◆ 2021 - 2023

with

genie music

Meteorological Admin.

빅데이터 분석

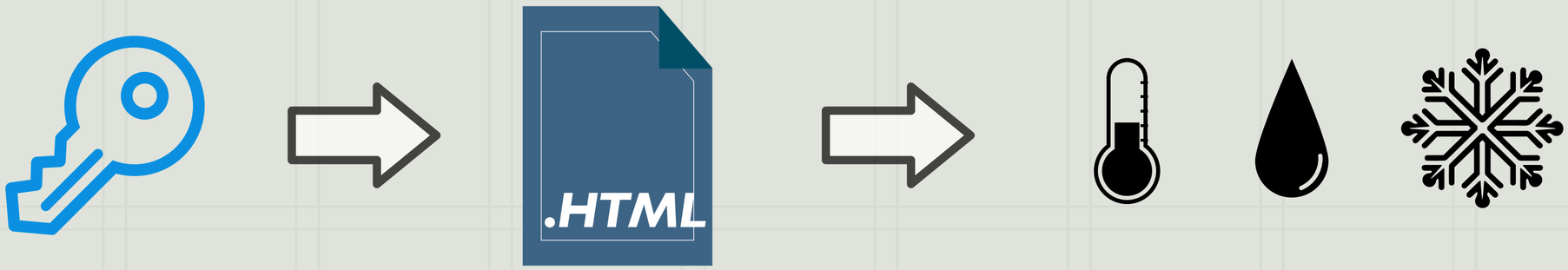
날씨와 음악 스트리밍 장르 간 연관성 분석

20192940 강준규

숭실대학교 IT대학 소프트웨어학부

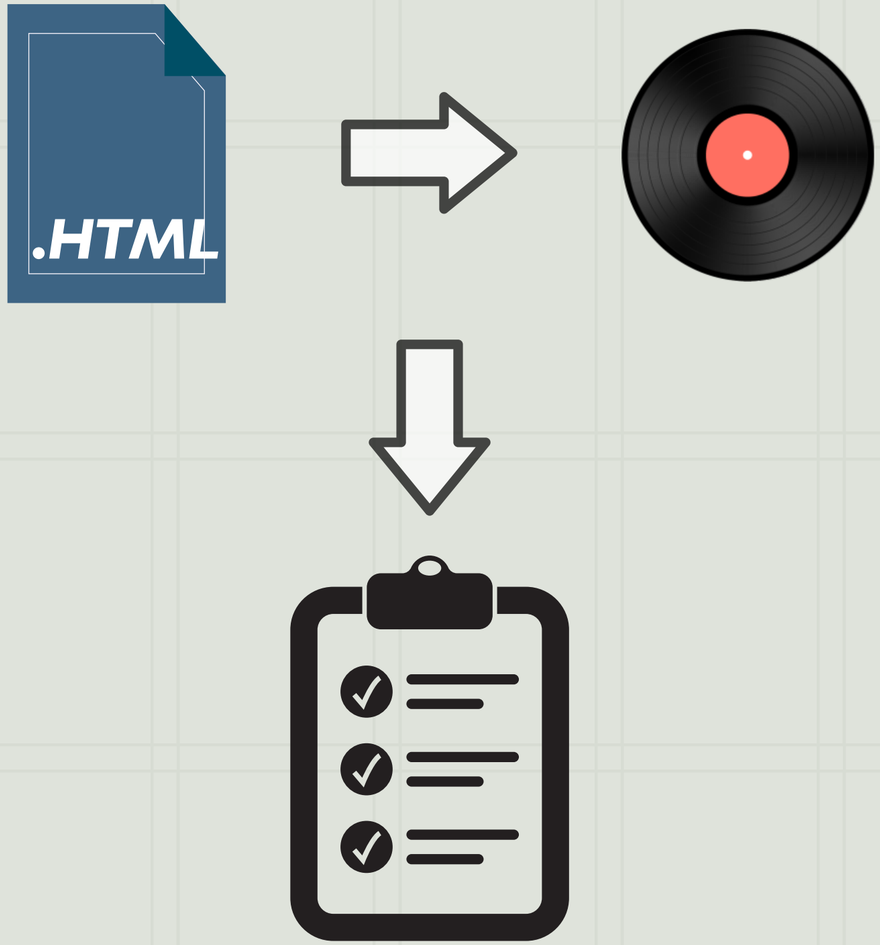
1. Data


기상청 API허브
국민을 위한 기상기후데이터를 직접 만나보세요
일상 생활에 꼭 필요한 날씨, 누구나 원하는 날씨데이터를 활용하여
새로운 가치를 창출할 수 있도록 기상청이 API허브를 통해
방대한 기상기후데이터를 손쉽게 제공합니다
+20°C
11:37
API


