+06
2010-01-24	2.480183e+06
2010-01-31	3.152892e+06
...	...
2022-08-28	1.673086e+06
2022-09-04	1.712906e+06
2022-09-11	1.216558e+06
2022-09-18	1.404903e+06
2022-09-25	1.356833e+06

[주 단위]



주단위예측값	
일자	
2010-01-03	1.816007e+06
2010-01-04	1.816007e+06
2010-01-05	1.816007e+06
2010-01-06	1.816007e+06
2010-01-07	1.816007e+06
...	...
2022-09-21	1.404903e+06
2022-09-22	1.404903e+06
2022-09-23	1.404903e+06
2022-09-24	1.404903e+06
2022-09-25	1.356833e+06

[일 단위]

✓ 주 단위 → 일 단위 변환

마지막 주의 경우 **누락된 일자** 존재 가능

→ 부족한 일수만큼 **추가 생성**

ex) 일 단위로 변환 시

마지막 주의 경우 25일까지만 변환됨

→ 09.26 ~ 09.30 추가 보충 필요

주단위예측값		주단위예측값	
일자		일자	
2010-01-03	1.816007e+06	2010-01-03	1.816007e+06
2010-01-04	1.816007e+06	2010-01-04	1.816007e+06
2010-01-05	1.816007e+06	2010-01-05	1.816007e+06
2010-01-06	1.816007e+06	2010-01-06	1.816007e+06
2010-01-07	1.816007e+06	2010-01-07	1.816007e+06
...
2022-09-21	1.404903e+06	2022-09-26	1.356833e+06
2022-09-22	1.404903e+06	2022-09-27	1.356833e+06
2022-09-23	1.404903e+06	2022-09-28	1.356833e+06
2022-09-24	1.404903e+06	2022-09-29	1.356833e+06
2022-09-25	1.356833e+06	2022-09-30	1.356833e+06

[수정 전]

[수정 후]

✓ 최종 일 단위 데이터 완성



[기존 일 단위 데이터에 주 단위 Auto ML 예측 값 추가]

데이터 컬럼

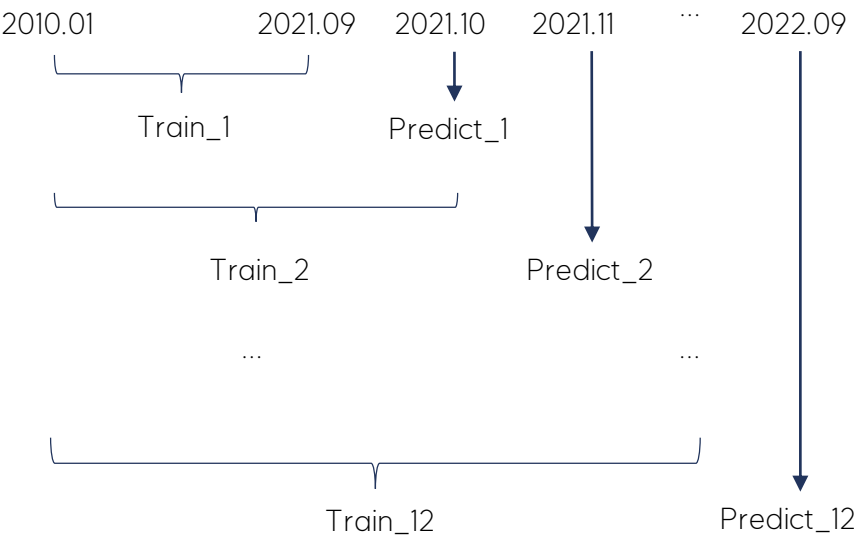
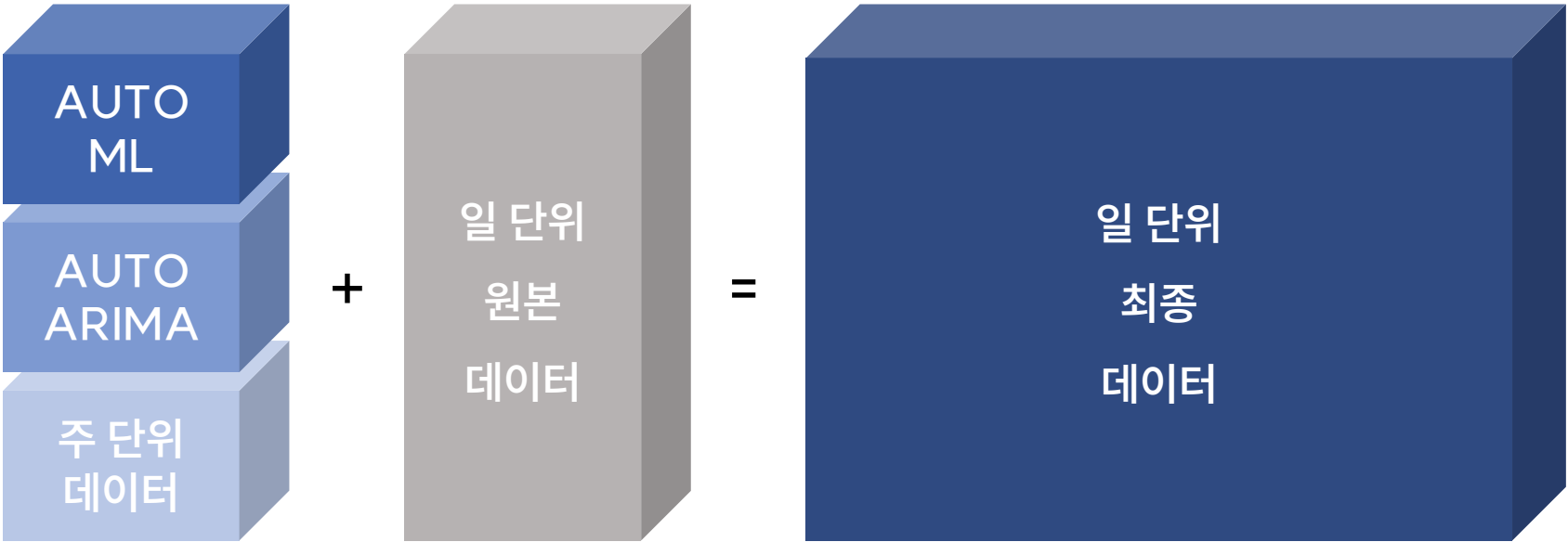
항공 화물 수요 예측 프로젝트

