

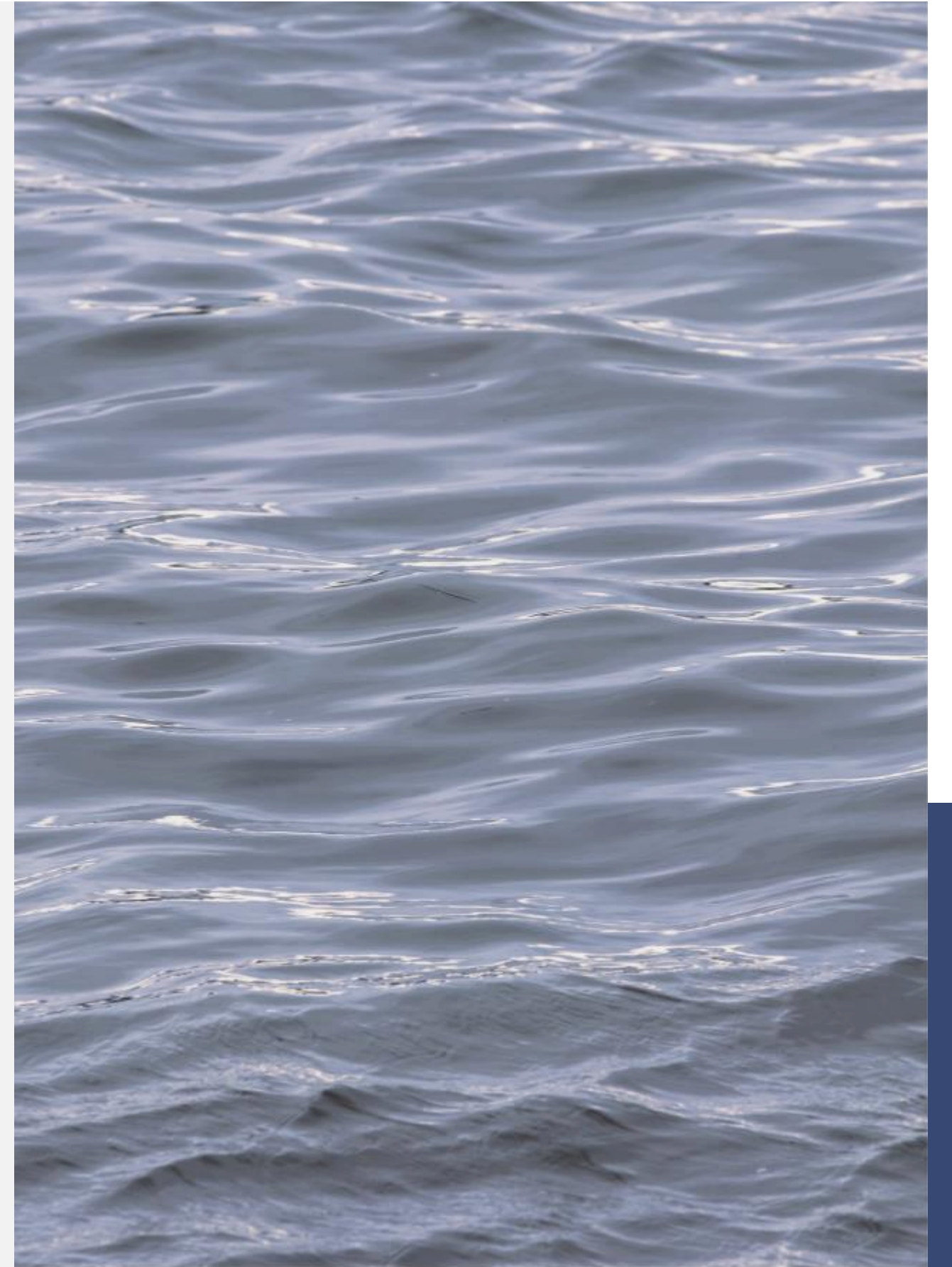
JOISS 해양과학 빅데이터 경진대회

# 남향 수송량의 장기추세 산출 및 예측

---



DS 수산시장



# 발표 순서



▶ 연구 주제 소개

▶ EDA

▶ Data Preprocessing

▶ Western UIG 수송량

▶ 상관관계식 산출

▶ U3 데이터 상관관계식 대입 및 수송량 장기 변화 추세 산출

▶ 2030년도 수송량 예측

# 연구 주제 소개



## 연구목적

2022년 서울 동작구에서 하루 동안 382mm의 비가 내려 1920년 이후 102년 만에 최고치를 기록했다.  
위와 같은 기상이변은 한국 뿐만 아니라 전 세계가 몸소 체험하고 있다.  
이렇게 전 세계가 겪고 있는 폭염과 가뭄, 홍수 등 극단적인 이상 기후의 주범으로 **지구온난화**가 지목되고 있다.

지구온난화가 기후를 조정하는 해류 순환체계를 완전히 붕괴시키고 있다는 분석이 나왔다.  
기온상승으로 기후가 변하면서 해수의 온도가 들쭉날쭉해졌고, 그 결과 바닷물의 흐름이 안정성을 잃은 탓이다.

이러한 분석결과로 50년 이내에 순환체계가 붕괴해버릴 수 있다는 가능성이 제기됐다.

➡ 따라서 남향으로 흐르는 유속과 수송량의 추세와 2030년의 수송량을 예측하고자 한다.

## 연구 주제 소개

---



### 심층순환이란?

심층순환은 수분과 염분의 차이에 의해서 생기는 밀도차로 생긴다.

해수의 밀도가 균일하게 분포하지 않을 경우, 밀도가 서로 다른 해수 사이에서 압력 기울기가 발생하여 해수가 이동한다.

또한 표층 해수의 온도가 낮아지거나 염분이 높아지게 되면 해수의 밀도가 커지게 되어 아래로 침강해 바닥을 따라 움직이는 느린 해류가 된다.

이와 같이 밀도 차이에 의한 순환은 염분과 온도에 영향을 받기 때문에 열염 순환이라고도 한다.

