

목

차

I. 머신러닝 기반 타이어 헬스 모니터링 영향 인자 연구

1. 탐색적 데이터 분석

----- 3

2. 상관관계 분석

----- 15

II. 머신러닝 예측기법을 통한 영향인자 관계 분석

----- 20

III. 결론

----- 25

I. 머신러닝 기반 타이어 헬스 모니터링 영향 인자 연구

1. 탐색적 데이터 분석

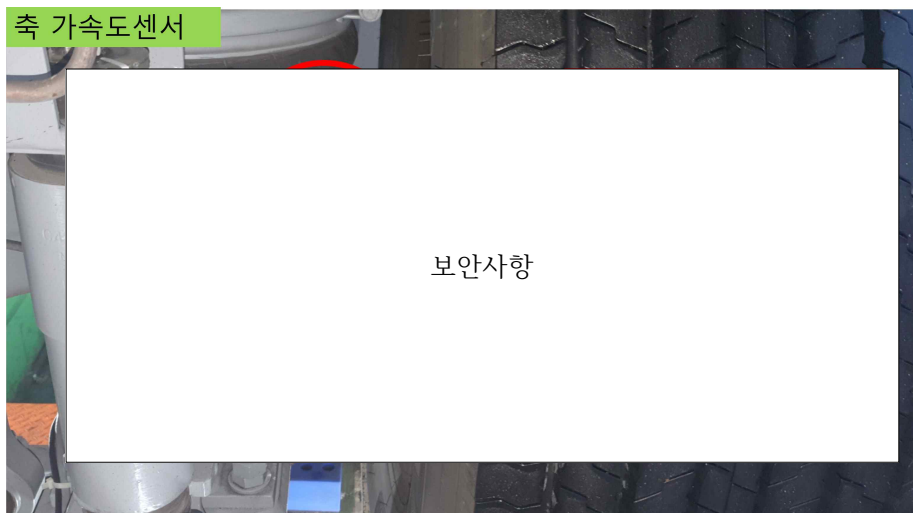
(1) 탐색적 데이터 분석

- 수집된 데이터 그대로를 시각화시켜 원인분석을 수행하는 접근법으로 데이터의 분포 (distribution) 및 값(value)을 여러 각도에서 관찰해보면서 데이터의 전체적인 양상과 보이지 않던 현상을 더 잘 이해할 수 있도록 해주기 때문에 내재하는 구조적인 관계를 파악할 수 있으며 아웃라이어와 비정상적인 값을 제거하기 위한 데이터 전처리과정 사용되는 기법
- 대표적인 기법으로 통계 값의 활용으로 분석, 시각화를 활용, 개별 데이터 관찰 등이 있음

(2) 10월 8일 데이터

□ 왕복 운행 데이터

○ 변수의 의미 파악



축 가속도 센서

- 위 사진과 같이 축 가속도 센서는 세 방향으로 되어 있음
- wheel axle X(g), wheel axle Y(g), wheel axle Z(g) 변수가 각각 X는 진행 방향 가속도, Y축은 횡 방향 가속도, Z축은 종 방향 가속도임을 판단
- Inside wheel Y, Inside wheel Z도 같은 의미로 각각 횡 방향 타이어 내부 가속도, 종 방향 타이어 내부 가속도를 의미

○ 데이터 병합

 Test_2020_10_08_151129 고촌_동부산대학.csv
 Test_2020_10_08_151129 고촌_안평.csv
 Test_2020_10_08_151129 금사_반여농산물시장.csv
 Test_2020_10_08_151129 금사_서동.csv
 Test_2020_10_08_151129 낙민_수안.csv
 Test_2020_10_08_151129 낙민_충렬사.csv
 Test_2020_10_08_151129 동래_미남.csv
 Test_2020_10_08_151129 동래_수안.csv
 Test_2020_10_08_151129 동부산대학_고촌.csv
 Test_2020_10_08_151129 동부산대학_영산대.csv

제공받은 데이터

- 역과 역 사이의 정보만 추출이 되고 이로 인해 전체적인 흐름 파악의 어려움이 존재
- 전체적인 흐름을 파악하기 위해 모든 데이터 통합 과정 필요
- 운행 흐름을 정렬하기 위해 운행 시간순으로 정렬

```

first_drive = os.listdir('1st_drive') # 저장 디렉토리에 접근
df = pd.DataFrame()
for i in first_drive:
    x = pd.read_csv('1st_drive/'+i, parse_dates=['Time (-)'])
    x['path'] = i[23:-4] # 시작역과 출발역 저장
    df = df.append(x)
df = df.sort_values('Time (-)') # 시간 순으로 정렬
  
```

데이터 병합 코드

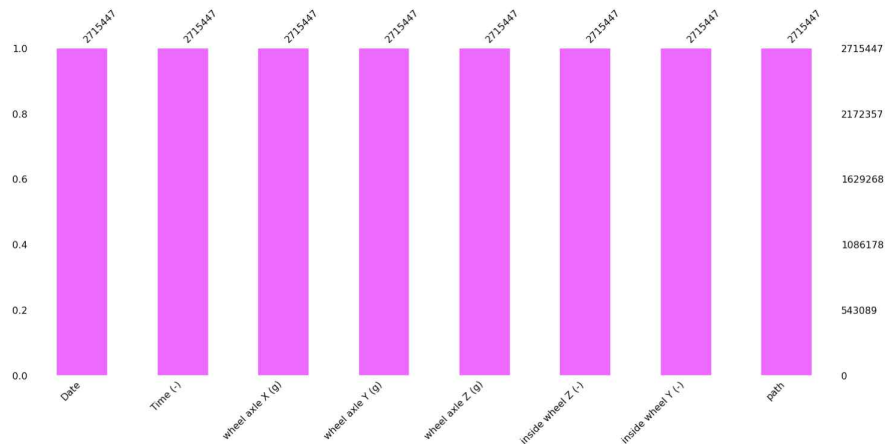
○ 데이터 통계적 탐색

	wheel axle X (g)	wheel axle Y (g)	wheel axle Z (g)	inside wheel Z (-)	inside wheel Y (-)
count	2.715447e+06	2.715447e+06	2.715447e+06	2.715447e+06	2.715447e+06
mean	-6.053452e-03	2.352695e-03	-2.336457e-03	-8.600297e-05	-3.685483e-04
std	1.620373e-01	2.275370e-01	2.706522e-01	5.535032e+00	1.631371e+00
min	-2.221534e+00	-2.160210e+00	-2.390376e+00	-7.932512e+01	-2.476840e+01
25%	-7.637989e-02	-9.687824e-02	-1.162973e-01	-1.974049e+00	-7.678845e-01
50%	-5.428748e-03	-8.884273e-04	-2.986378e-03	-5.387449e-02	-4.185638e-05
75%	6.233610e-02	1.006818e-01	1.120312e-01	1.318227e+00	7.772421e-01
max	2.069208e+00	2.271284e+00	2.336139e+00	6.174561e+01	2.343070e+01

데이터 변수의 통계 수치 확인

- 병합한 데이터 중 연속형 변수의 통계적 수치 확인

- 데이터의 개수가 모두 동일
- 분산이 가장 큰 변수는 내부 타이어의 Z(종 방향) 가속도, 가장 작은 변수는 X축(진행 방향) 가속도

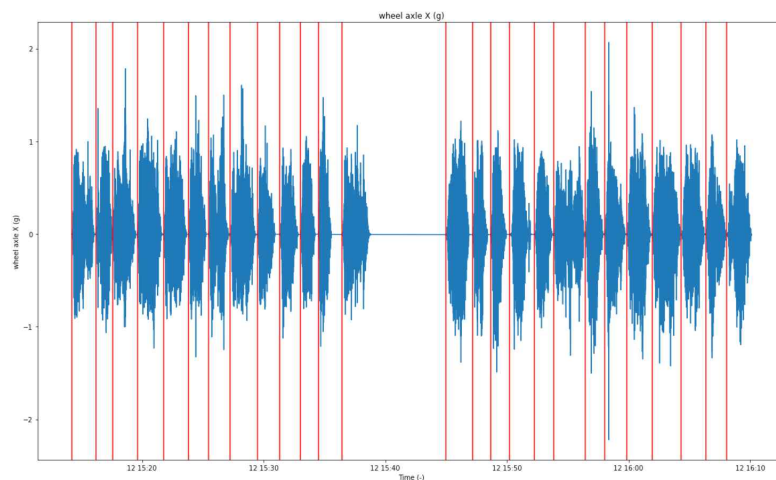


데이터 결측치 확인

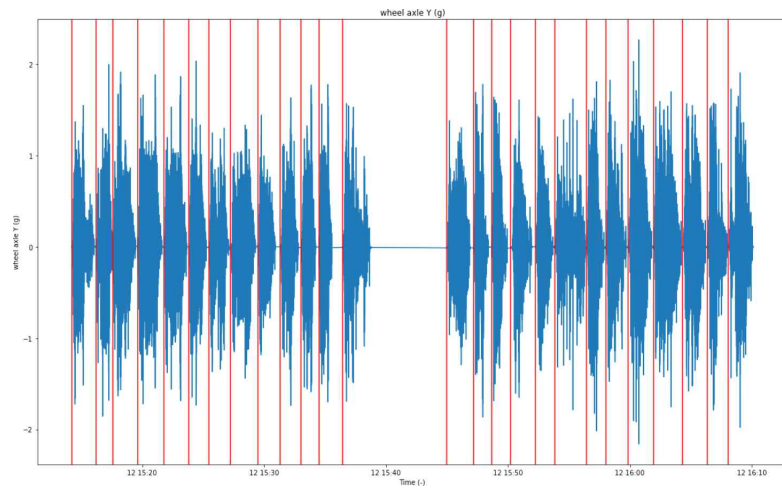
- 파이썬 `missingno` 모듈을 활용하여 데이터의 결측치 유무 확인 및 개수 확인
- 모든 바 그래프가 꽉 채워진 것으로 보아 데이터의 결측치는 존재하지 않음을 확인