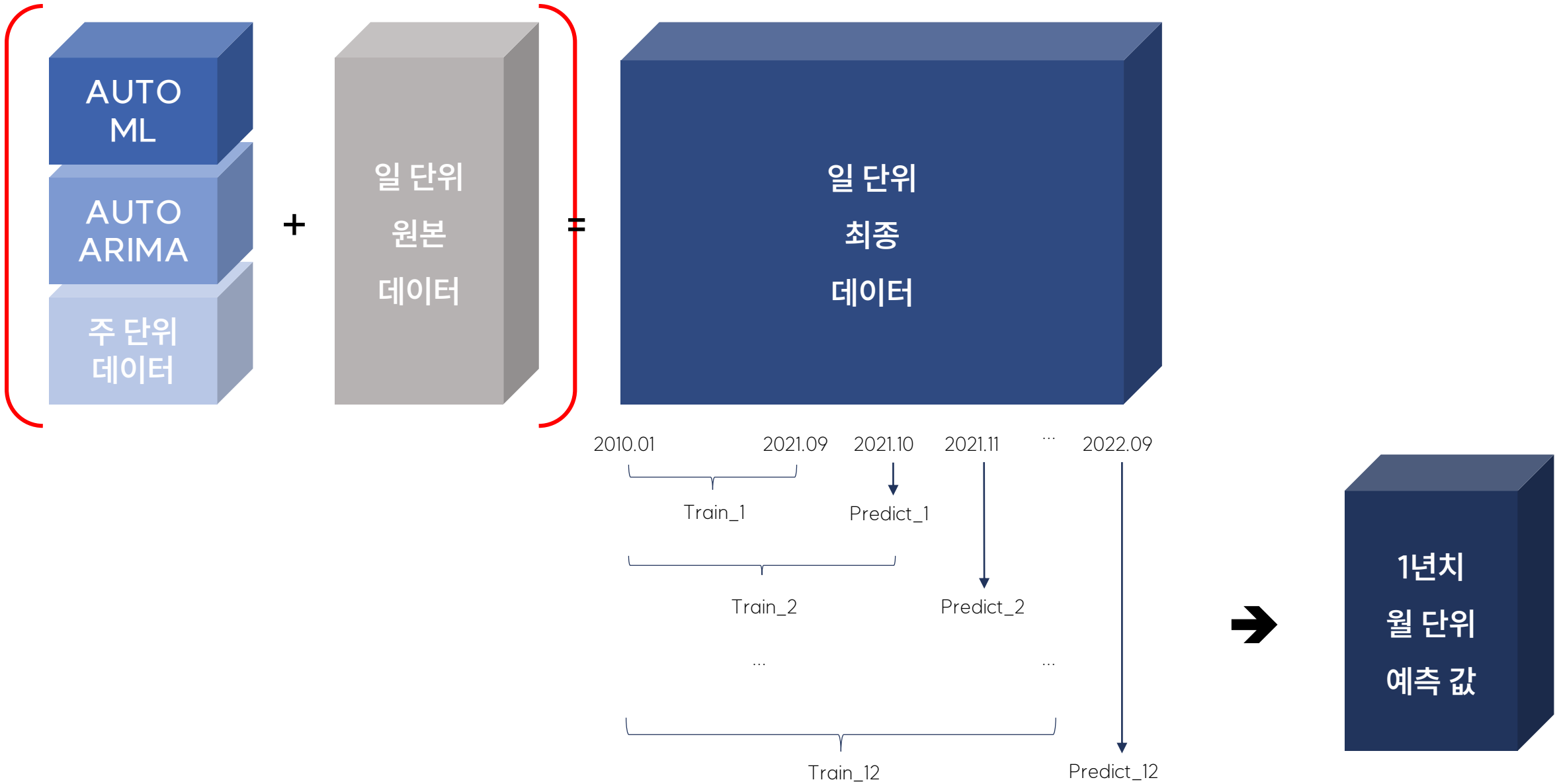
	컬럼명	데이터 타입	예시
시간	시작일자	Datetime	2010-01-03
	종료일자	Datetime	2010-01-09
	월	Categorical	1월
	계절	Categorical	겨울
운항 정보	운항횟수	Numerical	1758
	노선수	Numerical	138
화물량	arima_pred	Numerical	2967886.796
	총화물량	Numerical	3372722
휴일	공휴일(수)	Numerical	0
	공휴일(여부)	Categorical	1
	주말제외휴일	Categorical	1
	주말동시휴일	Categorical	0
	휴일 1주 전후	Categorical	0

