BIGDATA TERM PROJECT

잠 좀 자고 싶조

20141845 노우석
20144444 이 철
20142167 최현경

CONTENTS

TEAM MEMBER

20141845 노우석

20144444 이철

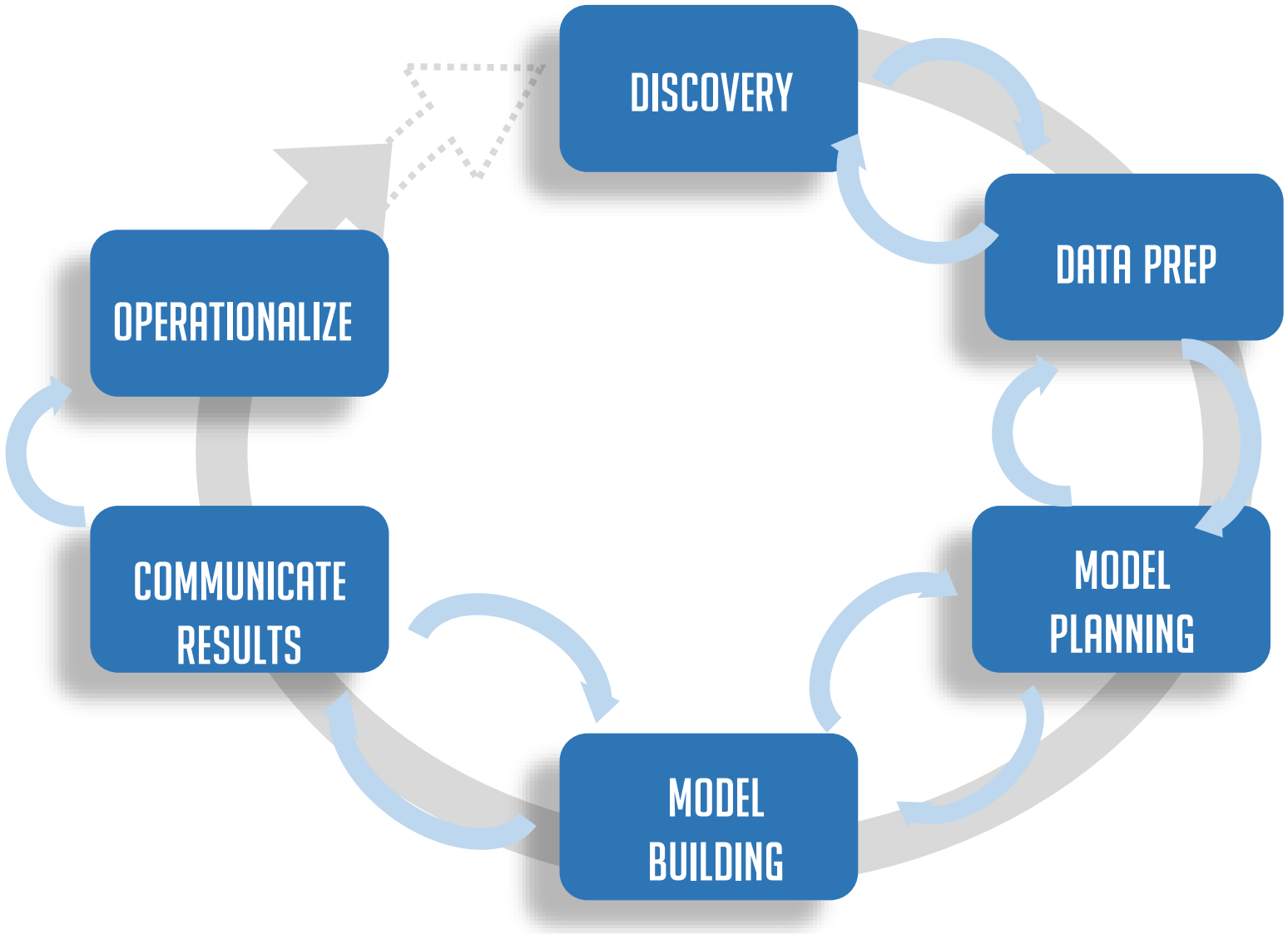
20142167 최현경

주제 선정

환율과 계절 변동에 따른 제주도 여행객 추이

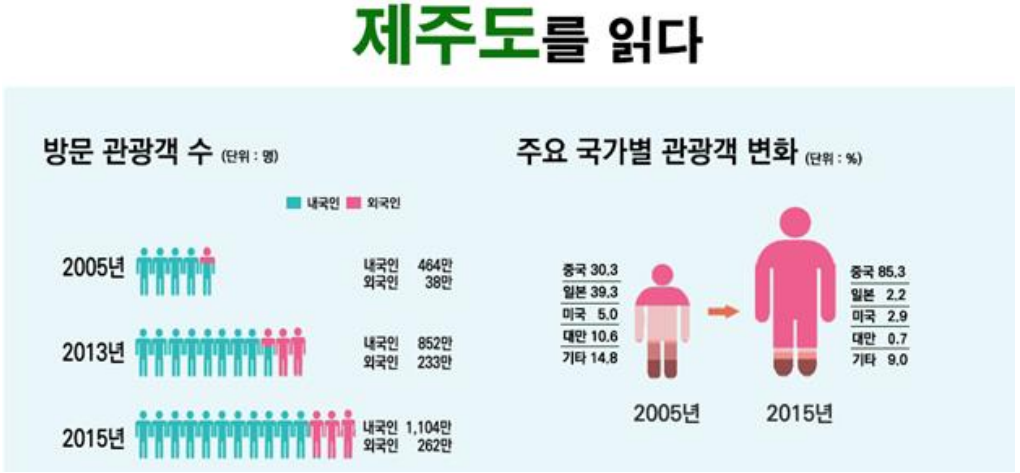
DATA ANALYTICS LIFECYCLE

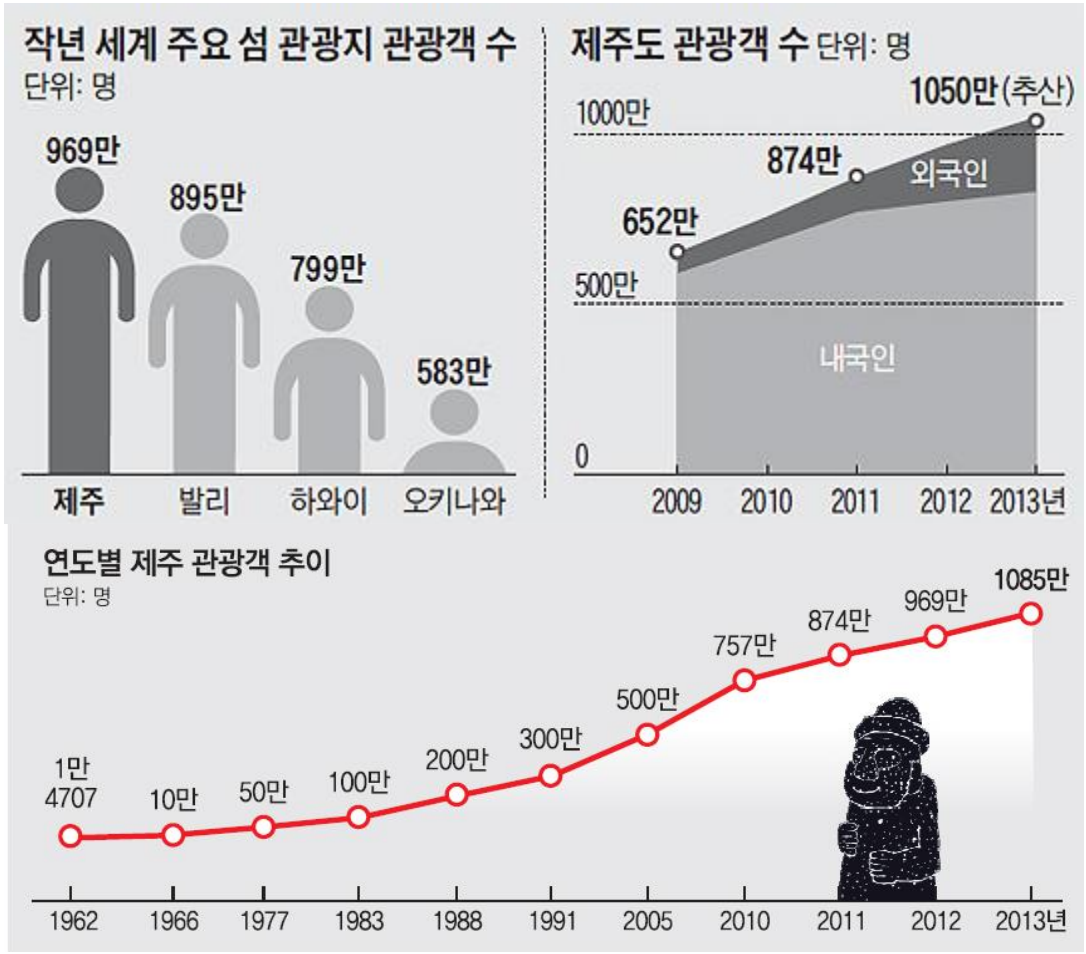
DA LIFE CYCLE



01.DISCOVERY

LEARN FROM THE PAST





RESOURCES :일일 환율정보, 일일 제주도 입도 정보,일일 날씨정보

01.DISCOVERY

FORMULATE INITIAL HYPOTHESES

Q1: 환율은 여행객에게 영향을 준다.

Q2: 날씨는 여행객에게 영향을 준다.

Q3. 환율과 날씨는 둘 다 여행객에 크게 영향을 준다.

01.DISCOVERY

FRAME THE PROBLEM