○ 시각화를 통한 흐름 파악

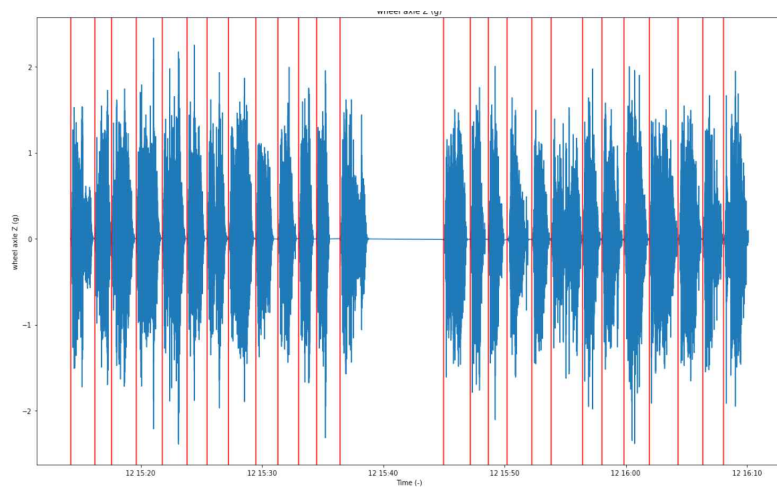
- 가속도의 흐름을 파악하기 위해 파이썬 `seaborn` 라이브러리의 `lineplot`을 이용하여 추세 파악
- 운행 노선의 역사 운행시간에 대한 가속도 변경 포인트 파악으로 패턴 분석



X축의 가속도

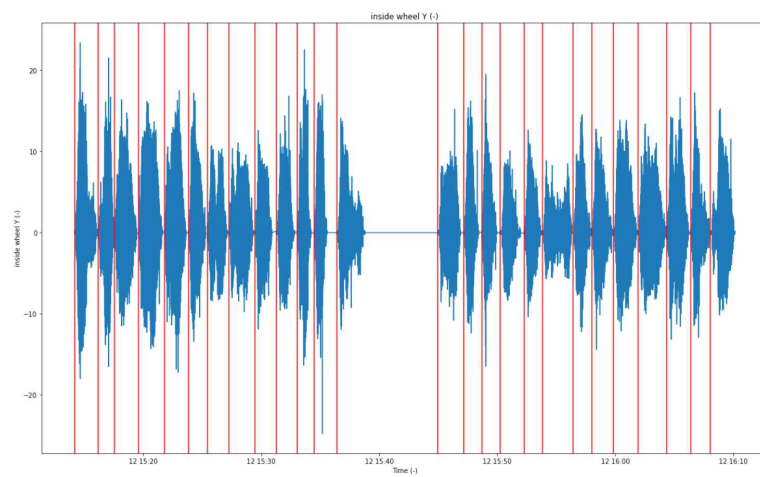


Y축의 가속도

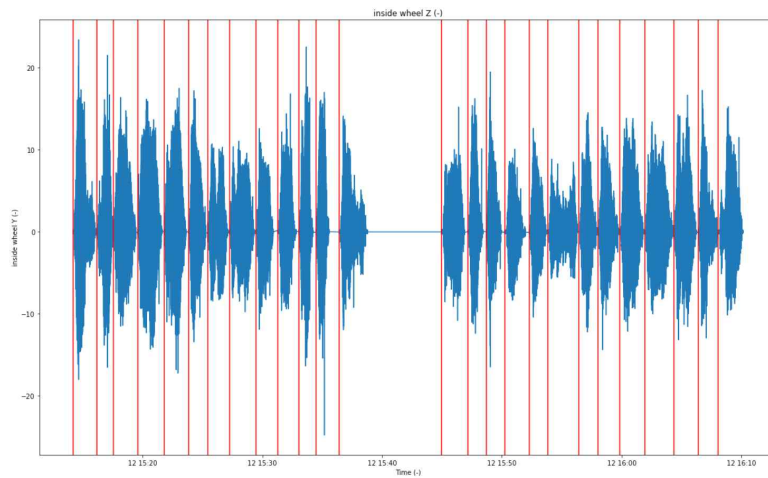


Z축의 가속도

- 10월 8일의 운행한 차축의 가속도 데이터(X,Y,Z 축) 분석결과 주행상태에 동일한 노선 운행이 비슷한 패턴을 보임을 확인할 수 있음



Y축의 내부 가속도

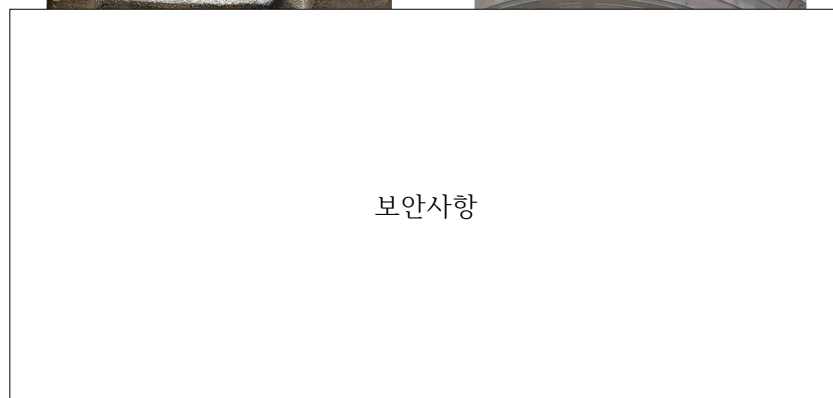


Z축의 내부 가속도

- 10월 8일의 운행한 휠 내부의 설치되어 계측된 가속도 데이터(Y,Z 축) 분석결과 주행상태에 동일한 노선 운행에서 비슷한 패턴을 보임을 확인하였으며 타이어 헬스모니터링을 위한 주요인자로 분류하게 적합하지 않음으로 판단됨

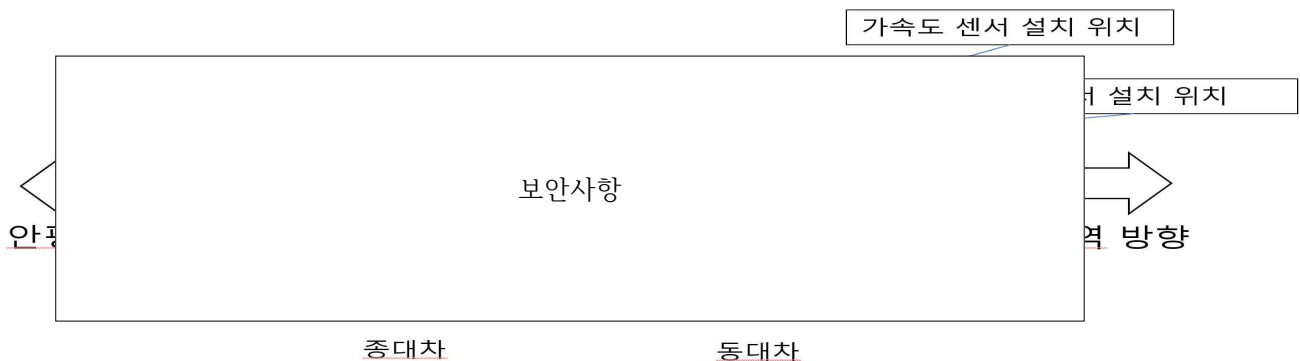
□ 타이어 내부 온도와 압력 데이터

○ 변수의 의미 파악



온도 압력 센서 부착 위치

- 위 사진과 같이 타이어 하나당 16개의 온도 채널과 한 개의 압력 채널 존재



타이어 위치

- 타이어위치_TTPMS_T채널번호 포맷으로 온도변수 구성
- 타이어위치_TTPMS_P 포맷으로 압력 변수 구성
- 왕복 데이터와 시간적 매핑 불가

○ 데이터 통계적 탐색

	LF_TTPMS_P ·(mbar)	LF_TTPMS_T1 ·(C)	LF_TTPMS_T2 ·(C)	LF_TTPMS_T3 ·(C)	LF_TTPMS_T4 ·(C)	LF_TTPMS_T5 ·(C)	LF_TTPMS_T6 ·(C)
count	58180.000000	58180.000000	58180.000000	58180.000000	58180.000000	58180.000000	58180.000000
mean	10854.482193	32.397539	33.311927	33.565105	34.541373	34.998249	36.220110
std	68.981752	34.931757	35.166052	35.229826	35.485619	35.605391	35.926142
min	10694.000000	-100.000000	-100.000000	-100.000000	-100.000000	-100.000000	-100.000000
25%	10811.000000	40.000000	41.099998	41.500000	42.599998	43.000000	44.299999
50%	10869.000000	41.599998	42.500000	42.799999	43.900002	44.400002	45.799999
75%	10899.000000	42.900002	43.799999	44.000000	45.000000	45.500000	46.700001
max	10974.000000	46.799999	46.900002	47.400002	48.700001	48.700001	50.099998