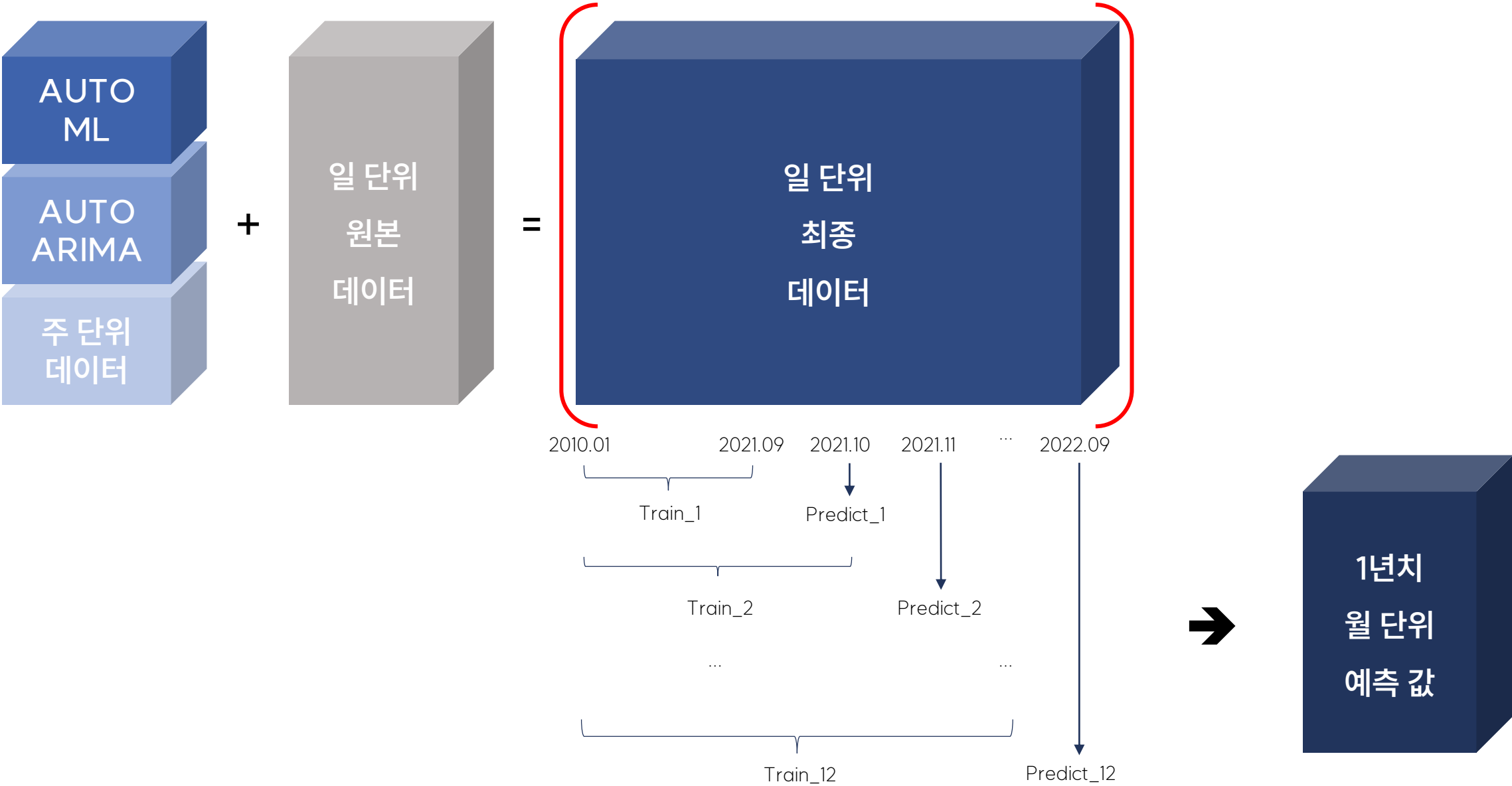
[최종 주 단위 데이터 컬럼]

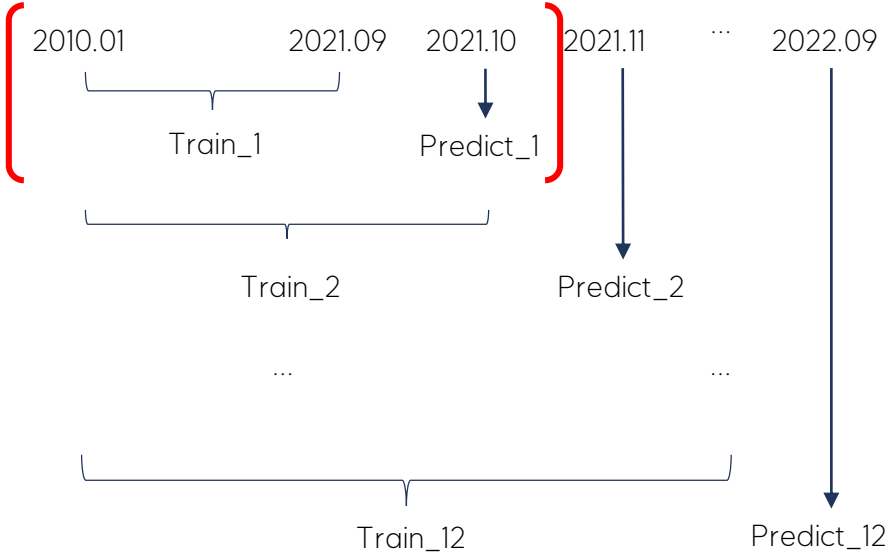
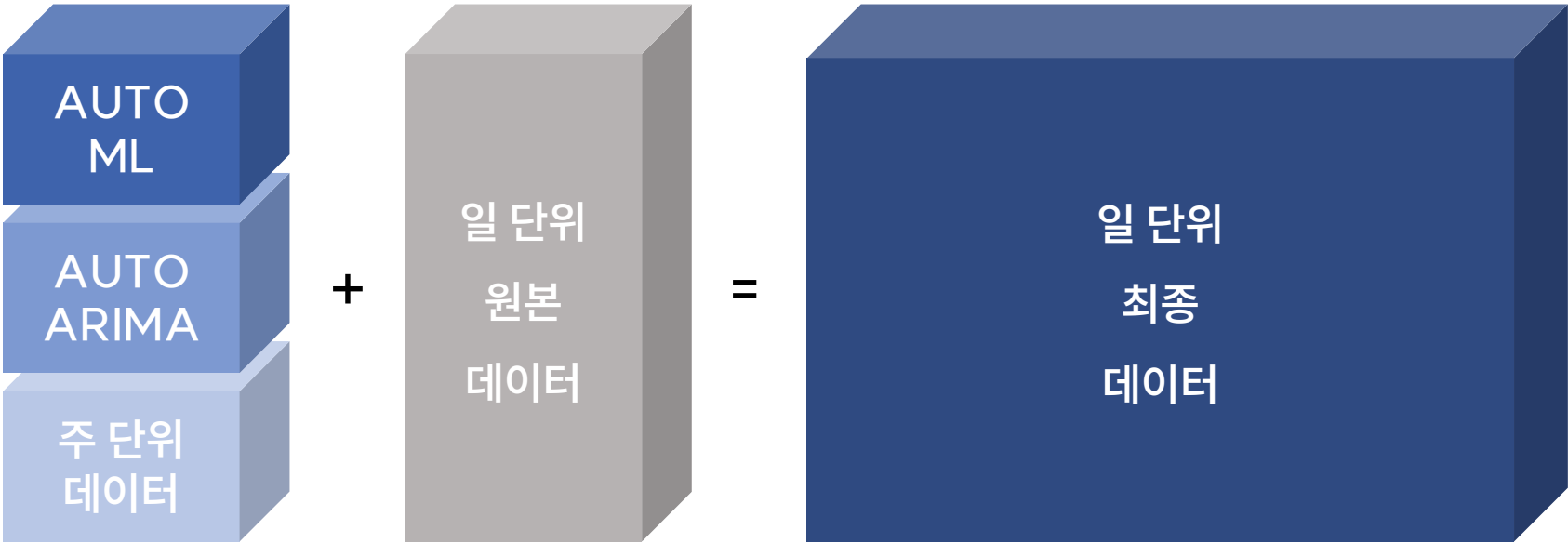
	컬럼명	데이터 타입	예시
시간	일자	Datetime	2010-01-03
	계절	Categorical	겨울
	요일별	Categorical	0
	주말	Categorical	1
운항 정보	운항횟수	Numerical	300
	노선수	Numerical	21
화물량	총화물량	Numerical	347016
	주단위예측값	Numerical	1816007
휴일	공휴일	Categorical	0
	주말제외휴일	Categorical	1
	주말동시휴일	Categorical	0
	휴일 1주 전후	Categorical	0

