

적시적지(適時適地)

날씨 기반 관광지 분석 추천

문화관광 빅데이터 분석대회 DATA102

Contents

01 분석 주제

- 주제 탐색 과정
- 주제 선정

02 데이터 전처리

- 데이터 구성 및 모델
- 데이터 정제

03 분석 및 결과

- 분석 모델링 및 성능 비교
- 분석 결과

04 결론 및 활용 방안

- 관광지 추천 제언
- 마케팅 활용 방안

■ 01 분석 주제 – 주제 탐색 과정

폭염이 피서철 관광패턴도 바꾸었다.

g+

햇볕 강한 대천해수욕장 관 뜨거운 한반도

서늘한 계곡, 휴양림, 냉풍! 폭염에 휴가철 주요 관광지 고속도로 통행량 감소



폭엽 피하고픈 **관광**객, 호텔·실내시설서 호캉스

트래블바이크뉴스 | 2018.08.13. | 🚅

사진/JW 메리어트 동대문 [트래블바이크뉴스=김지수 기자] 올 여름 사상 최고 수준의 폭 염으로 서울 등... 호텔이 폭염 특수를 누렸다. 이들 호텔은 6~7월 여기어때 예약 상위권에 자리 잡았다. 특급호텔은 물론이고...

폭설·한파 몰아친 관광의 섬 제주 '개점휴업' 한숨만

항공기 140여편 결항 지연

지난 5월 바다**날씨** 최악..울릉**관광** 타격 대구MBC | 2018.06.23. | 🚅

울릉도 **관광** 성수기인 지난 달에 바다 **날씨**가 좋지 않아 울릉도 **관광**객 증가에 약영향을 미친 것으로 분석됐습 니다. 지난 5월 동해 남부 먼바다의 풍랑 특보 발효 횟수는 7.2일로 최근 20년 중 가장 많았고, 같은 기간...

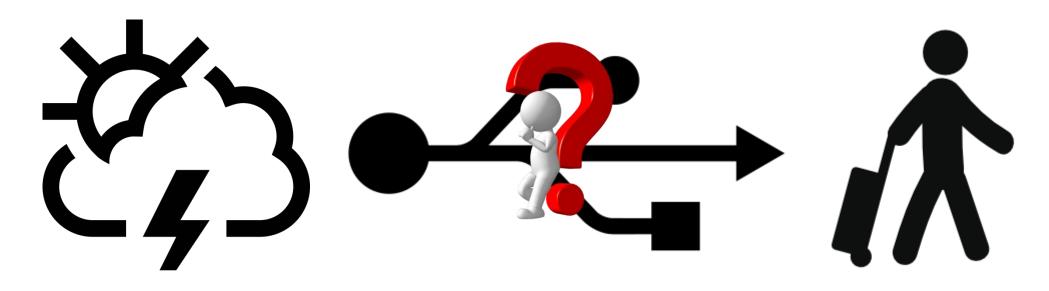
날씨 기반 관광지 분석 선택 이유

- ✓ 연일 지속된 폭염과 같이 날씨가 관광에 미치는 영향은 지대
- ✓ 그러나 관광객 또는 관광업에서 이를 활용하는 방안은 전무
- 관광객에게 새로운 패러다임 기반으로 관광지를 추천하고 관광업에서 날씨에 능동적으로 대처할 수 있는 분석이 필요

01 분석 주제 - 주제 선정

분석 주제

날씨에 따른 최적의 관광지 도출



분석 목표

- ① 회귀분석을 통해 날씨에 따른 관광지 특징 파악
- ② 분석결과에 따른 오늘의 날씨에 최적화된 관광지 추천

02 데이터 전처리 – 데이터 구성

데이터 선정

날씨 기반 관광지 추천을 위해 특정 날짜의 날씨와 그 날의 방문객 수가 필요



인스타그램은 전세계적으로 많은 사용자들이 이용하는 소셜네트워크서비스(SNS)

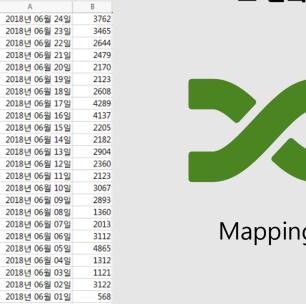
충분한 수의 관광객들이 인스타그램에서 사진, 동영상을 공유하므로 방문 경향성을 파악하는데 적합하다고 판단, 크롤링 통해 데이터 수집을 결정