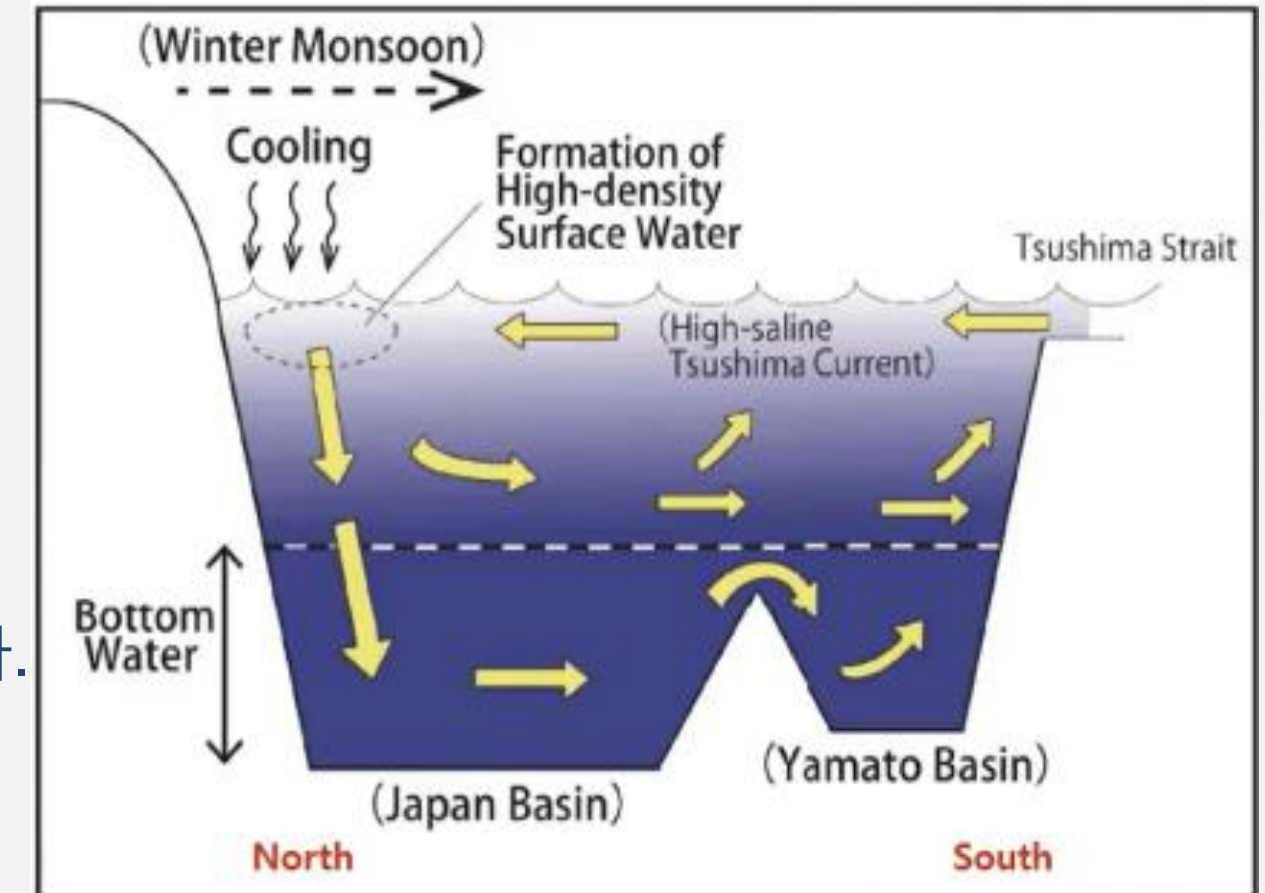


## 연구 주제 소개

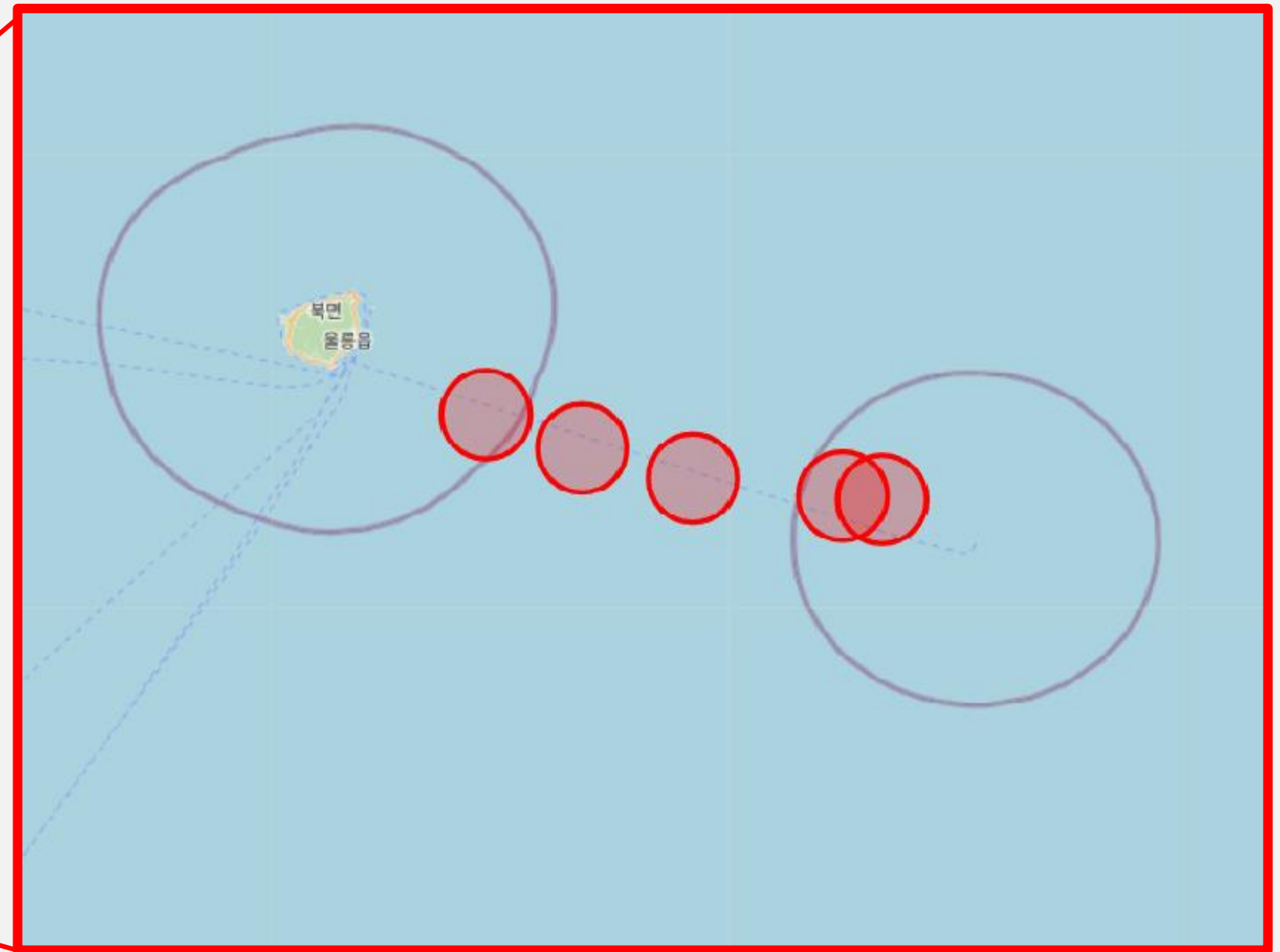
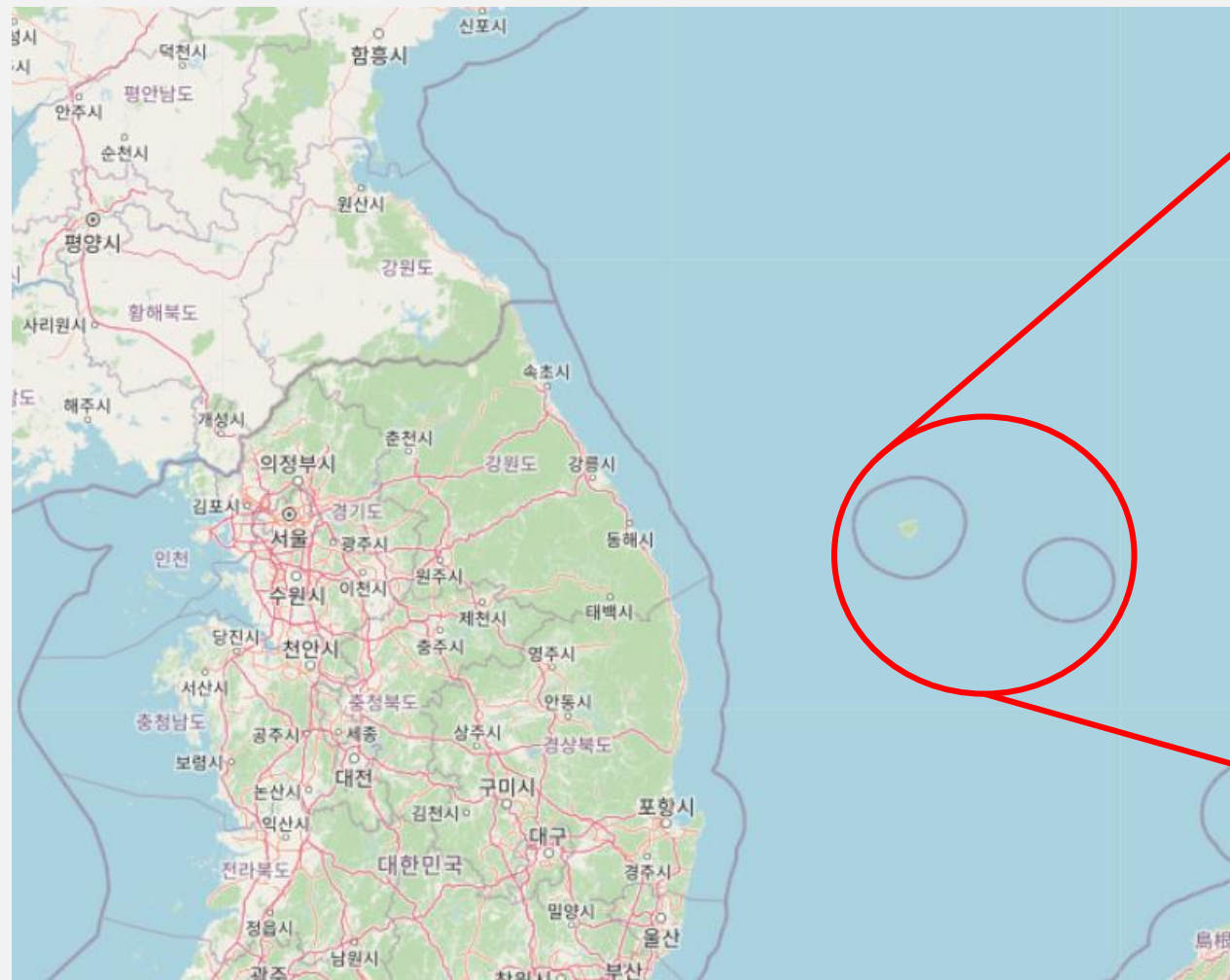


### 심층순환 발생과정

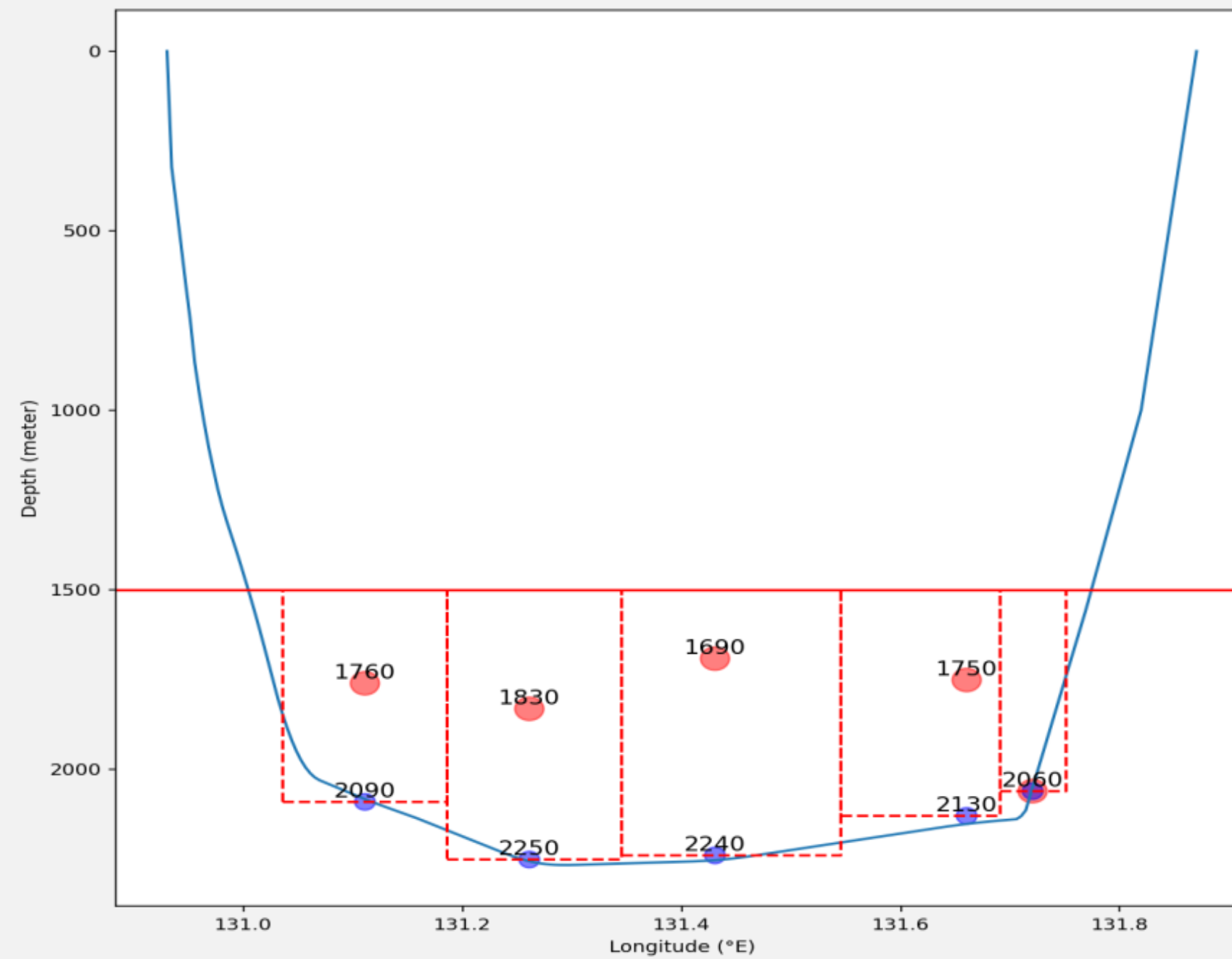
- ▶ 동해북부에서 차가운 공기를 만나 표층수가 냉각된다.
- ▶ 수온이 낮고 염분이 높아진 해수가 심층수로 가라앉는다.
- ▶ 해양 바닥을 따라 적도 쪽으로 서서히 이동하여 바다 전체를 순환한다.
- ▶ 따뜻해져 가벼워진 해수가 인도양과 태평양에서 떠올라 표층 순환과 이어진다.



