An abstract 3D composition featuring several geometric shapes: a large brown cube, a smaller blue cube, a yellow sphere, and a dark blue cylinder. The shapes are arranged in a dynamic, overlapping manner against a dark blue background. A thin white horizontal line is positioned above the main title text.

Prediction of subway usage by time according to weather

2023.07.28

San Gwon, Joeun Kim, Hyeonjeong Kim, Yejin Lee

Team C

Content

- 1 Multi Linear Regression
- 2 Data interpolation
- 3 Data Analysis

Part 1

Multiple Linear Regression

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

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

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

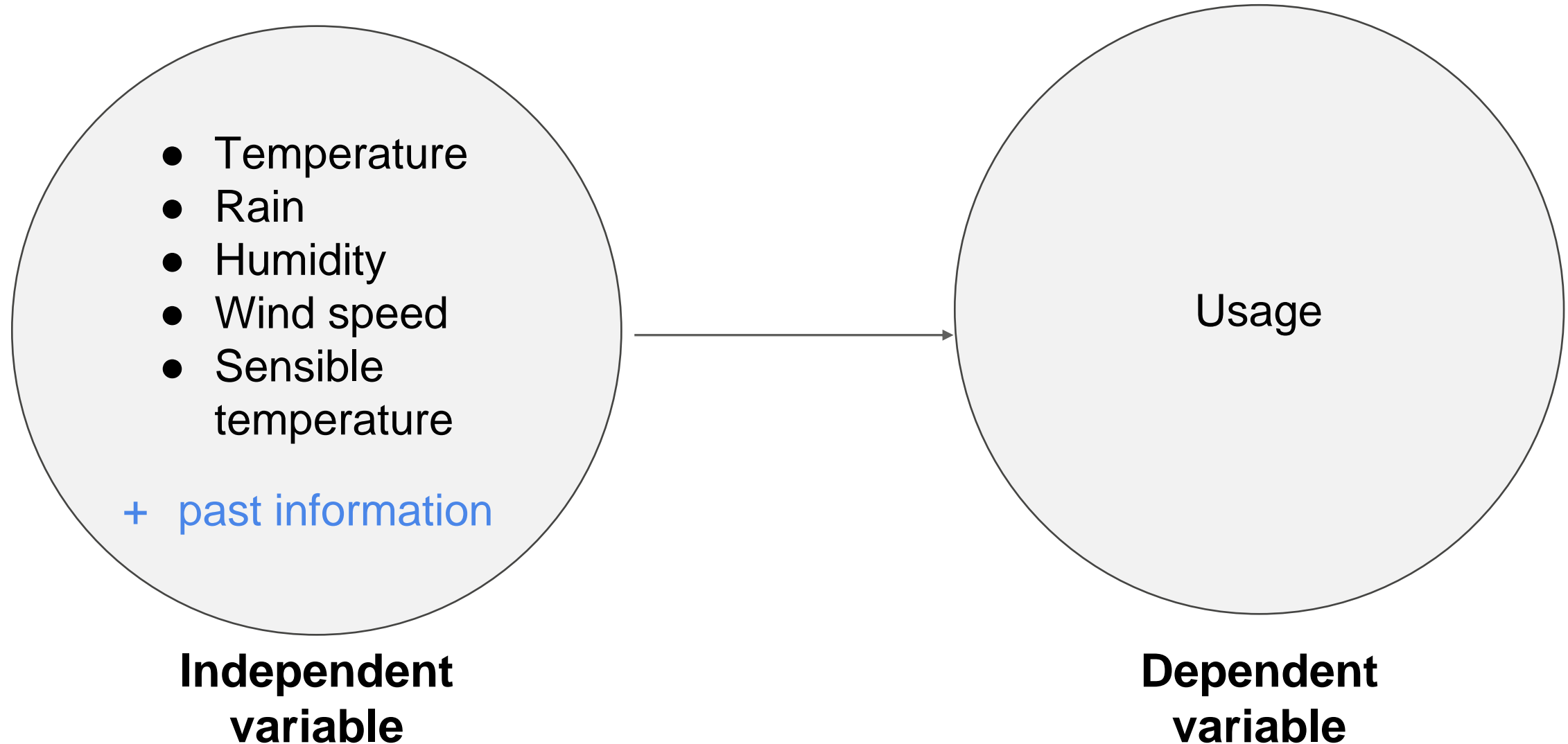
Variable	Coeff.	Std. Err.	t-value	P> t
precipitation	-2.5592	115.1345	-0.0222	0.9823
sensible_temp	27.1674	6.202	4.3804	1.22E-5
time	336.333	7.0898	47.4387	0.0
Intercept	-1,594.8166	111.0058	-14.367	0.0

