QI Summer AI project

#week_2

#Team_C

#HyeonjeongKim

#SanGwon

#JoeunKim

#YejinLee

Contents

Data Selection Criterion

Data Set

Data Analysis

Data Selection Criterion

Subject: A Study on the Prediction of Subway in Seoul by Weather

Date & Time

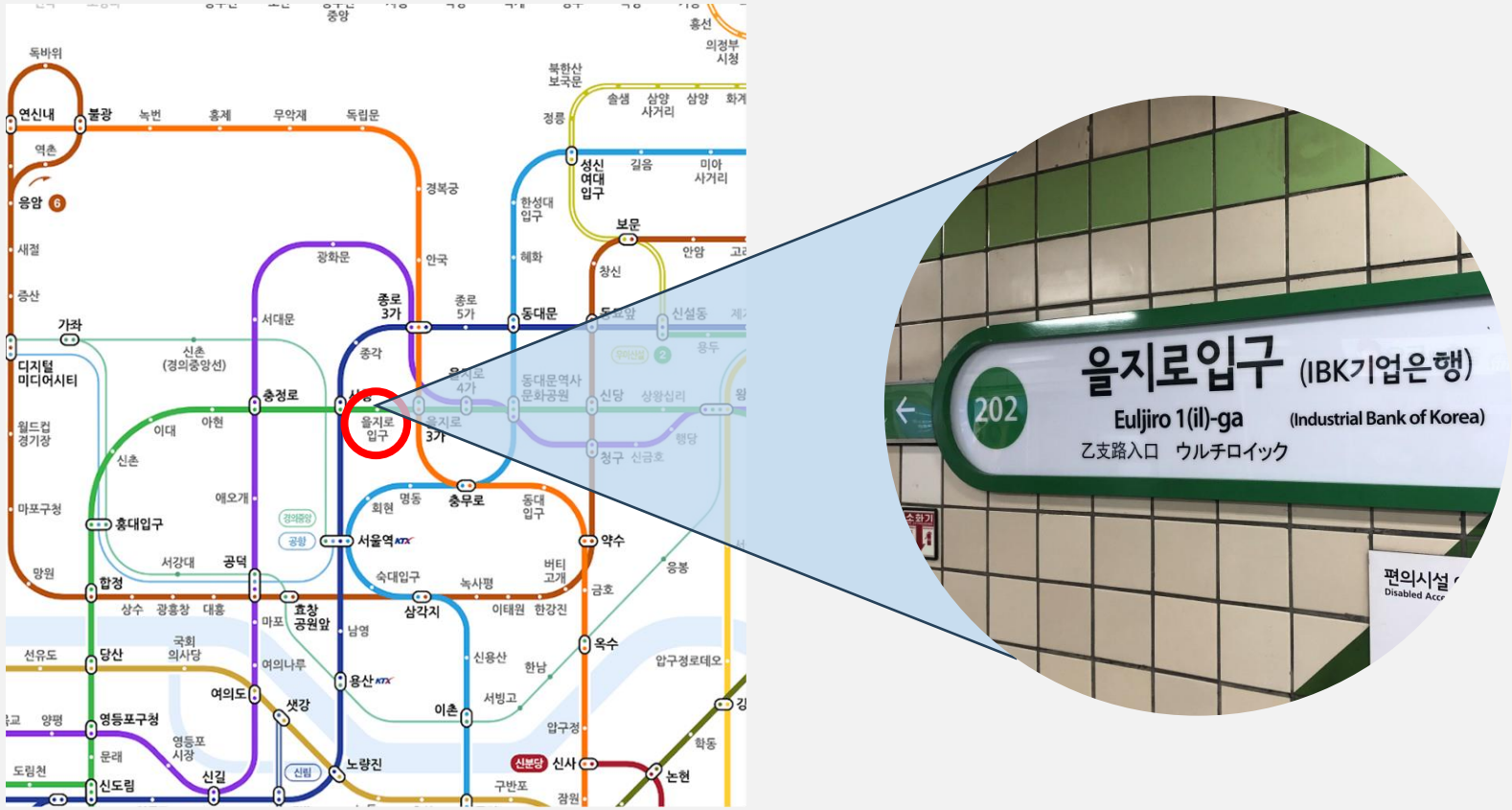
Rain