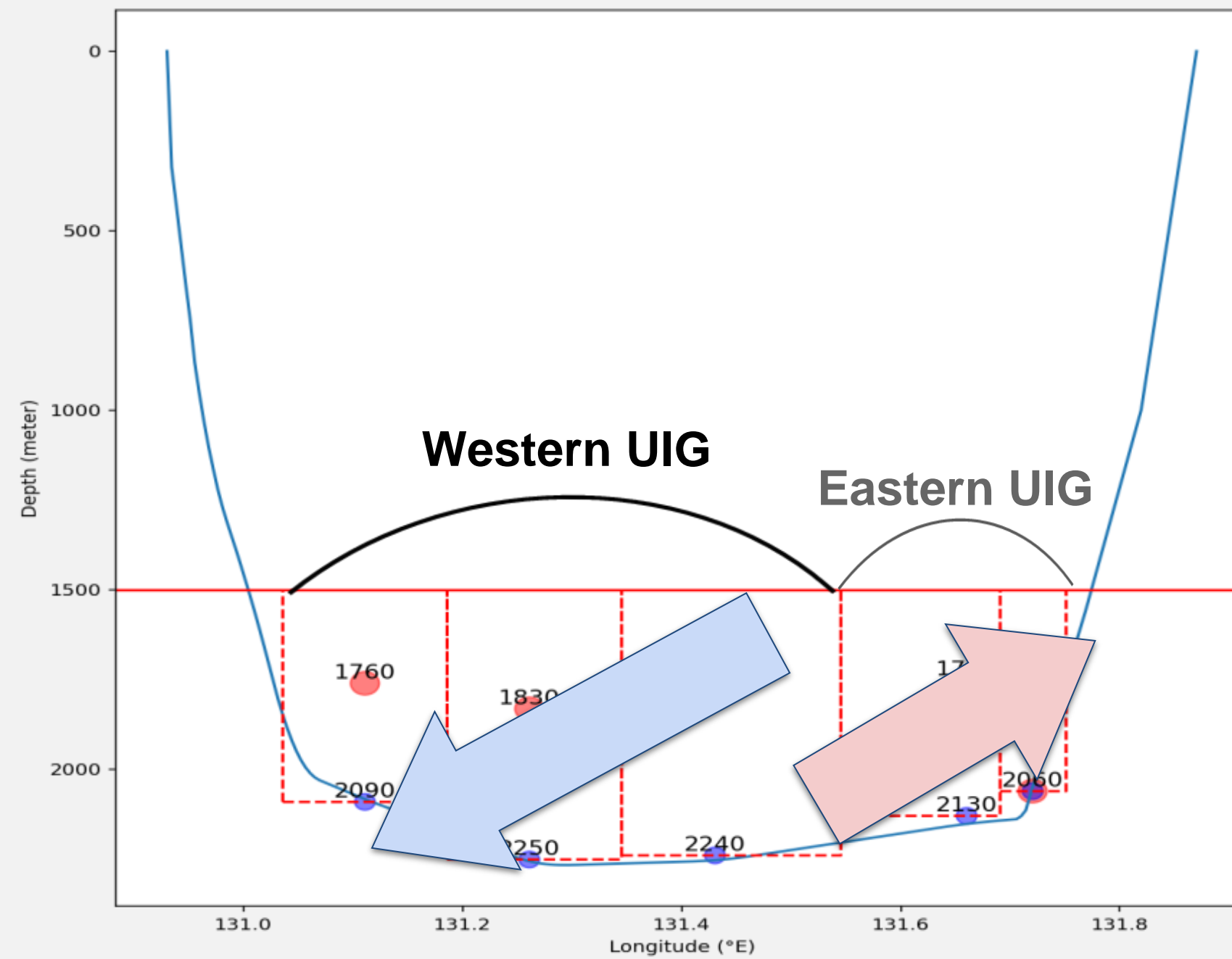
## U1 - U5 위치 확인



## 울릉도-독도 해저지형 단면도



## 울릉도-독도 해저지형 단면도





# Data Preprocessing



## 데이터 날짜

### 1) 데이터들의 첫 행의 시작 날짜와 마지막 날짜가 다름을 확인(위에 있는 수심 데이터)

Year	Month	Day	Hour
2002	11	30	11
2002	11	30	12
...	...	...	...
2004	4	7	10
2004	4	7	11

<U1\_1760>

Year	Month	Day	Hour
2002	11	30	13
2002	11	30	14
...	...	...	...
2004	4	7	1
2004	4	7	2

<U2\_1830>

Year	Month	Day	Hour
2002	11	29	17
2002	11	29	18
...	...	...	...
2004	4	7	10
2004	4	7	11

<U3\_1690>

- 기준
  - 가장 늦은 시작날짜를 기준으로 그 전의 날짜는 제거  
: (U2의 시작날짜를 기준으로 삼음 2002-11-30 13:00:00)
  - 가장 빠른 마지막날짜를 기준으로 그 후의 날짜는 제거  
: (U2의 마지막날짜를 기준으로 삼음 2004-04-07 02:00:00)

# Data Preprocessing



## 데이터 날짜

2) 유속은  $\text{cm/s}^{-1}$ , 수심은 m 단위이므로 단위 변환(위에 있는 수심 데이터)

- 유속/100 과정을 거쳐 m단위로 맞춰줌

Year	Month	Day	Hour	Minute	Second	Depth (meter)	Ur Current speed (cm/s)
2002	11	30	13	0	0	1760	0.000069
2002	11	30	14	0	0	1760	0.000218
2002	11	30	15	0	0	1760	0.000531
2002	11	30	16	0	0	1760	0.001077
2002	11	30	17	0	0	1760	0.001919
...	...	...	...	...	...	...	...
2004	4	6	22	0	0	1760	0.106519
2004	4	6	23	0	0	1760	-0.092469
2004	4	7	0	0	0	1760	-0.283257
2004	4	7	1	0	0	1760	-0.465926
2004	4	7	2	0	0	1760	-0.640375

