



수산물 수입 가격 예측

2021 빅콘테스트 챔피언리그 | M2B

팀장: 이동빈 (dbindi@naver.com)
팀원: 김선민 (seonmk12@gmail.com)
성민지 (sung.minji91@gmaill.com)

목차

1 프로젝트 소개

2 EDA & 데이터 전처리

3 모델 선정

4 결과 및 논의





Aerial photograph of a tropical coastline featuring several small, densely forested islands. A boat is visible on the turquoise-colored water, leaving a white wake. The surrounding water is shallow and sandy.

Part 1

프로젝트 소개

프로젝트 소개

분석 배경

- 1** 유엔 식량농업기구 (FAO) 의 2016년 세계수산양식현황 통계에 따르면, 2013-15년 기준. 우리나라 국민 1인당 연간 수산물 섭취량은 58.4kg으로, 세계 주요국 중 1위를 차지했다.
- 2** 특히 2020년, 한국인의 밥상에 가장 많이 오른 수산물은 중량기준 오징어였으며, 그 뒤를 새우, 멸치, 굴 순으로 나타났다.
- 3** 이 중. 새우, 오징어, 연어의 경우 수입 의존성이 높아 다른 어종에 비해 높은 변동성을 가지고 있으며, 이는 소비자에게 구매 부담을, 생산자에게 수입 불안정성을 유발한다.

프로젝트 소개

재편되는 수산물 유통. 그러나 준비되지 않은 수산업계

비대면 거래가 확대됨에도 불구하고,

국내 수산물 유통구조의 오프라인 유통구조에 머물러 있음



코로나19로 인한 국경의 차단. 공급망의 불확실성

전세계적 공급망에 의존하는 수산업

코로나 19의 확산과 보호무역의 심화, 미중갈등 등으로
정체된 글로벌 공급망



어업분야 국내 수산물의 비싼 원가

비싼 국내산 수산물의 원가에 대비하여, 수입원료 가격의
예측모델의 필요성 대두



프로젝트 소개

연구 목적

- 1 새우, 오징어, 연어의 품목별, 국가별 시계열 분석을 통한 수입 가격 예측**
- 2 위 분석을 토대로 다른 어종의 수입가격예측 모델 생성**

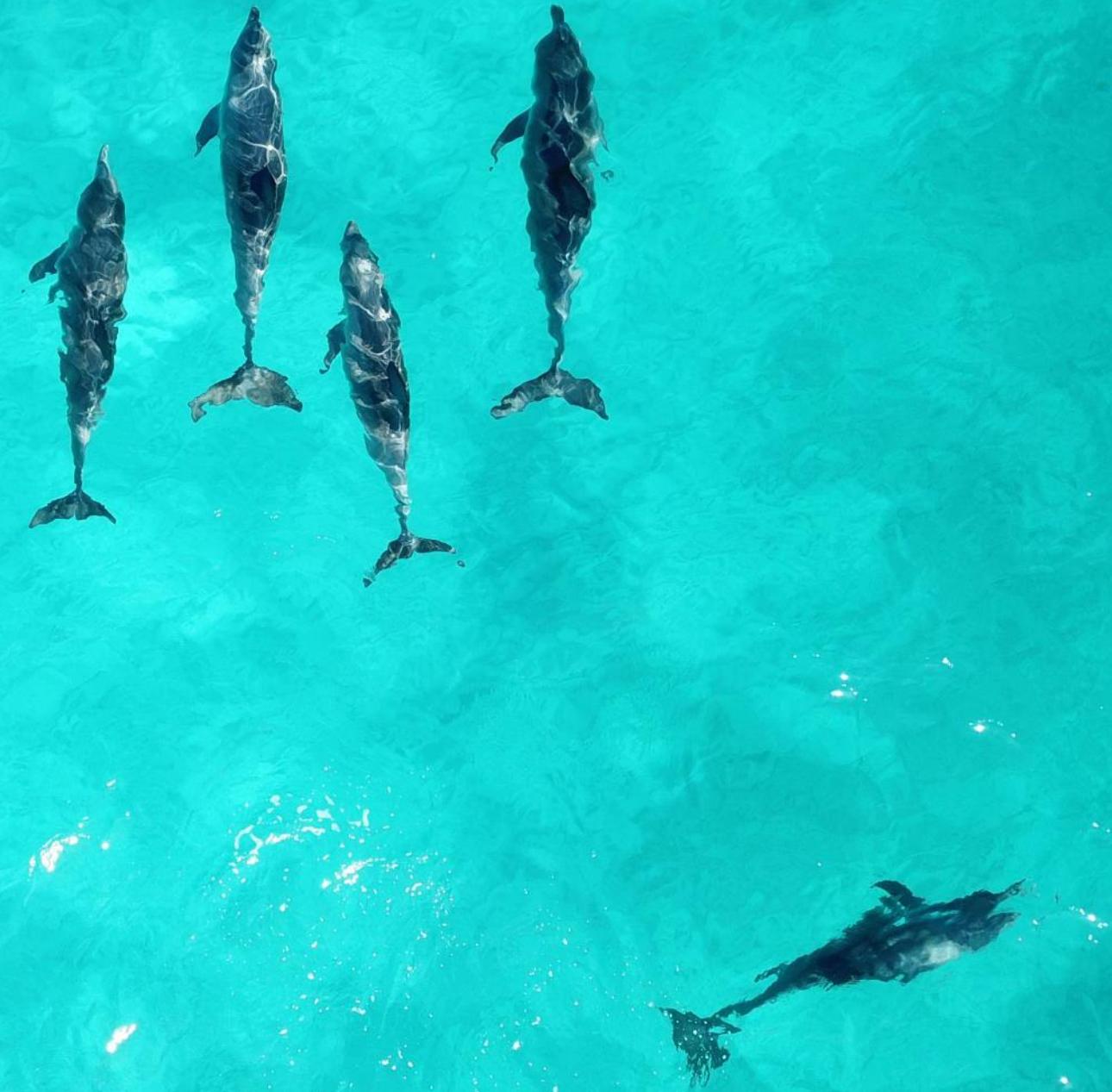
프로젝트 소개

기대 효과

- 1** 한국인의 밥상을 책임지는 세 수산물 - 새우, 오징어, 연어의 적절한 수입 가격 예측 모델의 개발을 통해, 불확실성이 팽배한 수산물 시장에서 소비자와 생산자 모두 보호 가능
- 2** 소비자의 가격정보 범위(price information scope)를 넓게 하여, 소비자로 하여금 준거가격 (reference price) 설정에 영향을 미침. 이는 소비자의 효용에 반영되어 구매 행위에 영향을 미친다.
- 3** 생산자들의 코로나 시국에 불확실성을 이겨낼 수 있는 데 도움을 줄 수 있다.