-IDENTIFY THE SUCCESS CRITERIA, KEY RISKS, AND STAKEHOLDERS

SUCCESS CRITERIA : 제주도 관광객(입도 수 기준)수는 봄과 여름에 확연히 높게 나온다. 이를 제주도 관광 활성화 사업을 이용한 집중화 홍보목적으로 사용할 수 있다고 생각한다.

KEY RISKS : 현재 중국에서 국가 차원에서 반한 감정을 조장하고 있다. 이 점이 제주도 관광객 수에 영향을 줄 것이라고 생각한다.

STAKEHOLDERS : 여행사,숙박업소들, 항공사 그리고 렌터카 직원들 관광사업에 밀접한 관련이 있는 사람들이 참고하기 좋은 자료

02.DATA PREPARATION

DATA CONDITIONING

날씨와 환율 데이터같은 경우 데이터 값 자체를 변경할 필요가 없이 필요한 정보만을 추출하였다.

날씨 같은 경우 온도 습도 기후상태,강수량 정보 외에 정보들은 삭제하였고,

환율데이터 역시 한국은행 기준율 데이터만을 추출하였다.

하지만 일별 제주도 입도 데이터 같은 경우는 제주 특별자치도 관광 협회에서 매일 결과를 공표하였다.

이 결과를 R을 이용하여 파싱하여 데이터를 추출하였다. 그 후 일별 관광객 데이터만을 추출하였다.

