

# DRIVER IDENTIFICATION BASED ON SCALOGRAM(SC)

2023.12.20

한림대학교 소프트웨어융합대학  
(Data-driven Cybersecurity  
Research Lab)

# 목차

---

» 연구 방향 및 구성

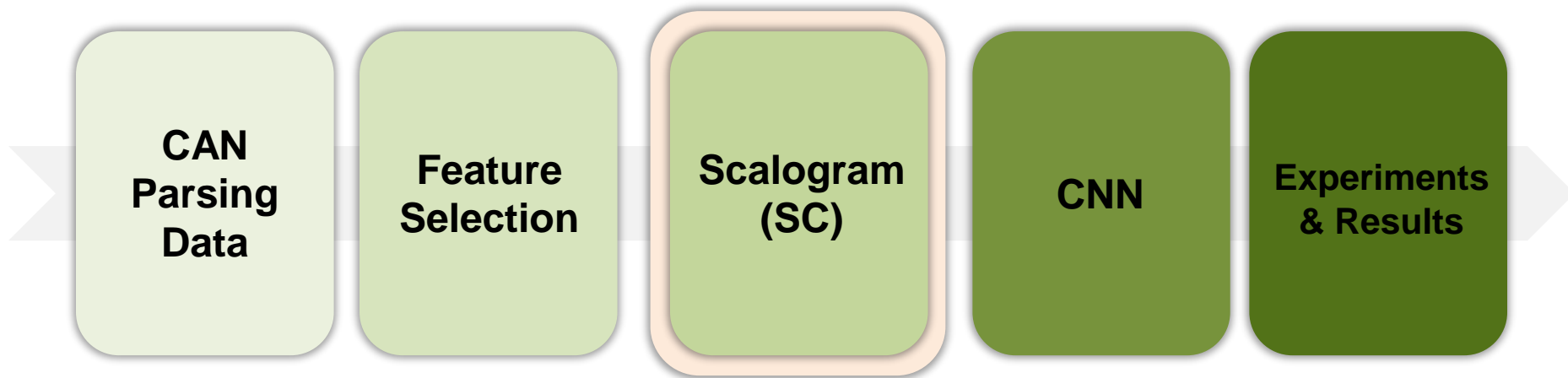
» 진행 사항

# 피드백

- 왜 Scalogram 을 사용하는지 이유
- 그렇다면 왜 dwt 가 아니라 cwt 를 사용하는지
- 이게 웨이블릿이란 뭔 차이가 있는지 => 효과면에서
- 많은 시계열 데이터 이미지화 방법이 있는데 왜 이것 사용했는지
- 데이터 전체 사용하는 게 아니라, **일정 구간**을 정해서 그거에 대한 걸 학습 시키는 것임. => **구간 선정해야 함.**
  - 그리고 천천히 운전하는 운전자랑 빠르게 운전하는 운전자에 대한 시간 문제 어떻게 통일화 시킬 건지
    - 피처에서 체크섬, 카운트 모두 제외, 필요 없음
    - 1초에 10번도 안 나오는 데이터는 모두 삭제
    - 전체적 데이터 처리를 이렇게 해야함.
- 사용할 피처 다 찾아보고 정의해야 함. 사용할 수 있는지 없는지
- 웨이블릿 처리할 때도 가장 좋은 성능 보이는 거 전체 다 실험해서 가장 좋은 걸로 사용
- 방법 제안에서 사전 공부가 더 필요할 것 같음. -> 이유에 대해서 더 명확하고 근거있게 찾아봐야 함.
- 윈도우 크기 어떻게 해서 할 지 등등 처리해야 할 부분이 매우 많고, 지금보다 공부를 깊게 해야할 것 같음.
- 그리고 같이 논문을 작성하게 되면, 1저자 2저자 부분을 명확하게 해야함.
  - 각자 어느 정도 기여하는지 명확하게 정해야 함.
  - 지도 교수님 두분 다 들어가야 함.
    - 합의 모든 이야기 다 해야함. 지도 교수님한테도 이게 안되면 혼자 쓰는 게 나을 수도...