데이터 변수의 통계 수치 확인

- 모든 변수의 개수가 동일하고 모든 변수가 표준편차가 큼
- 압력의 평균이 온도의 평균보다 높음

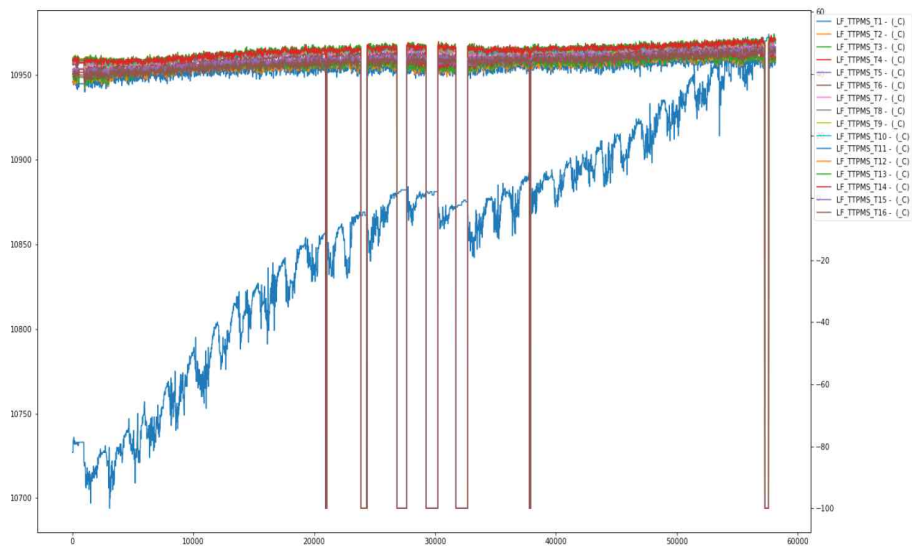


온도 압력 데이터의 결측치 파악

- 날짜와 시간 데이터만 결측치가 없고 나머지 변수의 결측치 비율은 80% 이상
- 압력과 온도 데이터의 변수의 개수는 모두 동일

○ 시각화를 통한 이상치 파악

- 4개의 타이어 중 왼쪽 전방에 있는 LF 타이어를 중심으로 분석



LF의 온도와 압력 데이터

- 기존의 데이터와 차이가 많이 나는 값 탐지
- Statistical-based Detection 방식을 기반으로 $Q3+3*IQR$ 보다 크거나 $Q1-3*IQR$ 작은 값들을 이상치라 정의 후 제거 $\ast IQR = Q3(3사분위 수) - Q1(1사분위 수)$

```

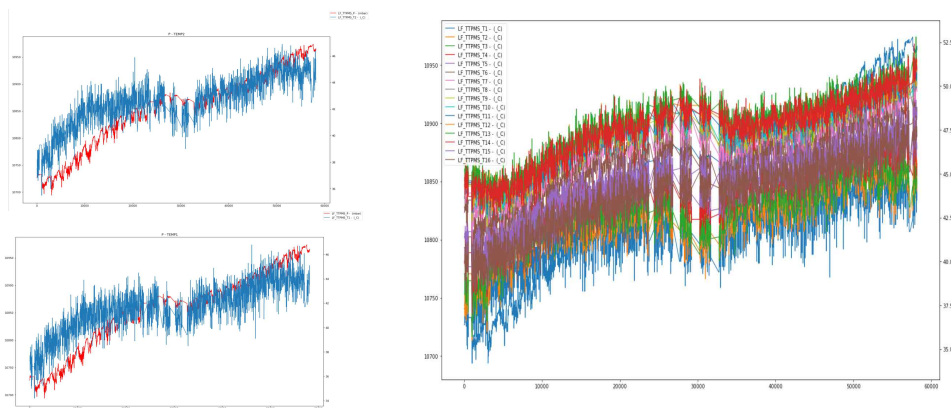
outlier_df = pd.DataFrame(columns=['col', 'max_outlier', 'low_outlier'])
for i in range(1,17):
    iqr = tmp['LF_TTPMS_T'+ (i)].quantile(0.75)-tmp['LF_TTPMS_T'+ (i)].quantile(0.25)
    max_out = tmp['LF_TTPMS_T'+ (i)].quantile(0.75)+3*iqr
    low_out = tmp['LF_TTPMS_T'+ (i)].quantile(0.25)-3*iqr
    outlier_df.loc[i, 'col'] = 'LF_TTPMS_T'+ (i)
    outlier_df.loc[i, 'max_outlier'] = max_out
    outlier_df.loc[i, 'low_outlier'] = low_out
outlier_df

```

	col	max_outlier	low_outlier
1	LF_TTPMS_T1 - (C)	51.6	31.3
2	LF_TTPMS_T2 - (C)	51.9	33
3	LF_TTPMS_T3 - (C)	51.5	34
4	LF_TTPMS_T4 - (C)	52.2	35.4
5	LF_TTPMS_T5 - (C)	53	35.5
6	LF_TTPMS_T6 - (C)	53.9	37.1
7	LF_TTPMS_T7 - (C)	55.1	37.6
8	LF_TTPMS_T8 - (C)	56.4	38.8
9	LF_TTPMS_T9 - (C)	56.9	38.6
10	LF_TTPMS_T10 - (C)	57.4	38.4
11	LF_TTPMS_T11 - (C)	57.5	38.5
12	LF_TTPMS_T12 - (C)	58	38.3
13	LF_TTPMS_T13 - (C)	58.3	38.6
14	LF_TTPMS_T14 - (C)	58.2	38.8
15	LF_TTPMS_T15 - (C)	58.3	31.8
16	LF_TTPMS_T16 - (C)	58.3	31.5

이상치 탐지 코드 및 결과

- 통계적 이상치 정의에 따라 이상치 값 제거



(좌측 상단)온도1 - 압력, (좌측 하단)온도2 -압력,
(우측) 온도1~16채널 - 압력 시각화 그래프

- 그래프 좌측 세로축은 압력 범위, 우측 세로축은 온도 범위, 가로축은 검출된 순서(시계열)를 의미

- 이상치 제거 후 분석 결과 타이어의 온도와 압력이 비슷한 추세를 이루고 있음

(2) 10월 24일 데이터

□ 왕복 운행 데이터(고촌~동부산대학, 동부산대학~고촌)

○ 데이터 구성

- 총 3번의 왕복 운행 데이터로 구성
- 첫 번째 왕복 운행 데이터와 두 번째 왕복 운행 데이터의 변수 구성
 - 시간, 온도(4개의 타이어, 각 16개의 온도 채널), X-Y-Z 축의 가속도, 거리, 차체 속도
- 세 번째 왕복 운행 데이터의 변수 구성
 - 시간, X-Y-Z 축의 가속도, 거리, 차체 속도
- 데이터의 의미는 10월 8일 데이터와 동일

○ 데이터 통계적 탐색

	RR_TTPMS_T13 - (_C)	RR_TTPMS_T14 - (_C)	RR_TTPMS_T15 - (_C)	RR_TTPMS_T16 - (_C)	X (G)	Y (G)	Z (G)	Velocity (Km/h)	Distance (m)
count	1512.000000	1512.000000	1512.000000	1512.000000	94095.000000	94095.000000	94095.000000	94095.000000	94095.000000
mean	12.265873	12.194643	11.879630	11.180886	0.009990	0.009539	0.044211	29.440595	3003.934504
std	0.709657	0.988024	0.948981	1.055958	0.134450	0.192676	0.242253	21.756369	296.075853
min	10.900000	10.200000	10.400000	9.400000	-1.251362	-1.392468	-1.591302	0.000000	2641.815000
25%	11.800000	11.300000	11.100000	10.500000	-0.039848	-0.074335	-0.035595	0.000000	2712.975000
50%	12.100000	12.300000	12.000000	11.000000	0.010245	0.007785	0.044209	27.365561	2959.070000
75%	12.700000	13.000000	12.500000	11.800000	0.055645	0.089426	0.123093	52.688860	3320.800000
max	13.800000	14.000000	13.900000	13.700000	1.161456	1.293992	1.643981	58.496320	3424.575000