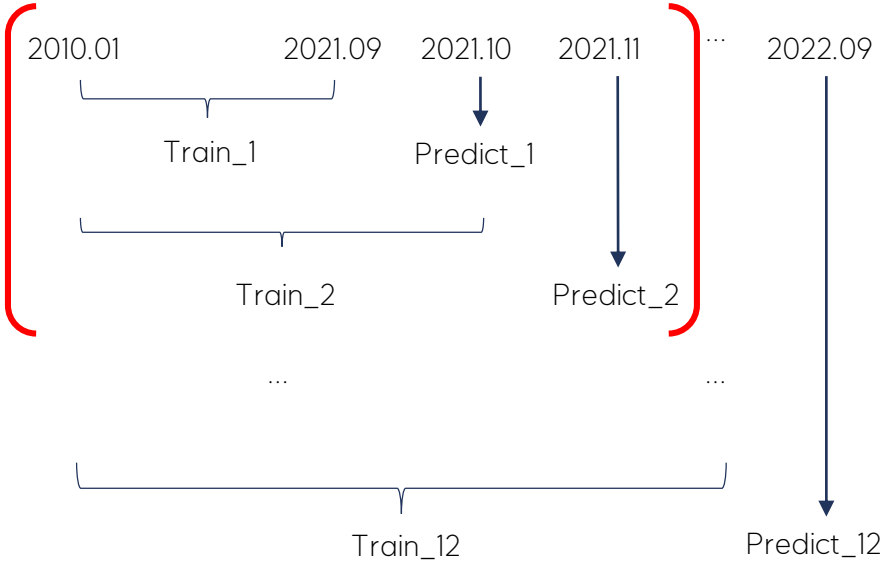
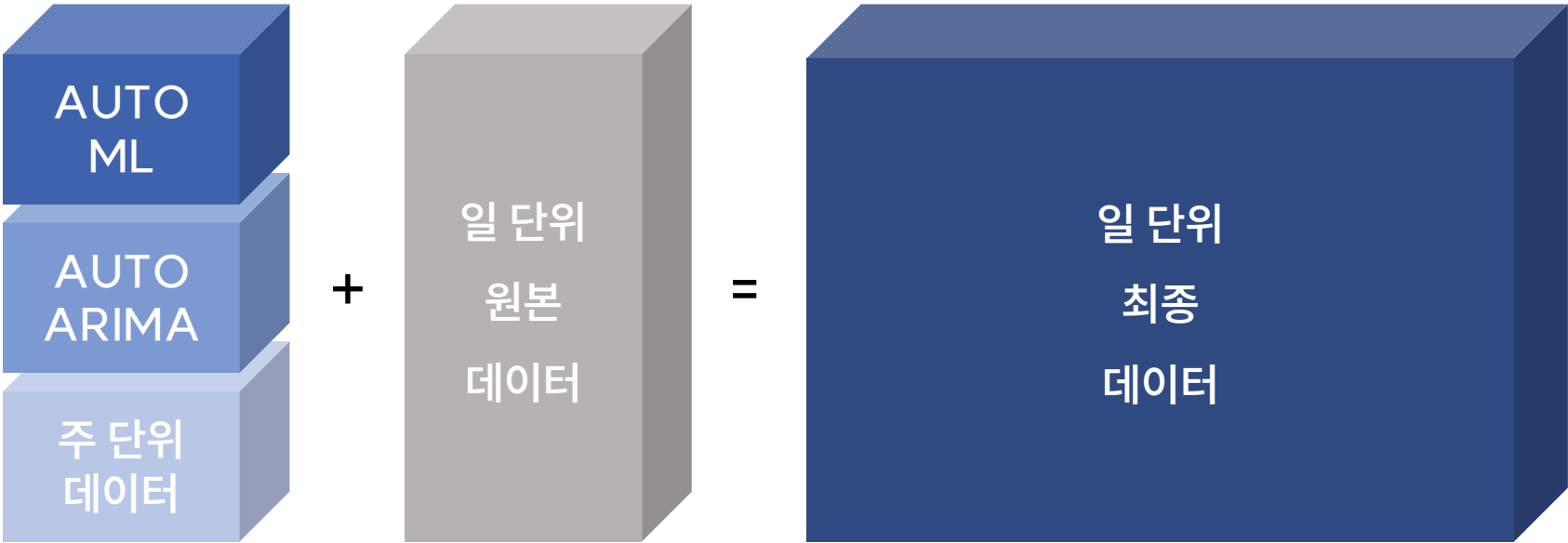
[최종 일 단위 데이터 컬럼]