Sensible Temperature

Subway usage

Data Selection Criterion

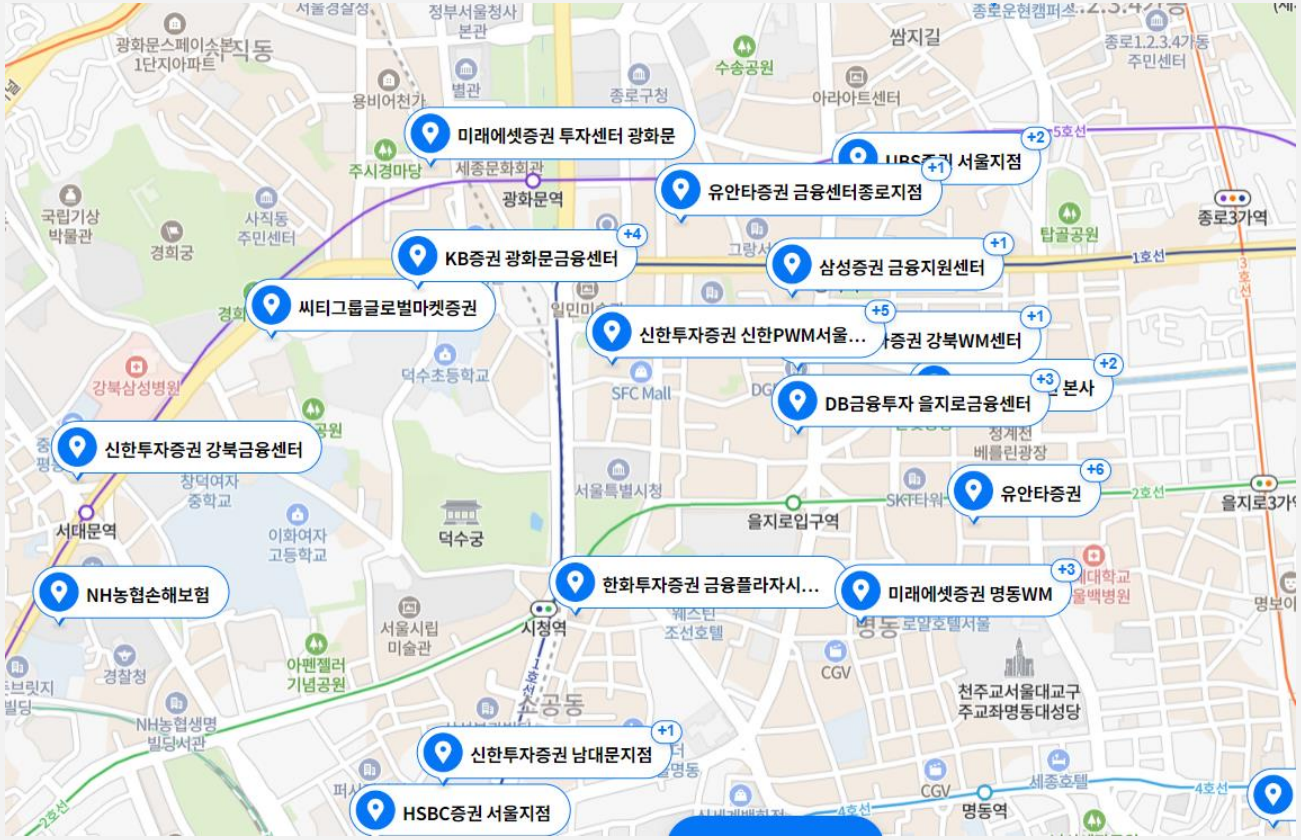
Q. Why did we select ‘Euljiro 1(il)-ga’ station



[Fig 1]. Subway Line by navermap

Data Selection Criterion

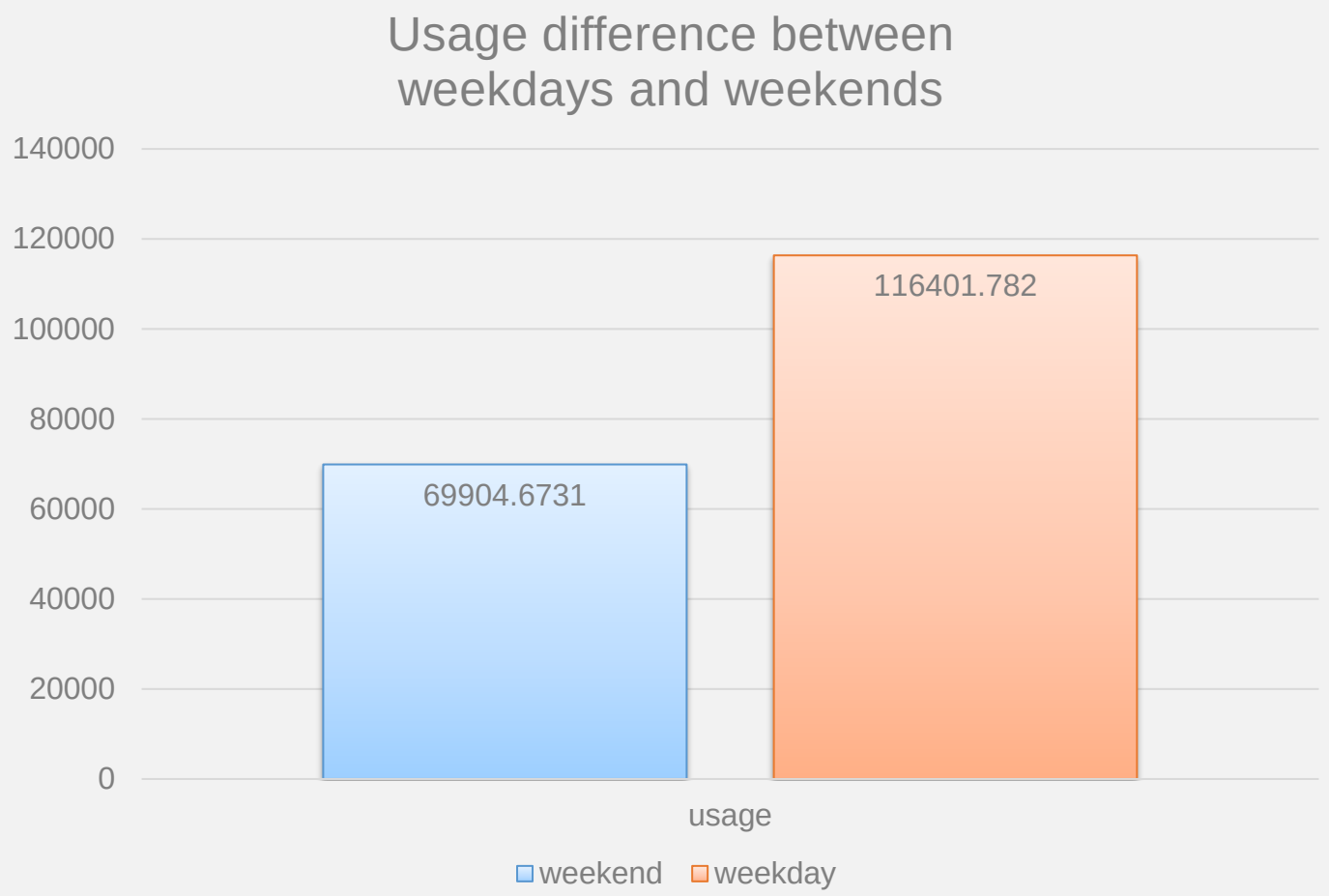
Q. Why did we select ‘Euljiro 1(il)-ga’ station