데이터 변수의 통계 수치 확인

- 온도 데이터의 개수와 가속도 및 속도 데이터 개수가 서로 다름
- 거리의 표준편차가 가장 크고 각 온도의 표준편차 차이는 크지 않음



데이터 결측치 확인

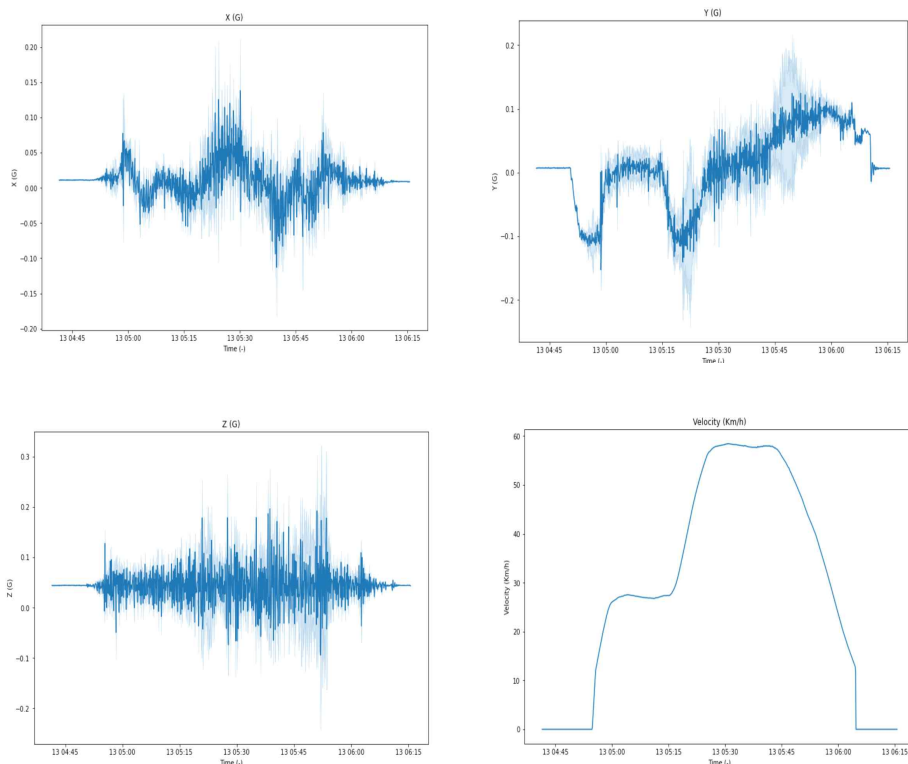
- 온도 데이터를 제외한 다른 데이터의 결측치는 존재하지 않음
- 온도 데이터의 너무 많은 결측치 비율로 인해 온도 데이터 사용이 어려움

○ 시각화를 통한 흐름 파악

- X,Y,Z 축 가속도와 차체 속도의 시각화를 통해 시간의 흐름에 따라 어떤 흐름이 나타나는지를 파악

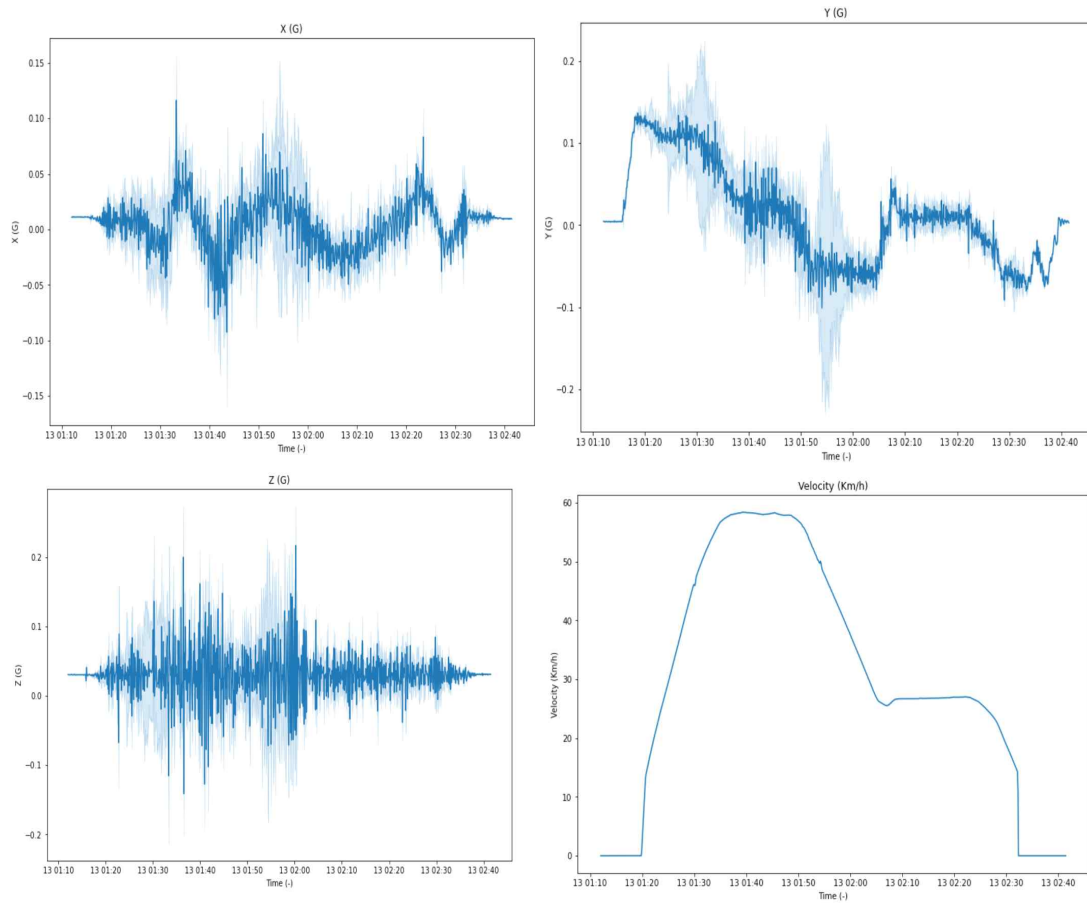
- 첫 번째 운행 데이터와 두 번째 운행 데이터를 활용해 시각화

1. 첫 번째 운행(고촌~동부산대학교 방향)



X,Y,Z 가속도 및 차량의 속도(첫번째 운행:고촌~동부산대학 방향)

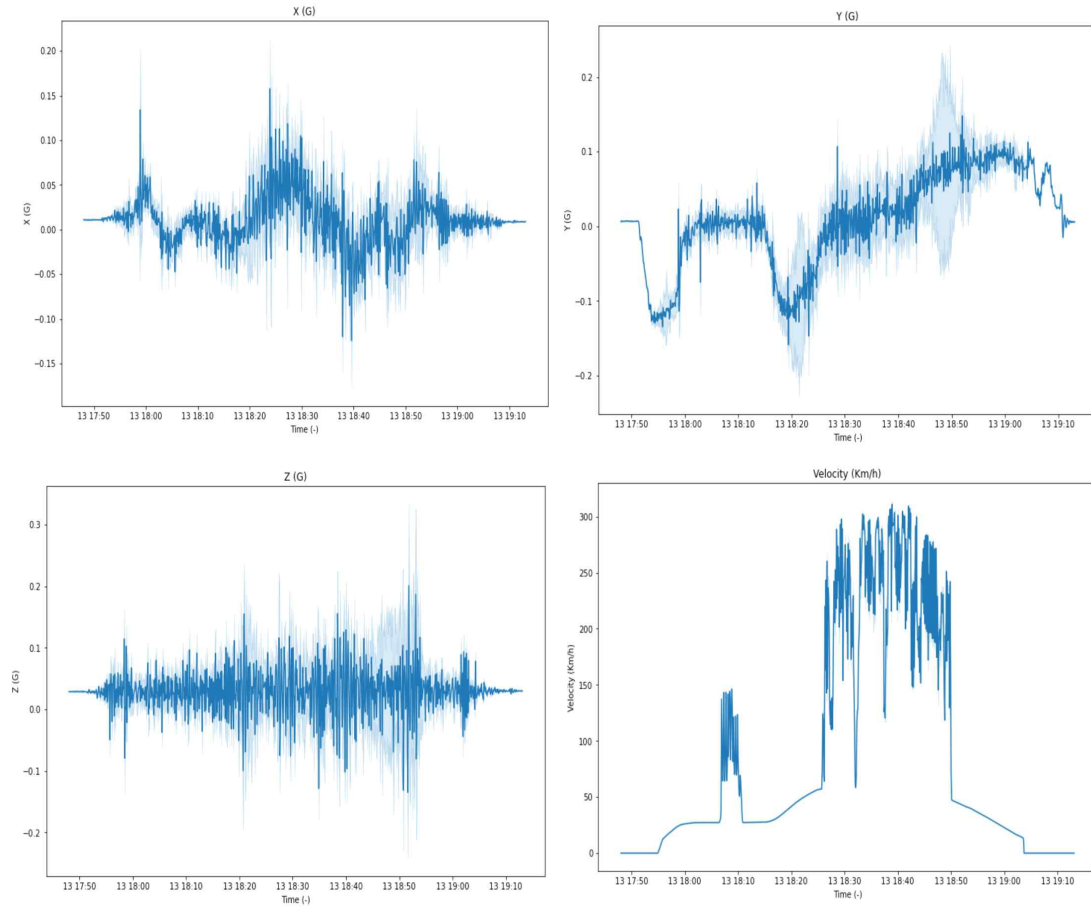
2. 첫 번째 운행(동부산대학교~고촌 방향)



X,Y,Z 가속도 및 차량의 속도(두번째 운행:동부산대학 ~ 고촌 방향)