02 데이터 전처리 - 데이터 모델

회귀분석 모델

일자별 여행지 게시물 개수 ~ 평균기온 + 강수량 + 습도 + …

일자별 여행지 방문객수를 반응변수, 기상데이터를 설명변수로 그 날의 날씨에 적합한 여행지 추천 모델 수립



2632 1009

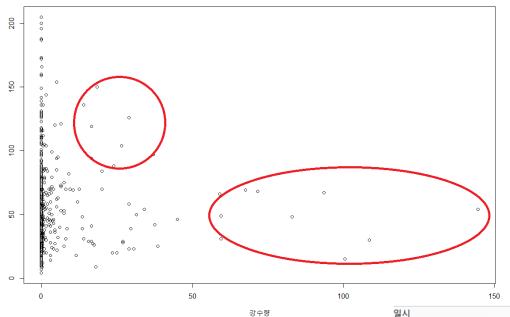
인스타그램 1일 관광객 게시글 수



Α	В	С	D	Е	F	G	Н	I	J	K	L	М	N
지점	일시	평균기온(최저기온(최고기온('	강수 계속	일강수량(෦	최대 풍속(평균 풍속	평균 상대	평균 증기	평균 현지	평균 해면	합계 일사(
108	2017-07-10	24	22.6	25.4	21.08	144.5	5.9	2.3	96.5	28.7	991.1	1000.8	2.01
108	2017-07-11	26	22.2	30.9	2.25	1	4.4	2.1	81.4	26.9	997.2	1006.9	13.94
108	2017-07-12	27.1	24.9	31.1	3.92	0	4.6	2.3	75.9	26.8	1000.2	1009.9	17.01
108	2017-07-13	27.6	23.1	32.1			4	1.5	71	25.7	998	1007.6	19.23
108	2017-07-14	28.7	26.1	33	0.5	0	6	2.1	68.4	26.6	998	1007.6	16.04
108	2017-07-15	25	23.4	27.3	15.67	42.5	5.2	2.2	89	28	999.5	1009.2	5.65
108	2017-07-16	26.2	23.4	28.8	7.08	22.5	4.3	1.4	83.4	27.9	1001.8	1011.5	10.5
108	2017-07-17	26.8	24	29	2.17	7.5	4.3	2	77.3	26.9	1001.9	1011.6	11.18
108	2017-07-18	26.5	23.9	30.7	8.83	2.5	3.5	1.6	80.6	27.5	999.5	1009.2	11.5
108	2017-07-19	28.3	24.5	34.1			4.6	1.8	71.4	26.9	997.1	1006.7	20.43
108	2017-07-20	29.9	26	34.9			4.8	2.2	66.1	27.4	996.6	1006.2	20.59
108	2017-07-21	29.8	27.3	32.4			3.9	2.5	73.1	30.5	998.2	1007.8	10.81
108	2017-07-22	29.3	27.5	32.5	6.5	0.5	4	2.3	75.8	30.6	999.1	1008.7	10.01
108	2017-07-23	27	26.1	28.1	8	133.5	4.1	2.1	89.5	31.7	994.9	1004.5	2.72
108	2017-07-24	26.2	25.6	27.3	3.75	3	2.7	1.2	82.3	27.9	992.9	1002.5	5.17
108	2017-07-25	29.4	24.6	35.4			4.6	1.9	64.1	25.4	993.9	1003.4	21.74
108	2017-07-26	28.7	24.2	32.8			4.8	2.7	55.6	21.5	997.6	1007.2	22.61
108	2017-07-27	27.1	24.9	29.7	3.33	3.5	3.6	1.9	62.5	22.1	998.2	1007.9	10
108	2017-07-28	25.5	24.8	26.9	13.25	8.5	3.1	1.5	92.1	29.9	998	1007.8	3.38
108	2017-07-29	27.2	25	30.1	0.75	0	3.9	2.3	72.3	25.8	997.6	1007.2	11.78
108	2017-07-30	27.9	24.7	32.6			4.2	1.9	64.4	23.8	997.2	1006.8	18.56
108	2017-07-31	25.7	24.8	27.1	10.08	3.5	3.5	1.4	82.3	27.1	998.9	1008.6	4.34
108	2017-08-01	28.6	24.1	33	3.25	2	3.3	1.7	73	28	1000.1	1009.7	18.08
108	2017-08-02	30	26.6	33			2.5	1.3	71.9	30.2	997.4	1006.9	11.98
108	2017-08-03	30.2	27.2	33.9			4.3	2.2	60.6	25.4	995.2	1004.8	21.79
108	2017-08-04	30 3	25 9	35 3			46	21	60 1	25.6	995 5	1005	22 18
714	L TI = 711 HL	₩ =1	1101		1 471	1 70	10	11 7	01 110	9 AL 3	Z 71 7	1 4 1 71	> □ 0