Part 2

EDA & 데이터 전처리

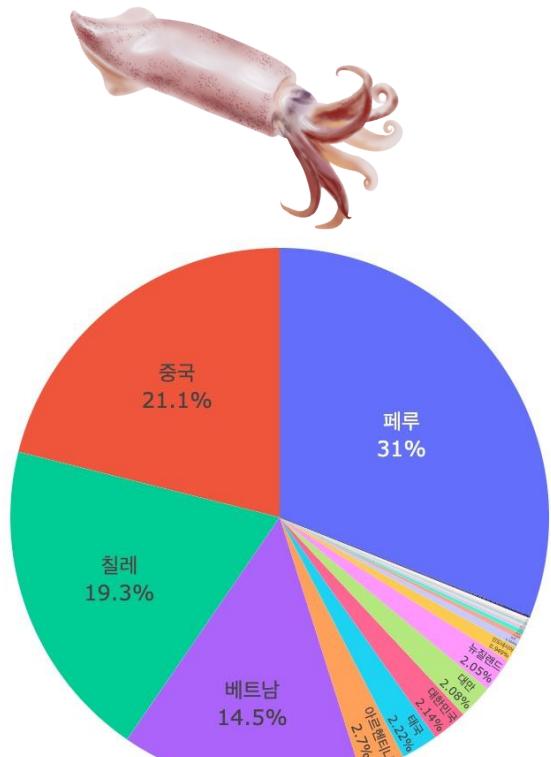


EDA & 테이터 전처리

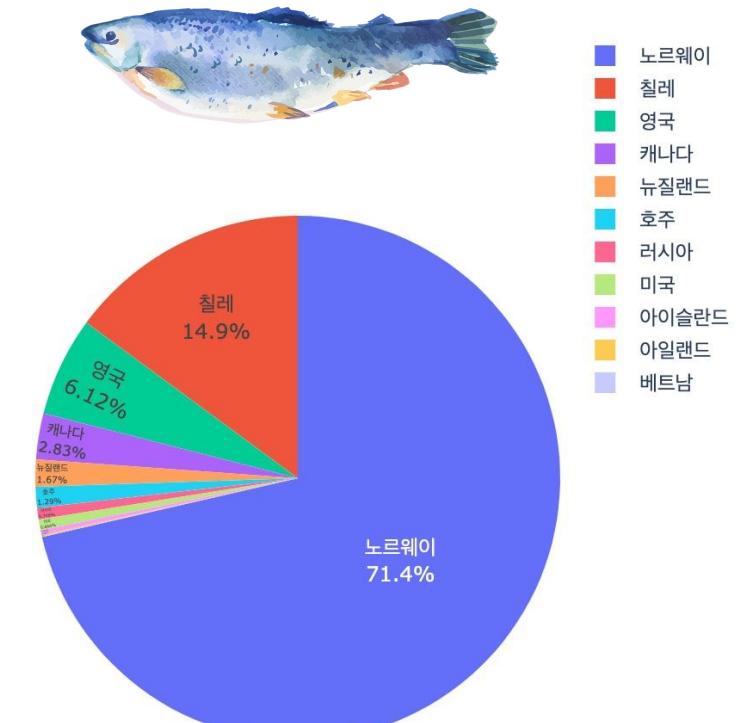
새우 제조국 별 수입 비중



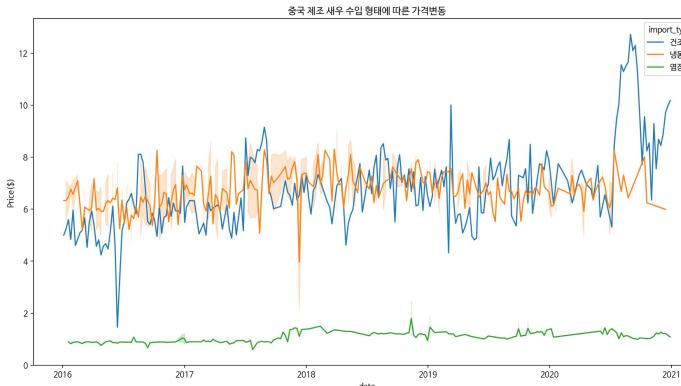
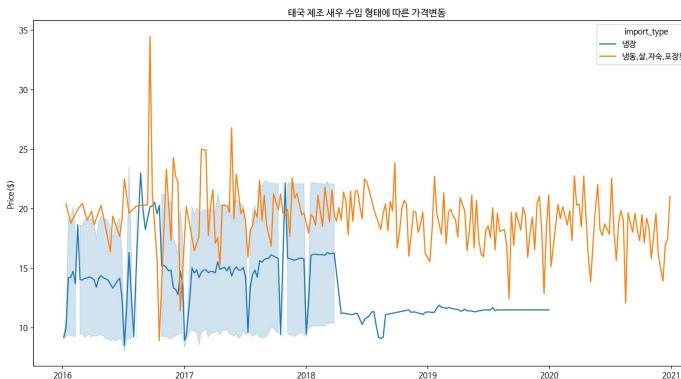
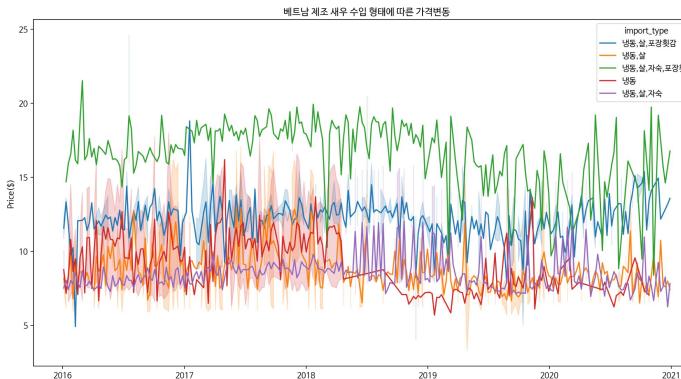
오징어 제조국 별 수입 비중



연어 제조국 별 수입 비중

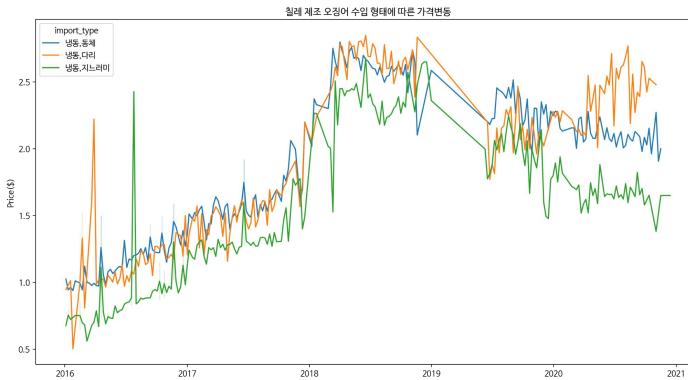
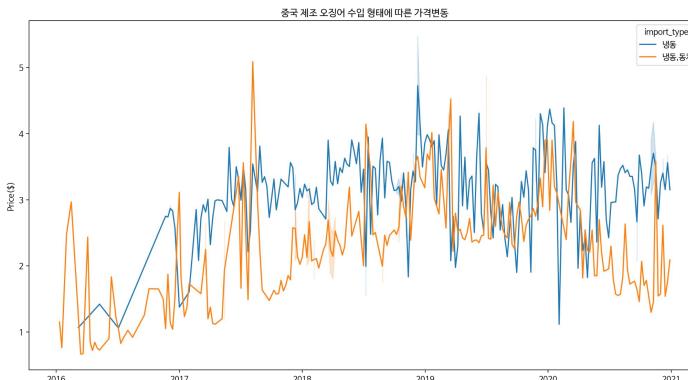
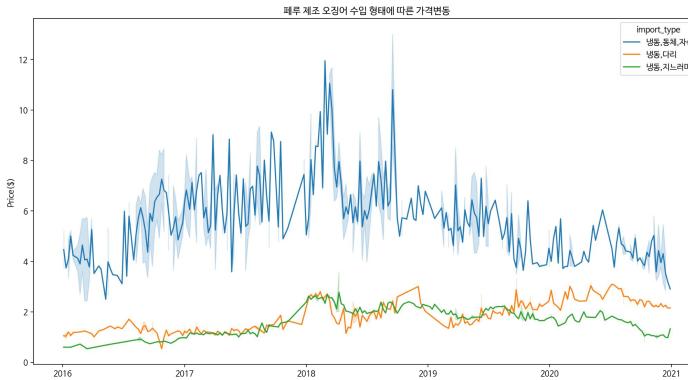


EDA & 데이터 전처리



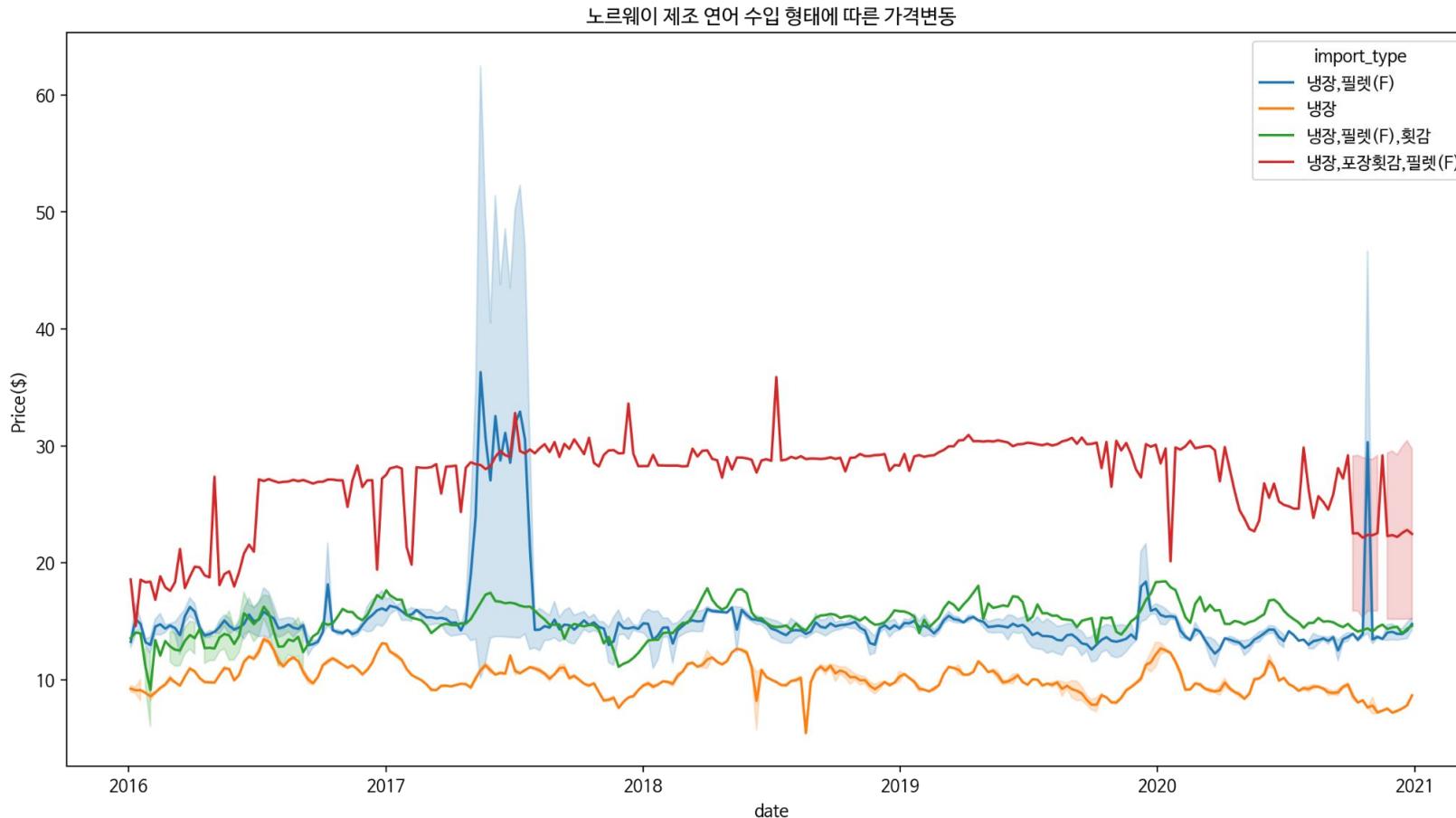
- 새우의 경우: 베트남, 중국, 태국
 - 새우 베트남: ['냉동,살,포장횟감', '냉동,살', '냉동,살,자숙,포장횟감', '냉동', '냉동,살,자숙']
 - 새우 태국: ['냉장', '냉동,살,자숙,포장횟감']
 - 새우 중국: ['건조', '냉동,살', '염장']
-
- 새우의 경우, 베트남에서의 수입이 가장 큰 비중을 차지하는데, 그 수입형태에 따라서도 가격의 다양성이 드러남
 - 특히 태국 제조 새우의 경우, 주기적 스파이크가 관측되는 바, 계절성이 있음이 유추됨