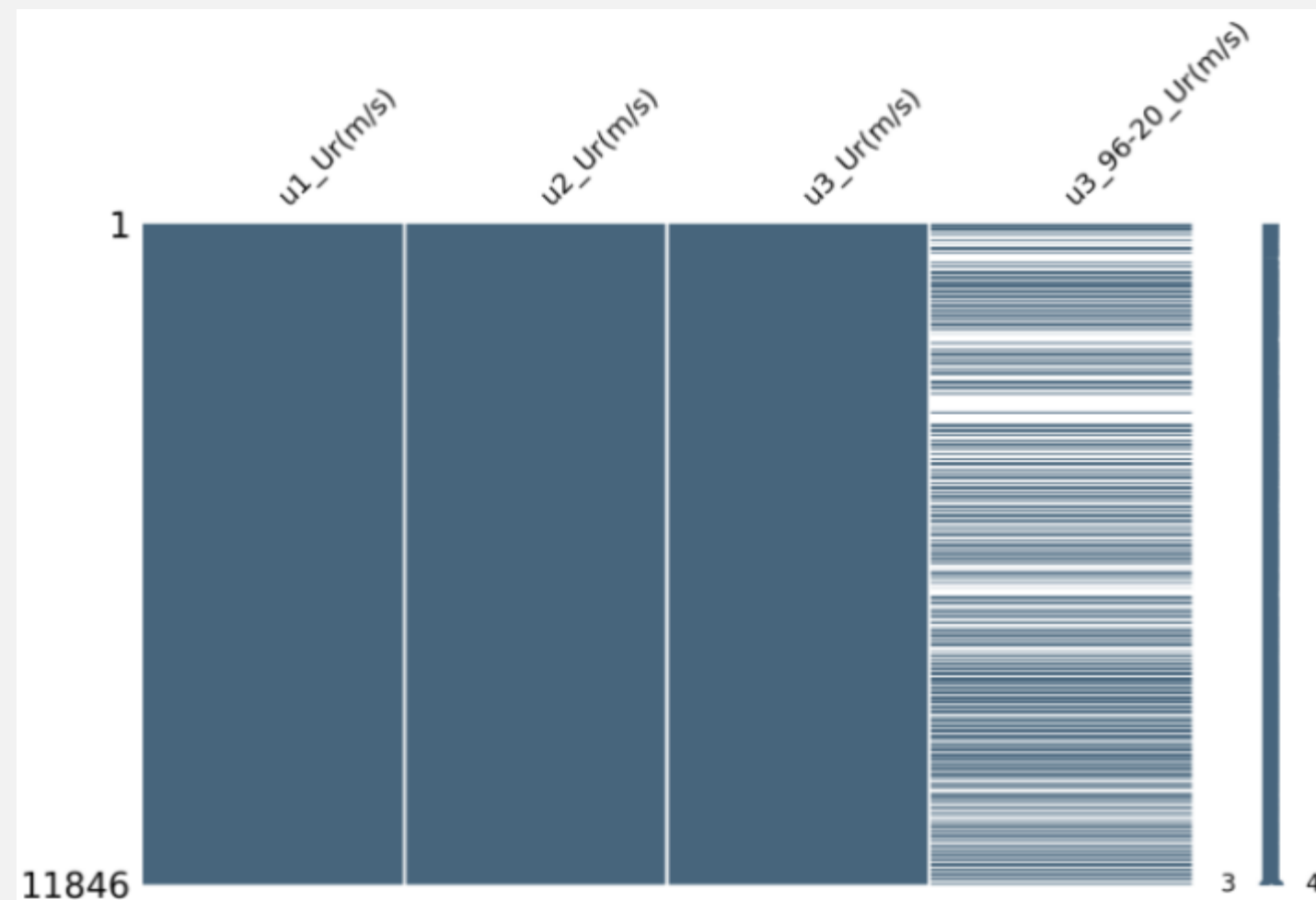


Ur Current speed (m/s)
6.916409e-07
2.182376e-06
5.308476e-06
1.077437e-05
1.918553e-05
...
1.065187e-03
-9.246915e-04
-2.832570e-03
-4.659258e-03
-6.403746e-03

# Data Preprocessing



## Interpolation (Linear)



u3(1996\_2020) 데이터의 유속 컬럼에 결측값(Nan)이 80134개로 많은 양의 결측값이 있는 것 확인

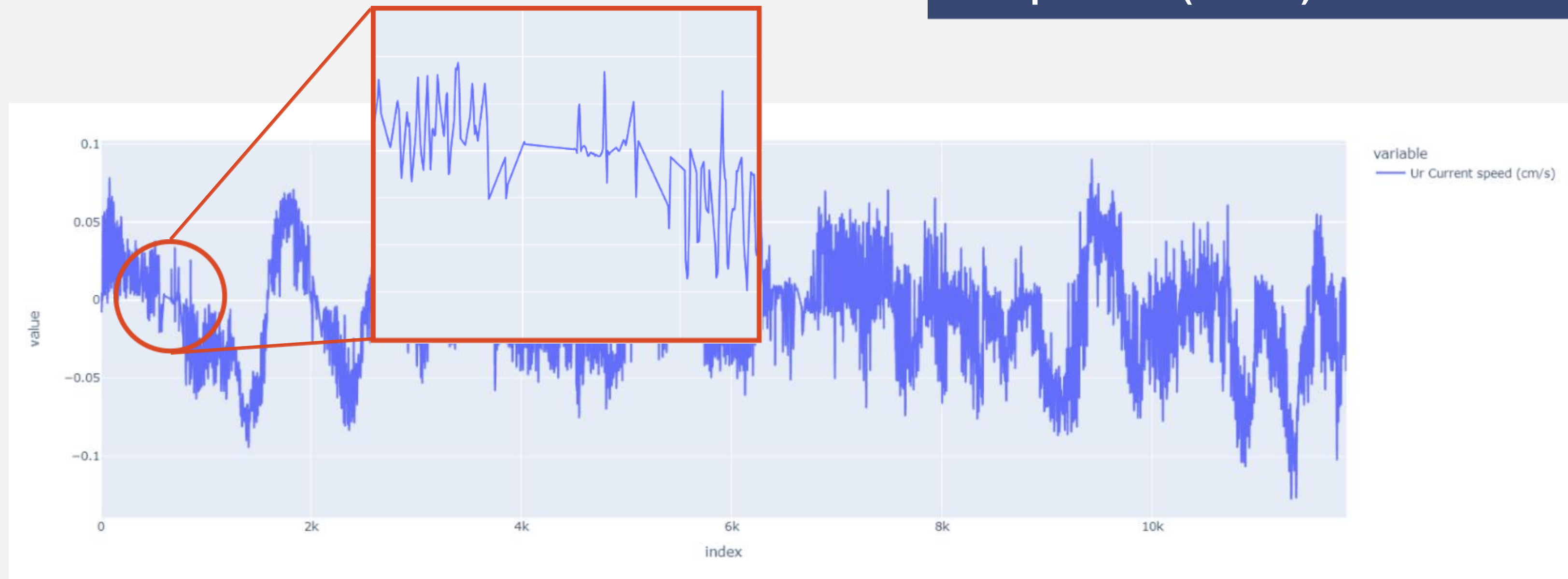
따라서 결측값을 제외하고 진행하는 것은 맞지 않다고 판단

따라서 데이터 보간법 중 하나인  
**Linear Interpolation**을 사용하기로 결정

# Data Preprocessing



## Interpolation (Linear)



# Data Preprocessing



습도, 기온, 수온 추세 (기상자료개방포털)

	일시	습도(%)	기온(°C)	수온(°C)
0	2011-12-28 00:00	45.0	6.2	11.9
1	2011-12-28 01:00	51.0	6.2	11.7
2	2011-12-28 02:00	42.0	6.5	12.4
3	2011-12-28 03:00	40.0	6.5	13.2
4	2011-12-28 04:00	41.0	6.7	13.3
...	...	...	...	...
78208	2020-12-31 19:00	NaN	2.8	13.7
78209	2020-12-31 20:00	NaN	3.2	13.6
78210	2020-12-31 21:00	NaN	3.2	13.6
78211	2020-12-31 22:00	NaN	3.6	13.6
78212	2020-12-31 23:00	NaN	3.5	13.6