기상자료개방포털 사이트에서 17년 7월 ~ 18년 7월 서울의 종관기상관측 데이터

이상치 처리



1. 특이값 처리

실질적 여행 시간에 따른 날씨 변수 재정의 :

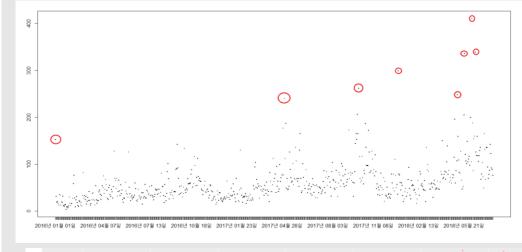
강수량이 많으면 게시물 수가 적을 것으로 예측됨 (북한산, 한강 등) 그러나 <mark>강수량이 많으나 게시물 개수가 많은 이상치</mark>가 존재 확인 결과 해당 일자의 강수 대부분이 새벽시간 혹은 소나기처럼 집중됨

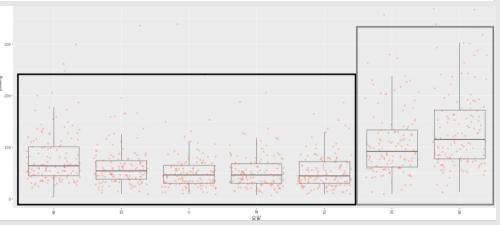


실질적 관광 시간인 8시~23시 사이의 강수량을 일 강수량으로 재정의 마찬가지로 적설량도 8시~23시 사이의 적설량을 일 적설량으로 재정의

일시	강수량(mm)	일시	강수량(mm)	일시	강수량(mm)	일시	강수량(mm)
2018-04-04 0:00	0.1	2018-06-14 5:00	0	2018-06-15 2:00	0.1	2018-06-28 2:00	0
2018-04-04 1:00	0	2018-06-14 6:00	1.5	2018-06-15 3:00	5.4	2018-06-28 3:00	0
2018-04-04 2:00	3.5	2018-06-14 7:00	13.5	2018-06-15 4:00	10.5	2018-06-28 4:00	0
2018-04-04 3:00	7	2018-06-14 8:00	11.5	2018-06-15 5:00	0.5	2018-06-28 5:00	0
2018-04-04 4:00	3	2018-06-14 9:00	1.5	2018-06-15 6:00	0	2018-06-28 6:00	0.1
2018-04-04 5:00	3	2018-06-14 10:00	1			2018-06-28 17:00	4.9
2018-04-04 6:00	2	2018-06-14 11:00	0			2018-06-28 18:00	21.5
2018-04-04 10:00	0	2018-06-14 12:00	0			2018-06-28 19:00	0
		2018-06-14 13:00	0			2018-06-28 22:00	0
		2018-06-14 20:00	0			2018-06-28 23:00	0
		2018-06-14 21:00	0				