EDA & 데이터 전처리



- 오징어의 경우: 페루, 중국, 베트남, 칠레
 - 오징어 페루: 3 가지 수입형태: ['냉동,동체,자숙', '냉동,다리', '냉동,지느러미']
 - 오징어 중국: 2 가지 수입형태: ['냉동', '냉동,동체']
 - 오징어 칠레: 3 가지 수입형태: ['냉동,동체', '냉동,다리', '냉동,지느러미']
- 각 국가별 수입 형태에 따라 가격 변동 및 가격의 다양성이 큰 것으로 나타남
- 특히 중국 제조 오징어의 경우, 매년 스파이크가 있는 것으로 보아 계절성이 있음이 유추됨

EDA & 데이터 전처리



- 총 수입의 72% 이상을 차지하는 노르웨이
- 그 중 200회 이상 관측된 수입형태인: ['냉장, 필렛(F)', '냉장회, 포장횟감, 필렛(F)', '냉장, 필렛(F), 횟감', '냉장'] 추출
- 다른 어종에 비해 연어는 비교적 큰 차이 없이 대체로 정상성을 보임

EAD & 데이터 전처리

주어진 데이터에서 오징어.
새우, 연어 구분



연도별 환율 및 유가
(한국은행 경제통계시스템
<https://ecos.bok.or.kr/EIndex.jsp>)

2015년 기준 물가 상승률

(국가지표체계.
<https://www.index.go.kr/unify/idx-info.do?idxCd=4226>)

동기간 내 대한민국 수온변화
(국립해양조사원.
<http://www.khoa.go.kr/oceangrid/khoa/koofs.do>)



Part 3

모델 선정:

계절성을 고려한 시계열 분석 (SARIMA)

계절적 시계열 분석 (SARIMA)

이론적 배경

- 1 ARIMA (Auto-Regressive Integrated Moving Average) 는 auto-regressive (AR) 모형과 moving average (MA) 모형을 동시에 포함하는 시간의 흐름에 따른 관측값들의 확률구조를 고려한 모형
- 2 SARIMA 는 Seasonal ARIMA 모형으로, 시계열이 계절성이나 주기적 특성을 지닌 경우 사용된다.

참조: 이동현, 정아현, 김진영, 김창기, 김현구, 이영섭. (2019). SARIMA 모형을 이용한 태양광 발전량 예보 모형 구축. 한국태양에너지학회 논문집, 39(3), 59-66.
doi:10.7836/kses.2019.39.3.059

계절적 시계열 분석 (SARIMA)

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)(1-B)^d Z_t,$$

where, $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$,

$$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q),$$

θ_0 : the deterministic trend

ARIMA

p: 자기회귀(AR)의 차수로서, + 시점의 값에
과거 시점 값이 영향을 주는 것

d: 차분, + 시점의 값에 (+d) 시점의 값을 빼서
비정상 시계열을 정상 시계열로 만들어준다.

q: 이동평균(MA) 차수로서, + 시점의 값이
연속적 오차항들의 영향을 받는 것

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)\epsilon_t$$

where, $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$

$$\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}, \quad \epsilon_t \sim iid \ N(0, \sigma_\epsilon^2)$$

SARIMA

P: 계절적 자기회귀(AR)의 차수

D: 계절적 차분

Q: 계절적 이동평균(MA) 차수로서, + 시점의
값이 연속적 오차항들의 영향을 받는 것

s: 계절주기

계절적 시계열 분석 (SARIMA)

현 데이터가 주간으로 이루어지고, 수산물은 특히 계절에 영향을 많이 받는 식품으로서 SARIMA의 활용이 적절
가장 최적의 파라미터를 찾기 위해 Stepwise model 을 활용

p와 q 모두 1부터 최대 3까지. 주간 데이터를 통해 분기 단위의 계절성 변화를 보기 위해 $m = 12$, $d = 1$, $D = 1$ 로 설정

- **p** : 추세 자기 회귀 순서
- **d** : 트렌드 차이 순서
- **q** : 추세 이동 평균 순서
- **P** : 계절적 자기 회귀 순서
- **D** : 계절별 차이 순서
- **Q** : 계절별 이동 평균 주문
- **m** : 단일 계절 기간 동안의 시간 단계 수

계절적 시계열 분석 (SARIMA)

적용 데이터

1

세분화된 데이터:

24개의 데이터 (어종별, 제조국가별, 수입형태별, 일부 수입목적별(*))

2

법주형 데이터:

흰다리새우, 오징어, 연어의 각 주별 수입 가격 평균값

계절적 시계열 분석 (SARIMA): 세분화된 데이터

1. 데이터 분류: 24개의 데이터 (어종별, 제조국가별, 수입형태별, 일부 수입목적별(*))



- 제조국: 베트남, 중국, 태국
- 베트남: ['냉동,살,포장횟감', '냉동,살', '냉동,살,자숙, 포장횟감', '냉동', '냉동,살, 자숙']
- 태국: ['냉장', '냉동,살,자숙, 포장횟감']
- 중국: ['건조', '냉동,살', '염장']



- 제조국: 페루, 중국, 칠레
- 페루: ['냉동,동체,자숙', '냉동,다리', '냉동,지느러미']
- 중국: ['냉동', '냉동,동체']
- 칠레: ['냉동,동체', '냉동,다리', '냉동,지느러미']



- 제조국: 노르웨이
- 노르웨이: ['냉장,필렛(F)*', '냉장회,포장횟감,필렛(F)', '냉장,필렛(F),횟감', '냉장*']