- 첫 번째 운행 데이터의 동부산대학에서 고촌 방향의 분석결과 가속도 Y축(횡방향), 차량속도(Velocity)축의 관측패턴은 고촌에서 동부산대학 방향의 반전 패턴을 보이며 2개의 변수(인자) 연관성을 보임

3. 두 번째 운행(고촌~동부산대학교 방향)



X,Y,Z 가속도 및 차량의 속도(두번째 운행:고촌~동부산대학 방향)

- 두 번째 운행 데이터의 분석결과 가속도 X축(진행방향), Z(종방향)축의 변화는 첫 번째 운행 데이터의 일정한 패턴과 유사한 것으로 분석됨
- 두 번째 운행 데이터의 분석결과 가속도 Y축(횡방향), 차량 속도(Velocity)축의 관측패턴이 유사하며 이는 2개의 변수(인자) 연관성이 있으므로 판단되나 차량 속도 계측 값의 변화량은 크나 추세분석을 통하여 첫 번째 운행데이터와 유사한 형태를 지님

4. 두 번째 운행(동부산대학교~고촌 방향)

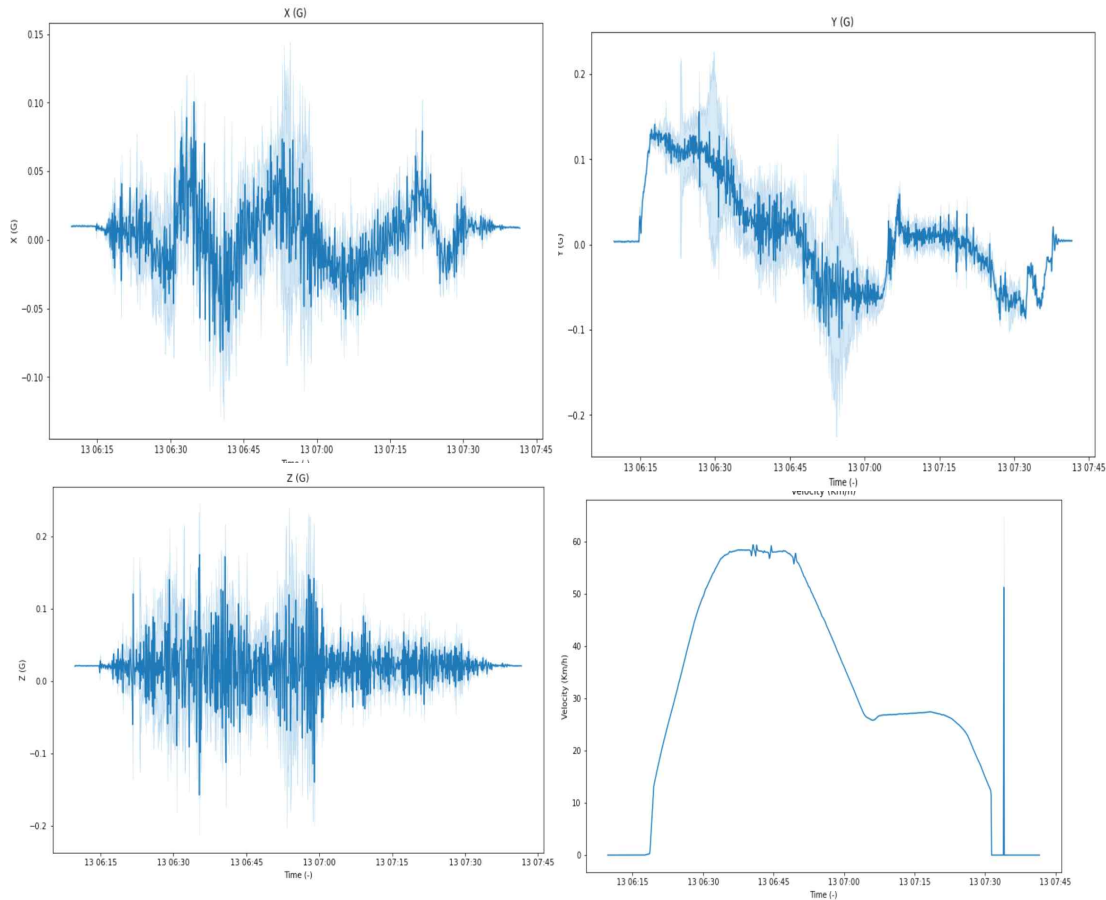


그림 26 가속도 및 차량의 속도(두번째 운행:동부산대학 ~고촌 방향)

- 첫 번째 운행에서는 차량의 속도가 급격하게 바뀌는 경우가 거의 없었지만 두 번째 운행에서는 속도가 급격하게 바뀌는 경우를 파악함. y축의 변화가 큼
- 따라서 두 번째 운행 시 속도 데이터의 이상값이 있는 것으로 판단
- x축 가속도와 z축의 가속도는 속도의 변화에 따라 크게 변화가 없는 것으로 분석됨
- 반면에 y축 가속도는 속도에 영향을 많이 받는 것으로 판단되며 속도와 Y축의 데이터는 고촌에서 동부산대학교 방향과 동부산대학교에서 고촌 방향이 서로 반전된 패턴으로 속도와 Y축의 영향이 있는 유효한 관계로 판단됨

2. 상관관계 분석

(1) 피어슨 상관관계 분석

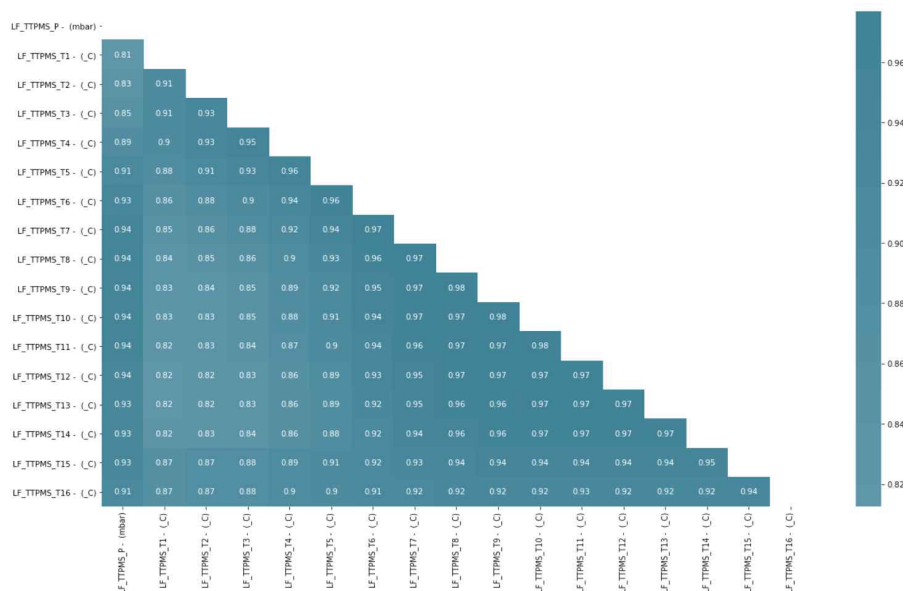
- 한 변수의 변화에 따른 다른 변수의 변화 정도와 방향을 예측하는 분석기법으로 간척도 이상의 두 변수 중에서 한 변수의 변화가 다른 변수의 변화에 따라 어떤 변화가 일어나는지를 보여주는 지표인 상관계수를 이용하고 각 변수 간의 상관관계 분석

- 피어슨 상관관계 분석(Pearson Correlation Coefficient)기법이 가장 일반적인 방법으로 두 변수가 서로 (선형) 상관 관계를 가지고 있는지를 확인하는 척도

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad -1 \leq \rho \leq 1$$

- 1이나 -1 에 가까우면 상관관계가 있다고 보고 ‘0’ 값이면 선형관계가 없음

○ 타이어 내부압력과 온도 변수(인자) 상관관계 분석

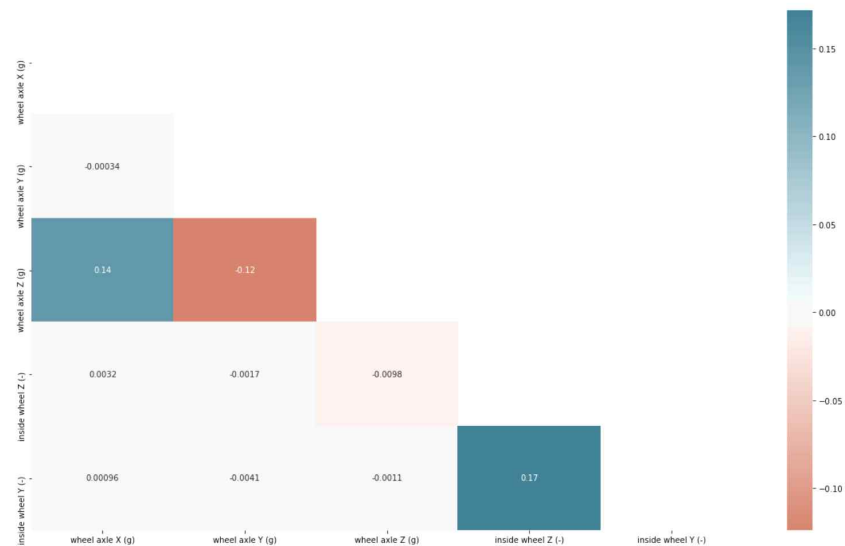


타이어 내부압력과 온도 인자의 상관계수 Matrix(경전철 차량 LF)

- 전방 왼쪽 타이어(LF)의 계측된 내부압력과 온도 인자(변수)의 상관관계지수가 1의 크기에 가까우며, 이는 온도와 압력의 선형 상관 관계성이 높음이 확인됨

- 대부분 0.8 이상의 높은 양의 선형 상관관계를 보임.