이상치 처리





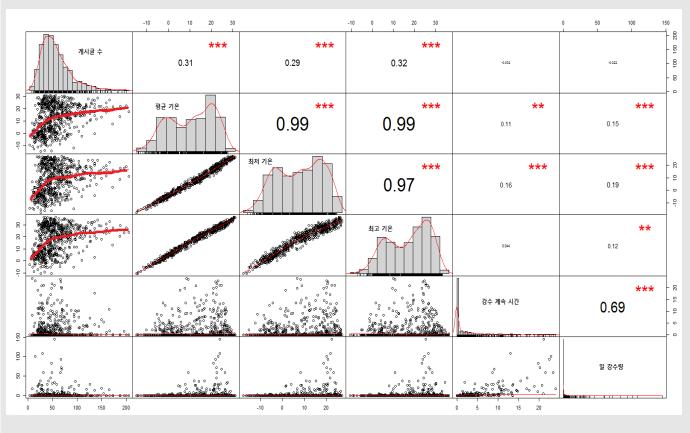
2. 공휴일 처리

공휴일의 경우 <mark>평일, 주말에 비해 다소 높은 관측치</mark> 모델 적합에 오차를 발생 시킬 가능성이 높아 이상치로 판단, 제거

3. 평일 주말 분할

주말의 경우 게시물의 수가 평일에 비해 많은 관광지가 존재 평일, 주말의 차이가 다소 있는 경우에는 데이터를 분할하여 분석을 실시

이상치 처리

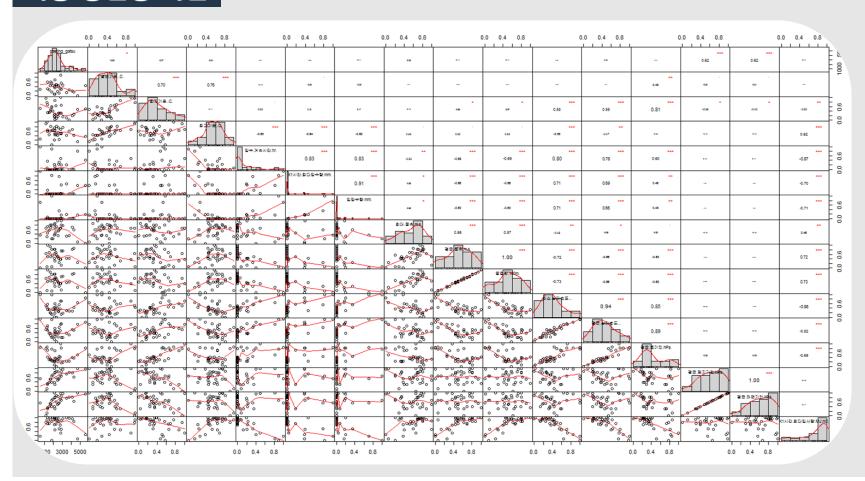


4. 계절별 분할

우리나라 지리 특성상 기온과 같은 날씨 변수는 계절적 특징을 지님

- 에시물의 개수와 변수간의 <mark>상관분석</mark>을 통해 이러한 변수들을 찾음
- 확인 결과 온도와 관련된 변수들은 대체로 상관관계가 급격히 변화하는 점이 존재
- 데이터가 계절성을 가지고 있는 것으로 판단, 이를 계절별로 나누어 분석 실시

다중 공선성 확인



선택된 설명변수들 중에서 <mark>공선성을 가지는 변수</mark>들을 제거하기 위해 설명변수들의 상관도를 분석

그 결과 (평균기온, 최저기온, 최고기온), (강수계속시간, 일강수량), (최대풍속, 평균풍속), (최소상대습도, 평균상대습도), (평균해면기압, 평균현지기압) 사이에 서로 높은 상관관계가 있음을 확인

다중 공선성 제거

```
cor(norbunsuckrain)
                    강수.계속시간.hr. X1시간.최다강수량.mm. 일강수량.mm.
강수.계속시간.hr.
                          1.0000000
                                              0.8299402
                                                          0.8262921
X1시간.최다강수량.mm.
                         0.8299402
                                                          0.9100976
                                             1.0000000
일강수량.mm.
                          0.8262921
                                              0.9100976
                                                          1.0000000
  cor(norbunsuckwind)
              최대.풍속.m.s. 평균.풍속.m.s. 풍정합.100m.
최대.풍속.m.s.
                 1.0000000
                               0.5574406
                                           0.5711354
평균.풍속.m.s.
                 0.5574406
                              1.0000000
                                           0.9972895
풍정합.100m.
                 0.5711354
                               0.9972895
                                           1.0000000
  cor(norbunsuckhumid_hpa)
                 최소.상대습도... 평균.상대습도... 평균.증기압.hPa. 평균.현지기압.hPa. 평균.해면기압.hPa.
최소.상대습도...
                     1.00000000
                                      0.9418788
                                                     0.8467741
                                                                      -0.1025273
                                                                                      -0.09907332
평균.상대습도...
                     0.94187885
                                                     0.8926155
                                                                      -0.1586487
                                                                                      -0.15506943
                                     1.0000000
평균.증기압.hPa.
                     0.84677406
                                      0.8926155
                                                     1.0000000
                                                                      -0.2765485
                                                                                      -0.27932418
평균.현지기압.hPa.
                    -0.10252731
                                    -0.1586487
                                                    -0.2765485
                                                                      1.0000000
                                                                                       0.99986244
평균.해면기압.hPa.
                    -0.09907332
                                                    -0.2793242
                                                                      0.9998624
                                     -0.1550694
                                                                                       1.00000000
  cor(norbunsuckelse)
                       X1시간.최다일사량.MJ.m2. 합계.일사.MJ.m2. 평균.지면온도..C. X1.5m.지중온도..C.
X1시간.최다일사량.MJ.m2.
                                                                   0.7328449
                                                                                   -0.5573984
                                   1.0000000
                                                  0.9543131
합계.일사.MJ.m2.
                                   0.9543131
                                                   1.0000000
                                                                   0.8136920
                                                                                    -0.5254284
평균.지면온도..C.
                                   0.7328449
                                                  0.8136920
                                                                   1.0000000
                                                                                   -0.3466765
X1.5m.지중온도..C.
                                  -0.5573984
                                                  -0.5254284
                                                                   -0.3466765
                                                                                     1.0000000
```