계절적 시계열 분석 (SARIMA): 세분화된 데이터

정상성 테스트: Augmented Dickey-Fuller test

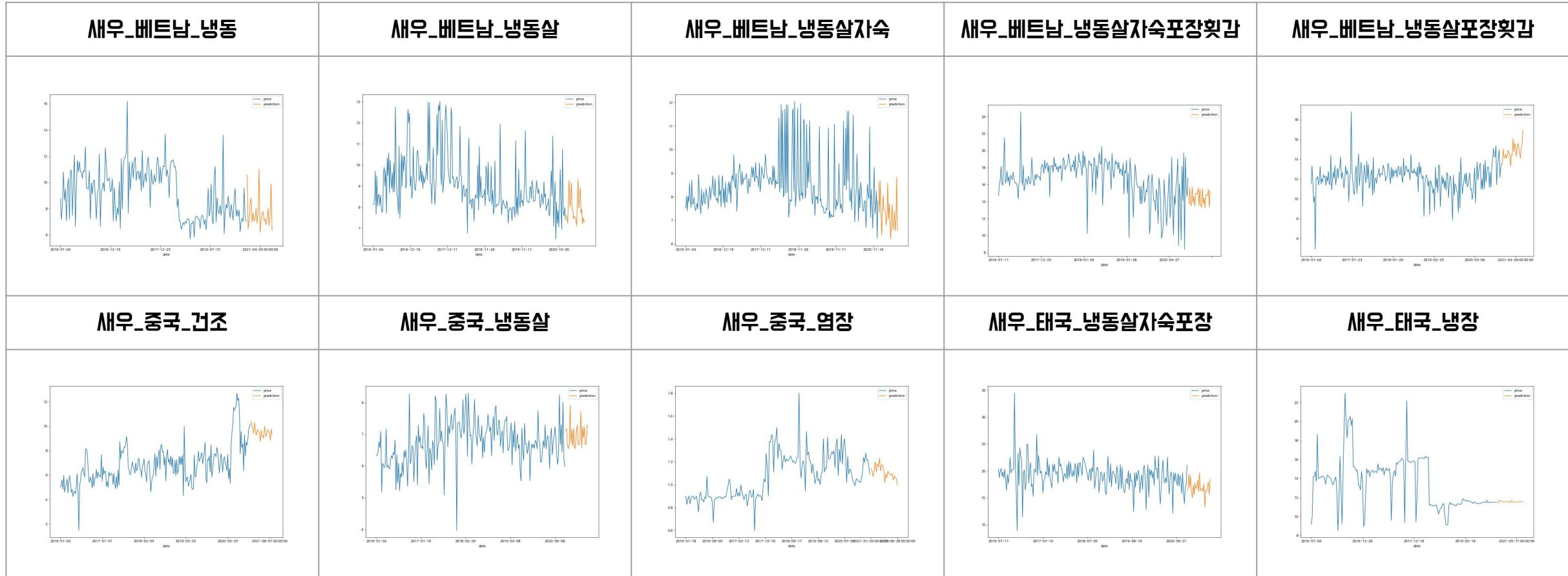
검정통계량 (ADF Statistics) 가 Critical Value 보다 작으면, 혹은 p-value가 통상적 신뢰수준 값 0.05 보다 작으면 정상성을 지니는 시계열 데이터로, 차분 혹은 이동평균 (rolling mean) 진행하지 않아도 ARIMA (SARIMA) 모형을 사용 가능

데이터 명	p-value	데이터 명	p-value	데이터 명	p-value	데이터 명	p-value
연어_노르웨이_냉장필렛 _자사제품제조용	0.001782	오징어_페루_냉동다리	0.194813	오징어_칠레_냉동다리	0.340685	새우_베트남_냉동살자속	0.045880
연어_노르웨이_냉장필렛 _판매용	0.000342	오징어_페루_냉동동체자 속	0.236275	오징어_칠레_냉동지느러 미	0.447094	새우_태국_냉장	0.026131
연어_노르웨이_냉장포장 횟감필렛	0.015991	오징어_페루_냉동지느러 미	0.237641	새우_베트남_냉동살포장 횟감	0.008648	새우_태국_냉동살자속포 장	0.079832
연어_노르웨이_냉장필렛 횟감	0.000089	오징어_중국_냉동동체	0.095850	새우_베트남_냉동살	0.492980	새우_중국_건조	0.470153
연어_노르웨이_냉장_자 사제품제조용	0.000750	오징어_중국_냉동	4.441825e-07	새우_베트남_냉동살자속 포장횟감	0.695588	새우_중국_냉동살	0.393652
연어_노르웨이_냉장_판 매용	0.002057	오징어_칠레_냉동동체	0.413070	새우_베트남_냉동	0.031955	새우_중국_염장	0.302487