R-Squared: 0.4097

Adjusted R-Squared: 0.4092

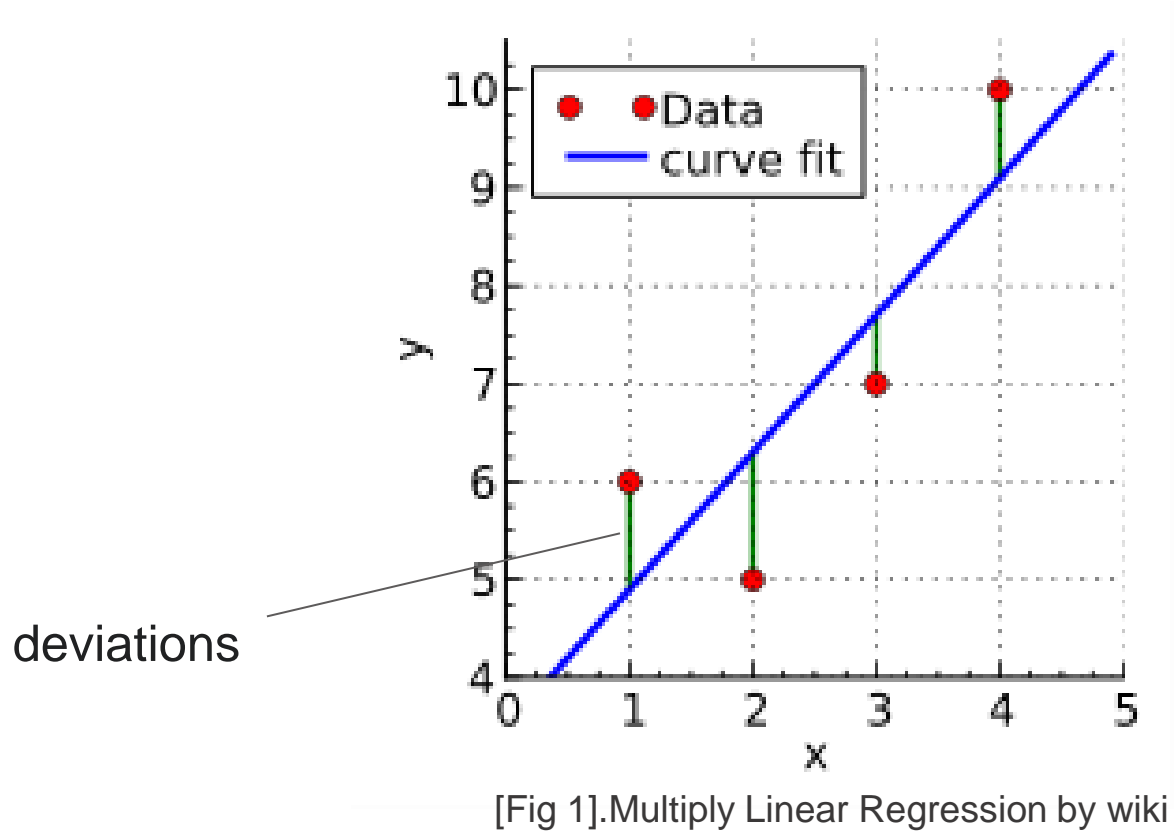
Part 1

Multiple Linear Regression



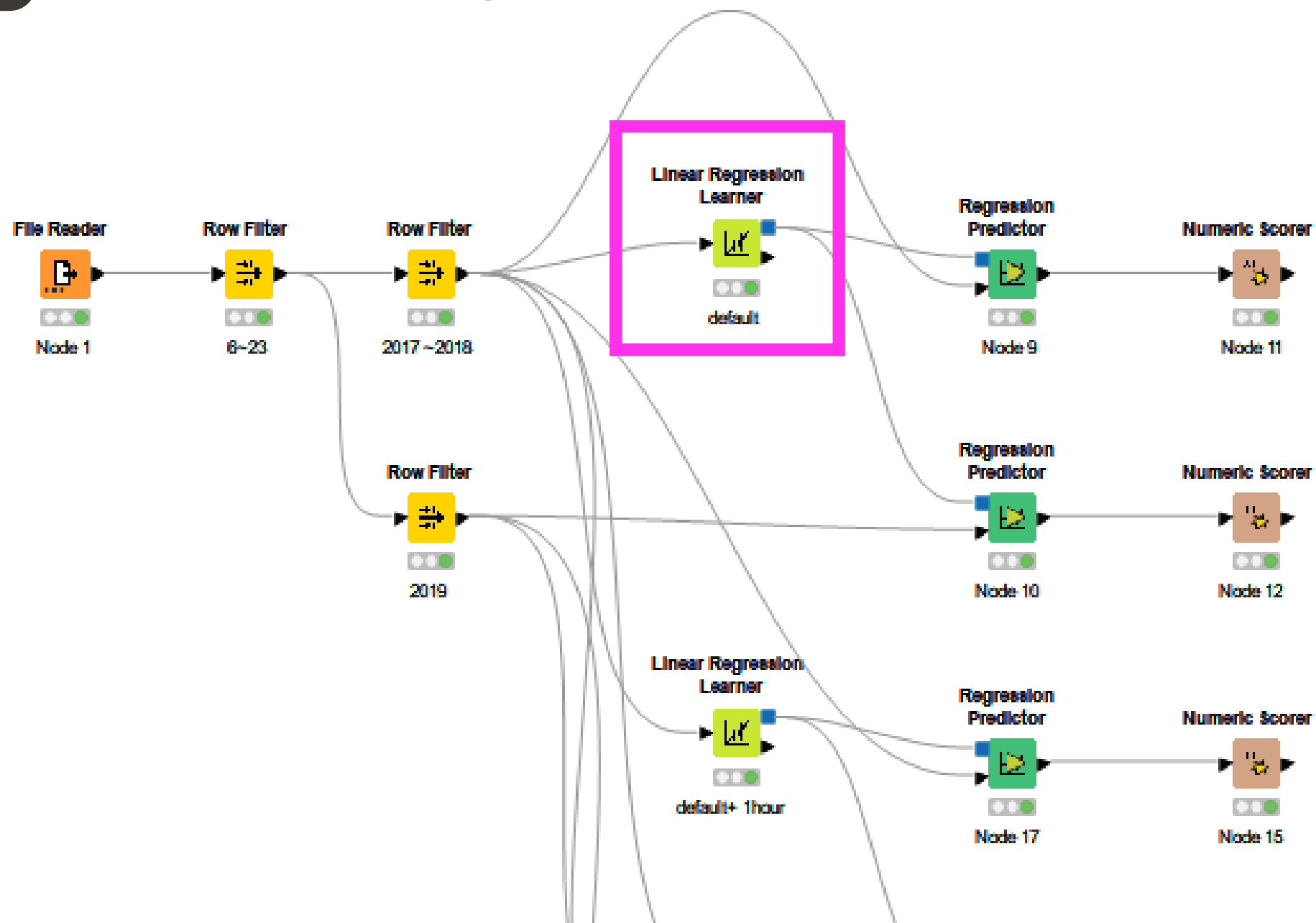
Part 1

Multiple Linear Regression



Part 1

Multiple Linear Regression



Part 1

Multiple Linear Regression

1	Date	Temp	Rain	Wind	Humidity	Sensible_tem	Hour	Usage
2	2017.6.1	20.6	0	0.9	82	22.615474	5	244
3	2017.6.1	20.9	0	2.1	80	22.7867256	6	1409
4	2017.6.1	20.9	0	1.3	80	22.7867256	7	6692
5	2017.6.1	21.9	0	2	75	23.4591084	8	15653
6	2017.6.1	23.2	0	1.8	66	24.1249323	9	6927
7	2017.6.1	23.7	0	3.3	65	24.5507164	10	3994
8	2017.6.1	24.5	0	3.9	59	24.8882158	11	4247
9	2017.6.1	25.7	0	3.9	52	25.5007043	12	2685
10	2017.6.1	26.6						
11	2017.6.1	26.2						
12	2017.6.1	26.9						
13	2017.6.1	26.5						

Add Column

1	Temp(1hour)	Rain(1hour)	Wind(1hour)	Humidity(1h)	Sensible_tem	Temp(1week)	Rain(1week)	Wind(1week)	Humidity(1w)	Sensible_tem
2										
3	20.6	0	0.9	82	22.615474	16.1	0	2.1	73	17.492534
4	20.9	0	2.1	80	22.7867256	15.8	0	2	73	17.192338
5	20.9	0	1.3	80	22.7867256	16.6	0	0.9	63	17.365293