엔화데이터샘플

2016-12-07	1,025.52
2016-12-06	1,031.37
2016-12-05	1,034.01
2016-12-02	1,027.47
2016-12-01	1,020.76
2016-11-30	1,038.81

02.DATA PREPARATION

달러,엔화,위안 데이터 샘플

D - DOLLAR환율

Y - 엔화 환율,

YU- 위안환율

R데이터 상 모습

RStudio Source Editor

d

Filter

	date	dr
1	2016-12-07	1169.5
2	2016-12-06	1173.8
3	2016-12-05	1170.5
4	2016-12-02	1172.6
5	2016-12-01	1168.0
6	2016-11-30	1168.5
7	2016-11-29	1170.3
8	2016-11-28	1178.9
9	2016-11-25	1182.4
10	2016-11-24	1174.9
11	2016-11-23	1175.5
12	2016-11-22	1183.6
13	2016-11-21	1182.3
14	2016-11-18	1174.4
15	2016-11-17	1167.2
16	2016-11-16	1168.8
17	2016-11-15	1169.8
18	2016-11-14	1164.4
19	2016-11-11	1151.0
20	2016-11-10	1145.8
21	2016-11-09	1137.5

RStudio Source Editor

y

Filter

	date	yr
1	2016-12-07	1025.520000
2	2016-12-06	1031.370000
3	2016-12-05	1034.010000
4	2016-12-02	1027.470000
5	2016-12-01	1020.760000
6	2016-11-30	1038.810000
7	2016-11-29	1046.640000
8	2016-11-28	1044.340000
9	2016-11-25	1044.020000
10	2016-11-24	1043.980000
11	2016-11-23	1057.680000
12	2016-11-22	1067.560000
13	2016-11-21	1065.280000
14	2016-11-18	1065.840000
15	2016-11-17	1072.600000
16	2016-11-16	1071.800000
17	2016-11-15	1080.500000
18	2016-11-14	1090.670000
19	2016-11-11	1077.110000
20	2016-11-10	1083.450000
21	2016-11-09	1081.480000

RStudio Source Editor

yu

Filter

	date	yur
1	2016-12-07	170.31
2	2016-12-06	170.52
3	2016-12-05	170.07
4	2016-12-02	170.01
5	2016-12-01	169.23
6	2016-11-30	168.98
7	2016-11-29	168.93
8	2016-11-28	169.80
9	2016-11-25	170.11
10	2016-11-24	169.78
11	2016-11-23	170.19
12	2016-11-22	171.23
13	2016-11-21	171.15
14	2016-11-18	170.47
15	2016-11-17	169.71
16	2016-11-16	170.30
17	2016-11-15	171.25
18	2016-11-14	170.44
19	2016-11-11	168.89
20	2016-11-10	167.69
21	2016-11-09	167.59

02.DATA PREPARATION

PARSING CODE

작성일 : 16-12-05 07:53 목록

2016년 12월 4일 (일) 관광객 입도현황 트랙백

글쓴이 : 제주특별자치도관광협회 조회 : 72

2016년 12월 4일 (일) 관광객 입도현황

구분		총계	내국인	외국인
2016년(명)	일계	37,419	29,552	7,867
	월계	153,139	124,105	29,034
	누계	14,834,881	11,429,310	3,405,571
2015년(명)	일계	34,167	29,115	5,052
	월계	116,172	97,803	18,369
	누계	12,697,111	10,237,215	2,459,896
증감(%)	일계	9.5	1.5	55.7
	월계	31.8	26.9	58.1
	누계	16.8	11.6	38.4

02.DATA PREPARATION

PARSING CODE

```
url <- "http://www.visitjeju.or.kr/APP/bbs/board.php?bo_table=total_day&wr_id=1488"
line <- readLines(url,encoding="UTF-8")
#DataParser <- "<td colspan=\"5\" height=\"22\" style=\"width: 381px; height: 22px; text-align: center;\">"
DataParser <- "2016"
title <- line[str_detect(line,DataParser)]
a <- substr(title,87,100)
final <- NULL
title <- NULL
for(i in 1:1490)
{
  url <- paste0("http://www.visitjeju.or.kr/APP/bbs/board.php?bo_table=total_day&wr_id=",i)
  line <- readLines(url,encoding="UTF-8")
  #DataParser <- "2016"
  #tmp <- line[str_detect(line,DataParser)]
  listTmp <- grep('201',line) #2016
  tmp <- line[listTmp[2]]
  tmp <- str_trim(tmp) #
  tmp <- gsub(" ","",tmp) #
  tmp <- gsub("?꿱","-",tmp)
  tmp <- gsub("?쌩","-",tmp)
```

일별 제주도 입도 데이터

jejunum

Filter

	number	date
1	17010	2013-01-01
2	22121	2013-01-02
3	22305	2013-01-03
4	20604	2013-01-04
5	20286	2013-01-05
6	17811	2013-01-06
7	21620	2013-01-07
8	17749	2013-01-08
9	17566	2013-01-09
10	21374	2013-01-10
11	22220	2013-01-11
12	25704	2013-01-12
13	18432	2013-01-13

03.MODEL PLANNING

DETERMINE METHODS

LINEAR REGRESSION & MULTIPLE LINEAR REGRESSION ANALYSIS

-> 환율과 여행객 상관관계 분석

-> 날씨와 여행객 상관관계 분석

VARIABLE SELECTION

앞서 얘기한 바와 같이 계절과 환율이 여행객 추이에 영향을 미칠 것으로 판단하여 계절과 환율을 독립변수(X 변수)로 지정하고 여행객 수를 종속변수(Y 변수)로 지정한 다음 선형회귀를 사용하여 데이터를 분석하였다

테스트는 1.환율 2.날씨 3.환율&날씨를 이용하므로 다중선형과 선형회귀 방법이 사용

04. MODEL BUIDING

기존의 데이터를 이용하여 분기(성수기,극성수기)로 나눈 새로운 데이터들을 이용하여 테스트하였다.