계절적 시계열 분석 (SARIMA): 세분화된 데이터



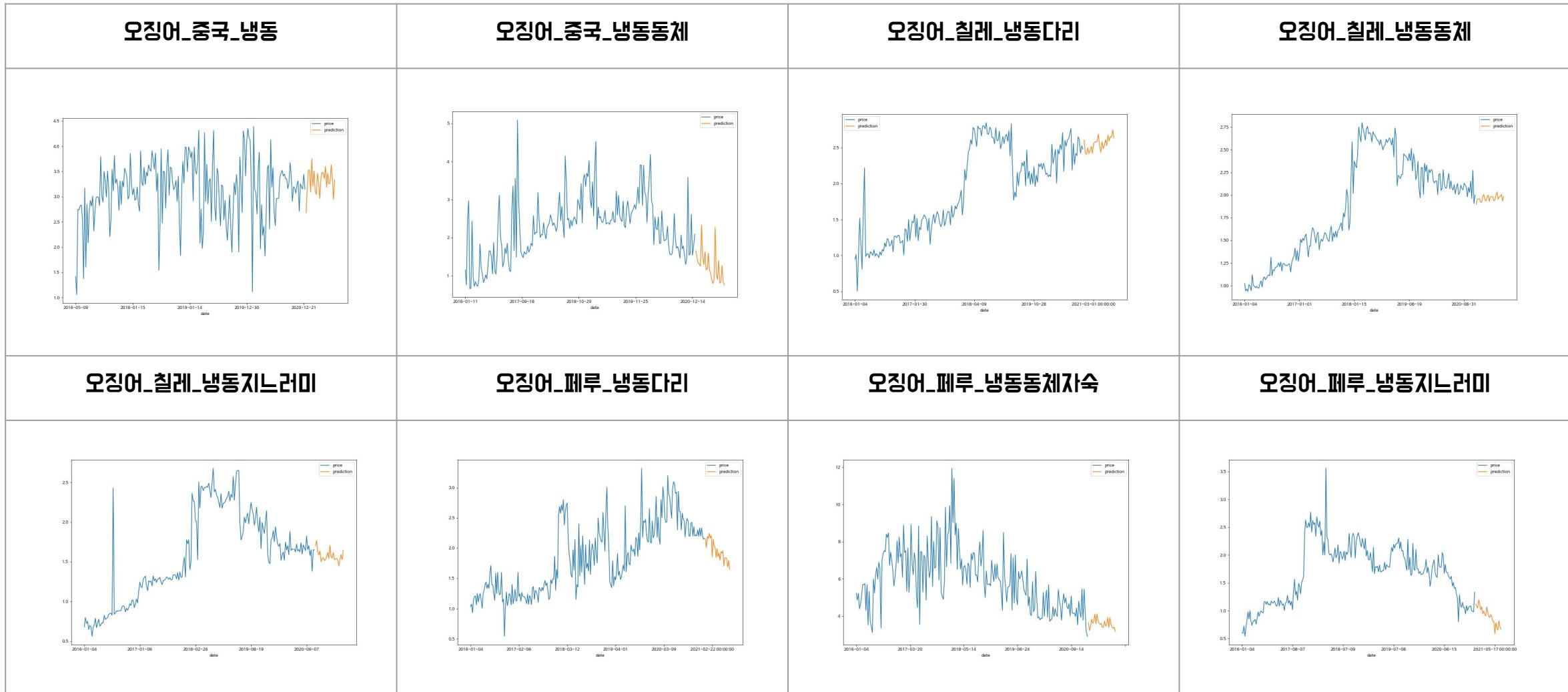
새우



계절적 시계열 분석 (SARIMA): 세분화된 데이터



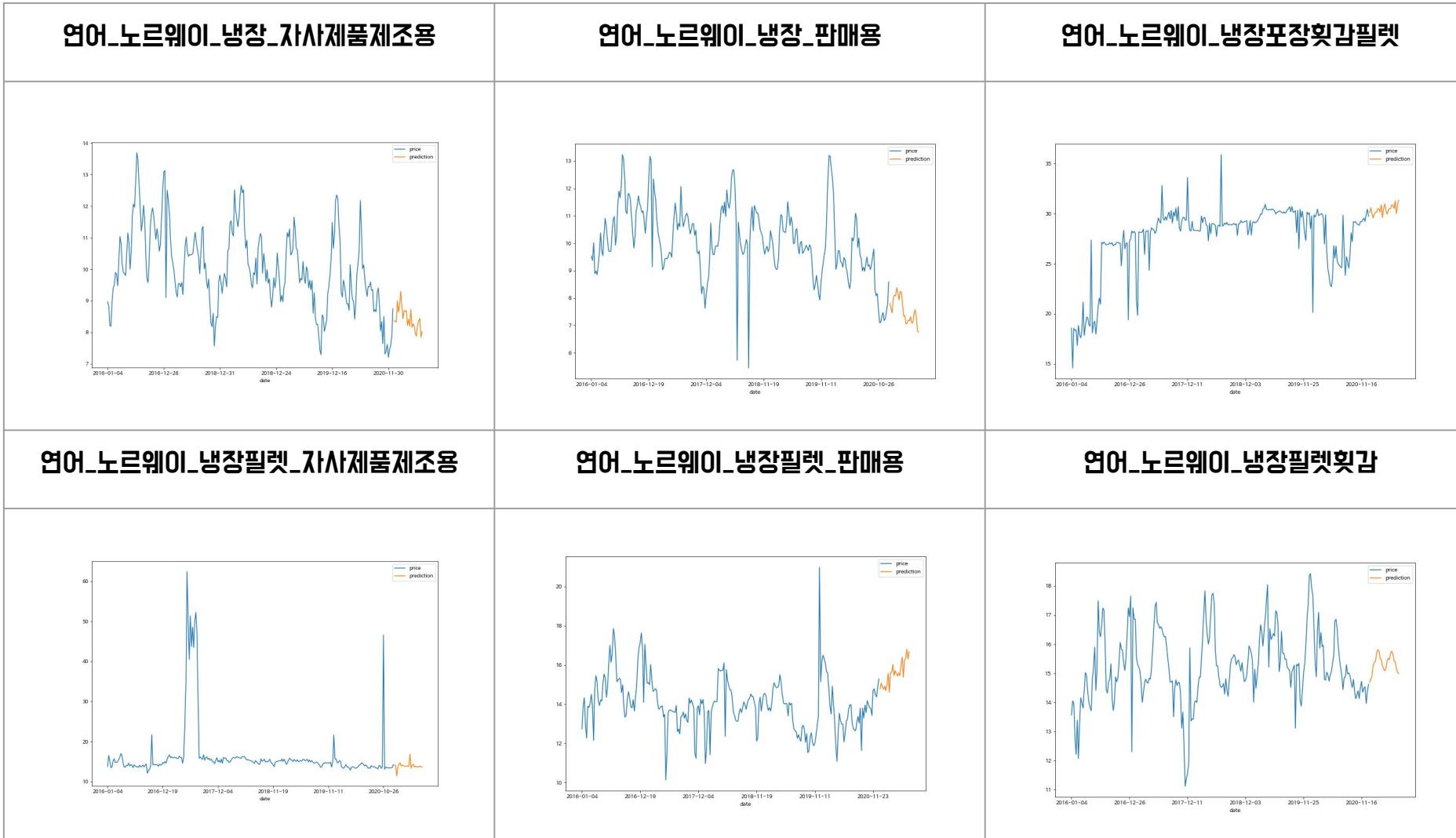
오징어



계절적 시계열 분석 (SARIMA): 세분화된 데이터



연어



계절적 시계열 분석 (SARIMA): 범주형 데이터

적용 모델

- 1 주어진 데이터 (2016.01-2020.12) 중 나중값을 20% 비율로 train, test set 분류. 이 후 계절적 시계열 분석 모형을 통해 나온 값과 실제 값 사이 MAE (Mean Absolute Error) 값 도출
- 2 미래 데이터 (2021.01-2021.06) 의 각 범주, 흰다리새우, 오징어, 연어의 예측 수입가격 도출

계절적 시계열 분석 (SARIMA): 범주형 데이터

정상성 테스트: Augmented Dickey-Fuller test

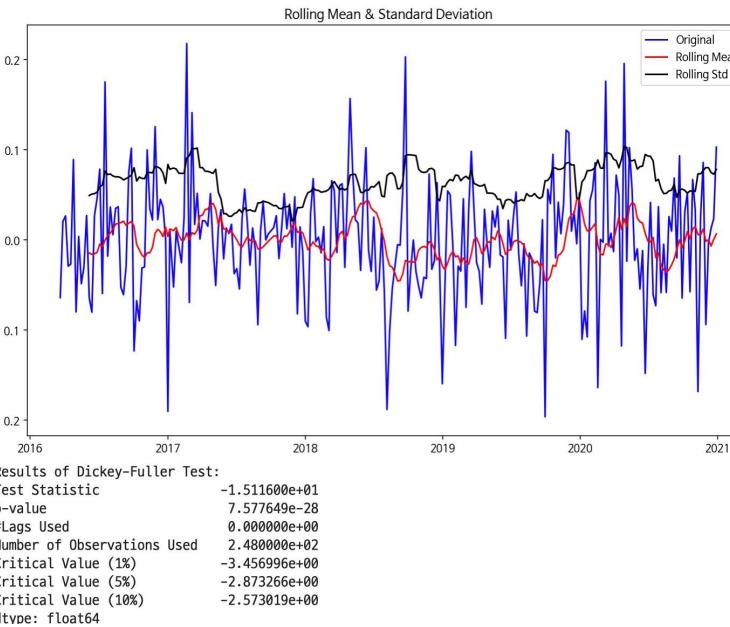
검정통계량 (ADF Statistics) 가 Critical Value 보다 작으면, 혹은 p-value가 통상적 신뢰수준 값 0.05 보다 작으면 정상성을 지니는 시계열 데이터로, 차분 혹은 이동평균 (rolling mean) 진행하지 않아도 ARIMA (SARIMA) 모형을 사용 가능