- 
- [Automobile Driver Fingerprinting: A New Machine Learning Based Authentication Scheme | IEEE Journals & Magazine | IEEE Xplore](#)
  - <https://homes.cs.washington.edu/~yoshi/papers/auto-fingerprint.pdf>
  - 이 두 논문 읽으면 감이 잡힐 거임.
  - 논문 방향성은 맞음

## ■ Over view



## ■ CAN Parsing Data

- Multi-Modal (Discussion)
  - 운전자 식별의 정확도는 높일 수 있지만, 실현 가능성이 떨어짐.
    - 실제 사용되기 위해서는 많은 센서를 추가해야한다는 단점과
      - 모델 학습에 사용된 센서 데이터를 모두 수집해야한다는 조건이 존재함.
  - 또한, privacy 문제도 존재함.
- 다양한 경로의 운전자 데이터를 가지고 있음.
  - A, B, C 코스로 4번씩 주행 된 데이터
    - 그 중 2-3개씩만 있는 운전자 데이터는 삭제 후, 4개씩 있는 운전자의 데이터만 처리 (9명 데이터)
      - choimingi, jungyubin, leegahyeon, leegihun, leejaeho, leekanghyuk, leeseunglee, leeyunguel, simboseok
  - 파일 관련
    - [Decode] : CAN으로 Parsing 데이터
    - [Norm] : CAN 원본 데이터

## ■ CAN Parsing Data

### □ 운전 경로



## ■ Feature Selection

- 데이터 전처리 - Variance, Entropy 사용
  - Variance : 데이터가 평균으로부터, 얼마나 퍼져 있는지 나타내는 값.
    - 데이터가 일정한 값은 0을 나타내므로, 0 에 근접한 값을 제거하여, 일정한 값을 가지는 데이터를 제거
  - Entropy : 일정 공간의 정보량을 나타내는 수치.
    - 엔트로피의 작은 값을 제거
      - 가능한 이벤트들이 얼마나 자주 발생하는지를 나타내는 데 사용
  - 추가적으로, 일정시간 동안 패턴을 나타내지 않는 데이터를 모두 제거함.



## ■ 시계열 데이터 이미지화

- Markov Transition Field : 이산화한 시계열 데이터의 전이 확률을 나타내는 알고리즘.
  - Gramian Angular Field 나 RP에 비해 일정한 시간 구간의 세부적인 정보를 잡아내는 데 유용함.
  - Deep-Learning-Based Automatic Modulation Classification Using Imaging Algorithm 논문에 따르면,
    - Gramian Angular Field, RP 이 두가지 이미지화 알고리즘 보다 더 자동 변조 분류 성능이 우수하였음.
- Gramian Angular Field : 각 시점 간의 시간적 상관관계를 극좌표를 기반으로 표현하는 알고리즘.
  - 극좌표 행렬은 시계열 데이터를 이미지로 변경할 때 시간 상관관계를 보존할 수 있음.
- Recurrence Plot : 시간 종속성 문제 해결
  - M차원의 각 궤도에서 점간의 거리를 구하기 때문에 현재(t)와 그 전(t-1)의 value 값을 참고하기 때문임.
- Gray Scale(GS) Encoding
  - 흑백 이미지 처리에 사용됨. 따라서, 색상 정보가 없기 때문에 컴퓨터 처리가 간단함.

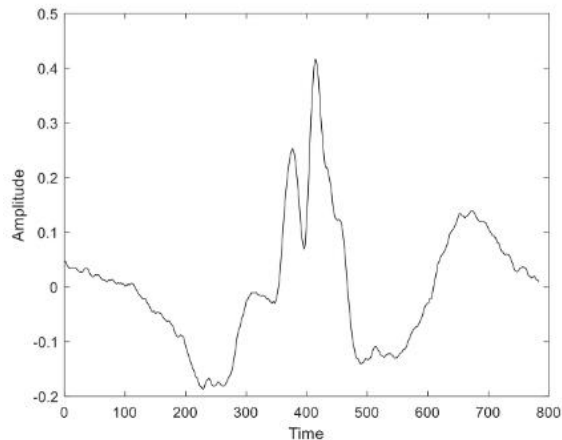
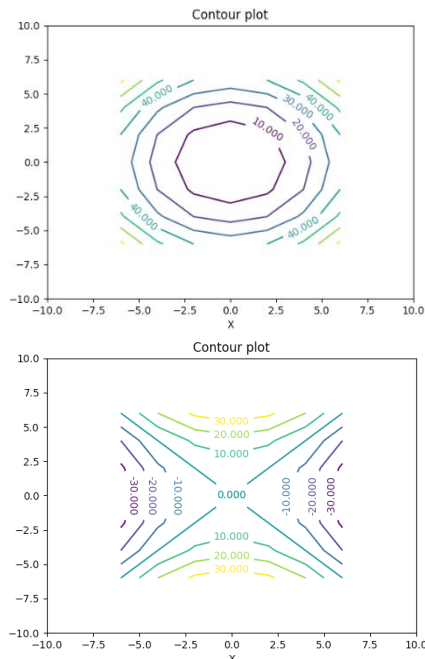
## ■ 시계열 데이터 이미지화