.HTML

1. Data

genie



1. Data

IgnoreList = [전체, 가요, POP(중복), OST(중복)]

WetList = [드라마, R&B/소울, 발라드, 인디]

WarmList = [드라마, R&B/소울, 발라드, 팝, 캐롤]

SolitaryList = [발라드, 인디, 블루스/포크]

BounceList = [댄스, 일렉트로니카, 랩/힙합, 트로트]

ExcitingList = [댄스, 락, 일렉트로니카, 트로트, 팝]

ETCList = [그외장르, 한국영화, 월드뮤직, ...]

1. Data

```
weather_df.head()
```

Out [45] :

	day	Temperature	Rainfall	Snowfall	warm	exciting	wet	solitary	bounce	etc
0	20210101	2.108421	0.047368	71.642105	91.0	71.0	90.0	63.0	75.0	1.0
1	20210102	1.929474	0.098947	33.957895	94.0	70.0	91.0	63.0	75.0	1.0
2	20210103	1.554737	0.000000	5.578947	92.0	68.0	93.0	64.0	73.0	1.0
3	20210104	4.453684	0.041053	99.000000	93.0	66.0	94.0	65.0	73.0	1.0
4	20210105	2.353684	0.054737	13.736842	94.0	66.0	95.0	65.0	72.0	2.0

2. Experiment Multiple Regression

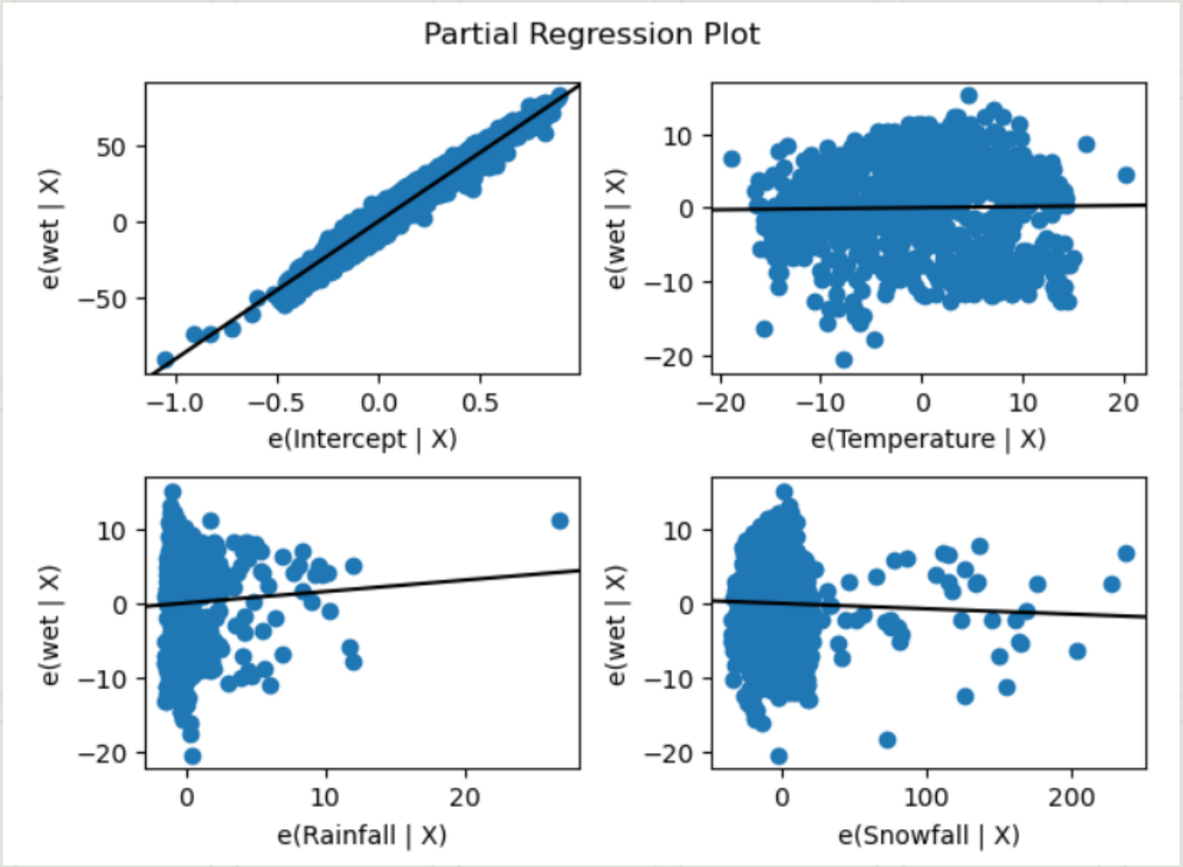
"다중 선형회귀 분석 모델"
(numeric to numeric)