[Fig 2]. The list about some of the companies nearby the station by navermap

Data Selection Criterion

Q. Why did we exclude holiday and weekend data?



Data Selection Criterion

Q. Why did we choc ata?

- 체감온도 = -0.2442 + 0.553)278TwTa + 3.0(기존 3.5)

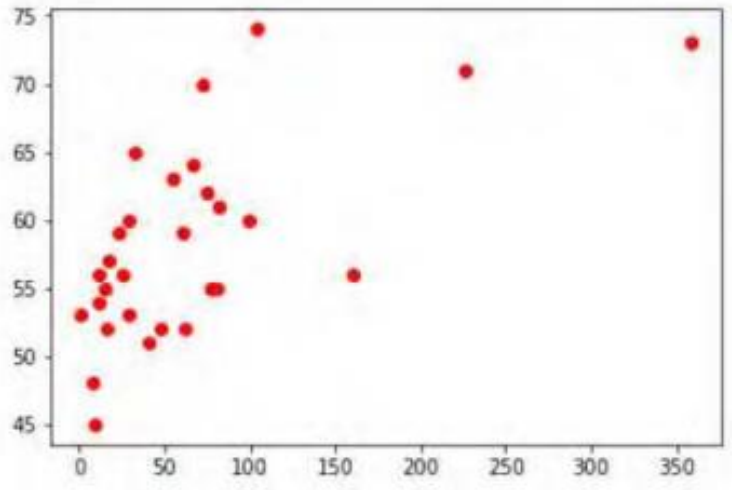
* Tw = TaATAN[0.151977(RH+8.313659) 8RH3/2ATAN(0.023101RH) - 4.686035

** Ta : 기온(℃), Tw : 습구온도(Stull의 추

[Fig 3.] Korea Meteorological Admin

2.2 날씨 데이터 처리

본 연구에서는 여러 가지 날씨 데이터 중 대중교통에 원인을 주는 요인으로 (평균)기온(℃), 강수량(mm), (평균) 상대습도(%), 미세먼지(㎍/㎥) 4가지의 날씨 데이터를 선정했다. 선정된 날씨 데이터 중 강수량과 상대습도의 경우 상관관계를 계산해 본 결과 (그림 1)에서도 알 수 있듯이 약 +0.641로 +0.3~+0.7 사이에 존재하여 뚜렷한 양적 선형 관계를 가진다.



(그림 1) x축 : 강수량, y축 : 상대습도

[Fig 4.] A Study on the Prediction of Public Transportation Consumption in Seoul by Weather by Hee-jin Kim, Sujin Oh, Ung-Mo Kim

Data Selection Criterion

Q. How can we get sensible data?

체감온도

▪ 자료설명

체감온도란 인간이 느끼는 더위나 추위를 수량적으로 나타낸 것으로 여름철(5~9월)과 겨울철(10~익년 4월)을 구분하여 제공합니다.

여름철 체감온도는 **일 최고 체감온도**를 제공합니다.

※ 2022.6.2.부터 여름철 체감온도 산출식 변경(기존 3.5 ⇒ 3.0)

- 체감온도 = -0.2442 + 0.55399Tw + 0.45535Ta – 0.0022Tw^2 + 0.00278TwTa + 3.0(기존 3.5)

* Tw = TaATAN[0.151977(RH+8.313659)1/2] + ATAN(Ta+RH) - ATAN(RH-1.67633) + 0.00391838RH3/2ATAN(0.023101RH) - 4.686035

** Ta : 기온(℃), Tw : 습구온도(Stull의 추정식** 이용), RH : 상대습도(%)

겨울철 체감온도는 **일 최저 체감온도**를 제공하며, **기온 10℃ 이하, 풍속 1.3 m/s 이상**일때만 산출합니다.