- Spectrogram(SP) : short-time Fourier Transform (STFT) 으로부터 나온 방법론.
  - 소리나 파동을 시각화하여 파악하기 위한 도구로, 파형(waveform)과 스펙트럼(spectrum)의 특징이 조합되어 있음.
    - 시간축과 주파수 축의 변화에 따라 진폭의 차이를 인색 농도 / 표시 색상의 차이로 나타냄.
  - STFT는 각 시간당 주파수가 가지는 값을 값의 크기에 따라 색으로 표현하여 3차원을 2차원으로 표현하게 됨.
    - 즉, 시간의 흐름을 가지는 푸리에 변환이라고 생각할 수 있음.
    - STFT의 단순히 magnitude 부분을 취해서 db 스케일로만 변환해주면 spectrogram이 됨.
- Scalogram(SC)
  - Continuous Wavelet Transform (CWT) 으로부터 나온 방법론.
    - 차원 평면에 CWT 변환값을 계조도나 색깔로 표현한 것.

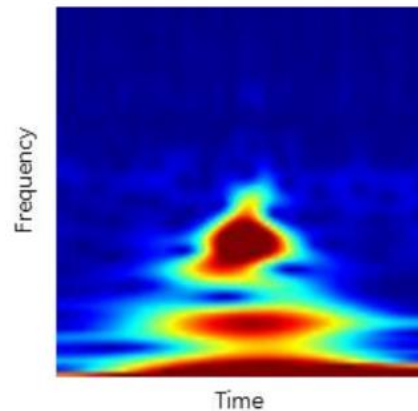
# 연구 방향

## ■ Scalogram(SC)

- 2차원 평면에 Continuous Wavelet Transform (CWT) 변환값을 계조도(contour plot)나 색깔로 표현한 것
  - contour plot(등고선) : 2차원 형식으로 그려 3차원 표면을 표현하는 그래픽 기술



(a) original signal



(b) scalogram by continuous wavelet transform

출처 : <https://min23th.tistory.com/42>

## ■ Scalogram(SC)

### □ Continuous Wavelet Transform (CWT)

#### • Wavelet Transform

- 웨이블릿 변환에 사용되는 웨이블릿 함수(mother wavelet)는
  - $\psi$ (\* 복소수 연산을 의미)로 표현됨.
- $a$ 와  $b$ 는 각각 웨이블릿의 스케일(scale)과 변화(shifting)를 나타내는 지표임.

- 즉, 웨이블릿 변환이란, 웨이블릿을 시간축 방향으로 이동시켜가며(Fig 1),
  - 원본 신호  $x(t)$ 와의 상관계수를 계산하고, 웨이블릿의 스케일을 조정해(Fig 2)
  - 다시 시간축 방향으로 이동시키며 상관계수를 계산하는 과정을 반복한 것임.