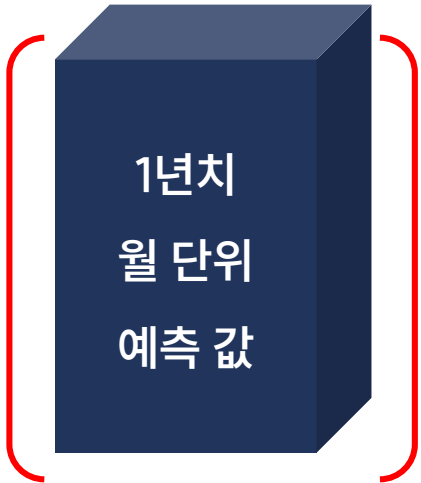
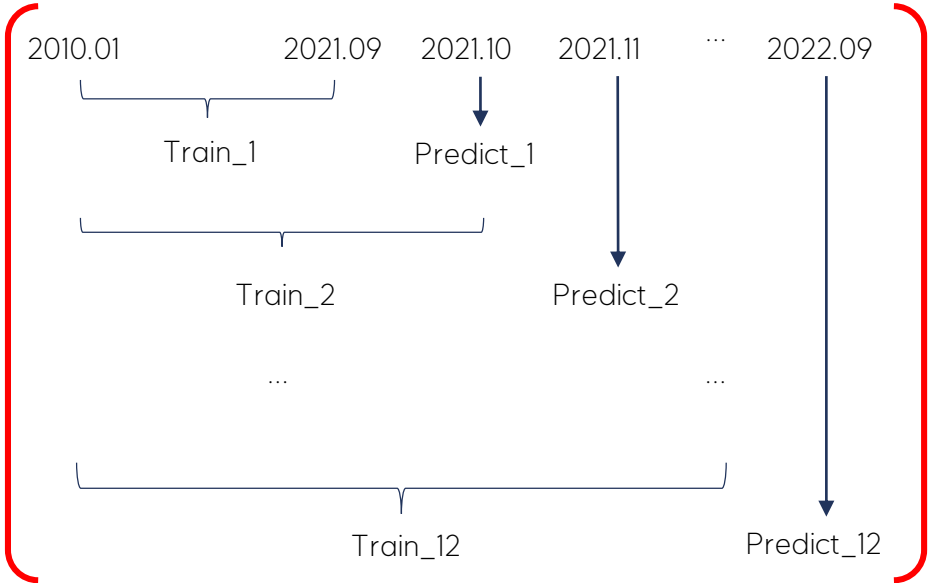
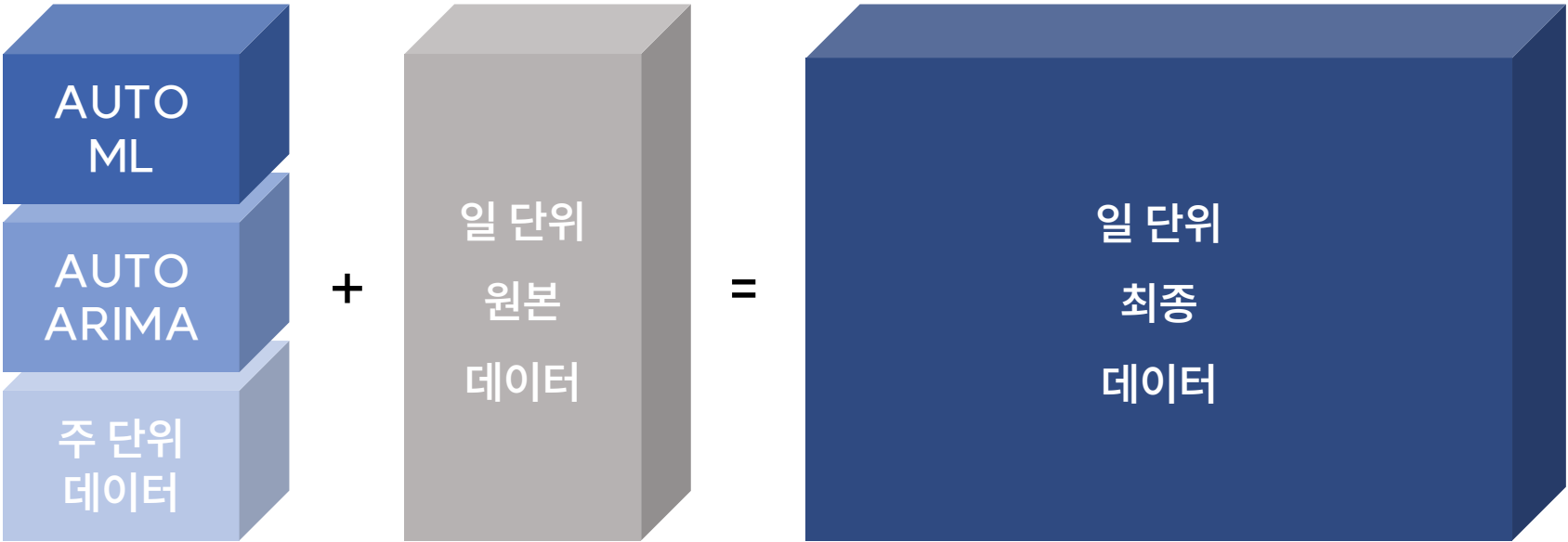