결론

02 데이터 전처리 - 데이터 정제

다중 공선성 제거 및 변수 선택

북한산	상관계수	한강	상관계수
평균기온	0.456	평균기온	-0.3463
최저기온	0.286	최저기온	-0.2723
최고기온	0.539	최고기온	-0.2433
강수계속시간	-0.641	강수계속시간	0.04
1시간최대강수량	-0.577	1시간최대강수량	0.0045
일강수량	-0.54	일강수량	-0.1111
최대풍속	-0.28	최대풍속	-0.26
평균풍속	-0.00595	평균풍속	-0.1095
풍정합	-0.0052	풍정합	-0.1141
최소상대습도	-0.36488	최소상대습도	0.04801575
평균상대습도	-0.6278	평균상대습도	0.0797
평균증기압	-0.1308	평균증기압	-0.071
평균현지기압	0.1458	평균현지기압	0.622
평균해면기압	0.1459	평균해면기압	0.624
합계일사량	0.187	합계일사량	-0.128
평균지면온도	0.1587	평균지면온도	-0.2045
안개계속시간	0.0001	안개계속시간	0

15개의 설명변수 중 공선성이 높은 변수 집합들과 나머지 변수 중에서 적절한 변수를 선택하기 위해 각 설명변수와 게시물 개수 사이의

상관도를 장소별로 측정

지역별로 상관도를 측정한 결과 세 군데 이상의 지역에서 상관도의 절대값이 0.3이상인 설명변수는 평균기온, 최고기온, 최저기온, 일강수량, 최대풍속, 평균풍속, 풍정합, 최소상대습도, 평균상대습도, 평균증기압, 평균현지기압, 평균해면기압, 평균지면온도로 선택

공선성을 가지는 변수는 여러 지역에서 평균적으로 상관도가 높은 변수를 선택

회귀모델로 적합하여 예측하기위한 설명변수를 평균기온, 최저기온, 최고기온, 평균풍속, 평균상대습도, 평균현지기압, 일강수량 7개의 변수로 최종 선택