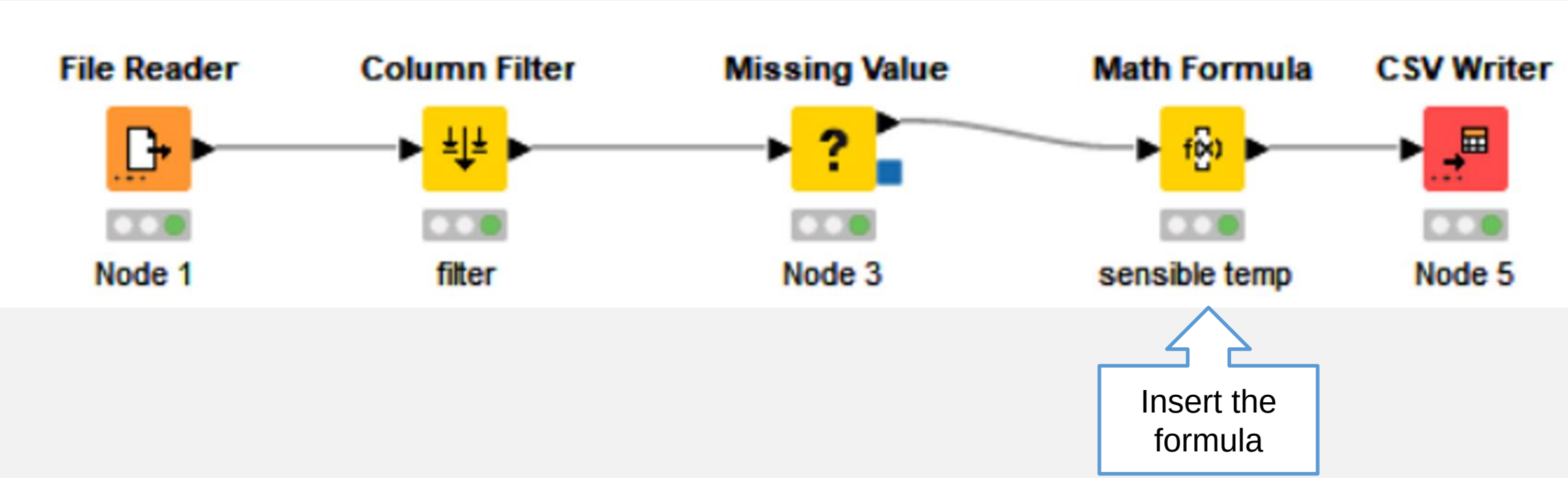
- 체감온도 = 13.12 + 0.6215Ta - 11.37 $V^{0.16}$ + 0.3965 $V^{0.16}$Ta

* T : 기온(℃), V : 10분 평균 풍속(km/h)

[Fig 5.] Korea Meteorological Administration

Data Selection Criterion

Q. How can we get sensible data?



Data Set

	Date	Station Num	Station Name	Category	05~06	06~07	07
2	2017-01-01	150	서울역	승차	470	286	
3	2017-01-01	150	서울역	하차	278	880	
4	2017-01-01	151	시청	승차	204	105	
5	2017-01-01	151	시청	하차	73	203	

	Date	Line	Station Num	Station Name	Category	05 ~ 06	06 ~ 07	
2	2018-01-01	1호선	150	서울역	승차	373	318	
3	2018-01-01	1호선	150	서울역	하차	205	1040	
4	2018-01-01	1호선	151	시청	승차	87	105	
5	2018-01-01	1호선	151	시청	하차	47	294	

	Date	Line	Station Num	Station Name	Category	06시 이전	06 ~ 07	
2	2019-01-01	1호선	150	서울역	승차	348	321	
3	2019-01-01	1호선	150	서울역	하차	222	821	
4	2019-01-01	1호선	151	시청	승차	87	98	
5	2019-01-01	1호선	151	시청	하차	48	237	
6	2019-01-01	1호선	152	종각	승차	669	318	
7	2019-01-01	1호선	152	종각	하차	68	179	
8	2019-01-01	1호선	153	종로3가	승차	227	198	
9	2019-01-01	1호선	153	종로3가	하차	38	180	
10	2019-01-01	1호선	154	종로5가	승차	48	80	
11	2019-01-01	1호선	154	종로5가	하차	25	102	
12	2019-01-01	1호선	155	동대문	승차	138	117	

◀ Before Preprocessing

- all subway stations managed by the Seoul Transportation Corporation

- the number of people getting on and off from 5 a.m. to 24 a.m. by station from 2017 to 2019.