데이터 명	p-value
오징어	0.022922
흰다리새우	0.177244
연어	0.003982

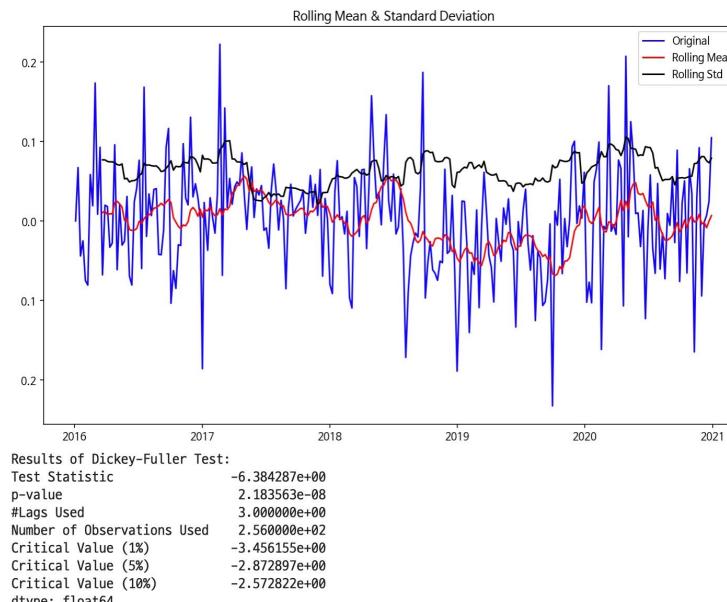
계절적 시계열 분석 (SARIMA): 범주형 데이터

흰다리새우의 정상성 도출 → p-value 비교 후, 기존 데이터 사용하기로 결정

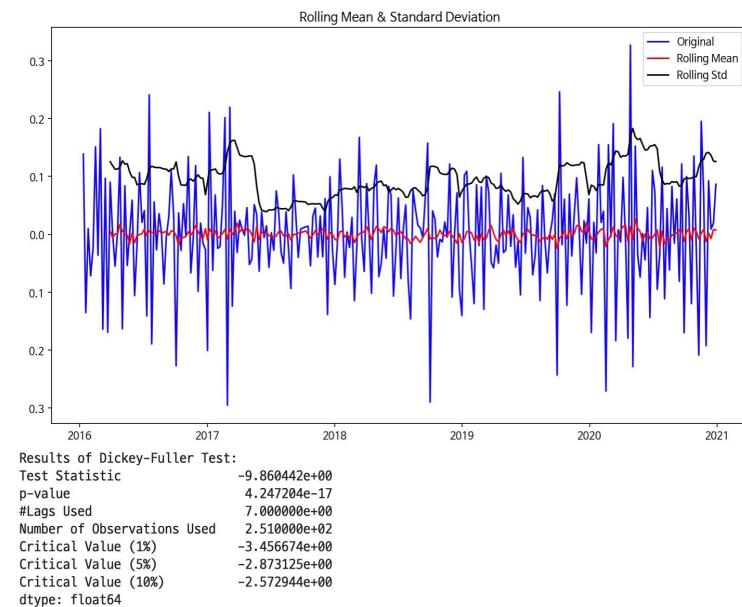
로그 후 이동평균 빼기



지수적 감소(exponential decay) 후 이동평균 빼기



타임 시프팅 (time shifting) 후 이동평균 빼기



계절적 시계열 분석 (SARIMA): 범주형 데이터

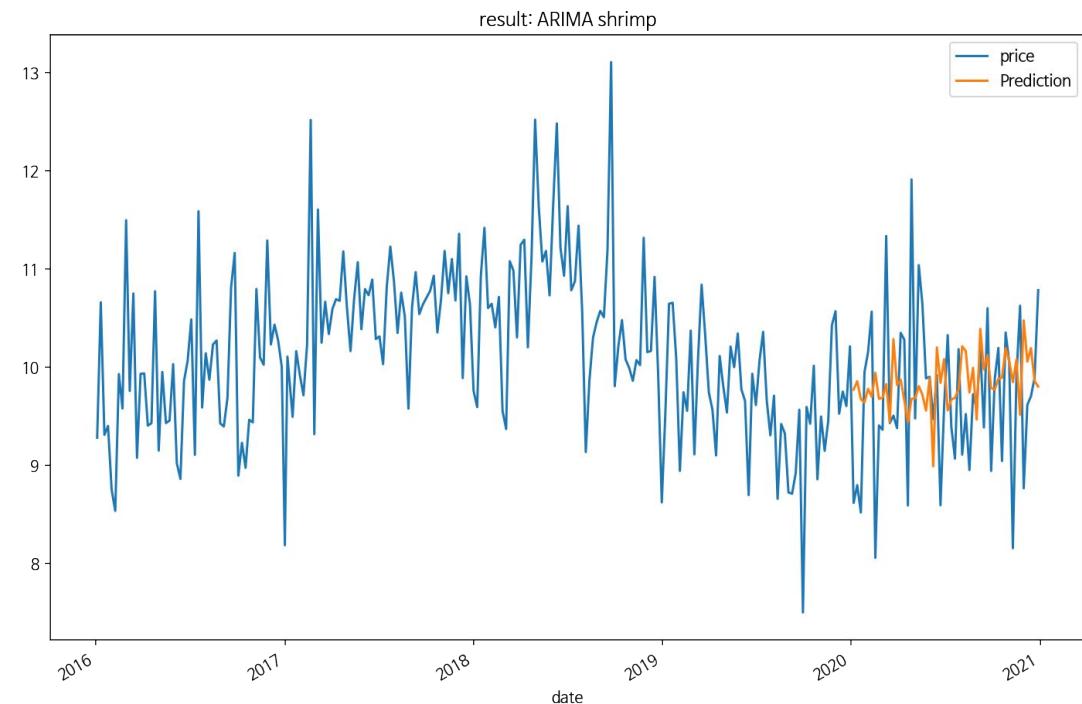


새우 경증력

새우 SARIMA (0, 1, 1) (2, 1, 0, 12)

SARIMAX Results						
Dep. Variable:	y	No. Observations:	208			
Model:	SARIMAX(0, 1, 1)x(2, 1, [], 12)	Log Likelihood	-236.495			
Date:	Tue, 14 Sep 2021	AIC	480.989			
Time:	15:08:14	BIC	494.081			
Sample:	0 - 208	HQIC	486.290			
Covariance Type:	opg					
coef	std err	z	P> z	[0.025	0.975]	
ma.L1	-0.8530	0.045	-18.908	0.000	-0.941	-0.765
ar.S.L12	-0.5663	0.060	-9.476	0.000	-0.683	-0.449
ar.S.L24	-0.3374	0.070	-4.853	0.000	-0.474	-0.201
sigma2	0.6396	0.062	10.319	0.000	0.518	0.761
Ljung-Box (L1) (Q):	0.97	Jarque-Bera (JB):	1.34			
Prob(Q):	0.32	Prob(JB):	0.51			
Heteroskedasticity (H):	0.79	Skew:	-0.05			
Prob(H) (two-sided):	0.33	Kurtosis:	3.39			

비교 그래프



계절적 시계열 분석 (SARIMA): 범주형 데이터

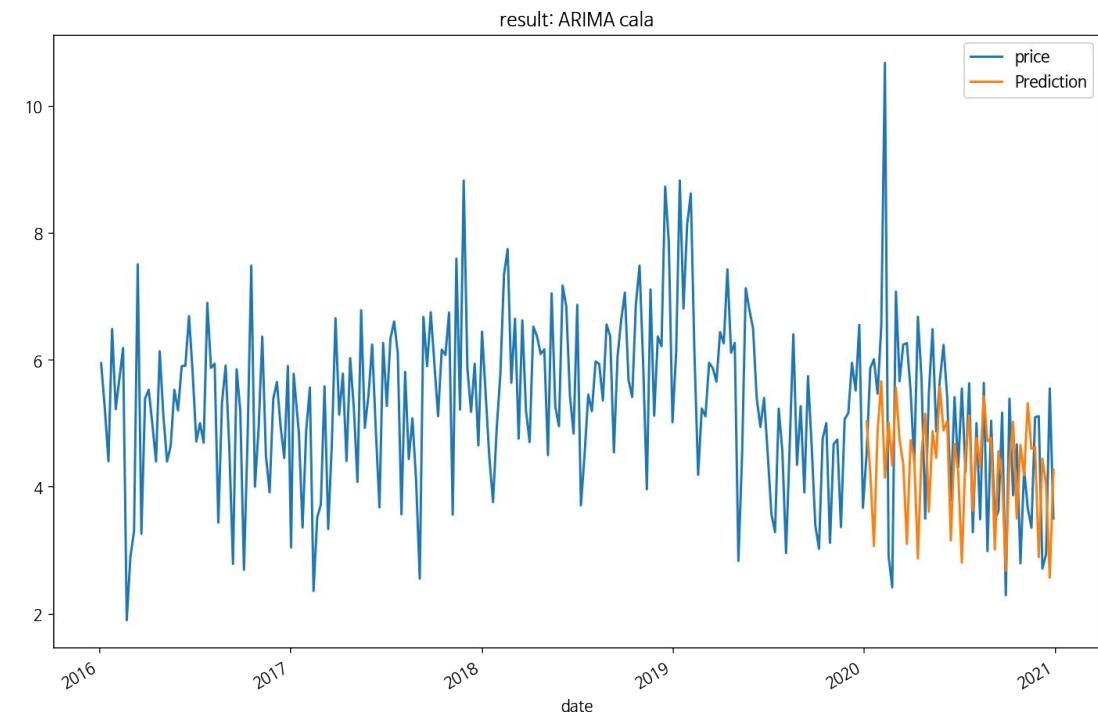


오징어 검증력

오징어 SARIMA (0, 1, 1) (2, 1, 0, 12)

```
SARIMAX Results
=====
Dep. Variable:                      y     No. Observations:      207
Model:             SARIMAX(0, 1, 1)x(2, 1, [], 12)   Log Likelihood   -343.529
Date:                Tue, 14 Sep 2021   AIC            695.057
Time:          15:09:37           BIC            708.129
Sample:                 0 - 207        HQIC           700.350
Covariance Type:            opg
=====
            coef    std err      z   P>|z|      [0.025      0.975]
-----
ma.L1     -0.8697    0.040  -21.661      0.000    -0.948     -0.791
ar.S.L12   -0.6374    0.078   -8.147      0.000    -0.791     -0.484
ar.S.L24   -0.1653    0.086   -1.921      0.055    -0.334      0.003
sigma2     1.9537    0.215    9.090      0.000     1.532     2.375
Ljung-Box (L1) (Q):            0.12  Jarque-Bera (JB):       0.52
Prob(Q):                  0.73  Prob(JB):            0.77
Heteroskedasticity (H):      1.14  Skew:                  -0.02
Prob(H) (two-sided):        0.61  Kurtosis:            2.75
=====
```

비교 그래프



계절적 시계열 분석 (SARIMA): 범주형 데이터



연어 검증력

연어 SARIMA (2, 1, 0) (2, 1, 0, 12)

SARIMAX Results

```
=====
Dep. Variable: y No. Observations: 208
Model: SARIMAX(2, 1, 0)x(2, 1, 0, 12) Log Likelihood: -318.631
Date: Tue, 14 Sep 2021 AIC: 647.263
Time: 15:37:11 BIC: 663.628
Sample: 0 HQIC: 653.889
                                                - 208
Covariance Type: opg
=====
```