다중 공선성 제거

```
> GLS <- gls(posting_gatsu ~평균기온..C.+최저기온..C.+최고기온..C.
+ +평균.풍속.m.s. + 평균.증기압.hPa.+일강수량.mm.
+ #평균.상대습도... , data = norbunsuck)
> AIC(GLS)
[1] 446.2735
> BIC(GLS)
[1] 457.5964
> GLS <- gls(posting_gatsu ~평균기온..C.+최저기온..C.+최고기온..C.
+ +평균.풍속.m.s. + 평균.현저기압.hPa.+일강수량.mm.
+ + #평균.상대습도... , data = norbunsuck)
> AIC(GLS)
[1] 434.4137
> BIC(GLS)
[1] 445.7366
> GLS <- gls(posting_gatsu ~평균기온..C.+최저기온..C.+최고기온..C.
+ #평균.풍속.m.s. + 평균.해면기압.hPa.+일강수량.mm.
+ #평균.상대습도... , data = norbunsuck)
> AIC(GLS)
[1] 434.4136
> BIC(GLS)
[1] 445.7365
> |
```

```
> GLS <- gls(posting_gatsu ~ 평균기온..C.+최저기온..C.+최고기온..C.

+ #평균.풍속.m.s. + 평균.해면기압.hPa.+일강수량.mm.

+ 최소.상대습도..., data = norbunsuck)

> AIC(GLS)

[1] 436.0692

> BIC(GLS)

[1] 447.3921

> GLS <- gls(posting_gatsu ~ 평균기온..C.+최저기온..C.

+ #평균.풍속.m.s. + 평균.해면기압.hPa.+강수.계속시간.hr.

+ 최소.상대습도..., data = norbunsuck)

> AIC(GLS)

[1] 437.5009

> BIC(GLS)

[1] 448.8237

> GLS <- gls(posting_gatsu ~ 평균기온..C.+최저기온..C.

+ #평균.풍속.m.s. + 평균.해면기압.hPa.+XI시간.최다강수량.mm.

+ #최소.상대습도..., data = norbunsuck)

> AIC(GLS)

[1] 440.4532

> BIC(GLS)

[1] 440.4532
```

```
> GLS <- gls(posting_gatsu ~평균기온..C.+최저기온..C.+최고기온..C.
+ +평균.풍속.m.s. + 평균.해면기압.hpa.+일강수량.mm.
+ #평균.상대습도..., data = norbunsuck)
> AIC(GLS)
[1] 434.4136
> BIC(GLS)
[1] 445.7365
> GLS <- gls(posting_gatsu ~평균기온..C.+최저기온..C.+최고기온..C.
+ +평균.풍속.m.s. + 평균.해면기압.hpa.+일강수량.mm.
+ *최소.상대습도..., data = norbunsuck)
> AIC(GLS)
[1] 436.0692
> BIC(GLS)
[1] 447.3921
> |
```

```
> RLM <-rlm(posting_gatsu ~ 최저기온..C. +평균기온..C.+최고기온..C.+ 일강수량.mm. + 최대.풍속.m.s. +
 최소.상대습도... + 평균.상대습도...+ 합계.일사.MJ.m2., data = norbunsuck)
> AIC(RLM)
[1] 566.9057
> BIC(RLM)
[1] 582.1693
> RLM <-rlm(posting_gatsu ~ 최저기온..C. +평균기온..C.+최고기온..C.+ 일강수량.mm. + 최대.풍속.m.s. +
 최소.상대습도... + 평균.상대습도...+ 평균.지면온도..C., data = norbunsuck)
> AIC(RLM)
[1] 573.1948
> BIC(RLM)
[1] 588.4584
```

03 분석 및 결과 – 회귀 분석 모델링

활용한 분석 모델

모델명	분석 사용처	날씨변수와 적합이유				
MLR	반응변수와 두 개 이상의 설명변수 간의 관계를 모형화 최소제곱법 또는 OLS에 의해 회귀계수를 추정	날씨 변수는 총 16가지로 MLR 회귀분석의 사용에 적합				
포아송 회귀	가산자료에 적합한 확률분포함수에 근거한	날씨변수에 대한 종속변수(여행지 인원수)가 가산자료이므로				
음이항 회귀	비선형 모델을 수립할 때 사용	포아송, 음이항 회귀 사용에 적합				
GLS 회귀	상관된 오차를 가지는 회귀모형의 적합에 유용한 방법 비정상적인 평균을 가지는 시계열자료에 대한 회귀를 수행하거나 오차항이 ARMA 과정을 따르는 경우의 회귀를 수행할 때 유용	날씨변수는 시계열 자료이므로 GLS회귀 사용에 적합				
로버스트 회귀	최소제곱회귀의 대안적인 방법으로 이상치 자료에 덜 민감하게 반응하는 회귀분석 방법 데이터 수가 작거나 이상치를 포함한 자료의 적합에 유용	날씨변수는 편향이 존재하고 분산이 크기 때문에 이상치가 많은 자료				

북한산	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.9681	0.8725	287.6229	319.0707	MLR	0.7997	0.7562	308.9276	318
포아송 회귀			283,2951	313.3756	포아송 회귀			367.0764	379.3821
음이항 회귀			285.2427	316.6905	음이항 회귀			295.7981	308.1037
GLS 회귀	0.9681	0.8725	75.3075	74.06344	GLS 회귀	0.67438	0.5658	276.6004	284.0993
로버스트 회귀	0.9329	0.7315	312.0709	343.5188	로버스트 회귀	0.6574	0.5639	309.4637	319.0348