#column 24

#2017 row: 200750

#2018 row: 200750

#2019 row: 200804

Data Set

◀ Before preprocessing

	area code	date and time	temperature	precipitation	wind speed	wind directio	humadity	vapor pressur
2	108	2017.1.1 0:00	0.2		2.2	20	79	4.9
3	108	2017.1.1 1:00	0		1.4	20	78	4.8

	area code	date and time	temperature	precipitation	wind speed	wind directio	humadity	vapor pressur
2	108	2018.1.1 0:00	-3.2		0.5	110	40	1.9
3	108	2018.1.1 1:00	-3.3		0.7	360	41	2
4	108	2018.1.1 2:00	-3.7		0.9	270	42	2

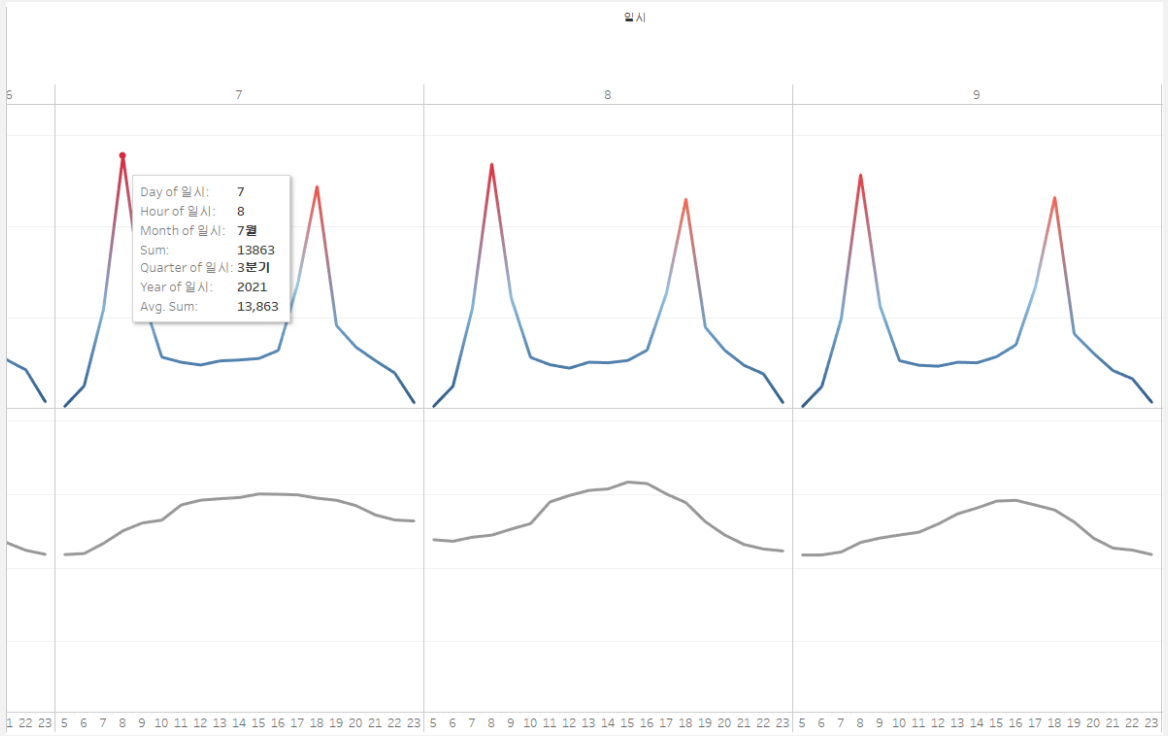
	area code	date and time	temperature	precipitation	wind speed	wind directio	humadity	vapor pressur
2	108	2019.1.1 0:00	-5.5		1	290	54	2.2
3	108	2019.1.1 1:00	-5.9		1.8	270	56	2.2
4	108	2019.1.1 2:00	-6.5		1.2	290	60	2.3
5	108	2019.1.1 3:00	-6.9		2.2	270	62	2.3
6	108	2019.1.1 4:00	-7.2		1.2	270	57	2
7	108	2019.1.1 5:00	-7.6		2.1	270	59	2
8	108	2019.1.1 6:00	-7.9		1.3	290	60	2
9	108	2019.1.1 7:00	-7.7		1.3	320	58	2
10	108	2019.1.1 8:00	-7.7		0.9	360	57	2
11	108	2019.1.1 9:00	-7		0.9	290	50	1.8
12	108	2019.1.1 10:00	-4.9		1.6	290	44	1.9