○ 차축 가속도와 내부 가속도 변수(인자) 상관관계 분석



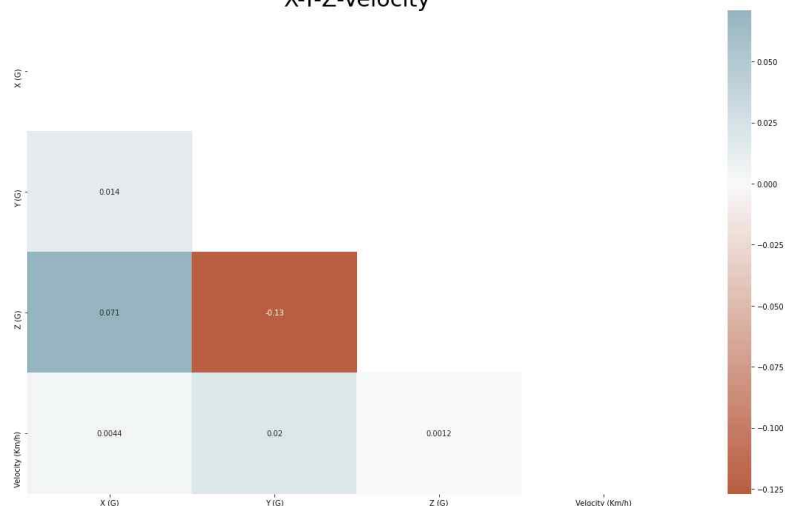
차축 가속도와 내부 가속도 변수(인자) 상관관계수 Matrix

- 피어슨 상관관계수 (Pearson Correlation Coefficient)를 이용한 선형 상관관계 분석결과 축상의 가속도와 휠 내부에서 측정된 가속도계의 값은 선형 상관관계는 작으므로 각 축과 내부 가속도의 상관관계는 거의 없는 것으로 분석됨

○ 차축 가속도와 차량 속도의 상관관계 분석

- 고촌에서 동부산대학교 방향 데이터의 선형 상관성

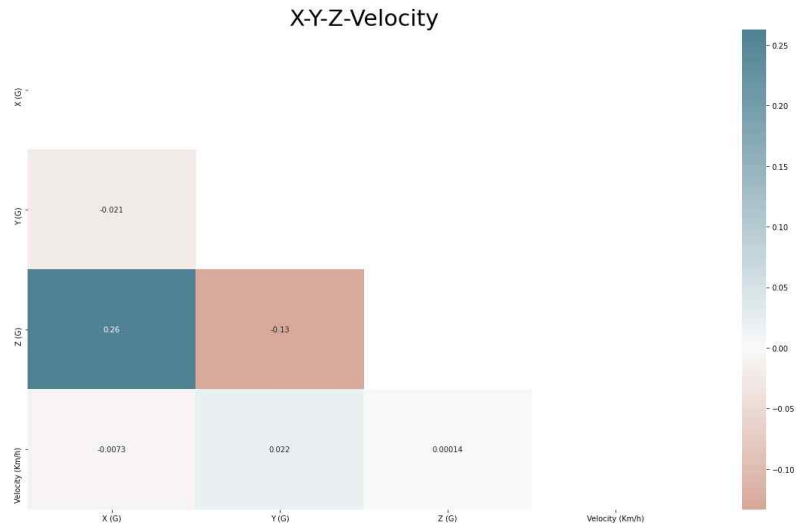
X-Y-Z-Velocity



가속도와 차량속도 상관관계수 Matrix

- 각 데이터의 선형 상관성은 거의 없는 것으로 분석되나 Y의 가속도와 Z의 가속도가 -0.13으로 약한 음의 선형 상관관계를 보임

- 동부산대학교에서 고촌 방향 데이터의 선형 상관성



가속도와 차량속도 상관계수 Matrix

- Y축 가속도와 Z축 가속도가 -0.13, X축 가속도와 Z축 가속도가 0.26으로 각각 약한 음의 상관관계 약한 양의 상관성이 보임.

- 각 데이터의 선형 상관성은 거의 없는 것으로 분석됨

(2) 교차 상관관계 분석(Cross Correlation)

- 교차 상관 분석(Cross Correlation) : 입력 시계열 자료와 출력 시계열 자료 사이의 연관성(linkage)을 알아보기 위하여 사용되는 것으로 입력 자료와 출력자료 간의 인과관계를 파악하는 분석기법

- 주행 중에 차량에서 측정되는 데이터에 대한 유효관계를 가진 변수(인자)를 찾기 위해서는 다양한 머신러닝 기법을 적용이 필요

- 수집되는 데이터의 특성상 시계열 자료로 구성되어 있어 이에 대한 타이어 헬스모니터링 인자로 사용할 수 있는 관련성에 대한 분석 필요

$$\rho_{XY}(k) = \text{Corr}(X_t, Y_{t+k}) = \frac{\gamma_{XY}(k)}{\sigma_X \sigma_Y} = \frac{\gamma_{XY}(k)}{\sqrt{\gamma_X(0) \gamma_Y(0)}}$$

- 파이썬 `scipy.signal` 라이브러리를 통해 교차 상관 관계를 계산할 수 있음

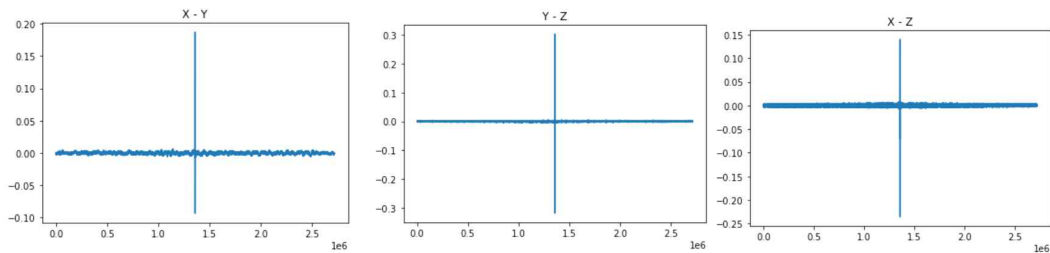

```

import matplotlib.pyplot as plt
from scipy import signal
import numpy as np
def lag_finder(y1, y2, sr):
    n = len(y1)
    corr = signal.correlate(y2, y1, mode='same') / np.sqrt(signal.correlate(y1, y1, mode='same')[int(n/2)] * signal.correlate(y2, y2, mode='same')[int(n/2)])
    delay_arr = np.linspace(-0.5*n/sr, 0.5*n/sr, n)
    delay = delay_arr[np.argmax(corr)]
    print('y2 is ' + str(delay) + ' behind y1')
    plt.figure()
    plt.plot(delay_arr, corr)
    plt.title('Lag: ' + str(np.round(delay, 3)) + ' s')
    plt.xlabel('Lag')
    plt.ylabel('Correlation coeff')
    plt.show()

```

그림 34 교차 상관 분석 코드

○ 차축 가속도와 내부 가속도 변수(인자) 교차 상관관계 분석

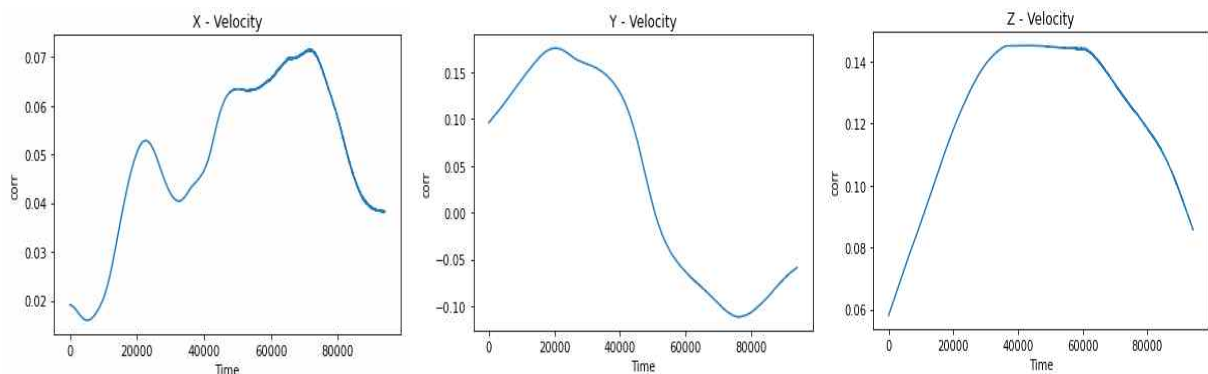


가속도계 X-Y, Y-Z, X-Z 축 간의 교차 분석 코드

- 가속도계의 계측된 값들 간의 교차 상관 분석 결과 특정 지점에서만 잠깐 높아지고 낮아지는 특징을 보이며 대부분 상관관계가 존재하지 않음으로 판단됨
- 피어슨 상관관계와 교차 상관관계 분석 모두 차축 가속도와 내부 가속도 변수(인자) 데이터에 대한 유의미한 결과를 얻을 수 없었음