6	21.9	0	2	75	23.4591084	18.5	0	3.8	58	18.905353
7	23.2	0	1.8	66	24.1249323	20.7	0	4.1	43	19.938244
8	23.7	0	3.3	65	24.5507164	22.7	0	4.7	41	21.671757
9	24.5	0	3.9	59	24.8882158	21.8	0	3.6	47	21.306649
10	25.7	0	3.9	52	25.5007043	22.3	0	3.7	49	21.947226
11	26.6	0	4.4	50	26.207865	23.2	0	6	42	22.232861
12	26.2	0	5.7	50	25.8176731	24.2	0	5	30	22.035366
13	26.9	0	5.1	46	26.1401343	23.9	0	6	22	20.871338
14	26.5	0	5.3	50	26.1102942	23.5	0	4.9	24	20.749833
15	26.6	0	5	42	25.4776088	22.9	0	5.5	29	20.744312

Originally Data

Part 2

Data Interpolation

Time	disappear
5	0
6	1

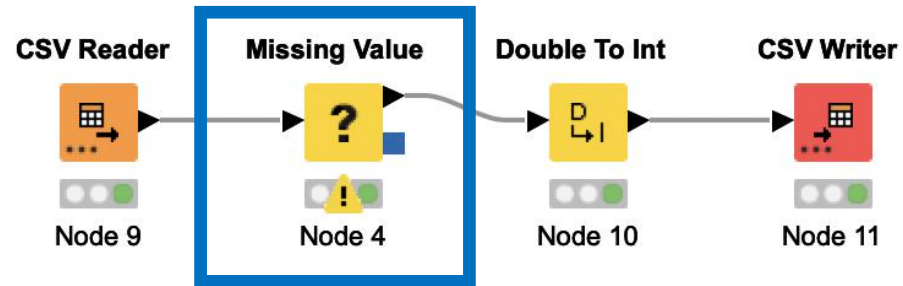
→ =IF(Current time +1=Next time,1,IF(Current time =23,1,0))

336	2017.9.26	30.4	0	1.2	39	28.80514185	16	6935	0
337	2017.9.26	27.9	0	0	45	27.0135227	18	17448	1

447	2017.10.12	13.7	0	3.3	61	12.81173512	14	4965	0
448	2017.10.12	12.8	0	2.3	66	12.24710923	18	16351	1

Part 2

Data Interpolation



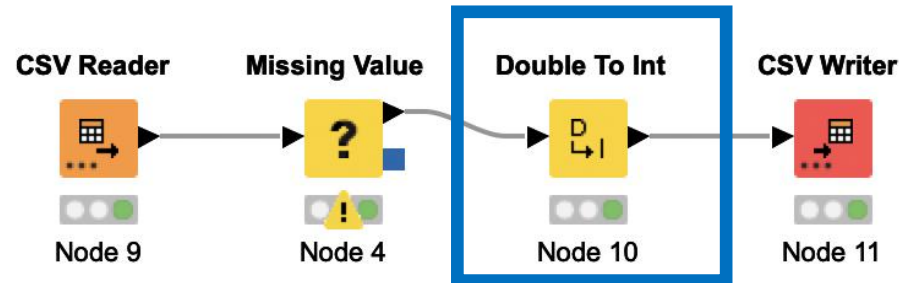
Temperature, wind speed, humidity and sensory temperature were interpolated.