유속에 영향을 미치는 요소인 ‘습도’, ‘기온’, ‘수온’의 열만 추출

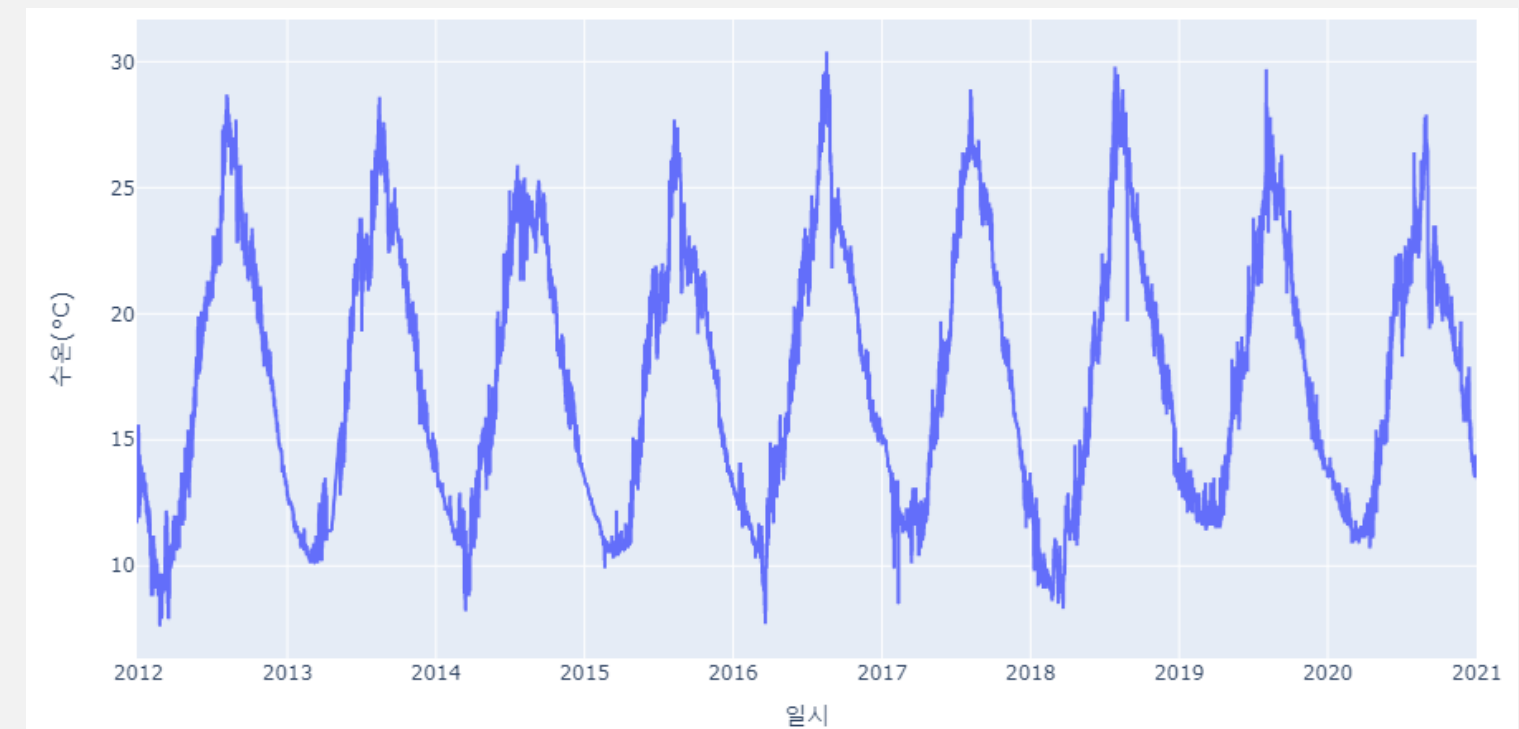
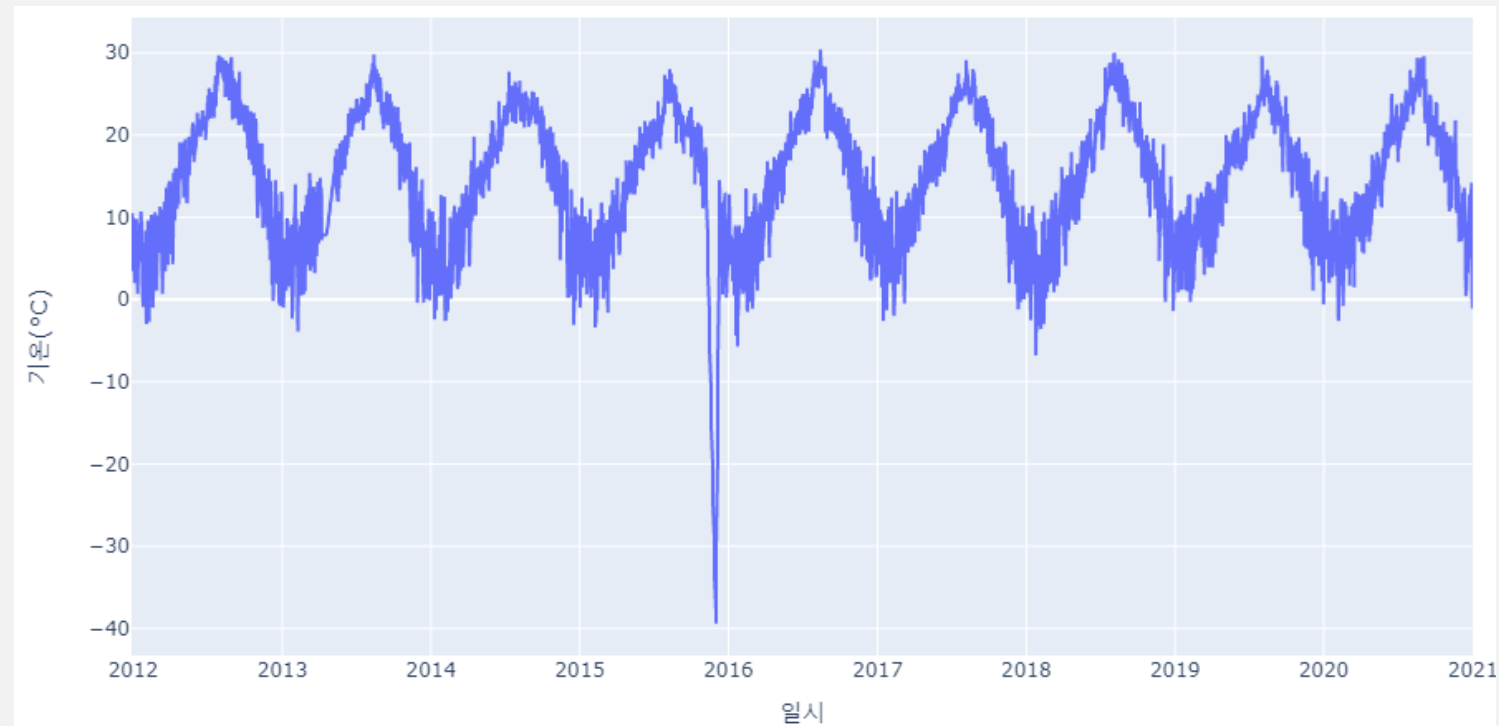
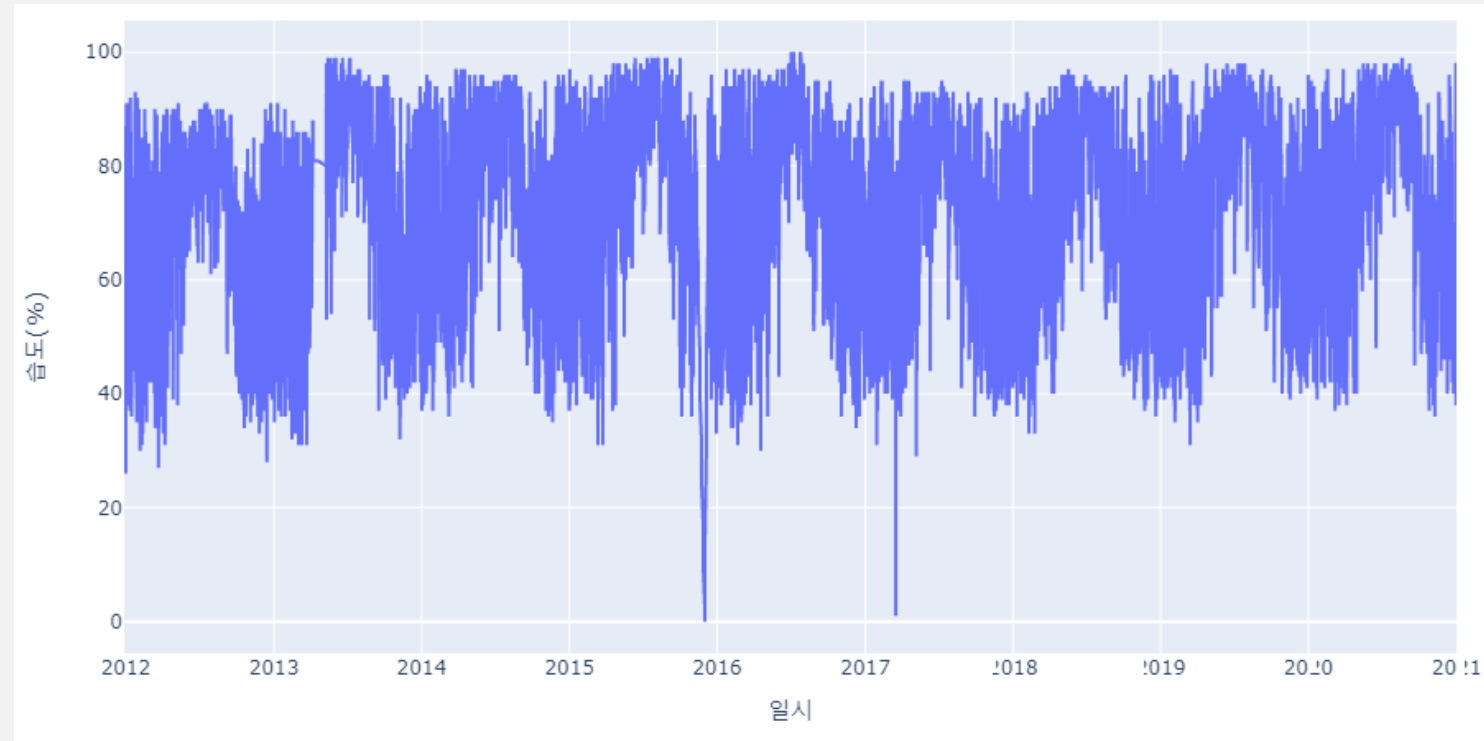
데이터의 결측치 확인 후 interpolation(Linear) 진행



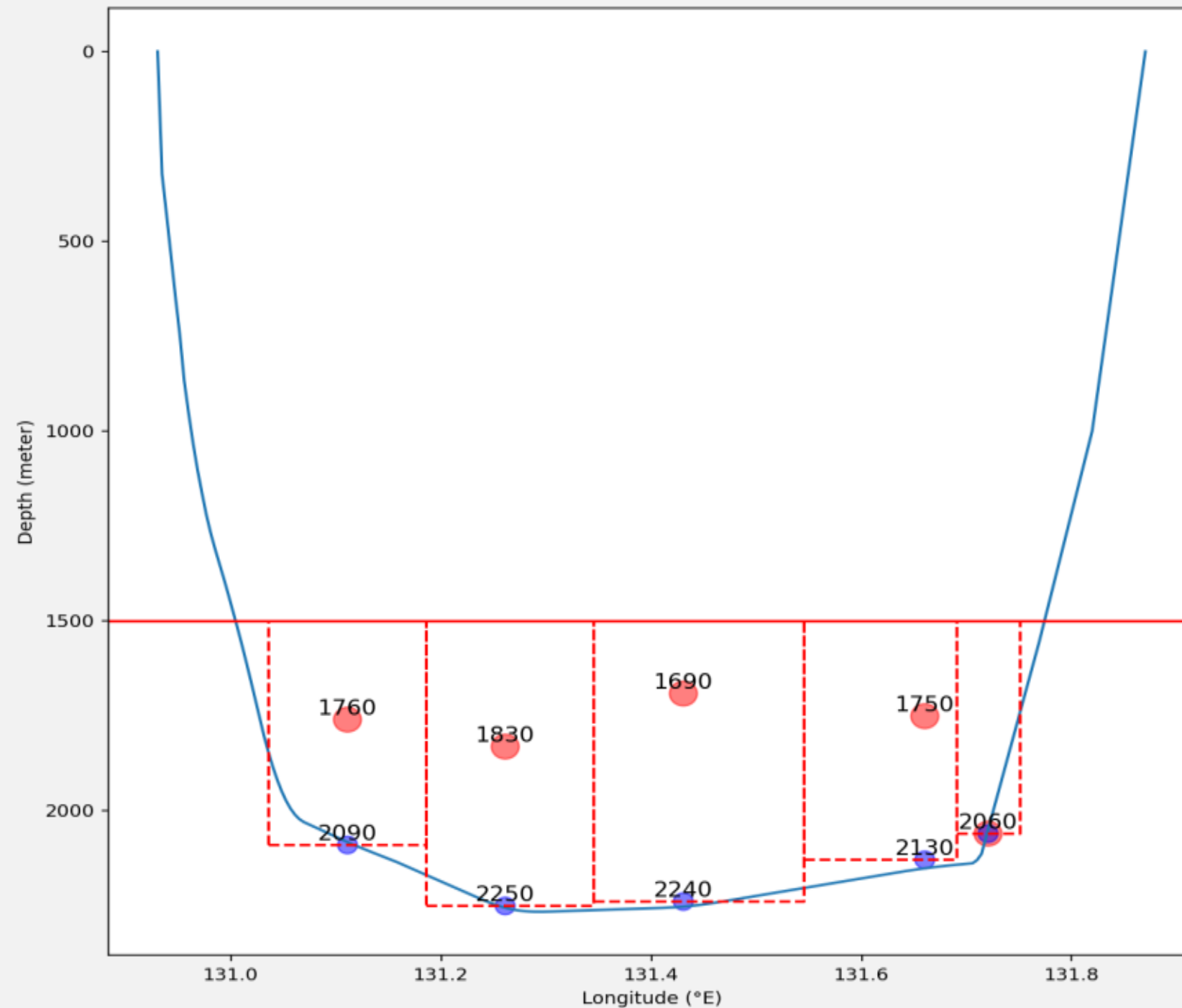
# Data Preprocessing



습도, 기온, 수온 추세 (기상자료개방포털)