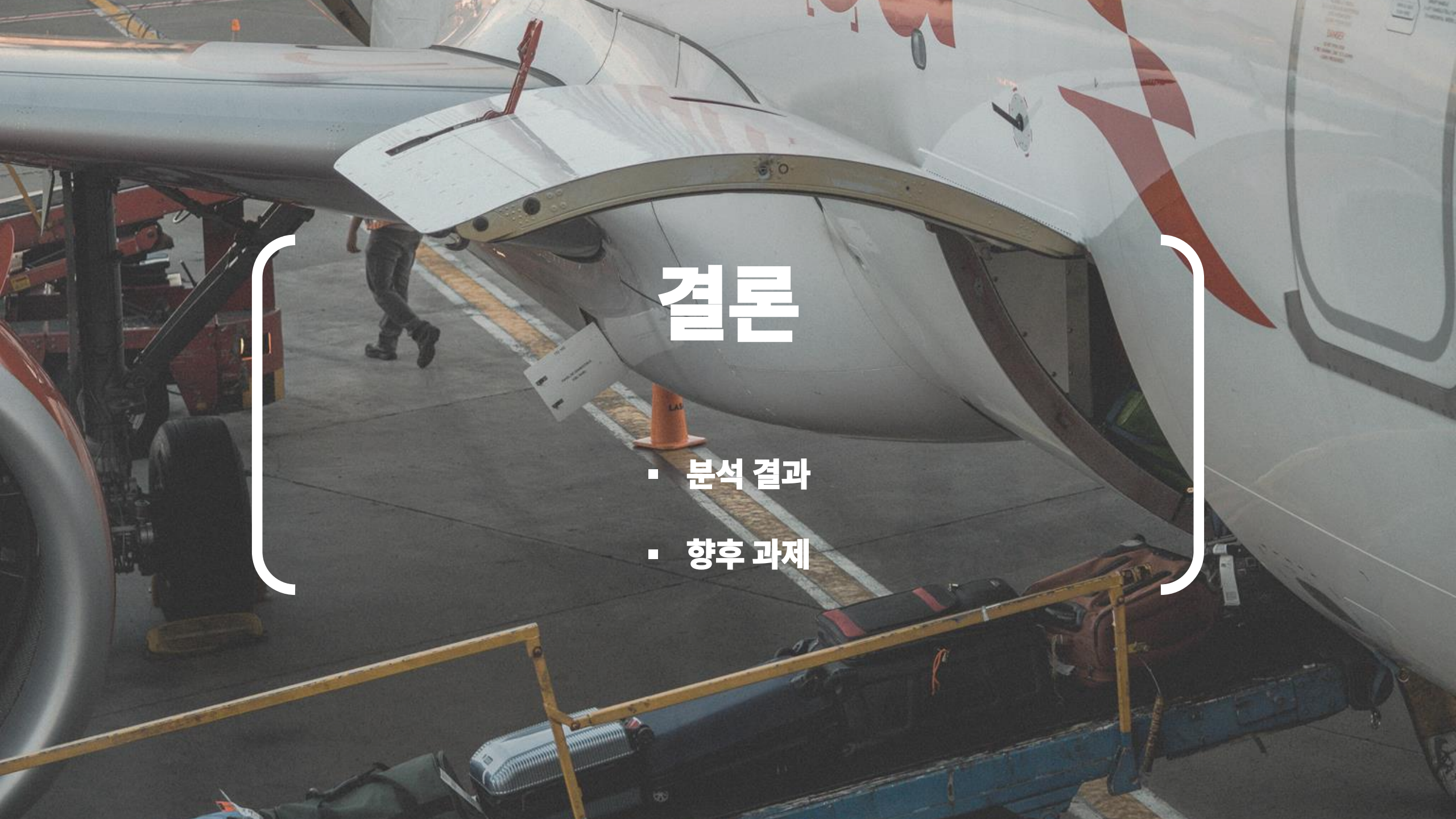










A photograph of an aircraft's open cargo door, showing the interior of the fuselage. The aircraft is white with red and white checkered patterns on the side. A person is visible walking on the tarmac in the background. A white text overlay is centered on the image, containing the word '결론' and a list of two items: '분석 결과' and '향후 과제'.

결론

- 분석 결과
- 향후 과제

1. 주 단위 Auto ARIMA



Train Data	2010.1.3 ~ 2022.8.31
Test Data	2022.9.1 ~ 2022.9.30
Auto Arima 추정 모델	ARIMA(0,1,1)
Test RMSE	220829
Total RMSE	288634