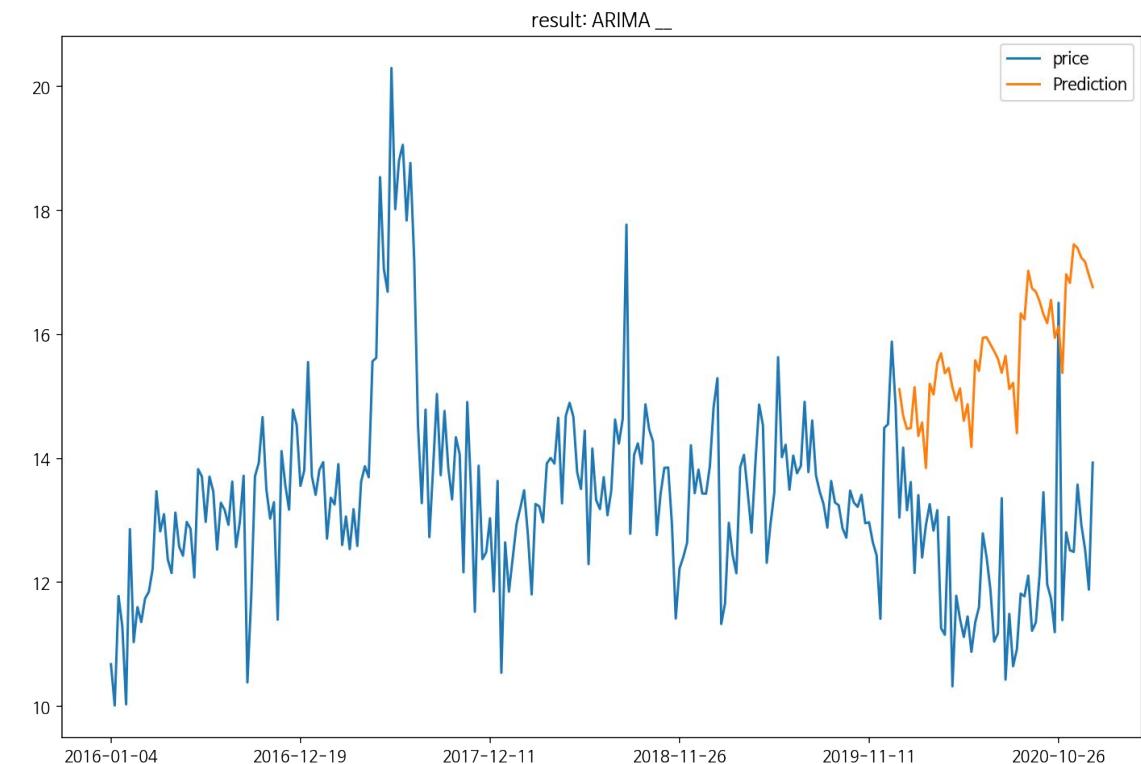
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.5282	0.061	-8.649	0.000	-0.648	-0.409
ar.L2	-0.2512	0.073	-3.428	0.001	-0.395	-0.108
ar.S.L12	-0.6115	0.069	-8.924	0.000	-0.746	-0.477
ar.S.L24	-0.3234	0.075	-4.332	0.000	-0.470	-0.177
sigma2	1.4919	0.132	11.293	0.000	1.233	1.751

=====

Ljung-Box (L1) (Q):	0.07	Jarque-Bera (JB):	9.32
Prob(Q):	0.80	Prob(JB):	0.01
Heteroskedasticity (H):	0.91	Skew:	-0.19
Prob(H) (two-sided):	0.71	Kurtosis:	4.00

=====

비교 그래프



계절적 시계열 분석 (SARIMA): 범주형 데이터 검증력 분석

MAE (Mean Absolute Error): 값이 낮을수록 모델의 설명력이 높다.

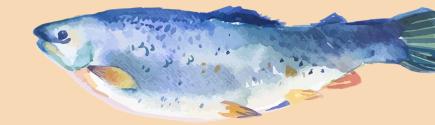
새우의 경우 우리의 모델은 실제값과 평균 1kg당 \$0.66, 오징어는 1kg당 \$1.28, 연어는 1kg당 \$3.5의 차이를 보인다.



corr	-0.0226
mae	0.66727
mape	0.07003
me	0.13480
minmax	0.06516
mpe	0.02048
rmse	0.84032



corr	0.05276
mae	1.28101
mape	0.27175
me	-0.5021
minmax	0.23079
mpe	-0.0202
rmse	1.71522



corr	0.10856
mae	3.50656
mape	0.29628
me	3.49194
minmax	0.22105
mpe	0.29540
rmse	3.75020

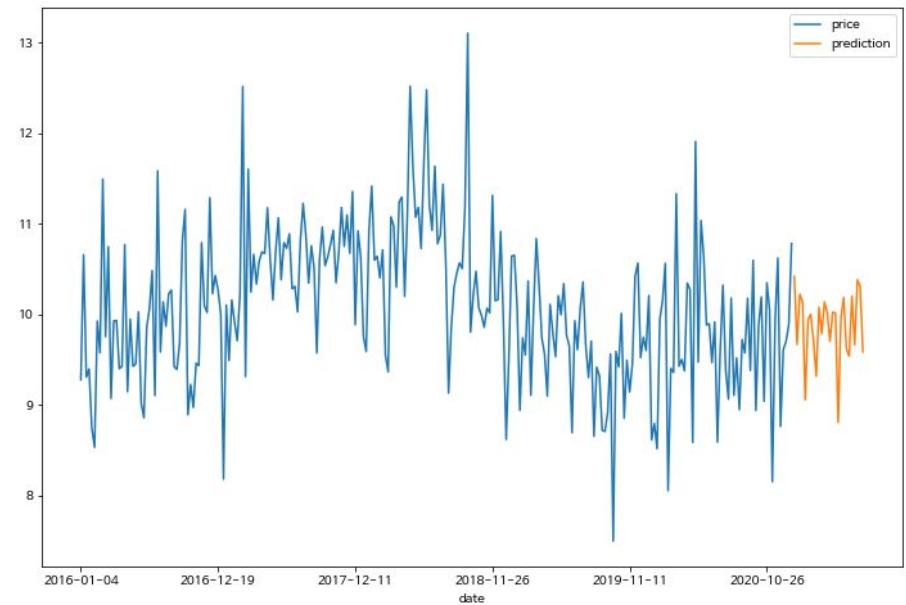
계절적 시계열 분석 (SARIMA): 범주형 데이터 새우 미래 예측



새우 SARIMA (0, 1, 1) (2, 1, 0, 12)

```
SARIMAX Results
=====
Dep. Variable:                      y   No. Observations:                  260
Model:             SARIMAX(0, 1, 1)x(2, 1, [ ], 12)   Log Likelihood:          -307.441
Date:                Tue, 14 Sep 2021   AIC:                         622.882
Time:                    12:39:49   BIC:                         636.919
Sample:                   0 - 260   HQIC:                        628.533
Covariance Type:            opg
=====
            coef    std err      z   P>|z|    [0.025    0.975]
ma.L1     -0.8875    0.033  -26.675   0.000    -0.953    -0.822
ar.S.L12   -0.6553    0.053  -12.397   0.000    -0.759    -0.552
ar.S.L24   -0.3057    0.064   -4.757   0.000    -0.432    -0.180
sigma2     0.6841    0.060   11.386   0.000     0.566     0.802
=====
Ljung-Box (L1) (Q):                 0.13   Jarque-Bera (JB):        0.94
Prob(Q):                           0.71   Prob(JB):                  0.63
Heteroskedasticity (H):              1.03   Skew:                     0.05
Prob(H) (two-sided):                0.88   Kurtosis:                 3.28
=====
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

미래값 예측



계절적 시계열 분석 (SARIMA): 범주형 데이터



오징어 미래 예측

오징어 SARIMA (0, 1, 1) (2, 1, 0, 12)

```
SARIMAX Results
=====
Dep. Variable:                      y
No. Observations:                  259
Model:             SARIMAX(0, 1, 1)x(2, 1, [1], 12)
Date:                Tue, 14 Sep 2021
Time:                12:38:32
Sample:                   0 - 259
Covariance Type:            opg
=====

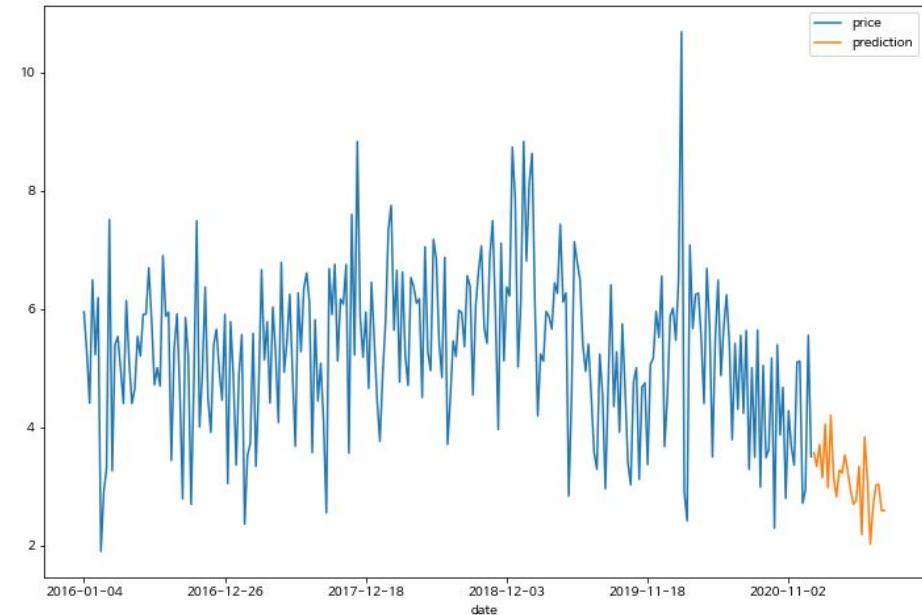
```

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.8618	0.029	-29.375	0.000	-0.919	-0.804
ar.S.L12	-0.6732	0.070	-9.642	0.000	-0.810	-0.536
ar.S.L24	-0.2524	0.071	-3.567	0.000	-0.391	-0.114
sigma2	2.1742	0.186	11.713	0.000	1.810	2.538