명동	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.5104	0.3348	916.5154	964.6150	MLR	0.3798	0.3234	907.7687	926
포아송 회귀			1465.106	1510.915	포아송 회귀			1609.623	1634.818
음이항 회귀			929.1557	977.2554	음이항 회귀			917.1578	930.9005
GLS 회귀	0.5104	0.33487	680.781	722.1571	GLS 회귀	0.2488	0.1927	840.0325	859.602
로버스트 회귀	0.4084	0.19631	927.3392	975.4389	로버스트 회귀	0.1746	0.08567	928.9132	949.5274

경복궁	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.2735	0.1705	1716.0270	1782.3310	MLR	0.2567	0.2393	1688.602	1,703
포아송회귀			5721.876	5785.298	포아송회귀			5929.305	5963.898
음이항회귀			1693.718	1760.022	음이항회귀			1667.382	1681.796
GLS회귀	0,2735	0.1705	1446.175	1508.286	GLS회귀	0.13138	0.08234	1728.631	1754.013
로버스트회귀	0.2558	0.1137	1727.811	1794.115	로버스트회귀	0.12671	0.0774	1718.745	1744.69

북촌한옥마을	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.3210	0.3058	5399.6620	5460.0010	MLR	0.3177	0.3089	5392.335	5,431
포아송회귀			8431.261	8487.29	포아송회귀			8429.116	8472.216
음이항회귀			5218.618	5278.956	음이항회귀			5209.884	5248.674
GLS회귀	0.32098	0.3058	5308.443	5368.447	GLS회귀	0.2268	0.2153	5400.654	5443.588
로버스트회귀	0.2468	0.2299	5424.815	5485.154	로버스트회귀	0.1915	0.1795	5473.189	5516.289

한강	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.8192	0.6021	551.8382	582.3654	MLR	0.7293	0.6692	541.5491	554.1871
포아송회귀			3040.146	3069.146	포아송회귀			3584.113	3605.482
음이항회귀			552.896	583.423	음이항회귀			553.8524	563.0106
GLS회귀	0.81916	0.6021	250.051	264.212	GLS회귀	0.5814	0.4687	434.4137	445.7366
로버스트회귀	0.7430	0.4346	586.7752	622.7602	로버스트회귀	0.3554	0.1491	610.3635	625.7857

가평	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.6740	0.5229	860.2215	904.5499	MLR	0.5833	0.5545	848.6378	863.1891
포아송회귀			2823.315	2865.532	포아송회귀			3406.189	3435.741
음이항회귀			836.9504	881.2787	음이항회귀			831.8723	846.6485
GLS회귀	0.6740	0.5229	588.8598	624.8448	GLS회귀	0.4927	0.4256	785.5826	799.637
로버스트회귀	0.4960	0.2624	883.54	927.8683	로버스트회귀	0.4564	0.4069	858,5548	873.3309

연남동	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.3210	0.3058	5399.6620	5460.0010	MLR	0.3177	0.3089	5392.335	5,431
포아송회귀			8431.261	8487.29	포아송회귀			8429.116	8472.216
음이항회귀			5218.618	5278.956	음이항회귀			5209.884	5248.674
GLS회귀	0.32098	0.3058	5308.443	5368.447	GLS회귀	0.2268	0.2153	5400.654	5443.588
로버스트회귀	0.2468	0.2299	5424.815	5485.154	로버스트회귀	0.1915	0.1795	5473.189	5516.289

종묘	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.5396	0.3967	860,2215	904.5499	MLR	0.5733	0.5345	848.6378	863
포아송회귀			11820.7	11837.4	포아송회귀			18076.4	18084.7
음이항회귀			830,6859	848.9851	음이항회귀			831,8723	846.6485
GLS회귀	0.5396	0.3967	657.1709	624.8448	GLS회귀	0.4926574	0.4256499	785.5856	799.637
로버스트회귀	0.1518	0.111489	897.1404	915.4395	로버스트회귀	0.1934572	0.1324186	882,2489	898.8845

창경궁	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.5869	0.4587	928.4194	946.7186	MLR	0.4997	0.4586	923.8845	932,2023
포아송회귀			30853.9	30870.6	포아송회귀			40269	40278
음이항회귀			904.3845	922.6837	음이항회귀			903.1828	911.5006
GLS회귀	0.5869	0.4587	702.7237	717.764	GLS회귀	0.575333	0.462088	728.3769	742.3889
로버스트회귀	0.496773	0.3406	932.4224	950.7216	로버스트회귀	0.557887	0.43999	928.0169	944.6525