○ 차축 가속도와 차량속도 변수(인자) 교차 상관관계 분석

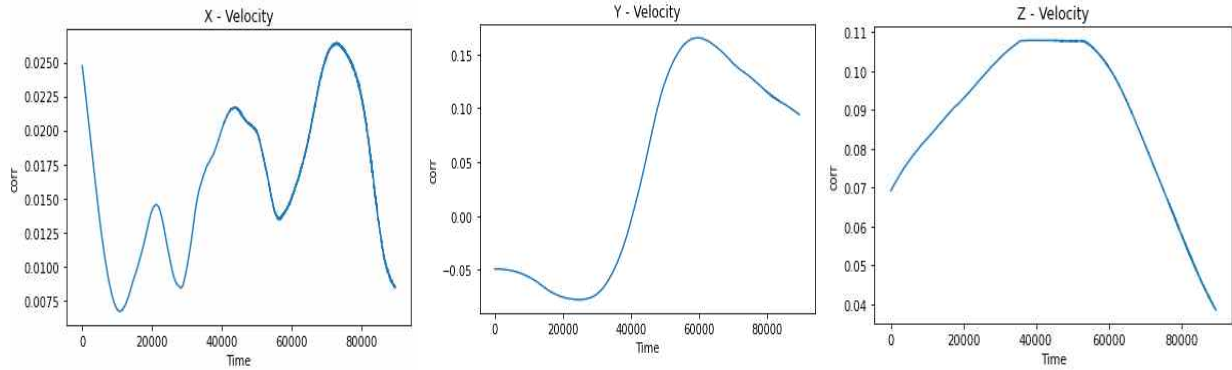
- 고촌에서 동부산대학교 방향 데이터의 교차 상관성



가속도와 차량속도 변수의 교차 상관관계(고촌~동부산대학교 방향)

- 속도와 X축의 가속도는 비교적 양의 상관관계를 보였지만 낮은 상관관계를 보임
- 속도와 Y축의 가속도는 약한 양과 음의 상관관계가 시간의 흐름에 따라 달라짐
- 속도와 Z축의 가속도는 약한 양의 상관관계가 보임

- 동부산대학교에서 고촌 방향 데이터의 교차 상관성



가속도와 차량속도 변수의 교차 상관관계(동부산대학교~고촌 방향)

- X축의 가속도와 속도의 상관관계는 매우 낮음
- 속도와 Y축의 가속도는 약한 양과 음의 상관관계가 시간의 흐름에 따라 달라지는 패턴을 보임
- 속도와 Z축의 가속도는 약한 양의 상관관계가 보임
- 서로 반대되는 노선의 데이터에서 패턴이 보인 것은 가속도와 차체 속도와의 교차 상관성이 Y축 가속도의 데이터가 가장 관련성 있음을 보임
- 고촌에서 동부산대학교 , 동부산대학교에서 고촌역의 차체 속도와 Y축 가속도의 교차 상관성이 반대되는 패턴을 가짐

II. 머신러닝 예측기법을 통한 영향인자 관계 분석

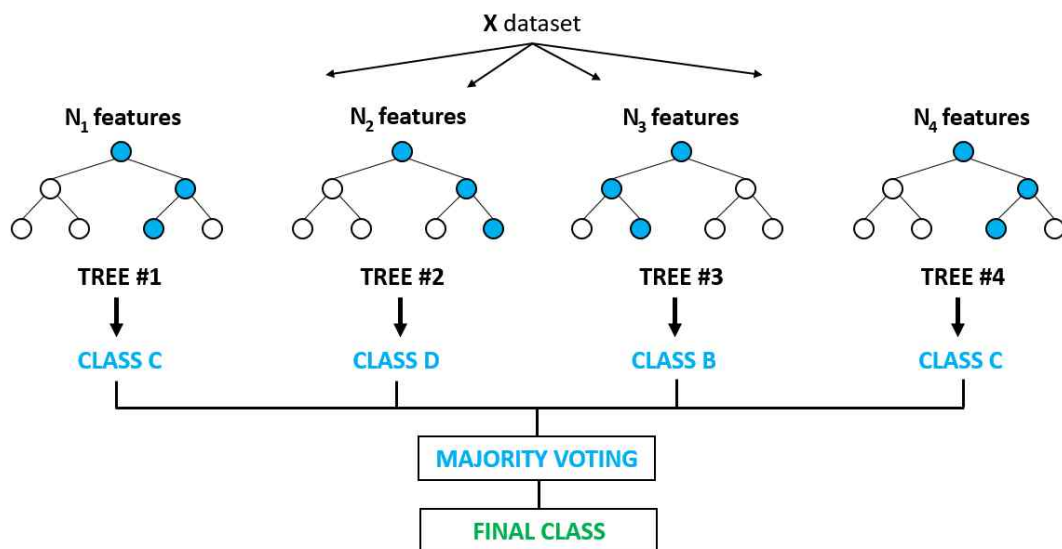
- 계측된 데이터를 기반으로 변수(인자)들 간의 관계분석 기법을 적용하여 타이어 내부온도와 압력이 가장 유효한 관계임으로 판단되며, 가속도 Y축과 차량속도와의 상관관계가 있음을 확인하였음

- 타이어 헬스모니터링 알고리즘에 적용하기 위한 인자로서의 관계 유효성을 파악하기 위해 머신러닝 기반의 예측기법을 적용을 수행함

○ 머신러닝 기법

① RandomForest 알고리즘

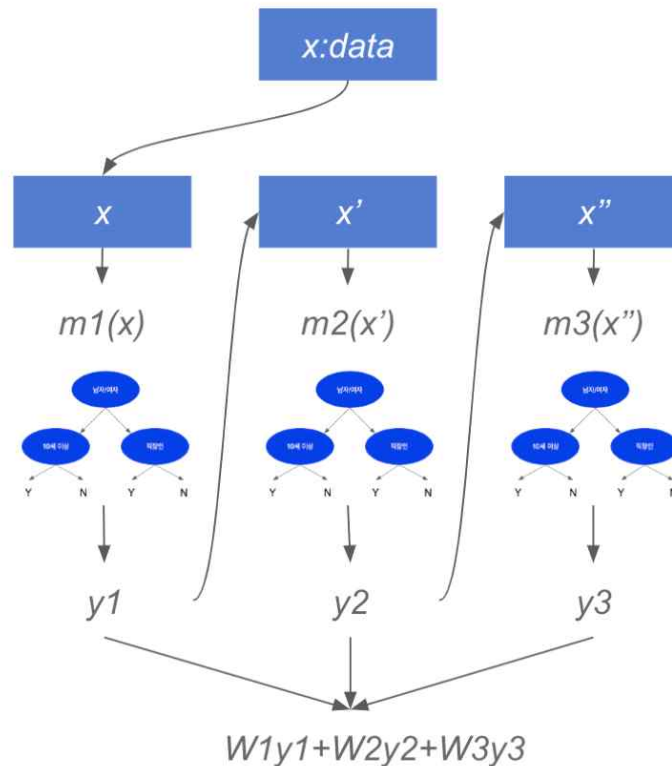
- 랜덤 포레스트는 데이터셋의 다양한 하위 샘플을 이용해서 다수의 의사결정 나무를 학습하는 메타 예측방식으로 과적합을 조절하고 예측의 정확도를 높이기 위해 예측값의 평균값을 사용
- 파이썬 모듈 : `from sklearn.ensemble import RandomForestRegressor` 사용
- 하이퍼파라미터 : `n_estimators : 100`, `criterion : 'mse'`, `max_depth=None`, `max_features : auto`



RandomForest 동작 원리

② Xgboost 알고리즘

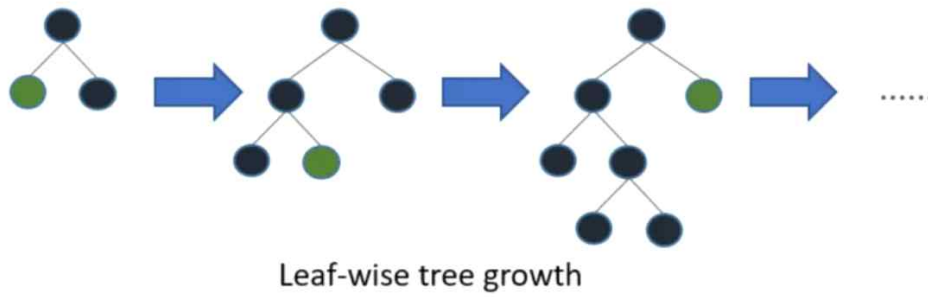
- xgboost는 여러 개의 Decision Tree를 조합해서 사용하는 Ensemble 알고리즘 기반 구조
- xgboost는 부스팅의 단점인 과대적합과 연산 속도의 저하를 규제 방식과 early stopping으로 극복하고 경사하강법에 가중치를 부여하여 학습하는 방식
- 파이썬 모듈 : `from xgboost import XGBRegressor`
- 하이퍼파라미터 : `n_estimators : 100, max_depth : 3, booster='dart', base_score=0.5`