# Western UIG 수송량



1. 각 지점 수심 계산
2. 각 지점 가로길이 계산
3. 각 지점 면적 계산
4. 각 지점 수송량 계산
5. Western UIG 수송량 계산

# Western UIG 수송량



## 각 지점 수심 계산

1500m 이하의 수심이므로  
제일 밑 수심 - 1500

	Station	Depth(m)	Width(m)	Size	Transport
0	U1	590	14029.736941	8.277545e+06	-152.800859
1	U2	750	14815.420639	1.111157e+07	-1292.757471
2	U3	740	18032.416611	1.334399e+07	-1477.061361

## 각 구간 가로길이 계산

haversine 함수를 이용해 지점간의 거리 계산

u1와 u2 거리를 u1\_u2  
u2와 u3의 거리를 u2\_u3  
u3와 u4의 거리를 u3\_u4 라 했을때

u1의 가로길이 : u1\_u2  
u2의 가로길이 :  $(u1\_u2 + u2\_u3) / 2$   
u3의 가로길이 :  $(u2\_u3 + u3\_u4) / 2$

	Station	Depth(m)	Width(m)	Size	Transport
0	U1	590	14029.736941	8.277545e+06	-152.800859
1	U2	750	14815.420639	1.111157e+07	-1292.757471
2	U3	740	18032.416611	1.334399e+07	-1477.061361

# Western UIG 수송량



## 각 구간의 면적 계산

각 지점의 수심 X 각 지점의 가로길이

	Station	Depth(m)	Width(m)	Size	Transport
0	U1	590	14029.736941	8.277545e+06	-152.800859
1	U2	750	14815.420639	1.111157e+07	-1292.757471
2	U3	740	18032.416611	1.334399e+07	-1477.061361

## 각 구간의 수송량 계산

수송량 = 지점의 면적 X 유속

Sverdrup 단위를 맞춰주기 위해  
수송량/1000000

	Station	Depth(m)	Width(m)	Size	Transport
0	U1	590	14029.736941	8.277545e+06	-152.800859
1	U2	750	14815.420639	1.111157e+07	-1292.757471
2	U3	740	18032.416611	1.334399e+07	-1477.061361

Western UIG 수송량

---



Western UIG 총 수송량

$(U1 + U2 + U3)$

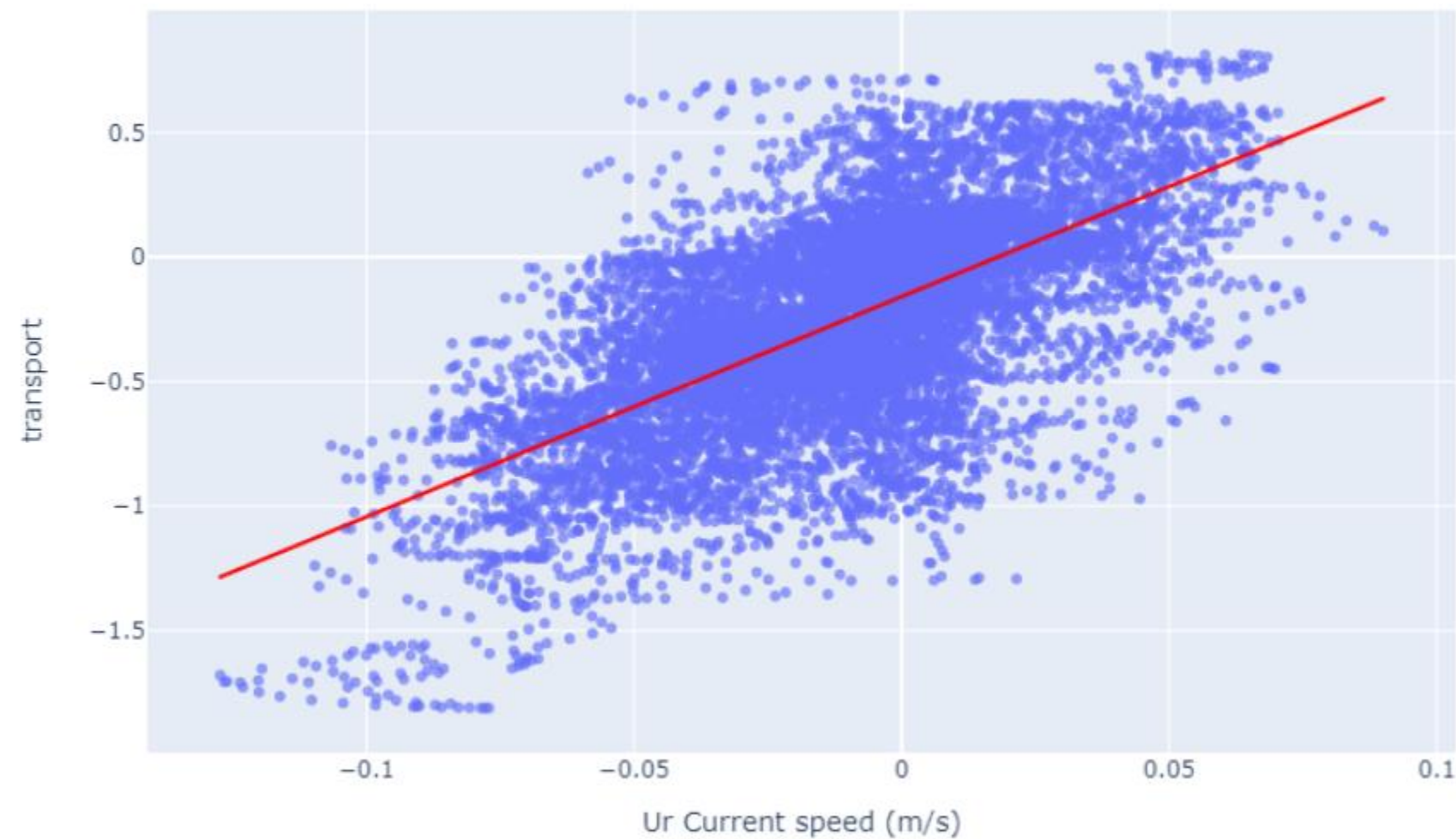
**= - 2922.619691060501 Sv**



## 상관관계식 산출



### 그래프 및 상관관계식 산출



polyfit 함수를 이용하여 회귀식을 산출

$$y=8.834922816434565x - 0.1593419390106443$$

$R^2$  (결정계수) : 0.42005482803748473

# U3 데이터 상관관계식 대입

U3 데이터 상관관계식 대입 및 수송량 장기 변화 추세 산출



	Station	date	Ur Current speed (m/s)
0	U3(EC1)	1996-11-04 22:00:00	-0.051551
1	U3(EC1)	1996-11-04 23:00:00	-0.049471
2	U3(EC1)	1996-11-05 00:00:00	-0.051176
3	U3(EC1)	1996-11-05 01:00:00	-0.053790
4	U3(EC1)	1996-11-05 02:00:00	-0.055360
...	...	...	...

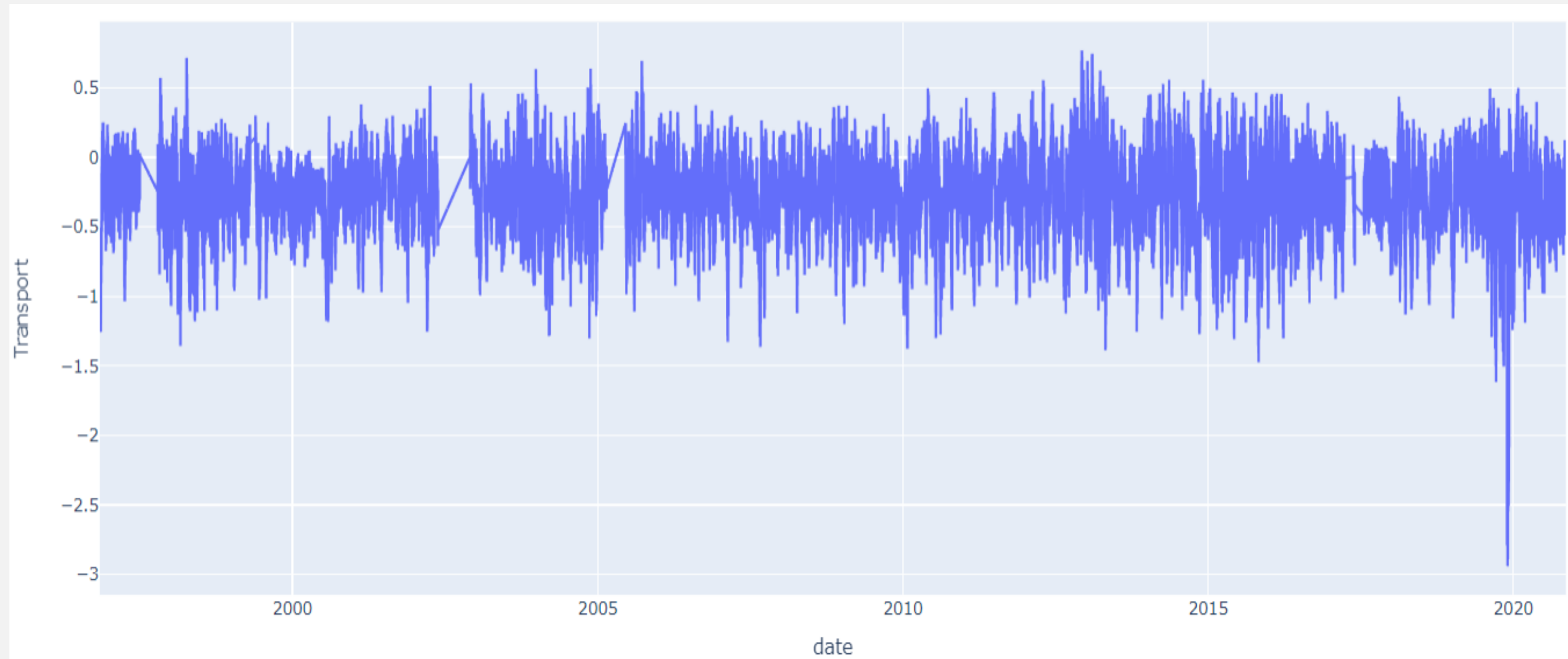
$$y = 8.834922816434565x - 0.1593419390106443$$