$$T(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \frac{t-b}{a} dt$$

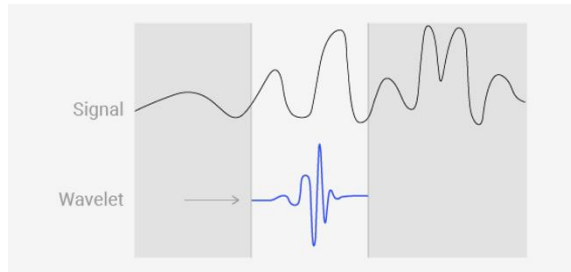


Fig. 1 Shifting 과정

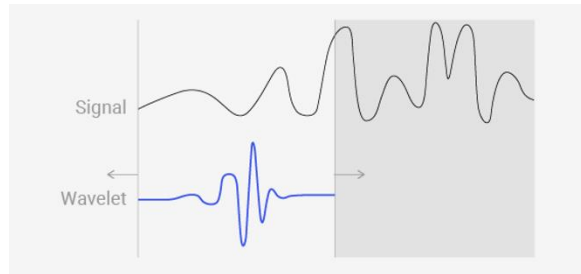


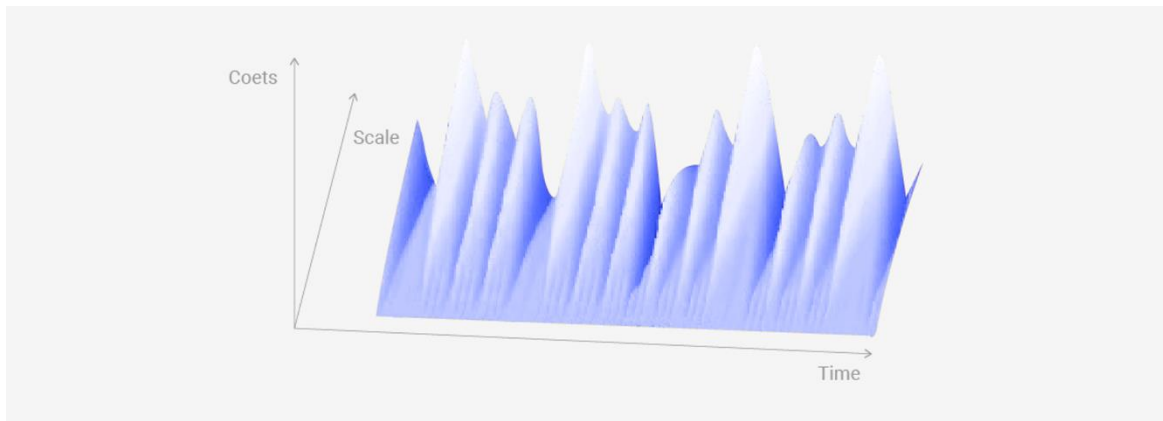
Fig. 2 Scaling 과정

## ■ Scalogram(SC)

### □ Continuous Wavelet Transform (CWT)

#### • Wavelet Transform

- 결과적으로 x축은 시간, y축은 스케일 지표인 2D 이미지 형태로 결과를 나타낼 수 있음.
  - 스케일을 웨이블릿의 중심주파수로 변환해 y축을 주파수 값으로 나타낼 수 있음.



시계열 데이터의 연속 웨이블릿 변환 결과

## ■ Scalogram(SC)

### □ Continuous Wavelet Transform (CWT)

- 웨이블릿은 다양한 주파수 대역을 가진 함수,
  - CWT는 웨이블릿(함수)를 시계열 데이터에 적용하여 시간에 따른 주파수 변화를 감지함.

### □ Scalogram(SC)

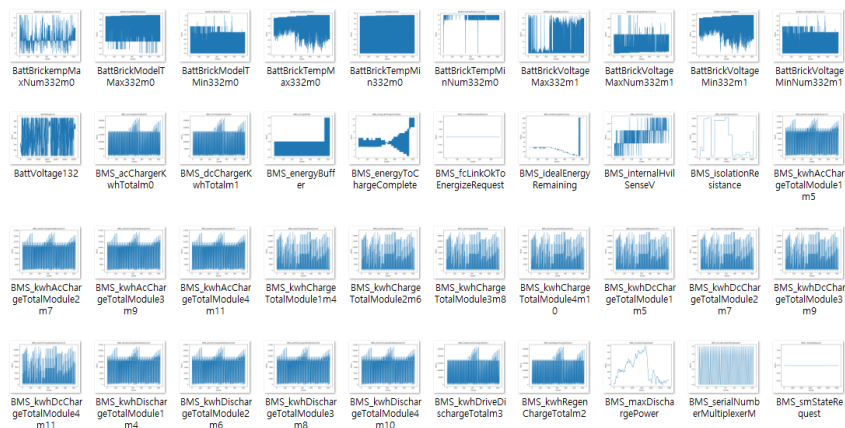
- 시계열 데이터의 주파수 특성을 시각적으로 표현하는 도구 중 하나.
- 시간은  $x$ , 주파수는  $y$  축에 나타내며, 각 점의 색상이나 밝기는 해당 시간과 주파수에서의 신호 강도를 나타냄.
- 시간 해상도와 주파수 해상도를 동시에 제공, 신호의 주파수 특성이 어떻게 시간에 따라 변하는 지 알 수 있음.
  - 특히, 빠르게 변하는 주파수 성분이나 특정 주파수 대역의 특징을 찾는 데 유용
  - => 따라서, 운전자 식별에 유용하다고 판단하였음.