이 과정에서 분석 방법은 선형회귀로 동일하였다.

04. MODEL BUIDING

	number	date	dr	yur	yr
18	17010	2013-01-01	1095.0	176.15	1141.46
84	22121	2013-01-02	1071.1	171.92	1236.05
88	22305	2013-01-03	1064.7	170.89	1218.96
56	20604	2013-01-04	1063.8	170.75	1216.33
50	20286	2013-01-05	1064.8	173.41	1033.94
26	17811	2013-01-06	1063.4	170.70	1216.01
72	21620	2013-01-07	1064.1	170.79	1205.78
25	17749	2013-01-08	1063.4	170.70	1216.01
24	17566	2013-01-09	1063.2	170.82	1222.14
71	21374	2013-01-10	1062.1	170.59	1206.25
85	22220	2013-01-11	1061.0	170.48	1193.41
170	25704	2013-01-12	1068.4	171.82	1182.64
32	18432	2013-01-13	1082.7	174.05	1188.47
70	21314	2013-01-14	1056.5	169.96	1180.58
44	19600	2013-01-15	1056.1	169.81	1180.53
33	18455	2013-01-16	1055.4	169.85	1188.05
97	23082	2013-01-17	1058.9	170.34	1195.01
164	25510	2013-01-18	1058.2	170.24	1174.93
234	27167	2013-01-19	1079.8	173.49	917.11
86	22255	2013-01-20	1061.0	170.48	1193.41
78	21885	2013-01-21	1056.1	169.92	1171.36
36	18707	2013-01-22	1060.0	170.38	1181.12
39	19093	2013-01-23	1064.4	171.13	1198.78
103	23237	2013-01-24	1065.4	171.34	1204.11
169	25609	2013-01-25	1068.4	171.82	1182.64
295	28453	2013-01-26	1115.8	181.37	931.27
60	20913	2013-01-27	1077.2	178.08	1056.54
64	21145	2013-01-28	1071.8	172.30	1176.77
30	18267	2013-01-29	1082.5	173.96	1194.09
13	15699	2013-01-30	1088.0	174.80	1198.70
31	18361	2013-01-31	1082.7	174.05	1188.47

Showing 1 to 31 of 1,437 entries

제주도입국객 데이터에서 NA.OMIT으로 NA제거
주말환율을 금요일 환율로 채워 2013.01.01~2016.12.06까지 완벽한 데이터 완성

04. MODEL BUIDING

Untitled1* × Untitled2* × Untitled3* × Untitled4* × nd ×

Filter

	number	date	dr	yur	yr	spot	avgtemp	avgrain	avgwv	avghum	avgc	sun
1	399	2013-12-23	1061.9	174.90	1020.57	<ec>젤ㄴ<bc>	7.1	0.0	4.4	62	8.8	1.10
2	4776	2014-06-02	1020.4	163.32	1002.06	<ec>젤ㄴ<bc>	22.7	2.4	6.5	68	8.0	13.55
3	11173	2014-07-09	1011.2	163.06	995.91	<ec>젤ㄴ<bc>	23.5	65.0	6.7	94	10.0	6.39
4	12824	2013-02-06	1086.8	174.46	1160.86	<ec>젤ㄴ<bc>	5.6	2.7	5.0	80	9.6	0.89
5	13572	2013-02-03	1086.8	174.46	1160.86	<ec>젤ㄴ<bc>	8.2	1.1	1.9	70	7.3	9.20
6	13841	2013-02-04	1092.9	175.50	1177.69	<ec>젤ㄴ<bc>	9.1	5.5	3.8	87	8.9	2.69
7	14434	2016-01-25	1203.2	182.23	1013.43	<ec>젤ㄴ<bc>	1.9	1.6	6.4	74	8.0	4.23
8	14481	2013-03-05	1091.4	175.32	1167.65	<ec>젤ㄴ<bc>	7.2	0.0	2.2	60	0.5	19.00
9	14711	2013-03-06	1087.0	174.73	1164.43	<ec>젤ㄴ<bc>	11.0	0.0	2.2	49	0.0	16.78
10	15699	2013-01-30	1088.0	174.80	1198.70	<ec>젤ㄴ<bc>	10.1	0.0	2.5	70	2.8	12.84
11	15749	2013-03-12	1098.3	176.63	1137.13	<ec>젤ㄴ<bc>	13.9	0.0	1.9	60	1.5	16.79
12	16056	2013-03-13	1095.0	176.15	1141.46	<ec>젤ㄴ<bc>	9.5	13.3	4.8	83	9.6	2.63
13	16722	2013-03-10	1095.0	176.15	1141.46	<ec>젤ㄴ<bc>	8.4	0.0	3.8	64	2.0	17.47
14	16774	2016-01-23	1095.0	176.15	1141.46	<ec>젤ㄴ<bc>	-0.5	12.2	7.2	80	8.3	1.23
15	17010	2013-01-01	1095.0	176.15	1141.46	<ec>젤ㄴ<bc>	7.3	0.0	4.5	54	5.8	7.90
16	17080	2014-05-11	1095.0	176.15	1141.46	<ec>젤ㄴ<bc>	20.6	1.4	5.5	76	8.3	11.18
17	17100	2013-03-03	1095.0	176.15	1141.46	<ec>젤ㄴ<bc>	5.0	0.0	3.2	64	1.8	16.84
18	17100	2014-03-03	1066.2	173.51	1051.06	<ec>젤ㄴ<bc>	9.5	0.0	4.4	47	1.1	19.70