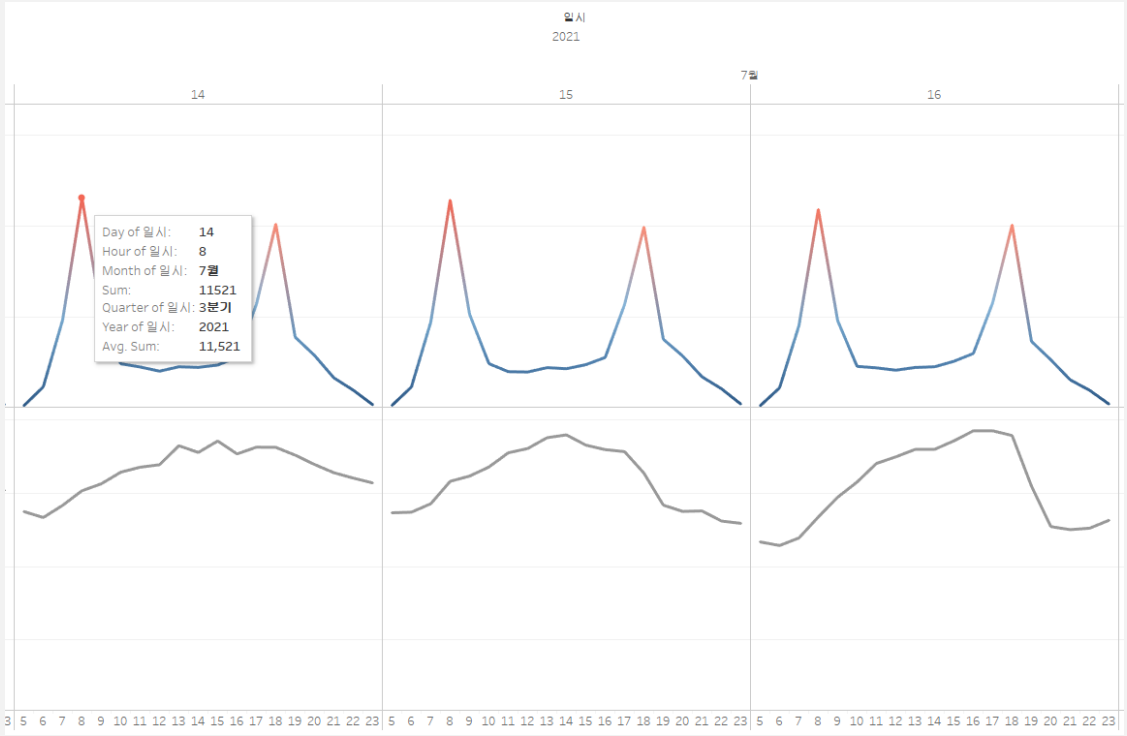
#total row 8760

#column 27

Data Selection Criterion

Q. Why did we use past data sets?



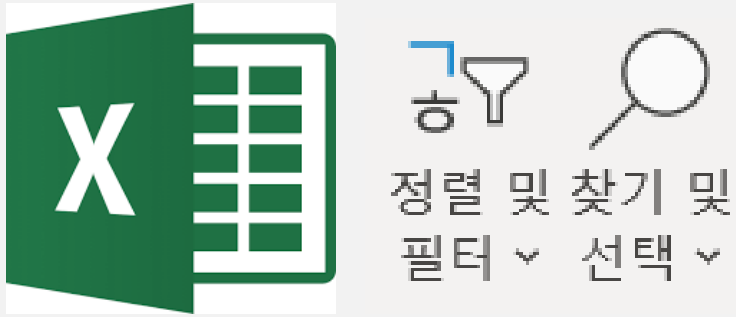


Data Set

일시	시간	기온(°C)	강수
2017-06-01	5	20.6	
2017-06-01	6	20.9	
2017-06-01	7	20.9	
2017-06-01	8	21.9	
2017-06-01	9	23.2	
2017-06-01	10	23.7	
2017-06-01	11	24.5	
2017-06-01	12	25.7	

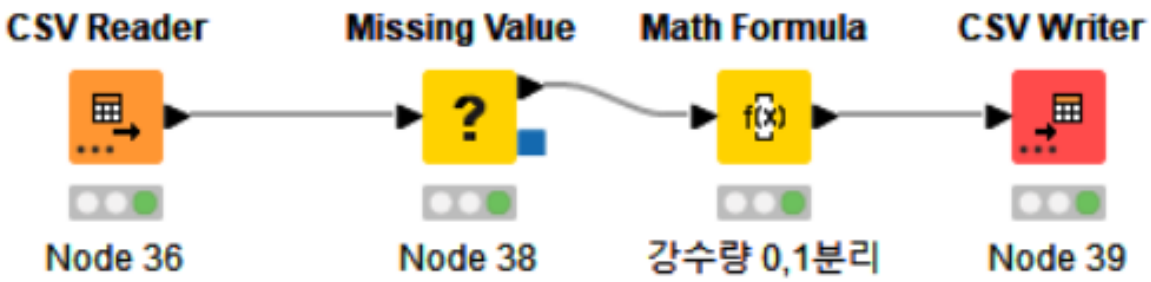
ex) 2018.04.11 7:00 → 2018-04-11 7

Data Set



2017.9.28	을지로입구	147	113	338	
2017.9.29	을지로입구	126	117	349	
2017.10.2	을지로입구	81	90	123	
2017.10.10	을지로입구	133	116	315	
2017.10.11	을지로입구	120	86	337	
2017.10.12	을지로입구	142	110	351	
2017.10.13	을지로입구	132	151	349	
2017.10.16	을지로입구	130	114	371	