x에 유속값 대입하여  
수송량값 y 산출

	Station	date	Ur Current speed (m/s)	Transport
0	U3(EC1)	1996-11-04 22:00:00	-0.051551	-0.614788
1	U3(EC1)	1996-11-04 23:00:00	-0.049471	-0.596412
2	U3(EC1)	1996-11-05 00:00:00	-0.051176	-0.611480
3	U3(EC1)	1996-11-05 01:00:00	-0.053790	-0.634573
4	U3(EC1)	1996-11-05 02:00:00	-0.055360	-0.648447
...	...	...	...	...

# U3 데이터 상관관계식 대입

U3 데이터 상관관계식 대입 및 수송량 장기 변화 추세 산출



## 2030년도 수송량 예측



습도, 기온, 수온  
예측



유속 예측

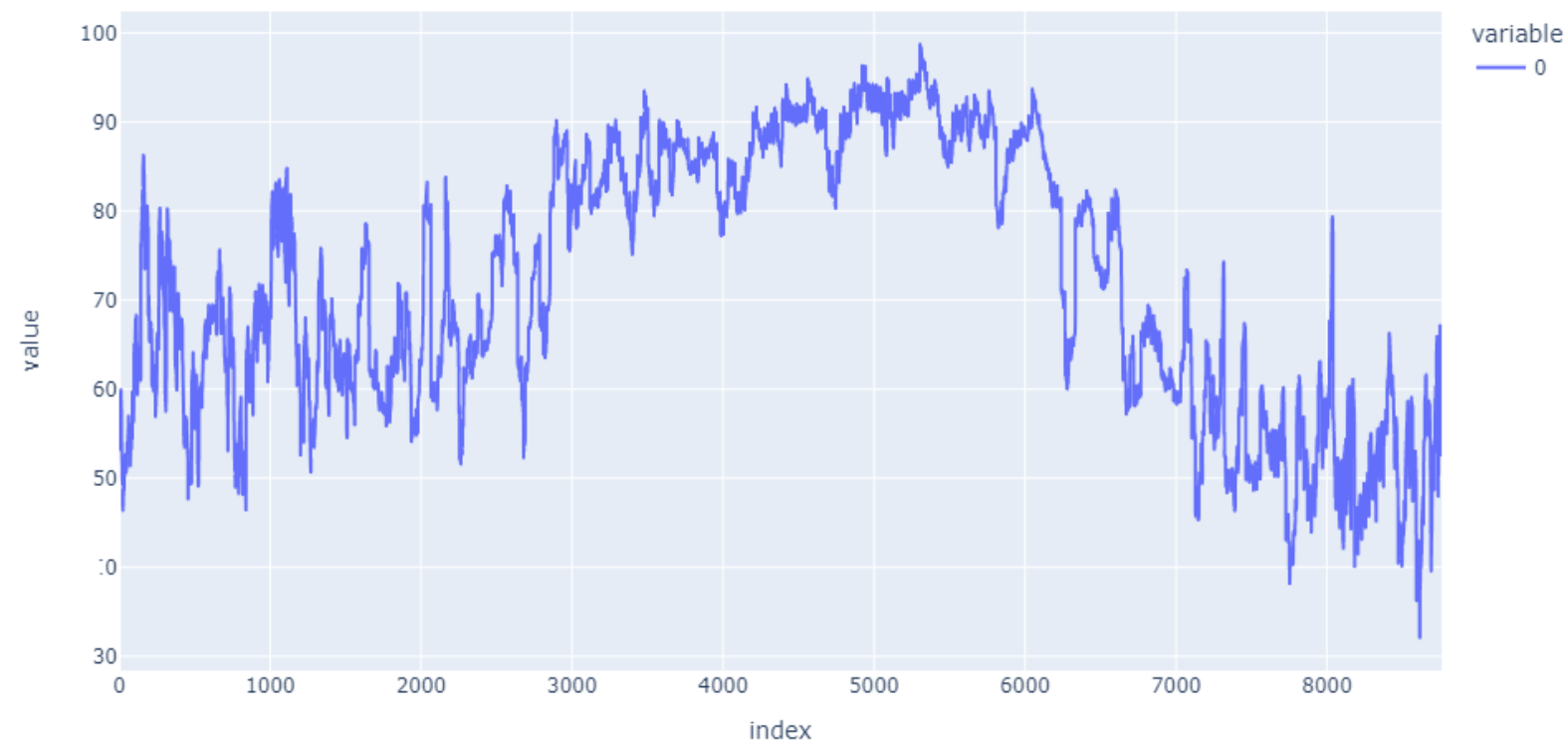


수송량 예측

# 2030년도 수송량 예측



## 습도, 기온, 수온 예측



<2030년의 습도 예측 그래프>

외부데이터 항목에

'Year', 'Month', 'Day', 'Hour', 'Minute', 'Second' 항목을 추가한 뒤  
독립변수에 넣고 종속변수에는 '습도(%)'를 넣어