날씨와 제주도 입국객 데이터를 합침

SPOT-지점,

DATE - 일자,

AVGTEMP-평균기온

AVGRAIN-일강수량,

AGVWV-평균풍속,

AVGHUM-평균습도

AVGC-평균전운량,

SUN-일사량

05. COMMUNICATE RESULTS

Q1) 환율이 여행객에게 영향을 준다

```
> linear <- lm(number ~ dr, data = newTable)
> summary(linear)

Call:
lm(formula = number ~ dr, data = newTable)

Residuals:
   Min     1Q Median     3Q    Max
-33294  -6036   -368   6300  26774

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -22892.685   4748.436  -4.821 1.58e-06 ***
dr              52.911      4.273  12.382  < 2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8315 on 1435 degrees of freedom
Multiple R-squared:  0.09653,   Adjusted R-squared:  0.0959
F-statistic: 153.3 on 1 and 1435 DF,  p-value: < 2.2e-16

>
```

```
> linear <- lm(number ~ yr, data = newTable)
> summary(linear)

Call:
lm(formula = number ~ yr, data = newTable)

Residuals:
   Min     1Q Median     3Q    Max
-35918  -5925   -456   6475  27196

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 42749.298   2946.350  14.509   <2e-16 ***
yr             -6.694      2.846  -2.352   0.0188 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8731 on 1435 degrees of freedom
Multiple R-squared:  0.00384,   Adjusted R-squared:  0.003146
F-statistic: 5.532 on 1 and 1435 DF,  p-value: 0.01881
```

```
> linear <- lm(number ~ yur, data = newTable)
> summary(linear)

Call:
lm(formula = number ~ yur, data = newTable)

Residuals:
   Min     1Q Median     3Q    Max
-35975  -5550     49   6932  25474

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 59509.75    6396.74   9.303  < 2e-16 ***
yur          -134.56      36.34  -3.703 0.000222 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8706 on 1435 degrees of freedom
Multiple R-squared:  0.009463,   Adjusted R-squared:  0.008773
F-statistic: 13.71 on 1 and 1435 DF,  p-value: 0.0002215
```

달러

엔화

위안

<0.05 → 유의미한 관계

05. COMMUNICATE RESULTS

Q1) 환율이 여행객에게 영향을 준다

```
> linear <- lm(number ~ dr+yr+yur, data = newTable)
> summary(linear)

Call:
lm(formula = number ~ dr + yr + yur, data = newTable)

Residuals:
     Min       1Q   Median       3Q      Max
-30802.6  -4994.5    795.1   5199.7  19841.4

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 59130.183   5928.586   9.974  < 2e-16 ***
dr            128.553      5.267  24.409  < 2e-16 ***
yr            -13.899      2.400  -5.791 8.58e-09 ***
yur          -862.112     42.632 -20.222  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7307 on 1433 degrees of freedom
Multiple R-squared:  0.3031,    Adjusted R-squared:  0.3017
F-statistic: 207.8 on 3 and 1433 DF,  p-value: < 2.2e-16
```

달러,엔화,위안 확률 비교

<0.05 → 유의미한 관계

다중공산성 확인 → 4보다 낮은 값 → 다중공산성 없음

```
> vif(linear)
      dr       yr      yur
1.966613 1.015093 1.953261
>
```

05. COMMUNICATE RESULTS

Q1) 환율이 여행객에게 영향을 준다

이상치테스트

```
> outlierTest(linear)
   rstudent unadjusted p-value Bonferonni p
4 -4.248141         2.2946e-05     0.032973
1 -4.225617         2.5335e-05     0.036406
2 -4.169977         3.2291e-05     0.046402
```

REMOVE OUTLIER

```
> newTable <- newTable[-1,]
> newTable <- newTable[-2,]
> newTable <- newTable[-4,]
> linear <- lm(number ~ dr+yr+yur, data = newTable)
> summary(linear)

Call:
lm(formula = number ~ dr + yr + yur, data = newTable)

Residuals:
     Min       1Q   Median       3Q      Max
-30893.8  -4976.7    791.8   5170.9  19715.3

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 59271.024   5852.103  10.128  < 2e-16 ***
dr            127.032      5.203  24.414  < 2e-16 ***
yr            -13.570      2.372  -5.722 1.28e-08 ***
yur          -854.930     42.091 -20.312  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7212 on 1430 degrees of freedom
Multiple R-squared:  0.3038,    Adjusted R-squared:  0.3024
F-statistic:   208 on 3 and 1430 DF,  p-value: < 2.2e-16
```

05. COMMUNICATE RESULTS

Q1) 환율이 여행객에게 영향을 준다

이상치 제거 후 환율로 선형회귀, SQUARED-R값 0.3024

```
> newTable <- newTable[-1,]
> newTable <- newTable[-2,]
> newTable <- newTable[-4,]
> linear <- lm(number ~ dr+yr+yur, data = newTable)
> summary(linear)

Call:
lm(formula = number ~ dr + yr + yur, data = newTable)

Residuals:
     Min       1Q   Median       3Q      Max
-30893.8  -4976.7    791.8   5170.9  19715.3

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 59271.024   5852.103  10.128  < 2e-16 ***
dr            127.032      5.203  24.414  < 2e-16 ***
yr            -13.570      2.372  -5.722 1.28e-08 ***
yur          -854.930     42.091 -20.312  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7212 on 1430 degrees of freedom
Multiple R-squared:  0.3038,    Adjusted R-squared:  0.3024
F-statistic:   208 on 3 and 1430 DF,  p-value: < 2.2e-16
```