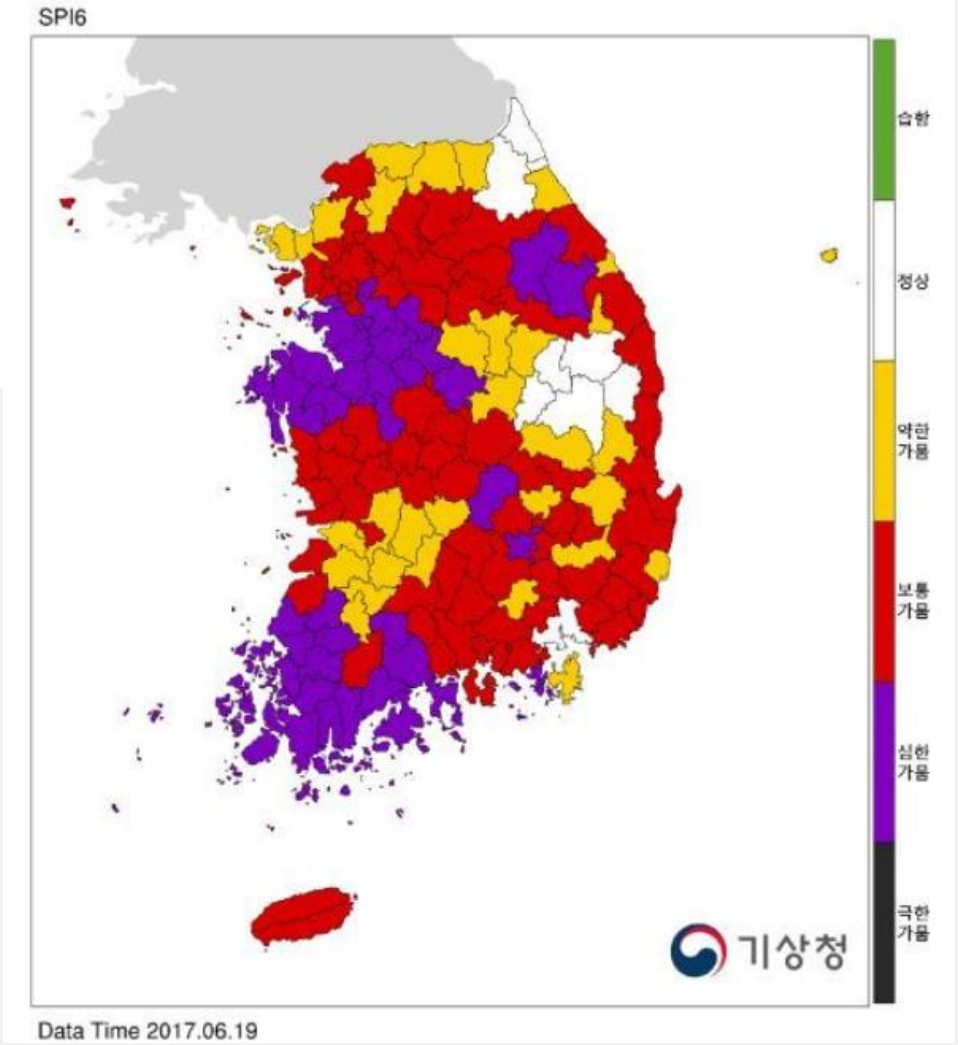
Data Set



강수량(mm)
0
0.3
0
0
0
0
0
0

Rain
0
1
0
0
0
0
0
0

Rainy day → 1
Not rainy day → 0



[Fig 6]. Korea Meteorological Agency

Data Set

Date	Temperature(°C)	Rain	Wind(m/s)	Humidity(%)	sensible_temp	Time	Usage
2017-03-02	1.1	0	3.7	93	-2.74237924	5	99

Date	Temperature	Rain	Wind(m/s)	Humidity(%)	sensible_temp	Time	Usage
2017-06-01	20.6	0	0.9	82	22.61547395	5	120
2017-06-01	20.9	0	2.1	80	22.78672561	6	115

Date	Temperature(°C)	Rain	Wind(m/s)	Humidity(%)	sensible_temp	Time	Usage
2017-09-01	17.9	0	1	70	19.10322166	5	122
2017-09-01	17.3	0	1.4	78	19.01013333	6	111

Date	Temperature(°C)	Rain	Wind(m/s)	Humidity(%)	sensible_temp	Time	Usage
2017-01-02	3.7	0	1.5	80	2.449322675	5	107
2017-01-02	3.6	1	1.1	89	3.6	6	122
2017-01-02	3.6	0	1.1	88	3.6	7	347
2017-01-02	3.8	0	1.4	87	2.705222674	8	601
2017-01-02	4.1	0	0.9	87	4.1	9	726
2017-01-02	5.2	0	2.4	86	3.20839456	10	1032
2017-01-02	6.3	0	2.3	82	4.592916045	11	1350
2017-01-02	7.6	0	2.1	70	6.293104325	12	1517
2017-01-02	8.1	0	3.1	69	6.152841377	13	1970