Missing Value -> Moving Average -> Column Settings

D Temperature	<div>Remove</div> <div>Moving Average*</div> <div>Lookbehind 4</div> <div>Lookahead 1</div>
D Wind speed	<div>Remove</div> <div>Moving Average*</div> <div>Lookbehind 4</div> <div>Lookahead 1</div>
I Humadity	<div>Remove</div> <div>Moving Average*</div> <div>Lookbehind 4</div> <div>Lookahead 1</div>
D Sensible_temp	<div>Remove</div> <div>Moving Average*</div> <div>Lookbehind 4</div> <div>Lookahead 1</div>

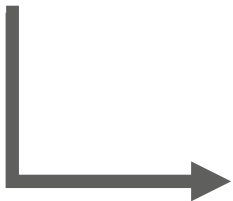
Part 2

Data Interpolation



D Temp...	D Rainfall	D Wind ...	D Huma...	D Sensib...	D Time	D Usage
17.9	0	1	70	19.103	5	211
17.3	0	1.4	78	19.01	6	1,280
17.4	0	2.3	75	18.922	7	6,409
19.8	0	1.6	61	20.389	8	17,072
22	0	3	52	21.89	9	7,755
24	0	2.9	46	23.341	10	4,703
25.7	0	2.6	40	24.429	11	4,698

Using the Missing Value Node ->
Integer type column changed
to Double type



D Temp...	I Rainfall	D Wind ...	I Huma...	D Sensib...	I Time	I Usage
17.9	0	1	70	19.103	5	211
17.3	0	1.4	78	19.01	6	1280
17.4	0	2.3	75	18.922	7	6409
19.8	0	1.6	61	20.389	8	17072
22	0	3	52	21.89	9	7755
24	0	2.9	46	23.341	10	4703
25.7	0	2.6	40	24.429	11	4698

Part 2

Data Interpolation result

S	Date	D	Temp...	D	Rainfall	D	Wind ...	D	Huma...	D	Sensib...	D	Time	D	Usage
	2017.9.26		29.88		0		1.26		40.4		28.446		17		9,906

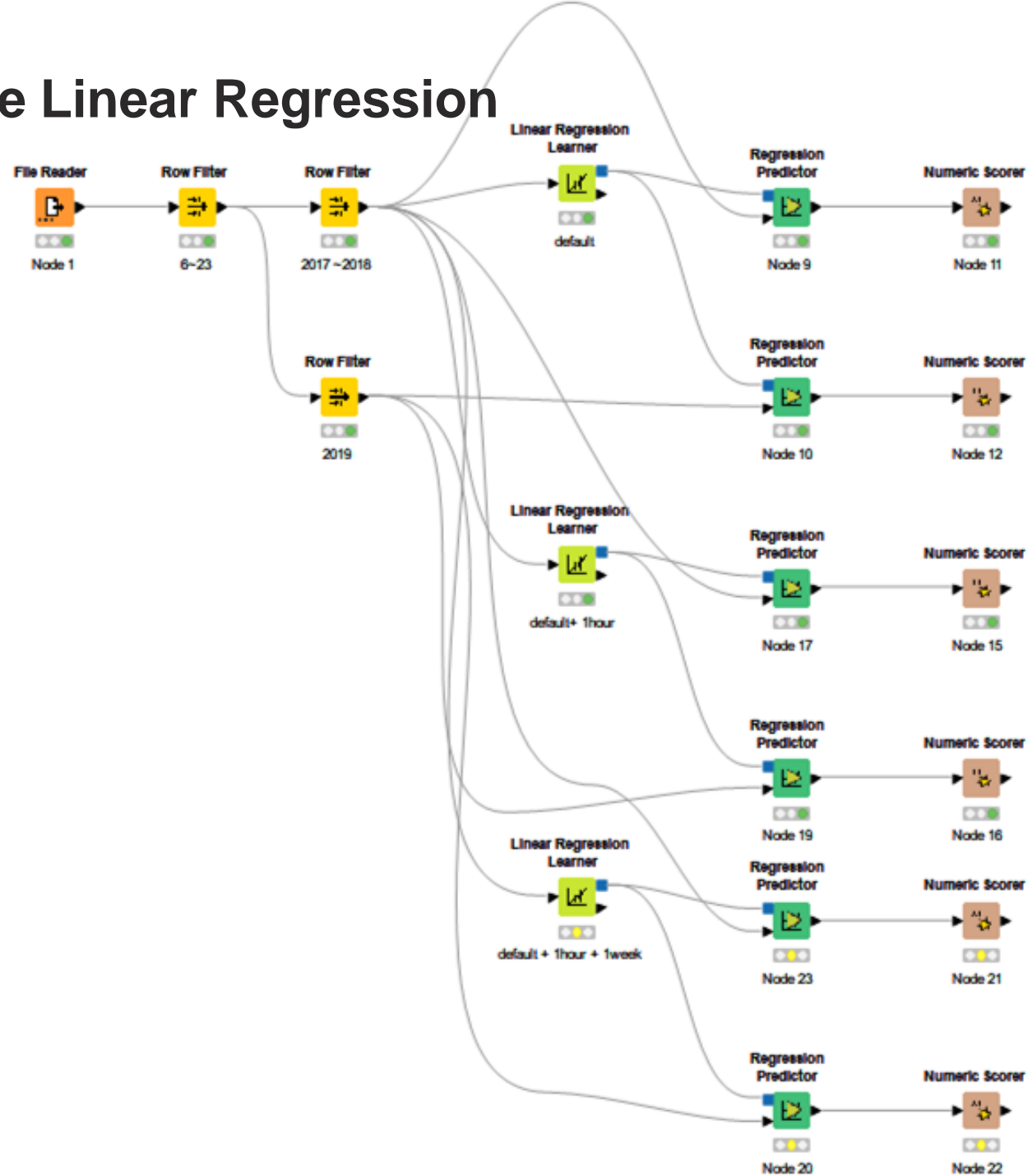
S	Date	D	Temp...	D	Rainfall	D	Wind ...	D	Huma...	D	Sensib...	D	Time	D	Usage
	2017.10.12		13.225		0		2.125		64.25		12.912		15		5,515
	2017.10.12		13.433		0		2.233		63.333		13.097		16		5,666
	2017.10.12		13.333		0		2.433		63.667		12.843		17		8,238