영향이 있긴 있으나 미미함

05. COMMUNICATE RESULTS

Q2) 날씨가 여행객에게 영향을 준다

```
> linear <- lm(number~avgtemp+avgrain+avgwv+avghum+avgc+sun, data=nd)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + avghum + avgc +
    sun, data = nd)

Residuals:
   Min     1Q Median     3Q    Max
-32049  -5353   -100   5545  24943

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 30480.12    1726.39  17.655  < 2e-16 ***
avgtemp       495.73      37.99  13.048  < 2e-16 ***
avgrain       -47.60      18.72  -2.542   0.0111 *
avgwv        -848.62     164.54  -5.158 2.85e-07 ***
avghum        -45.26      22.77  -1.988   0.0470 *
avgc          300.29     116.64   2.574   0.0101 *
sun           103.09      45.69   2.256   0.0242 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7421 on 1426 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.2644,    Adjusted R-squared:  0.2613
F-statistic: 85.44 on 6 and 1426 DF,  p-value: < 2.2e-16
```

<0.05 → 유의미한 관계

다중공산성 확인 → 4보다 낮은 값 → 다중공산성 없음

```
> vif(linear)
 avgtemp  avgrain    avgwv   avghum     avgc      sun
2.152058 1.266147 1.266383 2.089200 3.040471 3.417620
```

평균기온, 일강수량, 평균풍속,평균습도,전운량,일조량 과 입국객 수

05. COMMUNICATE RESULTS

02) 날씨가 여행객에게 영향을 준다

이상치테스트

```
> outlierTest(linear)
  rstudent unadjusted p-value Bonferonni p
2  -4.3637         1.3709e-05      0.019646
> |
```

2번 데이터 이상치 발견

REMOVE OUTLIER

```
> nd <- nd[-2,]
> linear <- lm(number~avgtemp+avgrain+avgwv+avghum+avgc+sun, data=nd)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + avghum + avgc +
    sun, data = nd)

Residuals:
     Min       1Q   Median       3Q      Max
-29871.5  -5353.5    -90.8   5508.3  24898.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30270.41    1716.25  17.638  < 2e-16 ***
avgtemp       502.22      37.78  13.292  < 2e-16 ***
avgrain       -48.65      18.61  -2.614  0.00903 **
avgwv        -796.64     163.94  -4.859 1.31e-06 ***
avghum        -46.73      22.63  -2.065  0.03908 *
avgc          309.34     115.93   2.668  0.00771 **
sun           104.26      45.40   2.296  0.02180 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7374 on 1425 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.2675,    Adjusted R-squared:  0.2644
F-statistic: 86.72 on 6 and 1425 DF,  p-value: < 2.2e-16
```

평균기온, 일강수량, 평균풍속,평균습도,전운량,일조량 과 입국객 수

05. COMMUNICATE RESULTS

Q2) 날씨가 여행객에게 영향을 준다

```
> nd <- nd[-2,]
> linear <- lm(number~avgtemp+avgrain+avgwv+avghum+avgc+sun, data=nd)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + avghum + avgc +
    sun, data = nd)

Residuals:
     Min       1Q   Median       3Q      Max
-29871.5  -5353.5    -90.8   5508.3  24898.0

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 30270.41    1716.25  17.638  < 2e-16 ***
avgtemp       502.22      37.78  13.292  < 2e-16 ***
avgrain       -48.65      18.61  -2.614  0.00903 **
avgwv        -796.64     163.94  -4.859 1.31e-06 ***
avghum        -46.73      22.63  -2.065  0.03908 *
avgc          309.34     115.93   2.668  0.00771 **
sun           104.26      45.40   2.296  0.02180 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7374 on 1425 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.2675,    Adjusted R-squared:  0.2644
F-statistic: 86.72 on 6 and 1425 DF,  p-value: < 2.2e-16
```

영향이 있긴 있으나 미미함

05. COMMUNICATE RESULTS

Q3) 환율과 날씨가 여행객에게 영향을 준다

```
> linear <- lm(number~avgtemp+avgrain+avgwv+sun+dr+yr+yur, data=nd)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + sun + dr +
    yr + yur, data = nd)

Residuals:
     Min       1Q   Median       3Q      Max
-25255.8  -3975.8    400.5   4201.5  18267.3

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 50412.391   4951.272  10.182  < 2e-16 ***
avgtemp       423.879     25.843  16.402  < 2e-16 ***
avgrain       -54.983     14.755  -3.726 0.000202 ***
avgwv        -618.150    129.099  -4.788 1.86e-06 ***
sun            76.638     24.678   3.106 0.001937 **
dr            115.950      4.362  26.583  < 2e-16 ***
yr            -15.476      1.976  -7.832 9.34e-15 ***
yur          -757.371     35.369 -21.413  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5997 on 1424 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5158,    Adjusted R-squared:  0.5134
F-statistic: 216.7 on 7 and 1424 DF,  p-value: < 2.2e-16
```