2. Experiment "Wet"

Dep. Variable:	wet	**R-squared:**	0.006
Model:	OLS	**Adj. R-squared:**	0.002
Method:	Least Squares	**F-statistic:**	1.706
Date:	Tue, 13 Jun 2023	**Prob (F-statistic):**	0.164
Time:	23:39:47	**Log-Likelihood:**	-2889.8
No. Observations:	894	**AIC:**	5788.
Df Residuals:	890	**BIC:**	5807.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	90.4367	0.544	166.264	0.000	89.369	91.504
Temperature	0.0142	0.026	0.539	0.590	-0.037	0.066
Rainfall	0.1531	0.108	1.414	0.158	-0.059	0.366
Snowfall	-0.0072	0.007	-1.029	0.304	-0.021	0.007

Omnibus:	39.078	**Durbin-Watson:**	0.110
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	30.520
Skew:	-0.363	**Prob(JB):**	2.36e-07
Kurtosis:	2.459	**Cond. No.**	91.9

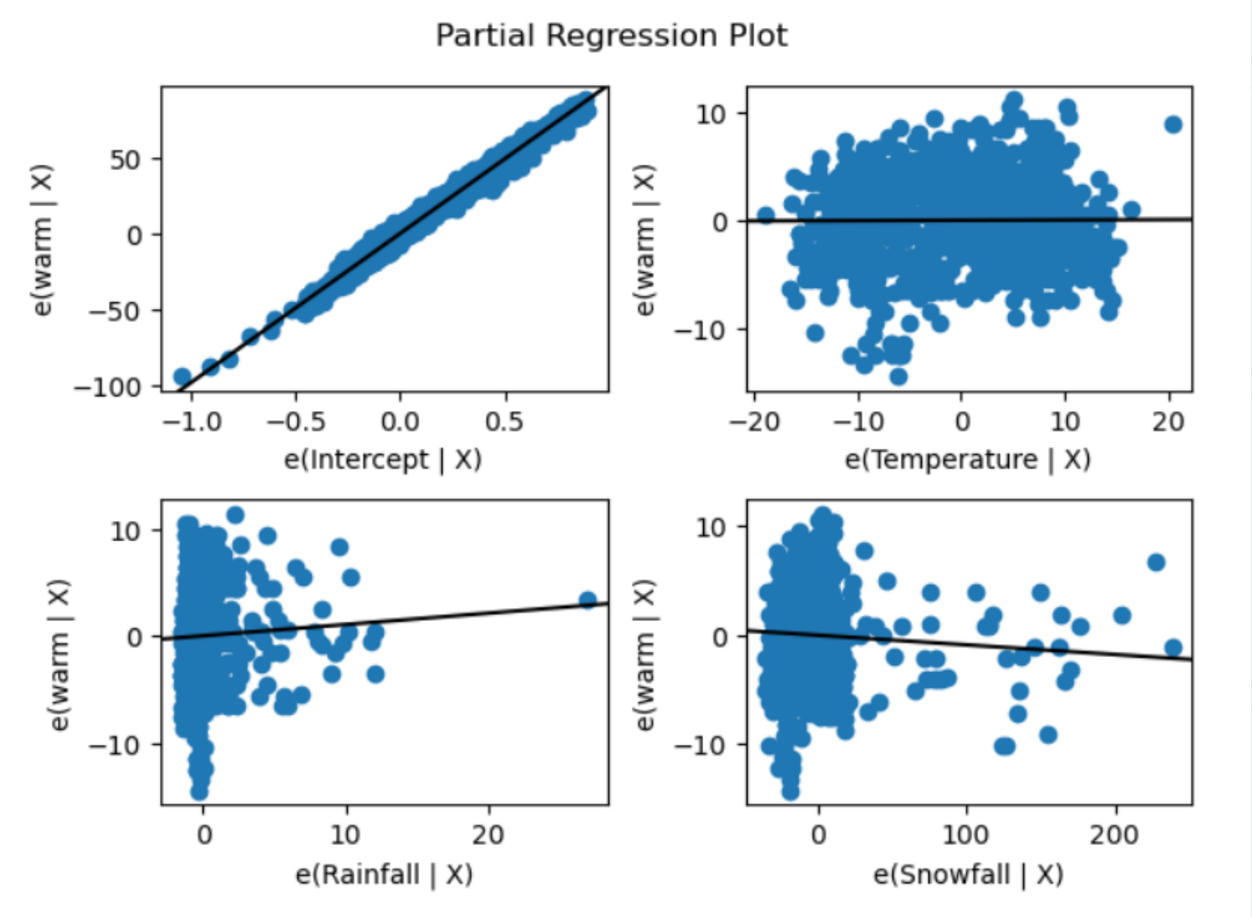


2. Experiment "Warm"

Dep. Variable:	warm	**R-squared:**	0.008
Model:	OLS	**Adj. R-squared:**	0.005
Method:	Least Squares	**F-statistic:**	2.379
Date:	Tue, 13 Jun 2023	**Prob (F-statistic):**	0.0684
Time:	23:41:22	**Log-Likelihood:**	-2568.2
No. Observations:	894	**AIC:**	5144.
Df Residuals:	890	**BIC:**	5164.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	98.3681	0.380	259.153	0.000	97.623	99.113
Temperature	0.0031	0.018	0.170	0.865	-0.033	0.039
Rainfall	0.1058	0.076	1.400	0.162	-0.042	0.254
Snowfall	-0.0089	0.005	-1.812	0.070	-0.018	0.001

Omnibus:	6.346	**Durbin-Watson:**	0.201
Prob(Omnibus):	0.042	**Jarque-Bera (JB):**	6.400
Skew:	-0.191	**Prob(JB):**	0.0408
Kurtosis:	2.840	**Cond. No.**	91.9