The data were supplemented on September 26th and October 12th, 2017

Rain data were found and entered on the weather data opening portal, and subway usage data was previously collected and entered.

Part 3

Multiple Linear Regression



Part 3

Data Analysis

Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P> t
Rain=1	-247.5901	388.8612	-0.6367	0.5244
Wind	-227.7038	86.9459	-2.6189	0.0089
Humidity	-68.2503	14.4063	-4.7375	2.25E-6
Sensible_temp	-656.9851	70.7605	-9.2846	0.0
Hour	-126.1501	15.4809	-8.1488	4.44E-16
Rain(hour)	-270.2489	383.3842	-0.7049	0.4809
Wind(hour)	417.1898	84.6912	4.926	8.80E-7
Humidity(hour)	66.575	14.0628	4.7341	2.29E-6
Sensible_temp(hour)	666.6604	70.5728	9.4464	0.0
Rain(1week)	-476.9798	332.6422	-1.4339	0.1517
Wind(1week)	193.4808	59.8137	3.2347	0.0012
Humidity(1week)	5.0098	3.5996	1.3918	0.1641
Sensible_temp(1week)	-8.2684	12.4892	-0.662	0.508
Intercept	7,113.3589	457.2509	15.5568	0.0

R-Squared: 0.0434

Adjusted R-Squared: 0.0397



**Spring R-Squared
< 0.1 : poor**

Part 3

Data Analysis

Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P> t
Temp	-1,346.1504	386.9357	-3.479	0.0005
Wind	202.029	92.7225	2.1789	0.0294
Humidity	-109.6467	34.8625	-3.1451	0.0017
Sensible_temp	888.808	400.5928	2.2187	0.0266
Hour	-138.7703	16.5879	-8.3658	0.0
Rain(1hour)	612.3988	260.756	2.3486	0.0189
Wind(1hour)	119.154	93.6621	1.2722	0.2034
Humidity(1hour)	-9.9918	14.9507	-0.6683	0.504
Sensible_temp(1hour)	474.9483	121.2446	3.9173	9.14E-5
Rain(1week)	227.3643	254.6447	0.8929	0.372
Wind(1week)	333.2989	68.8355	4.842	1.34E-6
Humidity(1week)	-3.6926	4.4426	-0.8312	0.4059
Sensible_temp(1week)	37.8404	23.8656	1.5856	0.1129
Intercept	12,285.7378	1,943.1321	6.3226	2.91E-10

R-Squared: 0.0411

Adjusted R-Squared: 0.0374



**Summer R-Squared < 0.1 :
poor**

Part 3

Data Analysis

Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P> t
Rain	104.1711	416.1141	0.2503	0.8023
Wind	-453.3336	97.2312	-4.6624	3.25E-6
Humadity	-28.2209	15.3891	-1.8338	0.0668
Sensible_temp	-856.5912	75.0828	-11.4086	0.0
Hour	-137.1297	15.607	-8.7864	0.0
Rain(1hour)	-136.6253	415.6677	-0.3287	0.7424
Wind(1hour)	452.6215	95.6562	4.7318	2.32E-6
Humadity(1hour)	24.6161	15.0804	1.6323	0.1027
Sensible_temp(1hour)	873.8102	74.911	11.6646	0.0
Rain(1week)	467.4926	318.5681	1.4675	0.1423
Wind(1week)	57.9501	66.6229	0.8698	0.3845
Humadity(1week)	-8.5807	4.2996	-1.9957	0.0461
Sensible_temp(1week)	-19.0016	16.8197	-1.1297	0.2587
Intercept	9,009.108	490.4832	18.3678	0.0

R-Squared: 0.0516

Adjusted R-Squared: 0.0478



**Fall R-Squared < 0.1
: poor**

Part 3

Data Analysis

Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P> t
Rain=1	-362.4117	2,730.3914	-0.1327	0.8946
Wind	-264.4804	454.3076	-0.5822	0.5613
Humidity	256.0464	72.1684	3.5479	0.0005
Sensible_temp	-513.7875	304.8263	-1.6855	0.094
Hour	-172.4925	65.6073	-2.6292	0.0095
Rain(hour)	1,491.1447	2,724.9758	0.5472	0.5851
Wind(hour)	796.4903	431.494	1.8459	0.0669
Humidity(hour)	-263.9472	75.2256	-3.5087	0.0006
Sensible_temp(hour)	610.1802	312.2821	1.9539	0.0526
Rain(1week)	-2,221.7975	1,149.7075	-1.9325	0.0552
Wind(1week)	331.4943	255.7451	1.2962	0.1969
Humidity(1week)	22.7984	22.4392	1.016	0.3113
Sensible_temp(1week)	-96.7631	103.8978	-0.9313	0.3532
Intercept	6,857.0499	1,804.788	3.7994	0.0002
R-Squared: 0.2116				
Adjusted R-Squared: 0.1424				