모든 독립적인 요소 < 0.05 유의미한 값

입국객 수와 날씨, 환율 통합적으로 다중 회귀분석 결과

05. COMMUNICATE RESULTS

Q3) 환율과 날씨가 여행객에게 영향을 준다

```
> linear <- lm(number~avgtemp+avgrain+avgwv+avghum+avgc+sun+dr+yr+yur, data=nd)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + avghum + avgc +
    sun + dr + yr + yur, data = nd)

Residuals:
     Min       1Q   Median       3Q      Max
-25337.0  -4003.3    382.1   4200.0  18394.7

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 50106.668   5057.763   9.907  < 2e-16 ***
avgtemp       417.437     30.943  13.490  < 2e-16 ***
avgrain       -54.893     15.158  -3.622 0.000303 ***
avgwv        -640.351    133.551  -4.795 1.80e-06 ***
avghum         -7.209     18.633  -0.387 0.698887
avgc           97.129     95.112   1.021 0.307332
sun           101.549     36.958   2.748 0.006077 **
dr            115.630      4.377  26.417  < 2e-16 ***
yr            -15.248      2.003  -7.614 4.83e-14 ***
yur          -756.219     35.409 -21.356  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6000 on 1422 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5162,    Adjusted R-squared:  0.5131
F-statistic: 168.6 on 9 and 1422 DF,  p-value: < 2.2e-16
```

전운량 ,평균습도 > 0.05 유의미한 값이 아님

입국객 수와 날씨, 환율 통합적으로 다중 회귀분석 결과

05. COMMUNICATE RESULTS

Q3) 환율과 날씨가 여행객에게 영향을 준다

다중 공산성 검사 다중공산성 < 4 존재하지 않음

```
> vif(linear)
 avgtemp  avgrain    avgwv      sun       dr       yr      yur
1.523706 1.203604 1.188284 1.526496 1.995556 1.017956 1.989183
```

이상치테스트

```
> outlierTest(linear)
   rstudent unadjusted p-value Bonferonni p
1 -4.244403         2.3336e-05      0.033417
```

1번 데이터 이상치 발견

REMOVE OUTLIER


```
> nd <- nd[-1,]
> linear <- lm(number~avgtemp+avgrain+avgwv+sun+dr+yr+yur, data=nd)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + sun + dr +
    yr + yur, data = nd)

Residuals:
     Min       1Q   Median       3Q      Max
-20660.1  -3992.0    389.7   4181.0  18249.6

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 50623.676   4922.206  10.285  < 2e-16 ***
avgtemp       423.323     25.691  16.478  < 2e-16 ***
avgrain       -56.133     14.670  -3.826 0.000136 ***
avgwv        -612.574    128.341  -4.773 2.00e-06 ***
sun            73.215     24.545   2.983 0.002904 **
dr            115.439      4.338  26.613  < 2e-16 ***
yr            -15.463      1.964  -7.872 6.89e-15 ***
yur          -755.087     35.164 -21.473  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5962 on 1423 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5161,    Adjusted R-squared:  0.5137
F-statistic: 216.8 on 7 and 1423 DF,  p-value: < 2.2e-16
```

05. COMMUNICATE RESULTS

03) 환율과 날씨가 여행객에게 영향을 준다

```
> nd <- nd[-1,]
> linear <- lm(number~avgtemp+avgrain+avgwv+sun+dr+yr+yur, data=nd)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + sun + dr +
    yr + yur, data = nd)

Residuals:
     Min       1Q   Median       3Q      Max
-20660.1  -3992.0    389.7   4181.0  18249.6

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 50623.676   4922.206  10.285  < 2e-16 ***
avgtemp       423.323     25.691  16.478  < 2e-16 ***
avgrain       -56.133     14.670  -3.826 0.000136 ***
avgwv        -612.574    128.341  -4.773 2.00e-06 ***
sun            73.215     24.545   2.983 0.002904 **
dr            115.439      4.338  26.613  < 2e-16 ***
yr            -15.463      1.964  -7.872 6.89e-15 ***
yur          -755.087     35.164 -21.473  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5962 on 1423 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.5161,    Adjusted R-squared:  0.5137
F-statistic: 216.8 on 7 and 1423 DF,  p-value: < 2.2e-16
```

영향이 있음을 알 수 있다.

05. COMMUNICATE RESULTS

Q3) 환율과 날씨가 여행객에게 영향을 준다

성수기 분석 (1,2,7,8,10월)

```
> linear <- lm(number~avgtemp+avgrain+avgwv+avghum+dr+yr+yur, data=Result)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + avghum + dr +
    yr + yur, data = Result)

Residuals:
     Min       1Q   Median       3Q      Max
-22210.8  -3105.1    106.3   3477.0  18329.0

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 77441.636   7235.965  10.702  < 2e-16 ***
avgtemp       529.810     28.659  18.487  < 2e-16 ***
avgrain       -55.964     16.678  -3.356 0.000842 ***
avgwv        -755.443    181.411  -4.164 3.58e-05 ***
avghum        -96.420     25.367  -3.801 0.000159 ***
dr            118.437      5.998  19.747  < 2e-16 ***
yr            -20.031      2.727  -7.346 6.70e-13 ***
yur          -861.314     47.899 -17.982  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5445 on 598 degrees of freedom
Multiple R-squared:  0.6501,    Adjusted R-squared:  0.646
F-statistic: 158.7 on 7 and 598 DF,  p-value: < 2.2e-16
```

REMOVE 일조량,전운량


```
> linear <- lm(number~avgtemp+avgrain+avgwv+avghum+dr+yr+yur, data=Result)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + avghum + dr +
    yr + yur, data = Result)

Residuals:
     Min       1Q   Median       3Q      Max
-22210.8  -3105.1    106.3   3477.0  18329.0

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 77441.636   7235.965  10.702  < 2e-16 ***
avgtemp       529.810     28.659  18.487  < 2e-16 ***
avgrain       -55.964     16.678  -3.356 0.000842 ***
avgwv        -755.443    181.411  -4.164 3.58e-05 ***
avghum        -96.420     25.367  -3.801 0.000159 ***
dr            118.437      5.998  19.747  < 2e-16 ***
yr            -20.031      2.727  -7.346 6.70e-13 ***
yur          -861.314     47.899 -17.982  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5445 on 598 degrees of freedom
Multiple R-squared:  0.6501,    Adjusted R-squared:  0.646
F-statistic: 158.7 on 7 and 598 DF,  p-value: < 2.2e-16
```