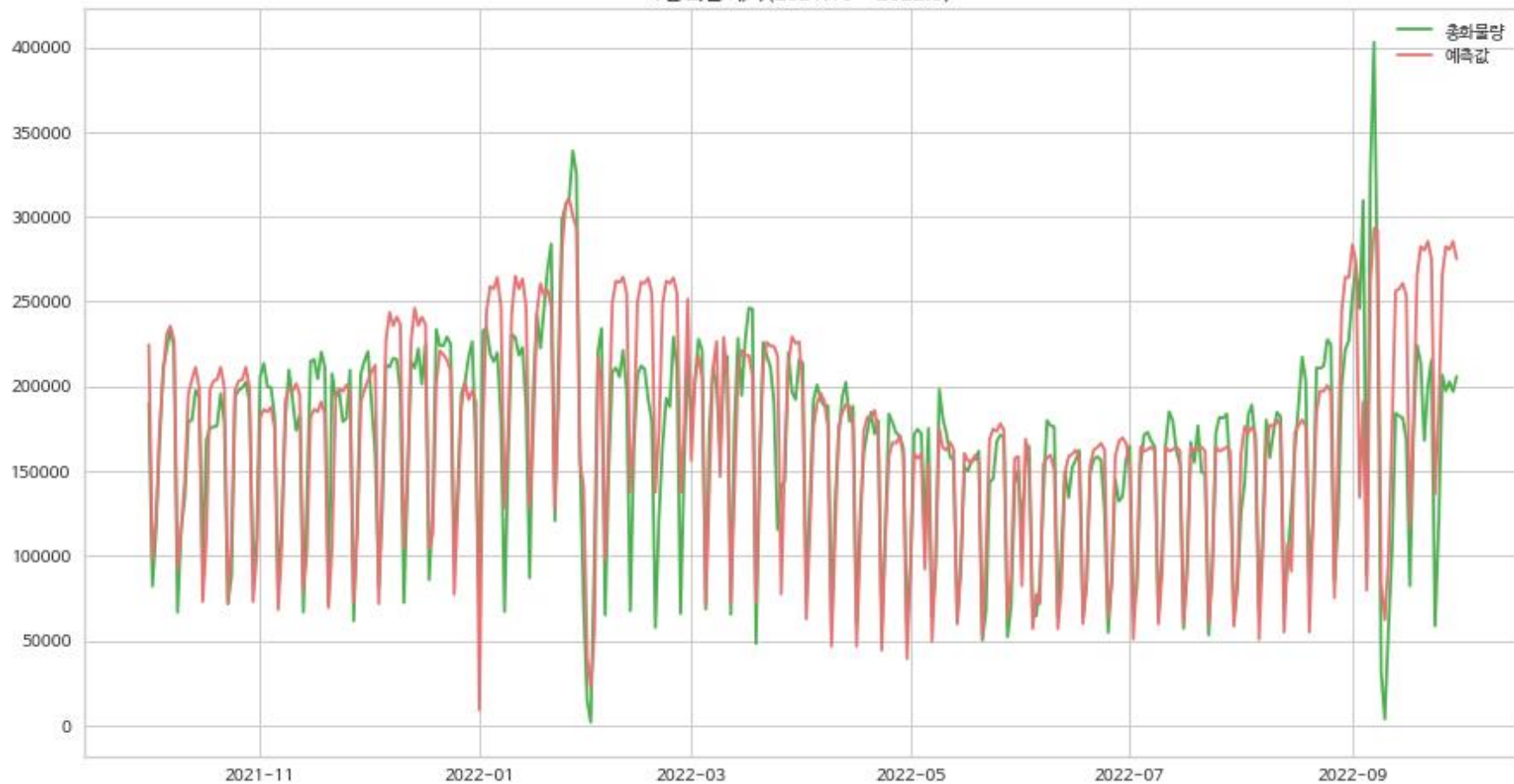
2. 주 단위 Auto ML



Train Data	2010.1.3 ~ 2022.8.31
Test Data	2022.9.1 ~ 2022.9.30
Auto ML 선정 모델	Catboost Regressor
Test RMSE	239294
Total RMSE	127990

3-1. 최종 예측 모델

1년 화물 예측 (2021.10 ~ 2022.9)



Train Data

2010.1.3 ~
(N-1)월

Test Data

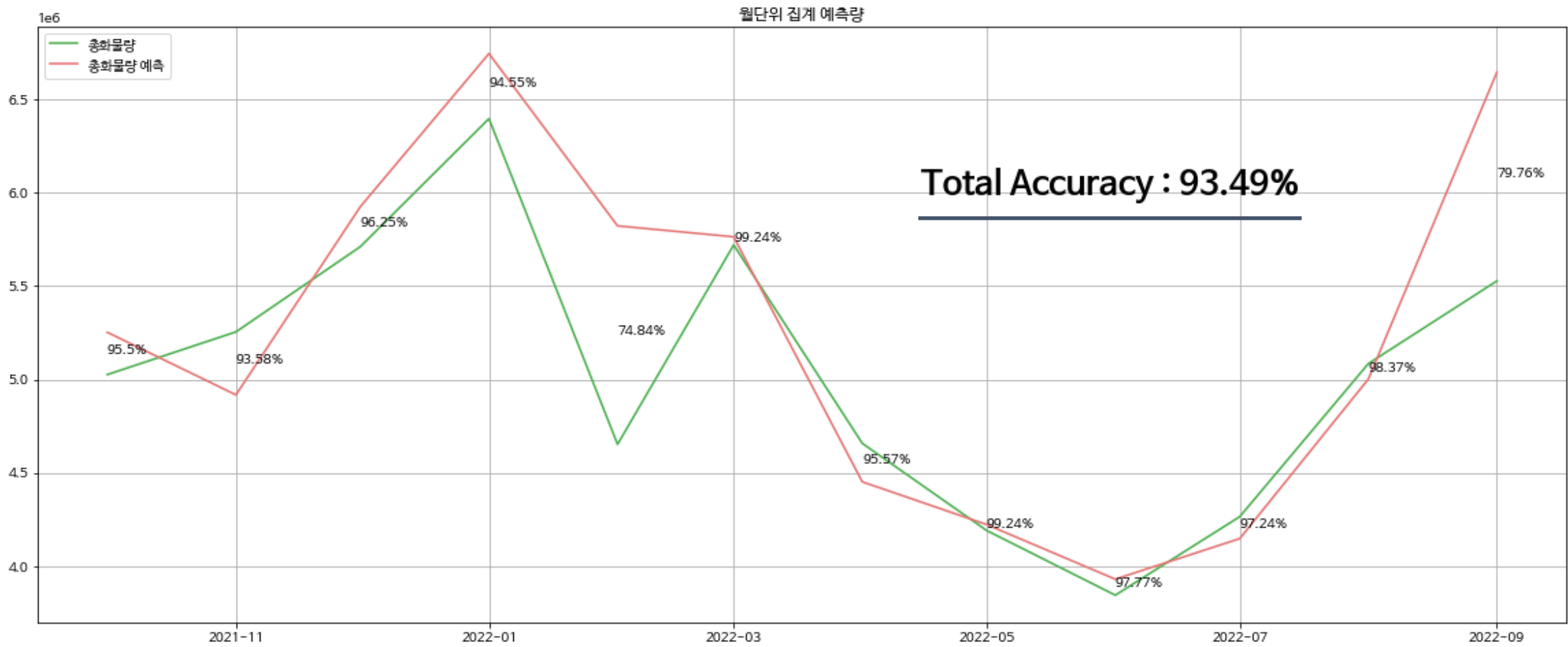
2021.10.1부터
한 달씩Auto ML
선정 모델Catboost
RegressorTest RMSE
일 단위 평균