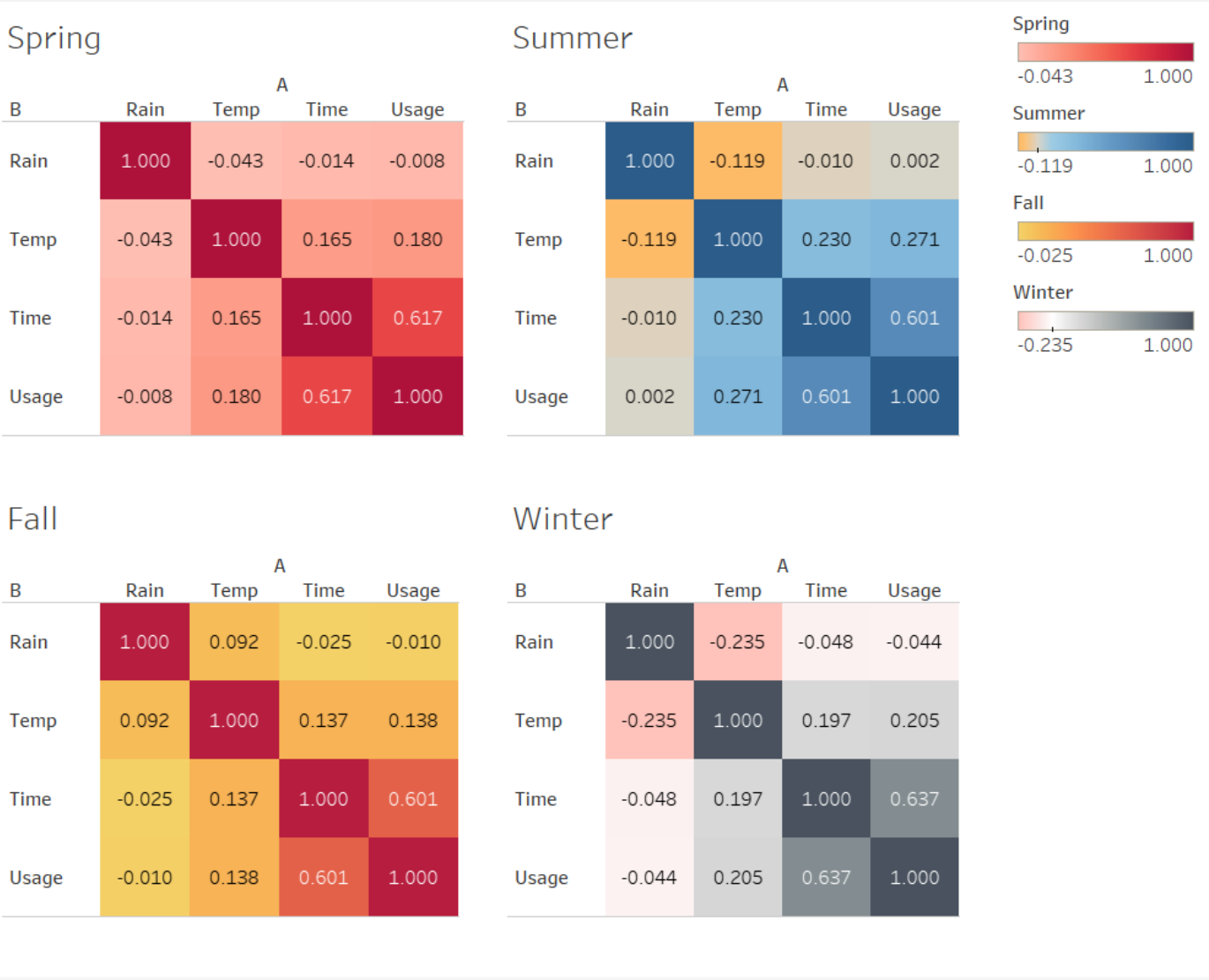
◀ **After Preprocessing**

- Divide into 4 seasons

#total row 14041

#column 8

Data Analysis



Data Analysis

Variable	Coeff.	Std. Err.	t-value	P>|t|
precipitation	29.2536	176.8146	0.1654	0.8686
sensible_temp	23.7006	5.17	4.5843	4.71E-6
time	316.4231	6.8791	45.9978	0.0
Intercept	-1,784.5719	116.8517	-15.2721	0.0

R-Squared: 0.3838
Adjusted R-Squared: 0.3833

Spring R-Squared < 0.5 : low relationship

Data Analysis

Variable	Coeff.	Std. Err.	t-value	P>|t|
precipitation	1.5572	130.1218	0.012	0.9905
sensible_temp	84.5179	10.6084	7.9671	2.22E-15
time	306.5327	7.1061	43.1364	0.0
Intercept	-3,676.0673	290.3004	-12.663	0.0

R-Squared: 0.3724
Adjusted R-Squared: 0.3719

Summer R-Squared < 0.5 : low relationship

Data Analysis

Variable	Coeff.	Std. Err.	t-value	P>|t|
precipitation	-5.4189	189.4627	-0.0286	0.9772
sensible_temp	15.6388	4.826	3.2406	0.0012
time	321.0237	7.4073	43.339	0.0
Intercept	-1,713.1421	124.6292	-13.7459	0.0

R-Squared: 0.3638
Adjusted R-Squared: 0.3632

Fall R-Squared < 0.5 : low relationship

Data Analysis

Variable	Coeff.	Std. Err.	t-value	P>|t|
precipitation	-2.5592	115.1345	-0.0222	0.9823
sensible_temp	27.1674	6.202	4.3804	1.22E-5
time	336.333	7.0898	47.4387	0.0
Intercept	-1,594.8166	111.0058	-14.367	0.0

R-Squared: 0.4097
Adjusted R-Squared: 0.4092

Winter R-Squared < 0.5 : low relationship

reference material

- Hee-jin Kim, Sujin Oh, Ung-Mo Kim. (2017). A Study on the Prediction of Public Transportation Consumption in Seoul by Weather
- Seoul Transportation Corporation. (2017-2019). Information on the number of people boarding and alighting by station, date, and time. *https://www.data.go.kr/data/15048032/fileData.do#layer_data_infomation*
- Open MET Data Portal. (2017-2019). Seoul Daily Weather Observation *https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36*
- Naver-map. Information around the station & Seoul subway map

Thank you