# 진행 사항

## ■ 데이터 전처리

### □ 데이터 전처리 - Variance, Entropy 사용

- 기존 시그널 수 : 1836
  - 분산과 엔트로피 0 값을 제외한 시그널 수 : 1028
  - 패턴 없는 데이터 제외 시그널 수 : 650



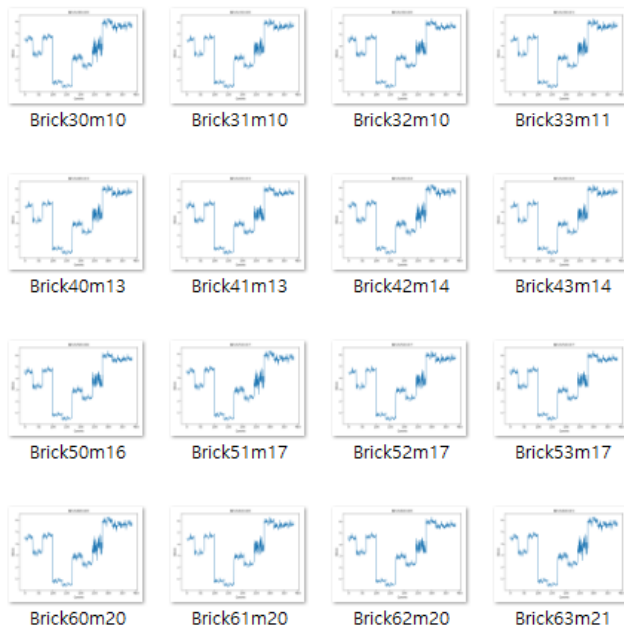
[ 각 피쳐마다의 plot 형태를 나타낸 그림]

Signal	Variance	Entropy
RearHeatPowerMax266	0.6083568959374887	5.00729962954089
RearHeatPower266	0.7964428387041174	5.1174245212121745
RearHeatPowerOptimal266	0.7659009320102266	5.058504999561284
RearExcessHeatCmd	0.0	-0.0
RearPower266	17044.2208962984	4.43996279679515
DI_accelPedalPos	158.08742797361586	4.407268435638761
DI_brakePedalState	0.0	-0.0
DI_driveBlocked	0.0	-0.0
DI_epbRequest	0.010791309738936255	0.04590290794044658
DI_gear	0.11027917400063286	0.096358982291103
DI_immobilizerState	0.0	-0.0
DI_keepDrivePowerStateRequest	0.012253241555625877	0.096358982291103
DI_proximity	0.0	-0.0
DI_regenLight	0.14527894137247718	0.6721238389803463
DI_systemState	0.11041580134406273	0.096457270710018