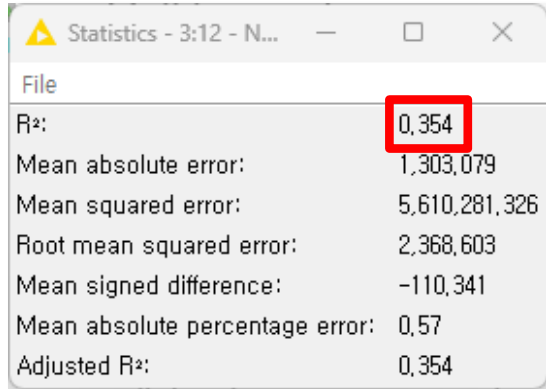


**Winter R-Squared
< 0.5 : poor**

Part 3

Data Analysis - Original Data

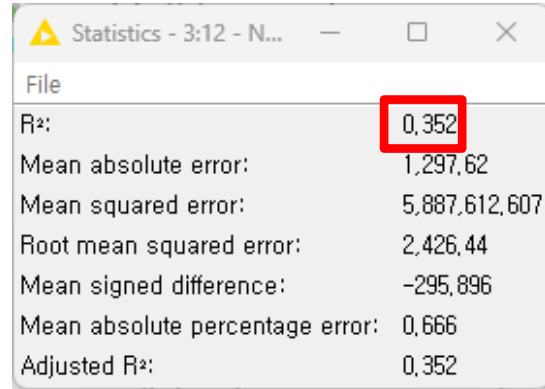
spring



Statistics - 3:12 - N...

File	
R ² :	0,354
Mean absolute error:	1,303,079
Mean squared error:	5,610,281,326
Root mean squared error:	2,368,603
Mean signed difference:	-110,341
Mean absolute percentage error:	0,57
Adjusted R ² :	0,354

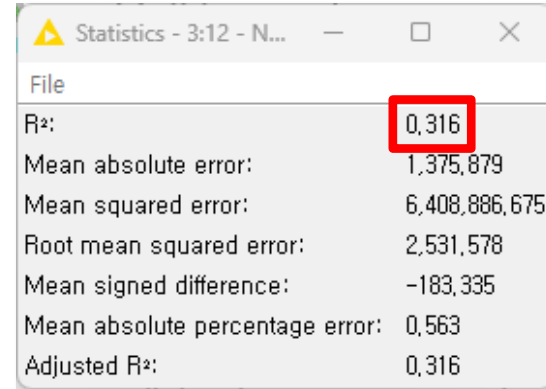
summer



Statistics - 3:12 - N...

File	
R ² :	0,352
Mean absolute error:	1,297,62
Mean squared error:	5,887,612,607
Root mean squared error:	2,426,44
Mean signed difference:	-295,896
Mean absolute percentage error:	0,666
Adjusted R ² :	0,352

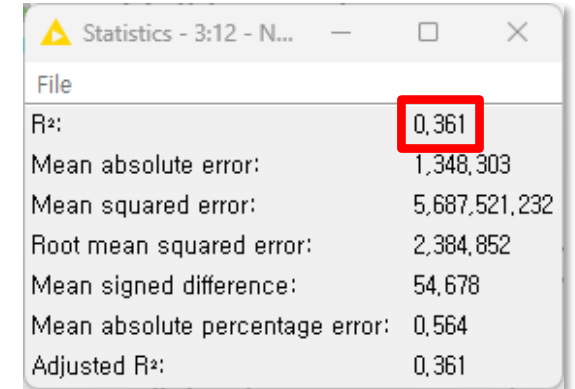
fall



Statistics - 3:12 - N...

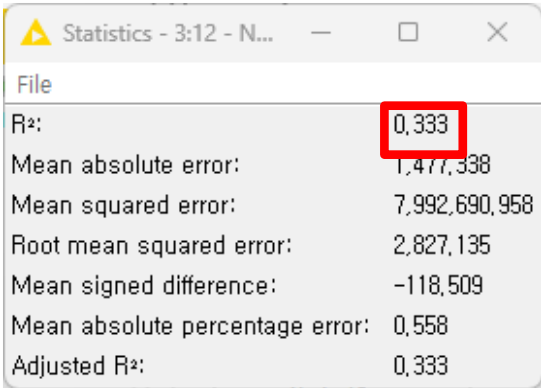
File	
R ² :	0,316
Mean absolute error:	1,375,879
Mean squared error:	6,408,886,675
Root mean squared error:	2,531,578
Mean signed difference:	-183,335
Mean absolute percentage error:	0,563
Adjusted R ² :	0,316

winter



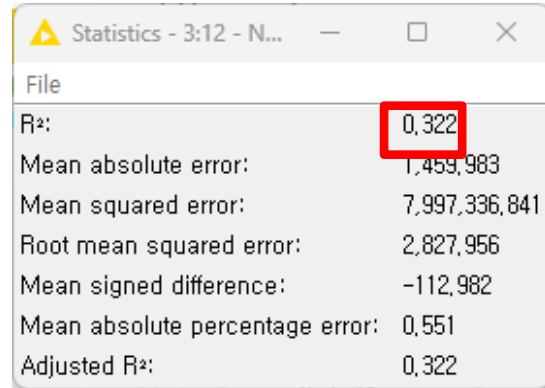
Statistics - 3:12 - N...