XGBoost 동작원리

③ LGBM(Light GBM) 알고리즘

- 예측의 트리를 수평적(level wise)으로 늘리는 기존 부스팅 모델과는 달리 수직적으로 늘리는 방식(leaf wise)
- 트리의 균형을 맞추지 않고 리프노드를 지속적으로 분할 이로 인해 비대칭적인 깊은 트리 생성되고 손실을 줄일 수 있음
- 파이썬 모듈 : `from lightgbm.sklearn import LGBMRegressor`
- 하이퍼파라미터 : `n_estimators : 100, learning_rate : 0.1, boosting_type : 'gbdt', min_child_samples : 20`



LGBM 동작 원리

○ 평가지표

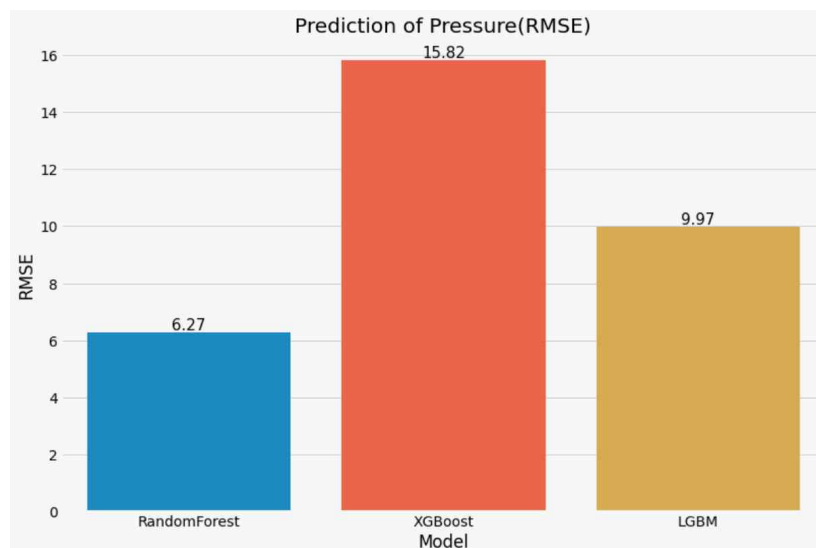
- RMSE(Root Mean Squared Error) : 제공된 에러를 다시 루트로 풀어주기 때문에 에러를 제공해서 생기는 값의 왜곡을 감소시킬 수 있음

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

- y 값은 실제 값, \hat{y} 값은 예측된 값, n 은 데이터의 개수

○ 타이어 내부 온도에 따른 타이어 내부 압력 예측

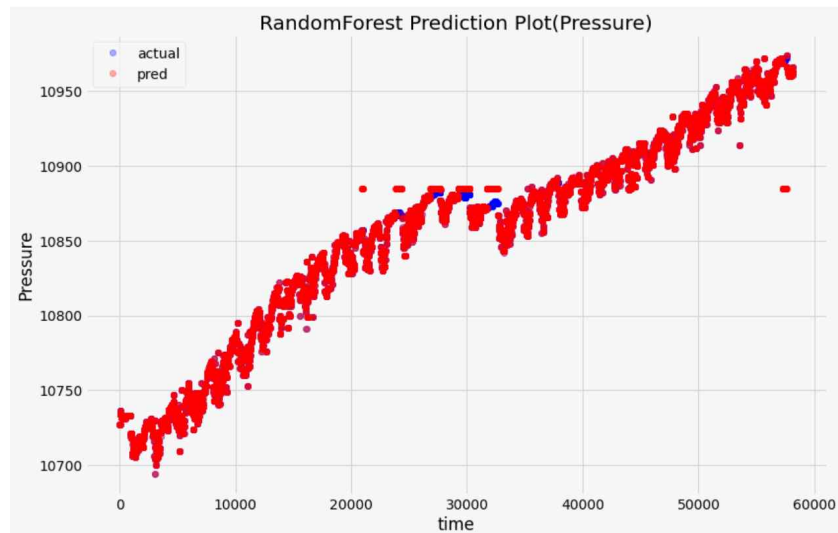
- 입력 데이터 : 타이어 내부 온도(LF_TTPMS_T1~LF_TTPMS_T16)
- 예측 데이터 : 타이어 내부 압력(LF_TTPMS_P)
- 온도와 압력의 선형 상관관계가 입증되었기 때문에 온도의 데이터로 압력을 예측



예측 모델별 RMSE(타이어 내부온도에 따른 압력 예측)

- RandomForestRegressor 모델의 RMSE가 6.27로 가장 우수했고 XGBoost 모델이 15.62로 오차가 가장 높았다.

- 종속변수(압력)의 범위는 10694~10974로 범위에 비해 오차가 6.27임에 따라 예측 성능이 가능함을 보였으며 두 변수 간의 관계성을 보일 수 있으므로 분석됨



예측 값과 실제 값 비교(파란: 실제값, 빨간: 예측값)

- 예측값과 실제값이 비슷한 패턴을 보여 예측이 잘 되었지만 오차가 큰 값도 보임

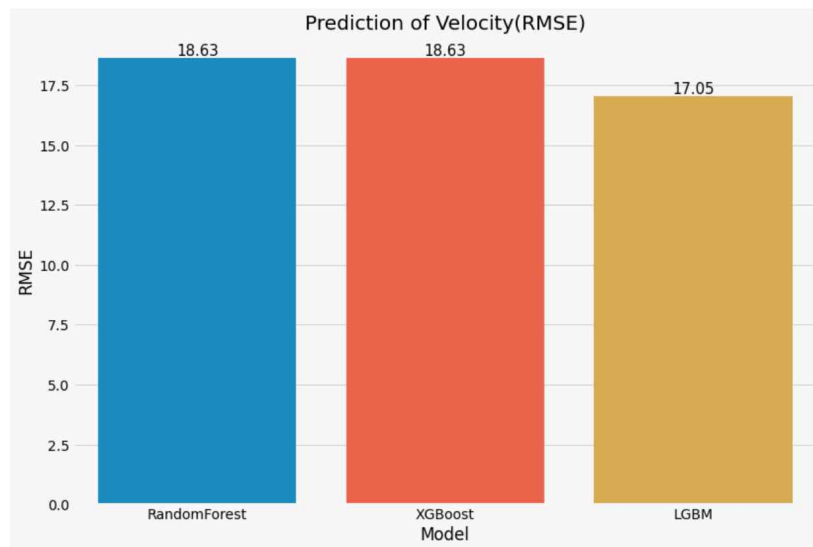
○ Y축 가속도에 따른 차체 속도 예측

- 입력 데이터: 차축 가속도(Y축)

- 예측 데이터: 차량 속도

- Y축의 가속도와 속도는 교차 상관성 분석에서 시간의 흐름에 따라 상관관계가 변화하는 패턴 파악

- 머신러닝 모델에 학습시켜 차체 속도를 예측



예측 모델별 RMSE(Y축 가속도와 차체속도)

- RandomForest와 XGBoost는 18.64로 같은 값이 나왔고, LGBM 알고리즘은 17.05로 가장 낮은 값으로 LGBM 알고리즘이 비교적 예측이 잘 됨
- 종속변수 차체 속도의 데이터 범위가 0~58.49631985로 범위에 비해 오차가 약간 큰 편임
- 현재 데이터를 기반으로 타이어 내부온도와 압력의 관계성만큼 유효한 인자로 사용하기는 어려울 것으로 판단됨

Ⅲ. 결론

- 축상 가속도와 차체 가속도 및 타이어 온도, 압력에 대해서 탐색적 데이터 분석과 상관 분석을 진행함
- 선형적 상관관계 분석인 피어슨 상관계수를 이용하여 분석한 결과 차축 타이어의 온도와 압력이 높은 양의 선형 상관관계가 나타남
- 차축 타이어의 온도와 압력 간의 데이터 유의성 판단을 하기 위해 머신러닝 기법을 이용하여 타이어의 온도 데이터로 타이어의 압력을 예측
- 예측 결과 실제 값과 예측의 오차가 작음. RandomForest 알고리즘이 가장 예측력이 뛰어남
- 시계열 데이터의 상관관계 분석 방법론인 교차 상관관계 분석 결과 Y축의 차축 가속도와 차체 속도의 패턴 파악
- Y축 가속도와 차체 속도의 유의성 패턴 검증을 위해 머신러닝 예측 알고리즘 이용하여 Y축 가속도를 이용하여 차체 속도를 예측
- LGBM 알고리즘이 가장 예측력이 좋았지만 오차가 커서 예측력의 결여
- Y축 가속도와 차체 속도의 패턴의 유의미성 입증 불가
- 추후 차체 속도 및 가속도, 타이어의 온도와 압력 데이터의 동시간대 자료가 요구됨
- 타이어의 마모도를 표현할 수 있는 데이터가 추가로 요구됨
- 타이어의 마모도 데이터도 가속도와 속도, 온도, 압력 데이터와 동시 수집 요구
- 마모도 데이터가 존재 시 유의미한 변수들을 추출하는데 용이