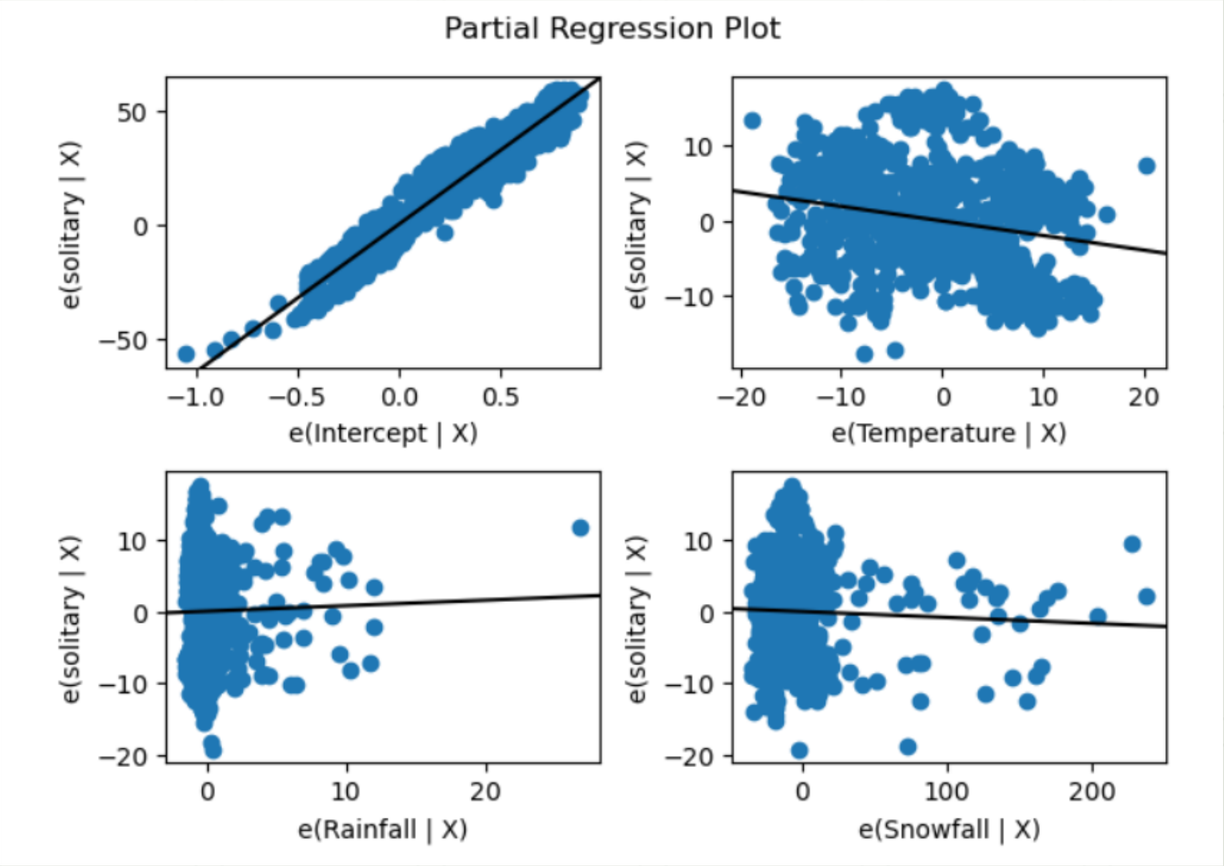


2. Experiment "Solitary"

Dep. Variable:	solitary	**R-squared:**	0.055
Model:	OLS	**Adj. R-squared:**	0.052
Method:	Least Squares	**F-statistic:**	17.37
Date:	Tue, 13 Jun 2023	**Prob (F-statistic):**	5.78e-11
Time:	23:42:14	**Log-Likelihood:**	-2975.9
No. Observations:	894	**AIC:**	5960.
Df Residuals:	890	**BIC:**	5979.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	65.0656	0.599	108.646	0.000	63.890	66.241
Temperature	-0.1946	0.029	-6.722	0.000	-0.251	-0.138
Rainfall	0.0768	0.119	0.644	0.520	-0.157	0.311
Snowfall	-0.0083	0.008	-1.071	0.284	-0.023	0.007

Omnibus:	16.040	**Durbin-Watson:**	0.073
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	12.085
Skew:	0.185	**Prob(JB):**	0.00238
Kurtosis:	2.566	**Cond. No.**	91.9