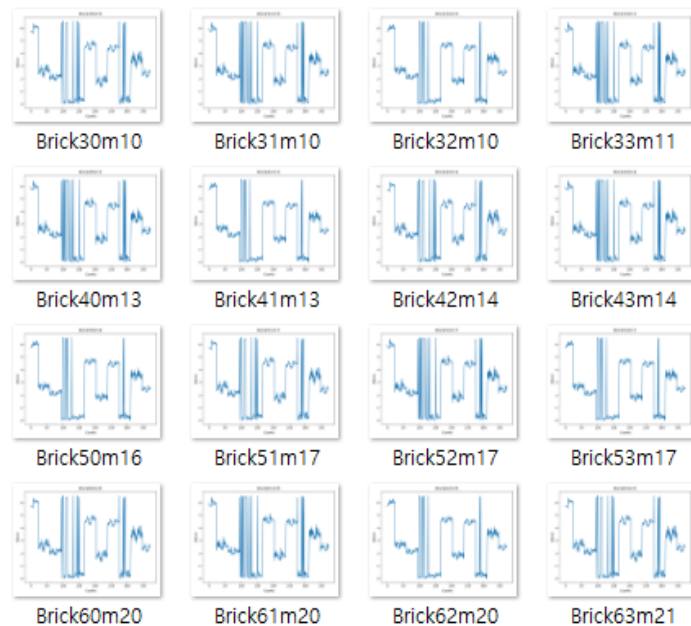
# 진행 사항

## ■ 데이터 비교

- Choimingi – plot vs jungyubin – plot
  - 같은 피쳐에서 명확하게 다른 패턴이 확인됨.



[ Choimingi ]



[ jungyubin ]



# 진행 사항

## ■ 데이터 크기 (수) 비교

canID		Signal	DLC	Timestamp	Physical_value	Unit	Comment
0	266	RearHeatPowerMax266	8	2023.08.22.16:12:31.393419	2.96	"kW"	NaN
1	266	RearHeatPower266	8	2023.08.22.16:12:31.393419	0.00	"kW"	Rear_Waste_Heat_Power
2	266	RearHeatPowerOptimal266	8	2023.08.22.16:12:31.393419	0.08	"kW"	NaN
3	266	RearExcessHeatCmd	8	2023.08.22.16:12:31.393419	0.00	"kW"	NaN
4	266	RearPower266	8	2023.08.22.16:12:31.393419	0.00	"kW"	Rear_Motor_Power
...	...	...	...	...	...	...	...
18924418	42A	VCSEC_TPMSConnectionTypeDesired0	8	2023.08.22.16:23:13.690442	2.00	""	NaN
18924419	42A	VCSEC_TPMSConnectionTypeDesired2	8	2023.08.22.16:23:13.690442	1.00	""	NaN
18924420	42A	VCSEC_TPMSConnectionTypeDesired3	8	2023.08.22.16:23:13.690442	2.00	""	NaN
18924421	42A	VCSEC_TPMSSensorState0	8	2023.08.22.16:23:13.690442	3.00	""	NaN
18924422	42A	VCSEC_TPMSSensorState2	8	2023.08.22.16:23:13.690442	0.00	""	NaN

18924423 rows × 7 columns

[ 원본 데이터 수 - 18,924,423 ]

canID		Signal	DLC	Timestamp	Physical_value	Unit	Comment
0	266	RearHeatPowerMax266	8	2023.08.22.16:12:31.393419	2.96	"kW"	NaN
1	266	RearHeatPower266	8	2023.08.22.16:12:31.393419	0.00	"kW"	Rear_Waste_Heat_Power
2	266	RearHeatPowerOptimal266	8	2023.08.22.16:12:31.393419	0.08	"kW"	NaN
3	266	RearPower266	8	2023.08.22.16:12:31.393419	0.00	"kW"	Rear_Motor_Power
4	118	DI_accelPedalPos	8	2023.08.22.16:12:31.393419	0.00	"%"	Pedal_Position
...	...	...	...	...	...	...	...
7431045	1D8	Checksum1D8	8	2023.08.22.16:23:13.688593	58.00	""	NaN
7431046	1D8	RearTorqueRequest1D8	8	2023.08.22.16:23:13.688593	0.00	"NM"	Rear_Motor_Torque_Request
7431047	1D8	RearTorque1D8	8	2023.08.22.16:23:13.688593	0.00	"NM"	Rear_Motor_Torque
7431048	129	SteeringSpeed129	8	2023.08.22.16:23:13.689601	-4080.00	"D/S"	Steering_Speed
7431049	42A	VCSEC_TPMSsensorState2	8	2023.08.22.16:23:13.690442	0.00	""	NaN

7431050 rows × 7 columns

[ 피쳐 선택한 수 - 7,431,050 ]

# Q&A

# Thank You