05. COMMUNICATE RESULTS

Q3) 환율과 날씨가 여행객에게 영향을 준다

다중 공산성 검사 다중공산성 < 4 존재하지 않음

```
> vif(linear)
 avgtemp  avgrain    avgwv   avghum       dr       yr      yur
1.495865 1.226317 1.201854 1.594651 2.055465 1.024673 2.071483
```

이상치테스트

```
      rstudent unadjusted p-value Bonferonni p
3   -4.284777         2.1315e-05      0.012981
473 -4.072895         5.2686e-05      0.032086
```

REMOVE OUTLIER


```
> Result <- Result[-473,]
> Result <- Result[-3,]
> linear <- lm(number~avgtemp+avgrain+avgwv+avghum+dr+yr+yur, data=Result)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + avghum + dr +
    yr + yur, data = Result)

Residuals:
     Min       1Q   Median       3Q      Max
-23358.2  -3087.8    119.2   3519.0  18233.8

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 80224.497   7346.211  10.921  < 2e-16 ***
avgtemp       533.797     28.897  18.472  < 2e-16 ***
avgrain       -67.222     16.332  -4.116 4.40e-05 ***
avgwv        -784.953    181.280  -4.330 1.75e-05 ***
avghum        -97.247     25.464  -3.819 0.000148 ***
dr            120.367      6.077  19.806  < 2e-16 ***
yr            -20.967      2.764  -7.585 1.27e-13 ***
yur          -883.430     48.628 -18.167  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5522 on 599 degrees of freedom
Multiple R-squared:  0.6555,    Adjusted R-squared:  0.6515
F-statistic: 162.8 on 7 and 599 DF,  p-value: < 2.2e-16
```

05. COMMUNICATE RESULTS

Q3) 환율과 날씨가 여행객에게 영향을 준다

```
> Result <- Result[-473,]
> Result <- Result[-3,]
> linear <- lm(number~avgtemp+avgrain+avgwv+avghum+dr+yr+yur, data=Result)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + avghum + dr +
    yr + yur, data = Result)

Residuals:
     Min       1Q   Median       3Q      Max
-23358.2  -3087.8    119.2   3519.0  18233.8

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 80224.497   7346.211  10.921  < 2e-16 ***
avgtemp       533.797     28.897  18.472  < 2e-16 ***
avgrain       -67.222     16.332  -4.116 4.40e-05 ***
avgwv        -784.953    181.280  -4.330 1.75e-05 ***
avghum        -97.247     25.464  -3.819 0.000148 ***
dr            120.367      6.077  19.806  < 2e-16 ***
yr            -20.967      2.764  -7.585 1.27e-13 ***
yur          -883.430     48.628 -18.167  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5522 on 599 degrees of freedom
Multiple R-squared:  0.6555,    Adjusted R-squared:  0.6515
F-statistic: 162.8 on 7 and 599 DF,  p-value: < 2.2e-16
```

성수기에 영향있음

05. COMMUNICATE RESULTS

03) 환율과 날씨가 여행객에게 영향을 준다

```
> linear <- lm(number~avgtemp+avgrain+avgwv+avghum+dr+yr+yur, data=Result)
> summary(linear)

Call:
lm(formula = number ~ avgtemp + avgrain + avgwv + avghum + dr +
    yr + yur, data = Result)

Residuals:
     Min       1Q   Median       3Q      Max
-27592.1  -4106.9    396.6   4699.5  15828.6

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 38964.922  10266.097   3.795 0.000171 ***
avgtemp       222.387     71.080   3.129 0.001889 **
avgrain       -64.792     30.046  -2.156 0.031664 *
avgwv       -1021.930    245.851  -4.157 3.98e-05 ***
avghum         24.867     33.910   0.733 0.463808
dr            110.391      9.362  11.791  < 2e-16 ***
yr             -8.025      4.127  -1.945 0.052528 .
yur          -680.496     75.745  -8.984  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6397 on 387 degrees of freedom
  (432 observations deleted due to missingness)
Multiple R-squared:  0.3964,    Adjusted R-squared:  0.3855
F-statistic: 36.31 on 7 and 387 DF,  p-value: < 2.2e-16
```

비수기

비 성수기에 영향
성수기보다 없음

06.CONCLUDE

비 성수기에는 날씨의 영향이나 환율의 영향을 받기보다 **우연적인 요소**, 즉 개인의 휴가일정이나 즉흥적인 여행을 통하여 제주도 여행을 떠나게 된다.

하지만 성수기시즌에는 날씨와 환율 등의 영향을 받는다는 것을 알 수 있다.
이는 방학,즉 휴가철 계절이라는 성수기의 특징으로 보아
성수기에는 **더 좋은 날씨, 환율이 좋은 날에 여행을 계획한다는 것**을 알 수 있다.

참고 사이트 목록

기상청 : HTTP://STS.KMA.GO.KR/
네이버 금융 : HTTP://FINANCE.NAVER.COM/
우리은행 : HTTPS://WWW.WOORIBANK.COM/
통계청 : HTTP://KOSIS.KR/
제주 특별 자치도 관광 협회 : HTTP://WWW.VISITJEJU.OR.KR/

END