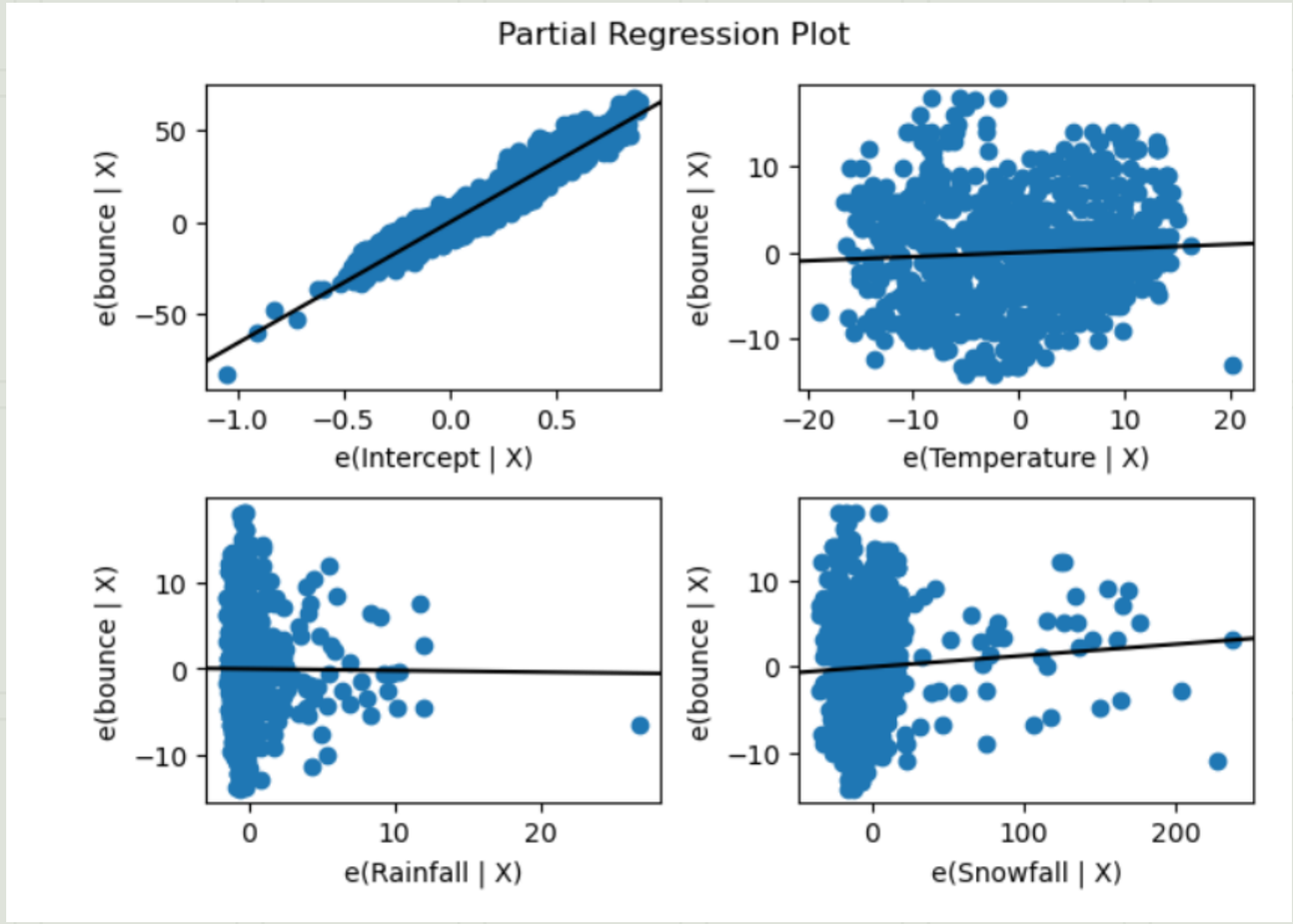


2. Experiment "Bounce"

Dep. Variable:	bounce	**R-squared:**	0.005
Model:	OLS	**Adj. R-squared:**	0.002
Method:	Least Squares	**F-statistic:**	1.487
Date:	Tue, 13 Jun 2023	**Prob (F-statistic):**	0.217
Time:	23:42:21	**Log-Likelihood:**	-2905.1
No. Observations:	894	**AIC:**	5818.
Df Residuals:	890	**BIC:**	5837.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	66.2202	0.553	119.678	0.000	65.134	67.306
Temperature	0.0474	0.027	1.773	0.077	-0.005	0.100
Rainfall	-0.0189	0.110	-0.171	0.864	-0.235	0.197
Snowfall	0.0132	0.007	1.845	0.065	-0.001	0.027

Omnibus:	13.432	**Durbin-Watson:**	0.104
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	13.065
Skew:	0.264	**Prob(JB):**	0.00146
Kurtosis:	2.731	**Cond. No.**	91.9