File	
R ² :	0,361
Mean absolute error:	1,348,303
Mean squared error:	5,687,521,232
Root mean squared error:	2,384,852
Mean signed difference:	54,678
Mean absolute percentage error:	0,564
Adjusted R ² :	0,361



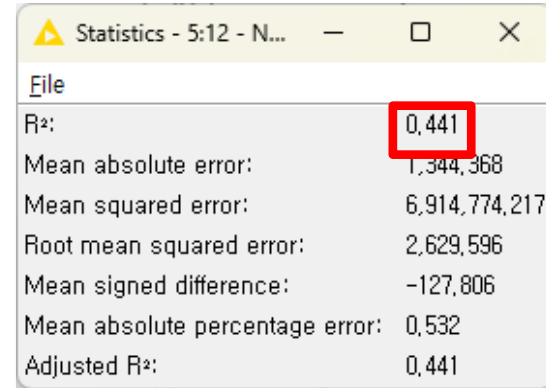
Statistics - 3:12 - N...

File	
R ² :	0,333
Mean absolute error:	1,477,338
Mean squared error:	7,992,690,958
Root mean squared error:	2,827,135
Mean signed difference:	-118,509
Mean absolute percentage error:	0,558
Adjusted R ² :	0,333



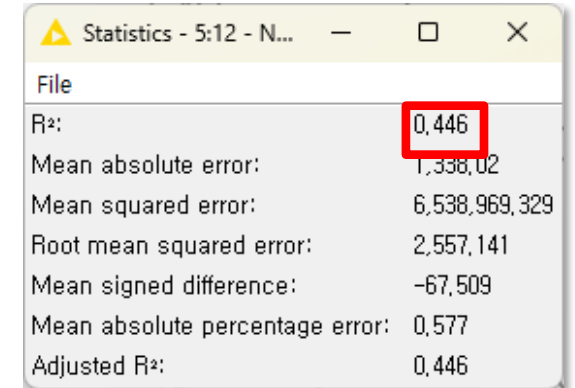
Statistics - 3:12 - N...

File	
R ² :	0,322
Mean absolute error:	1,459,983
Mean squared error:	7,997,336,841
Root mean squared error:	2,827,956
Mean signed difference:	-112,982
Mean absolute percentage error:	0,551
Adjusted R ² :	0,322



Statistics - 5:12 - N...

File	
R ² :	0,441
Mean absolute error:	1,344,368
Mean squared error:	6,914,774,217
Root mean squared error:	2,629,596
Mean signed difference:	-127,806
Mean absolute percentage error:	0,532
Adjusted R ² :	0,441



Statistics - 5:12 - N...

File	
R ² :	0,446
Mean absolute error:	1,338,02
Mean squared error:	6,538,969,329
Root mean squared error:	2,557,141
Mean signed difference:	-67,509
Mean absolute percentage error:	0,577
Adjusted R ² :	0,446

R-Squared < 0.5 : poor

Part 3

Data Analysis - Original Data+1hour

spring

File	
R²:	0,497
Mean absolute error:	1,164,179
Mean squared error:	4,364,465,006
Root mean squared error:	2,089,13
Mean signed difference:	-31,215
Mean absolute percentage error:	0,633
Adjusted R²:	0,497

summer

File	
R²:	0,488
Mean absolute error:	1,133,611
Mean squared error:	4,651,310,12
Root mean squared error:	2,156,69
Mean signed difference:	-234,72
Mean absolute percentage error:	0,526
Adjusted R²:	0,488

fall

File	
R²:	0,5
Mean absolute error:	1,216,87
Mean squared error:	4,682,901,728
Root mean squared error:	2,164,001
Mean signed difference:	-151,082
Mean absolute percentage error:	0,752
Adjusted R²:	0,5

winter

File	
R²:	0,534
Mean absolute error:	1,241,055
Mean squared error:	4,142,346,942
Root mean squared error:	2,035,276
Mean signed difference:	113,992
Mean absolute percentage error:	0,876
Adjusted R²:	0,534

File	
R²:	0,412
Mean absolute error:	1,404,997
Mean squared error:	7,044,646,591
Root mean squared error:	2,654,175
Mean signed difference:	-86,799
Mean absolute percentage error:	0,489
Adjusted R²:	0,412

File	
R²:	0,441
Mean absolute error:	1,344,368
Mean squared error:	6,914,774,217
Root mean squared error:	2,629,596
Mean signed difference:	-127,806
Mean absolute percentage error:	0,532
Adjusted R²:	0,441