```
Ljung-Box (L1) (Q):      0.53    Jarque-Bera (JB):     2.71
Prob(Q):                 0.47    Prob(JB):          0.26
Heteroskedasticity (H):  1.31    Skew:              0.09
Prob(H) (two-sided):    0.22    Kurtosis:         3.48
=====
```

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

미래값 예측



계절적 시계열 분석 (SARIMA): 범주형 데이터

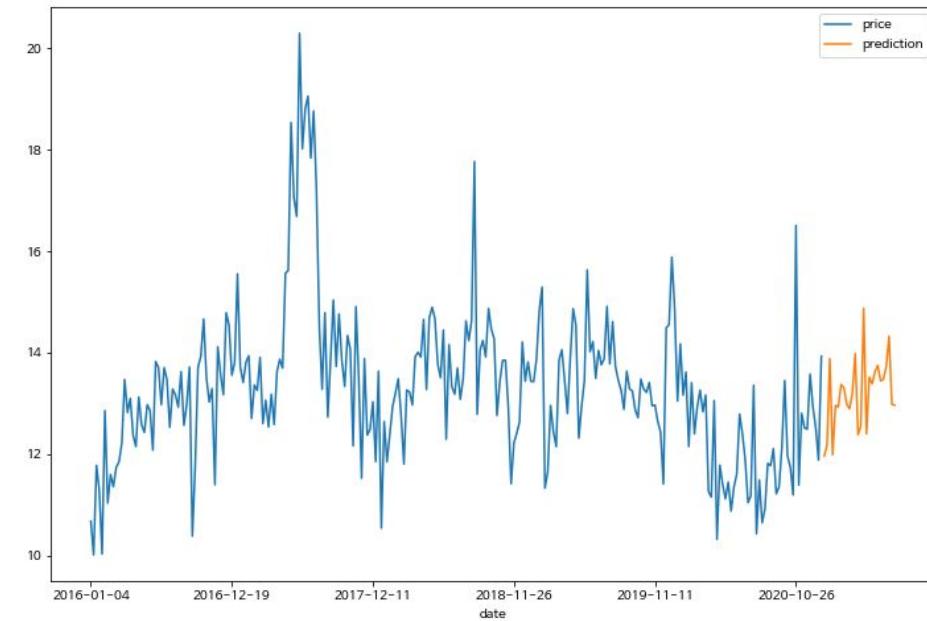


연어 미래 예측

연어 SARIMA (2, 1, 0) (2, 1, 0, 12)

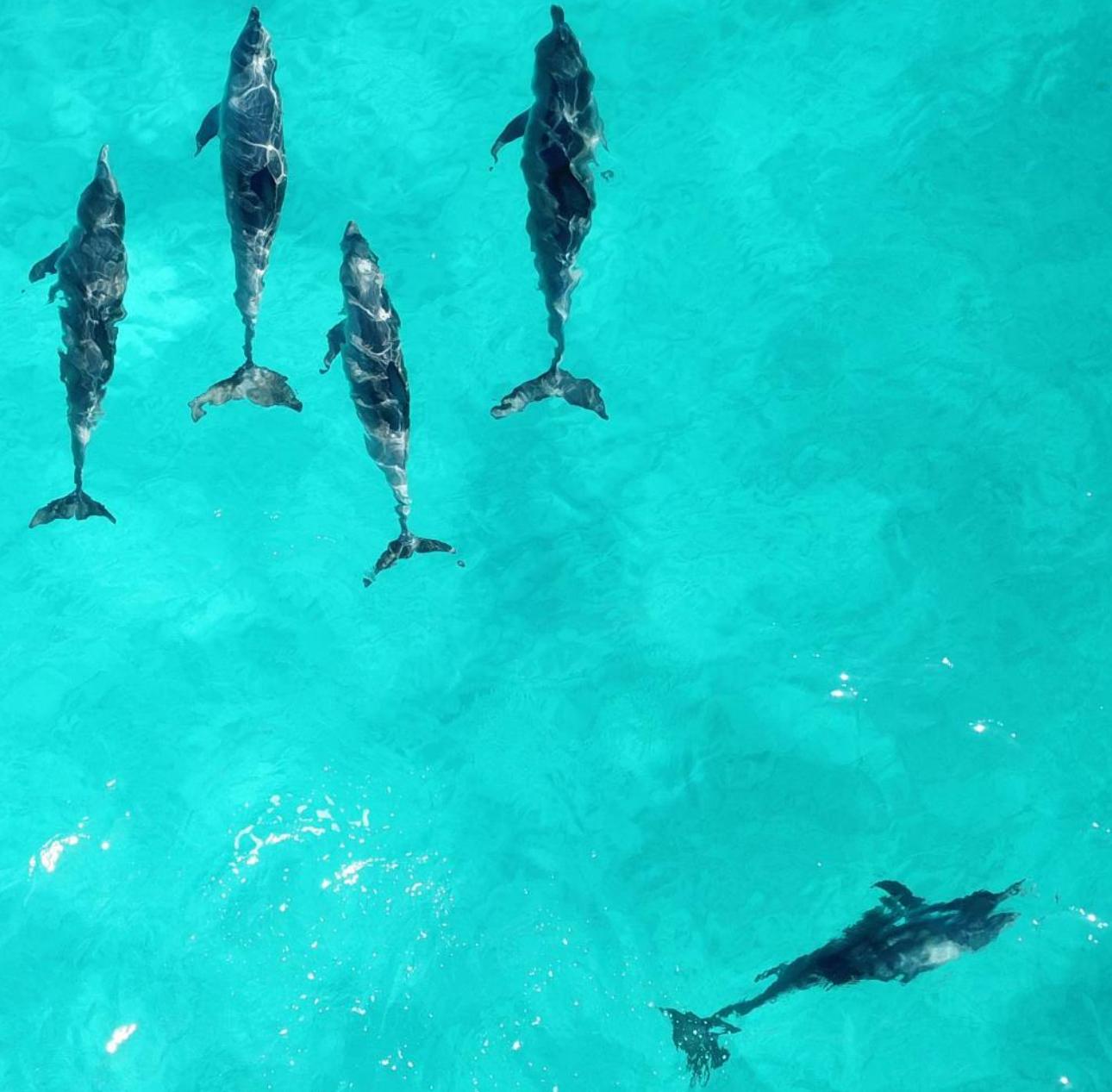
```
SARIMAX Results
=====
Dep. Variable: y No. Observations: 260
Model: SARIMAX(2, 1, 0)x(2, 1, 0, 12) Log Likelihood: -406.104
Date: Tue, 14 Sep 2021 AIC: 822.209
Time: 12:40:15 BIC: 839.756
Sample: 0 HQIC: 829.273
                                                - 260
Covariance Type: opg
=====
            coef    std err      z   P>|z|   [0.025    0.975]
-----
ar.L1     -0.5926   0.053  -11.223   0.000   -0.696   -0.489
ar.L2     -0.2511   0.064   -3.900   0.000   -0.377   -0.125
ar.S.L12   -0.6593   0.064  -10.296   0.000   -0.785   -0.534
ar.S.L24   -0.2872   0.068   -4.198   0.000   -0.421   -0.153
sigma2    1.5307   0.106   14.385   0.000    1.322    1.739
-----
Ljung-Box (L1) (Q): 0.01 Jarque-Bera (JB): 34.69
Prob(Q): 0.94 Prob(JB): 0.00
Heteroskedasticity (H): 0.95 Skew: 0.20
Prob(H) (two-sided): 0.83 Kurtosis: 4.79
-----
Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

미래값 예측



Part 4.

결과 및 논의



결과 및 논의



새우의 MAE는 0.66으로
예측값은 실제값과는 평균
1kg당 \$0.66 차이를
보이며, 0.07의 MAPE로
93%의 예측도를 보인다.



오징어의 MAE는 1.28로
예측값은 실제값과는 평균
1kg당 \$1.28 차이를 보이며.
0.27의 MAPE로 새우보다는
낮은 73%의 예측도를
보인다.



연어의 MAE는 3.5로
예측값은 실제값과는 평균
1kg당 \$3.5 차이를 보이며.
이는 약 0.29의 MAPE로
다른 어종에 비해 가장 낮은
71%의 예측도를 보인다.

결과 및 논의

재편되는 수산물 유통, 보다 다양한 어종의 예측모델 필요

새우, 오징어, 연어 뿐 아닌, 다양한 어종의 다양한 세분화된 분야로 수입가격을 예측한다면 유통업계의 불안정성 완화에 도움을 주리라 기대됨

수입원료 가격을 예측

국내 자영업자 및 소비자의 권익 보호

코로나19로 인한 국경의 차단, 공급망의 불확실성

정체된 글로벌 공급망에 대비할 수 있는 예측모델

감사합니다

BANANA BEACH