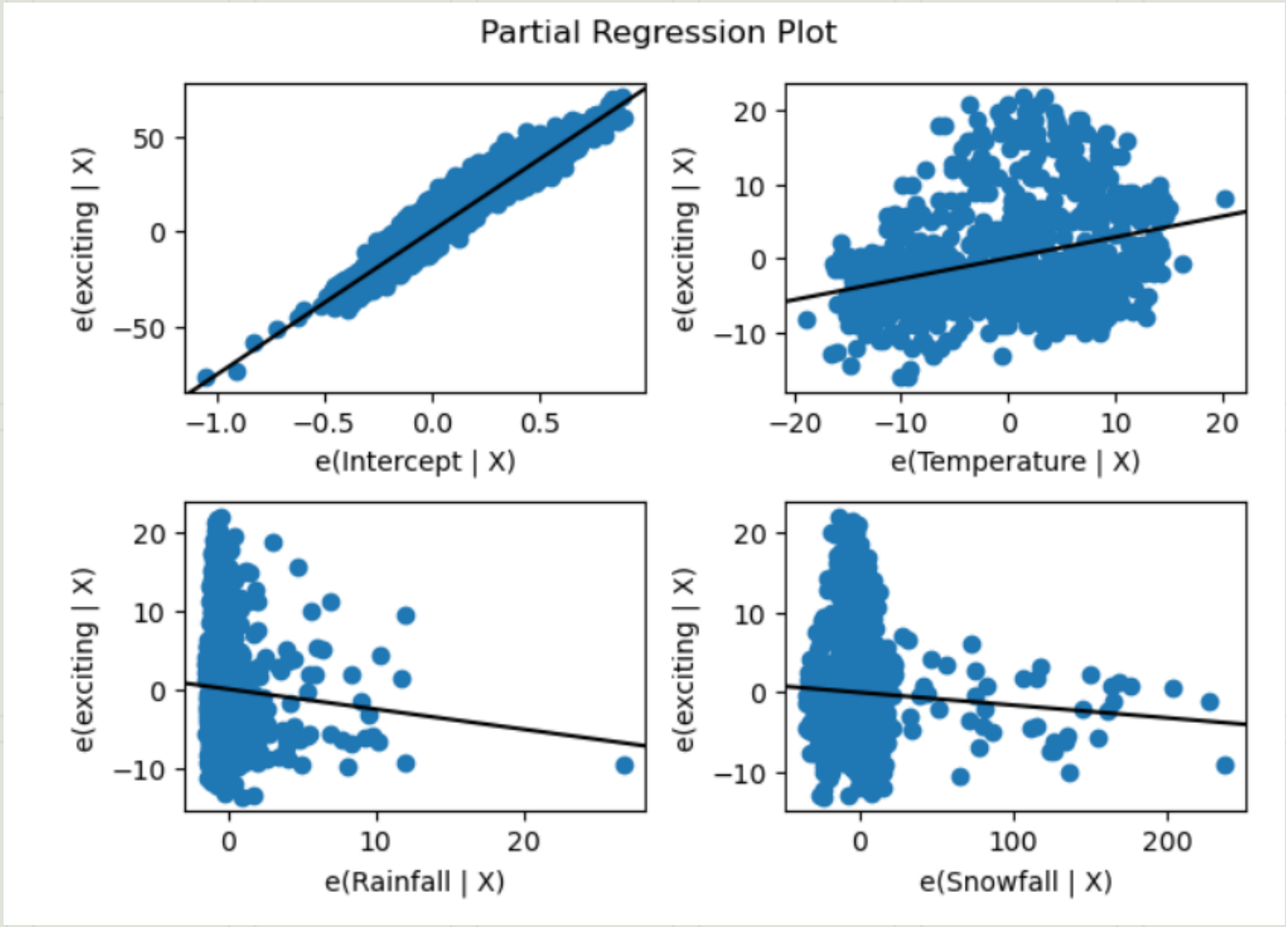


2. Experiment "Exciting"

Dep. Variable:	exciting	**R-squared:**	0.145
Model:	OLS	**Adj. R-squared:**	0.142
Method:	Least Squares	**F-statistic:**	50.29
Date:	Tue, 13 Jun 2023	**Prob (F-statistic):**	4.96e-30
Time:	23:41:33	**Log-Likelihood:**	-2984.2
No. Observations:	894	**AIC:**	5976.
Df Residuals:	890	**BIC:**	5996.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	75.8388	0.604	125.459	0.000	74.652	77.025
Temperature	0.2821	0.029	9.654	0.000	0.225	0.339
Rainfall	-0.2558	0.120	-2.126	0.034	-0.492	-0.020
Snowfall	-0.0159	0.008	-2.043	0.041	-0.031	-0.001

Omnibus:	95.017	**Durbin-Watson:**	0.082
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	123.579
Skew:	0.883	**Prob(JB):**	1.46e-27
Kurtosis:	3.447	**Cond. No.**	91.9



3. Verification

"날씨와 음악 스트리밍 장르 간에는 연관성이 있다."

3. Findings & QnA