28217

Total RMSE
일 단위 평균

38180

3-1. 최종 예측 모델



[1년치 화물 예측 값 정확도]

3-2. 최종 예측 모델 vs ARIMA 성능 비교

	3-Part Model	ARIMA
	Accuracy(%)	
Month*	95.50	90.71
Month + 1	93.58	98.53
Month + 2	96.25	97.62
Month + 3	94.55	85.51
Month + 4	74.84	92.64
Month + 5	99.24	96.58
Month + 6	95.57	85.07
Month + 7	99.24	66.14
Month + 8	97.77	58.72
Month + 9	97.24	68.48
Month + 10	98.37	89.54
Month + 11	79.79	99.59
Total	93.49	85.78

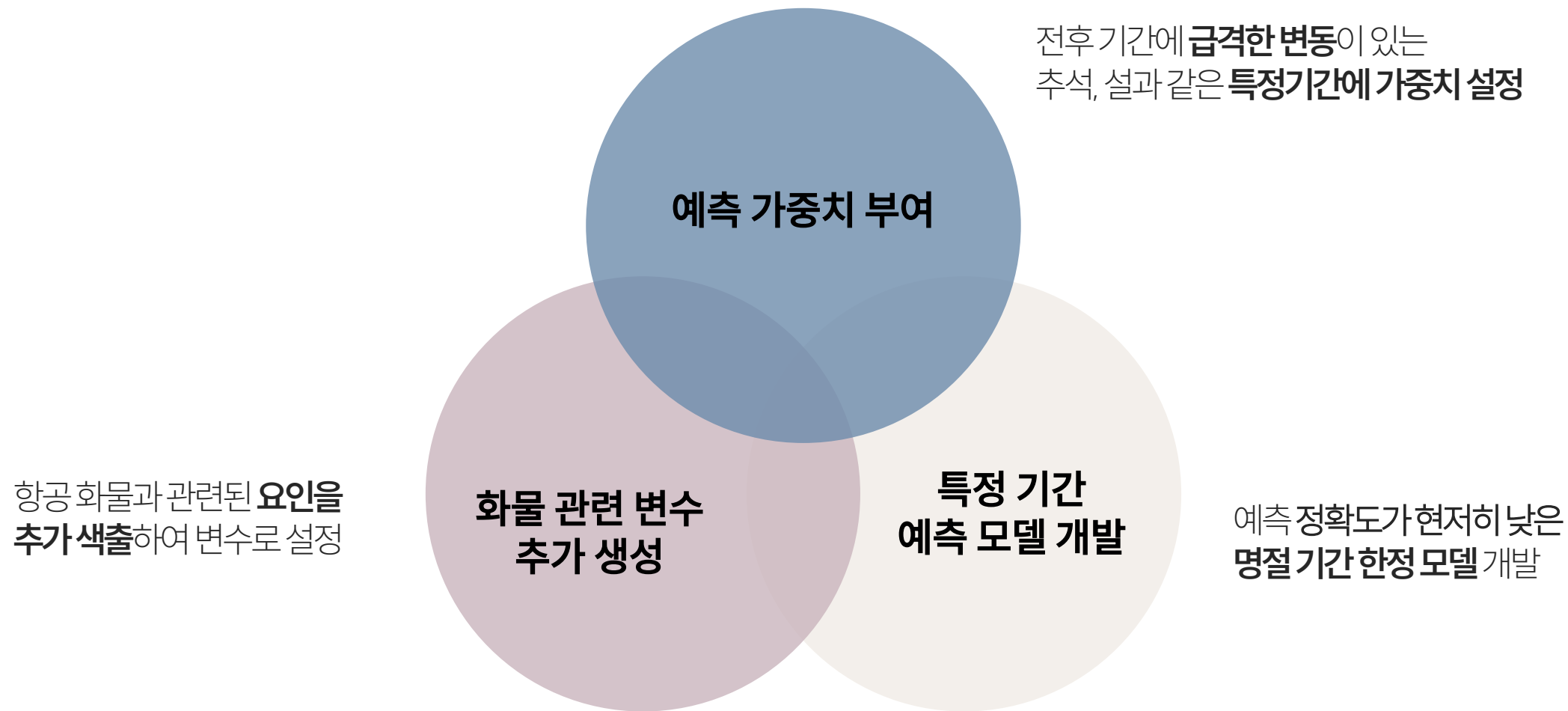
* Month: 2021.10

** 월 단위 RMSE 차이는 약 33만 정도

3-2. 최종 예측 모델 vs ARIMA 성능 비교

	3-Part Model	ARIMA	
	Accuracy(%)		
Month*	95.50	90.71	
Month + 1	93.58	98.53	
Month + 2	96.25	97.62	
Month + 3	94.55	85.51	
Month + 4	74.84	92.64	2월
Month + 5	99.24	96.58	
Month + 6	95.57	85.07	
Month + 7	99.24	66.14	
Month + 8	97.77	58.72	
Month + 9	97.24	68.48	
Month + 10	98.37	89.54	
Month + 11	79.79	99.59	9월
Total	93.49	85.78	

* Month: 2021.10
 ** 월 단위 RMSE 차이는 약 33만 정도





Thank you