XGBRegressor 모델을 학습

학습한 모델을 사용하여

2030-01-01-00:00:00 부터 2030-12-31-23:00:00 까지의 습도 예측

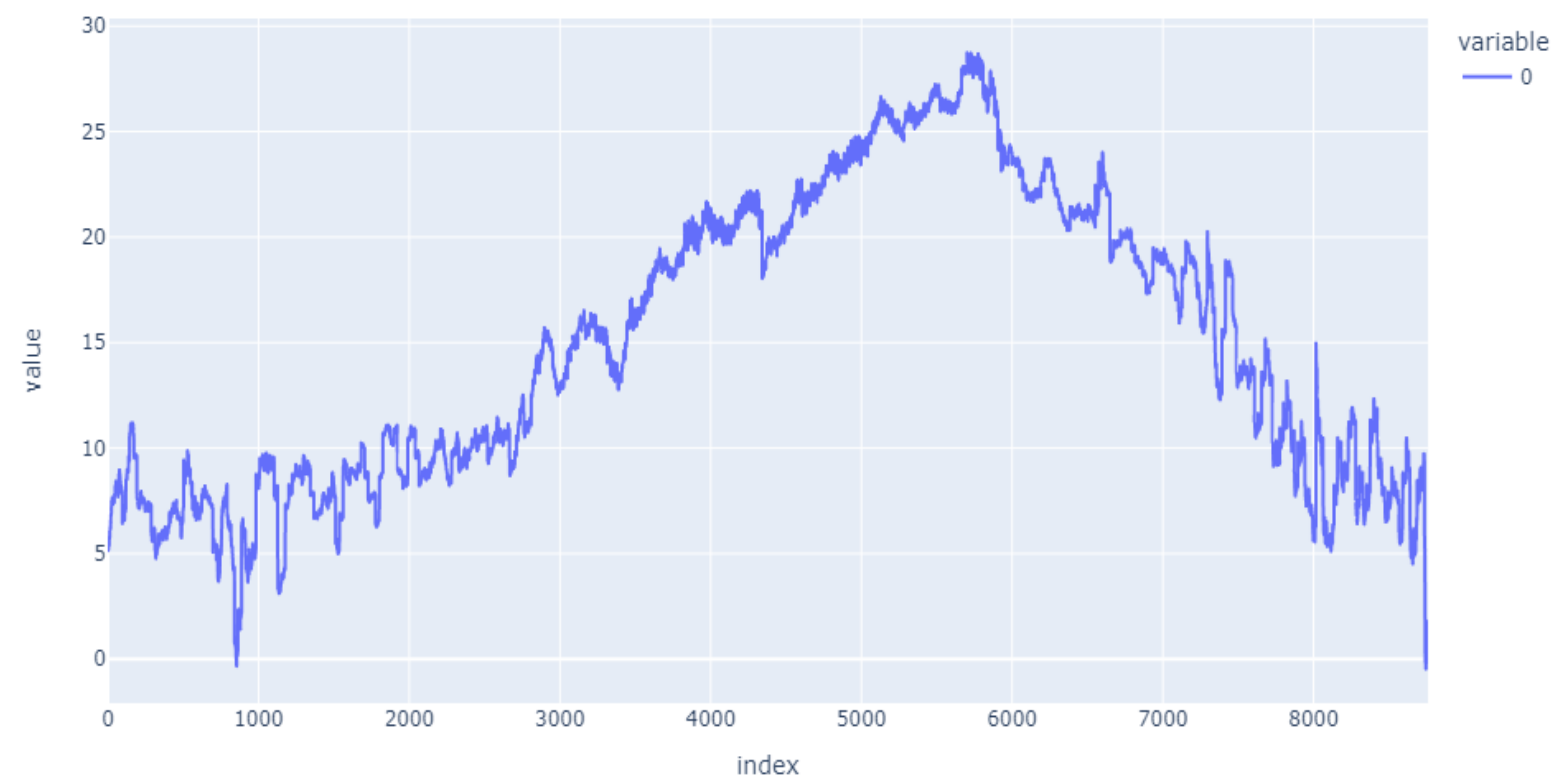


# 2030년도 수송량 예측

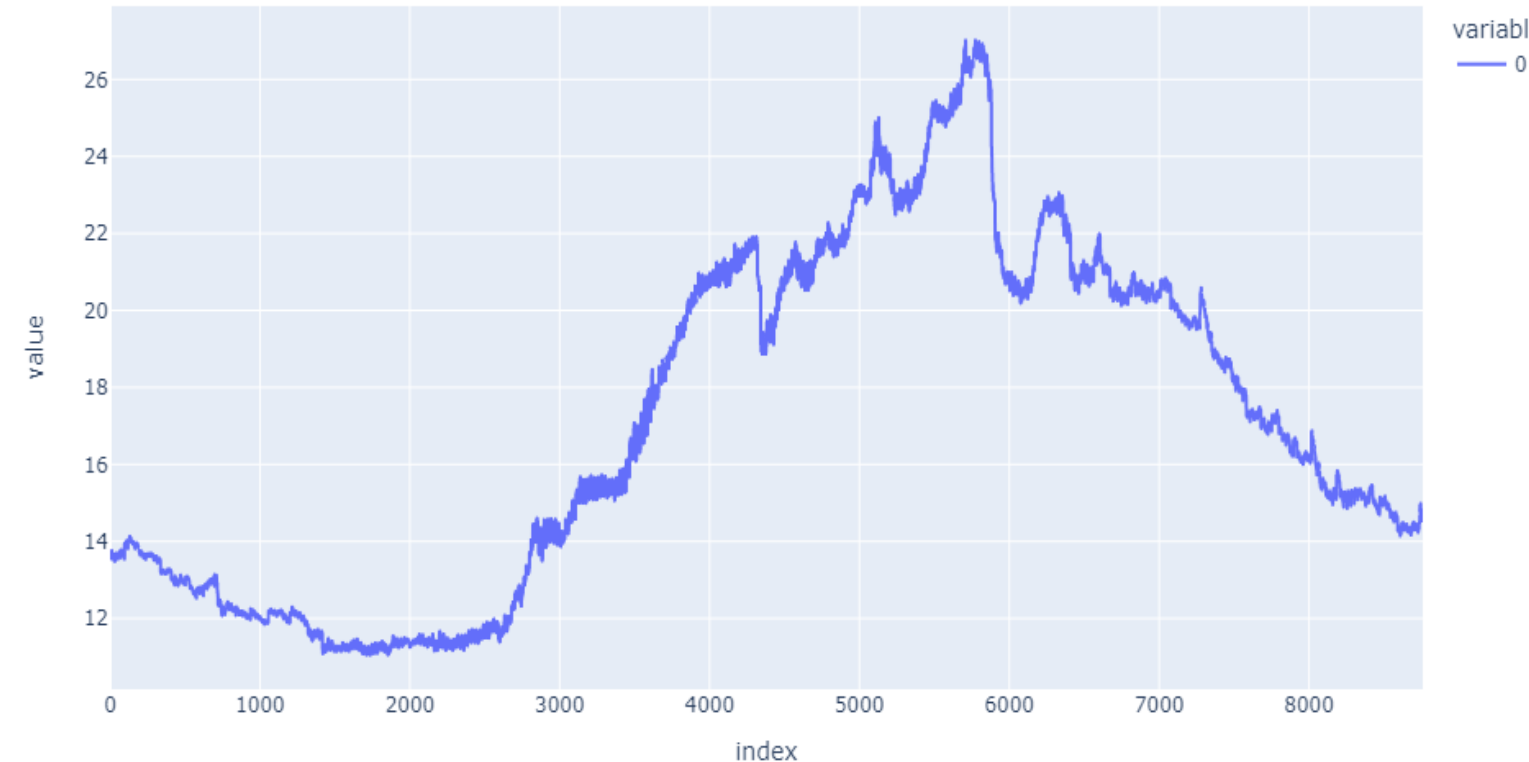


## 습도, 기온, 수온 예측

기온과 수온에도 앞서 학습한 모델을 동일하게 적용



<2030년의 기온 예측 그래프>



<2030년의 수온 예측 그래프>

# 2030년도 수송량 예측



## 유속 예측

Model	RMSE
Random forest	0.010104
BaggingRegressor	0.010791
DecisionTreeRegressor	0.013668
XGBRegressor	0.016346
CatBoostRegressor	0.016875
GradientBoostingRegressor	0.026989
AdaBoostRegressor	0.029548
LinearRegressor	0.031200

외부데이터에서 가져온 ‘습도’, ‘기온’, ‘수온’ 데이터와  
U3 데이터를 병합해 총 8개의 모델을 학습해본 결과  
**RMSE가 가장 낮게 나온 Random forest를 최종 모델로 선정**

앞서 예측한 ‘습도’, ‘기온’, ‘수온’ 데이터에서 2030년에 해당하는  
행만 가져와 U3 데이터의 날짜와 함께 2030년의 **유속**을 예측

# 2030년도 수송량 예측



## 수송량 예측

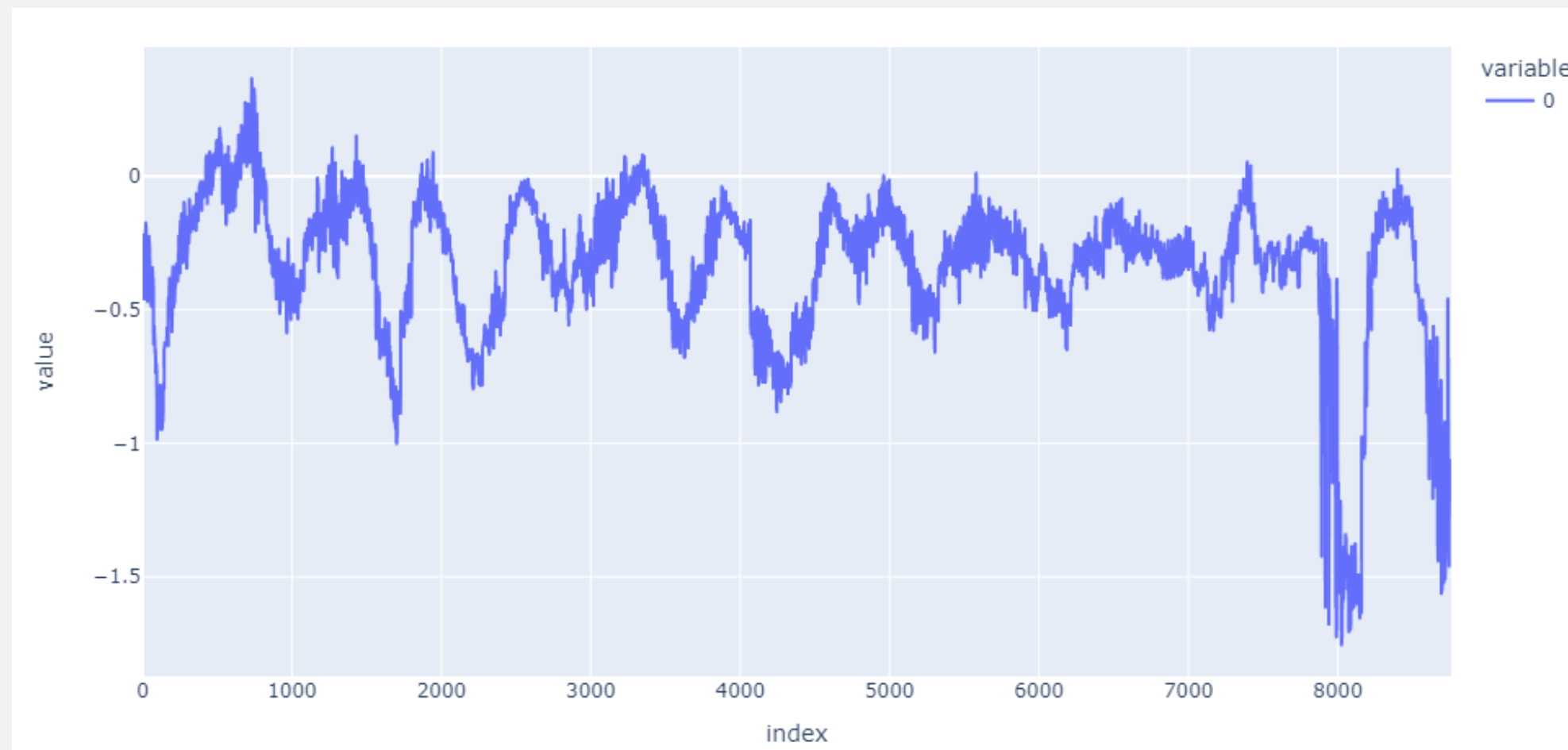
Date	Ur Current speed (m/s)
2030-01-01 00:00:00	-0.006911
2030-01-01 01:00:00	-0.007913
2030-01-01 02:00:00	-0.009164
...	...
2030-12-31 20:00:00	-0.106167
2030-12-31 21:00:00	-0.105872
2030-12-31 22:00:00	-0.103196
2030-12-31 23:00:00	-0.102386

상관관계식 대입



Date	Transport
2030-01-01 00:00:00	-0.220397
2030-01-01 01:00:00	-0.229251
2030-01-01 02:00:00	-0.240303
...	...
2030-12-31 20:00:00	-1.097323
2030-12-31 21:00:00	-1.094709
2030-12-31 22:00:00	-1.071069
2030-12-31 23:00:00	-1.063911

## 2030년도 수송량 예측



<2030년의 수송량 예측 그래프>

**2030년 남향 수송량 : - 2979.209393557897 Sv**

# 감사합니다

DS 수산시장

---