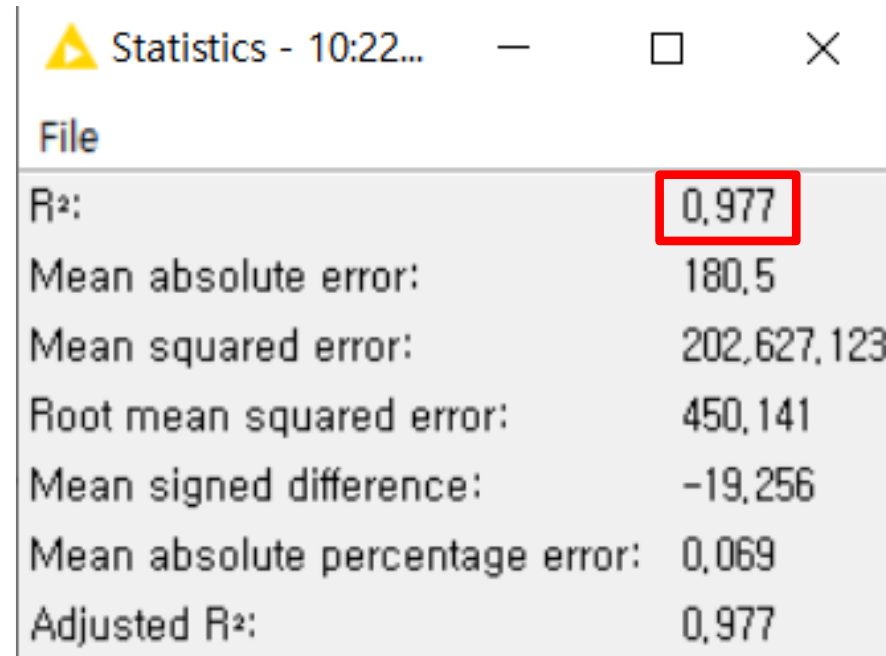
File	
R²:	0,441
Mean absolute error:	1,344,368
Mean squared error:	6,914,774,217
Root mean squared error:	2,629,596
Mean signed difference:	-127,806
Mean absolute percentage error:	0,532
Adjusted R²:	0,441

File	
R²:	0,446
Mean absolute error:	1,338,02
Mean squared error:	6,538,969,329
Root mean squared error:	2,557,141
Mean signed difference:	-67,509
Mean absolute percentage error:	0,577
Adjusted R²:	0,446



R-Squared < 0.6 : poor

board



File	
R ² :	0,977
Mean absolute error:	180,5
Mean squared error:	202,627,123
Root mean squared error:	450,141
Mean signed difference:	-19,256
Mean absolute percentage error:	0,069
Adjusted R ² :	0,977

**Spring R-Squared < 1 : poor**

Part 3

Data Analysis

A	Temp	Rain	Wind	Humidity	Hour	Usage	Temp(1hour)	Rain(1hour)	Wind(1hour)	Humidity(1ho..	Usage(1hour)	Temp(week)	Rain(week)	Wind(week)	Humidity(we..	Usage(week)
Temp	1.000	-0.053	0.140	-0.260	0.115	0.152	0.985	-0.058	0.112	-0.268	0.069	0.734	0.068	0.173	-0.143	0.153
Rain	-0.053	1.000	-0.027	0.410	-0.011	-0.005	-0.038	0.604	-0.009	0.368	-0.013	0.028	-0.034	-0.015	-0.005	-0.003
Wind	0.140	-0.027	1.000	-0.341	0.202	0.295	0.172	-0.045	0.756	-0.372	0.190	0.180	0.032	0.331	-0.200	0.296
Humidity	-0.260	0.410	-0.341	1.000	-0.108	-0.165	-0.237	0.422	-0.310	0.963	-0.082	-0.132	-0.084	-0.278	0.211	-0.158
Hour	0.115	-0.011	0.202	-0.108	1.000	0.586	0.218	-0.011	0.334	-0.224	0.719	0.120	0.003	0.195	-0.109	0.589
Usage	0.152	-0.005	0.295	-0.165	0.586	1.000	0.246	-0.017	0.397	-0.260	0.669	0.152	0.017	0.287	-0.143	0.988
Temp(1hour)	0.985	-0.038	0.172	-0.237	0.218	0.246	1.000	-0.049	0.165	-0.277	0.177	0.726	0.071	0.185	-0.137	0.247
Rain(1hour)	-0.058	0.604	-0.045	0.422	-0.011	-0.017	-0.049	1.000	-0.031	0.405	-0.009	0.025	-0.048	-0.027	-0.009	-0.014
Wind(1hour)	0.112	-0.009	0.756	-0.310	0.334	0.397	0.165	-0.031	1.000	-0.364	0.321	0.162	0.031	0.318	-0.177	0.398
Humidity(1hour)	-0.268	0.368	-0.372	0.963	-0.224	-0.260	-0.277	0.405	-0.364	1.000	-0.195	-0.139	-0.089	-0.296	0.207	-0.253
Usage(1hour)	0.069	-0.013	0.190	-0.082	0.719	0.669	0.177	-0.009	0.321	-0.195	1.000	0.074	0.002	0.185	-0.065	0.666
Temp(week)	0.734	0.028	0.180	-0.132	0.120	0.152	0.726	0.025	0.162	-0.139	0.074	1.000	-0.037	0.126	-0.232	0.151
Rain(week)	0.068	-0.034	0.032	-0.084	0.003	0.017	0.071	-0.048	0.031	-0.089	0.002	-0.037	1.000	-0.015	0.401	0.008
Wind(week)	0.173	-0.015	0.331	-0.278	0.195	0.287	0.185	-0.027	0.318	-0.296	0.185	0.126	-0.015	1.000	-0.319	0.290
Humidity(week)	-0.143	-0.005	-0.200	0.211	-0.109	-0.143	-0.137	-0.009	-0.177	0.207	-0.065	-0.232	0.401	-0.319	1.000	-0.155
Usage(week)	0.153	-0.003	0.296	-0.158	0.589	0.988	0.247	-0.014	0.398	-0.253	0.666	0.151	0.008	0.290	-0.155	1.000

Thank you