창덕궁	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.6288	0.5136	966.1627	984.4618	MLR	0.5799	0.5439	958.9931	967.3109
포아송회귀			33986.1	33992.8	포아송회귀			40676.3	40684.6
음이항회귀			939.52	957.8192	음이항회귀			933,9281	942.2459
GLS회귀	0,6288	0.5136	730.7892	745.8295	GLS회귀	0.60815	0.5037	758,2104	772.2224
로버스트회귀	0.4955	0.3389	974.5531	992.8523	로버스트회귀	0.5398	0.4170	968.3653	985.001

덕수궁	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.5285	0.3822	980.4655	998.7647	MLR	0.4594	0.4131	973.795	982.1128
포아송회귀			34694.9	34711.6	포아송회귀			45547.5	45555.8
음이항회귀			929.7655	948.0647	음이항회귀			929.2789	937.5967
GLS회귀	0.5285	0.3822	741.4246	756.4649	GLS회귀	0.5249	0.3982	767.8125	781.8245
로버스트회귀	0.2175307	-0.0253	994.8533	1013.152	로버스트회귀	0.2573	0.0593	988.9961	1005.632

서울시립미술관	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.5167	0.3667	935.4664	953.7656	MLR	0.2991	0.2601	935.9655	942.6198
포아송회귀			35369.3	35385.9	포아송회귀			49221.5	49229.8
음이항회귀			926.5828	944.882	음이항회귀			927.377	934.0312
GLS회귀	0.5167	0.3667	707.9638	723.004	GLS회귀	0.5113	0.4010	733,3032	747.3152
로버스트회귀	0.4788	0.3171	939,261	957.5601	로버스트회귀	0.4975	0.3635	934.494	951.1296

태릉, 강릉, 조선왕릉전시관	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.7069	0.6159	770.3583	788.6575	MLR	0.6883	0.6616	760.7569	769.0747
포아송회귀			24443.78	24460.41	포아송회귀			27074.57	27082.89
음이항회귀			706.0351	724.3343	음이항회귀			698.718	708.6994
GLS회귀	0.7069	0.6159	585.1911	600.2314	GLS회귀	0.7066	0.6283	605.9957	620.0077
로버스트회귀	0.4083	0.2247	790.3275	808.6267	로버스트회귀	0.6080	0.5035	722.3245	788.9601

트릭아이미술관	R square	Adjust R square	AIC	BIC		변수선택 R^2	변수선택 Adj R^2	변수선택 AIC	변수선택 BIC
MLR	0.3551	0.155	823.8113	842.1105	MLR	0.3210	0.1621	827.2288	837.2102
포아송회귀			73262.49	73279.12	포아송회귀			102139	102147
음이항회귀			823,877	842,1762	음이항회귀			822.0051	825,3323
GLS회귀	0.3551	0.1550	624.9382	639.9785	GLS회귀	0.3448	0.1700	647.5604	661.5723
로버스트회귀	0.4749	0.3119	840.6596	858.9587	로버스트회귀	0.3754	0.3287	823.2954	839.931

03 분석 및 결과 - 분석 결과

회귀 분석 결과

- 연남동, 경복궁, 북촌한옥마을, 명동, 트릭아이미술관, 덕수궁, 서울시립미술관은 AIC 값이 높고 R square 값이 낮아 날씨로 모형을 만들기에 부적합
- 2. 북한산, 가평, 한강, 종묘, 창경궁, 창덕궁, 조선왕릉전시관은 R square, adjust R square 값이 유의미하게 높아 날씨로 모형을 만들기에 적합

결론적으로 날씨에 따라 적합한 관광지를 선택하여 최적의 관광지를 추천한다.

03 분석 및 결과 – 분석 결과

회귀 분석 결과

1. 유의미한 장소들의 회귀분석식은 다음과 같다.

place	base	avg_tmp	min_tmp	max_tmp	rain	wind	humid	hpa	vapor	avg_wind	snow	ground_ temp
북한산	269.7804	630,236	-145.464	-423,106	-201,238	-90,2604	-180,534	-5.53941	0	0	0	0
한강	1069,297	-2221,73	311,4358	602,395	-3487,51	-985.144	2159,441	2636,126	0	0	0	0
가평	248.6889	-1436,37	1461,317	1069.371	-161,452	-329,275	-91.5244	306.5732	0	0	0	0
조선왕릉전 시관	-16417,7	54569.55	0	0	6537,286	-252,761	1155,144	13246,76	-44705.8	-2096.56	3128,587	6079.493
창경궁	-89409.8	562694	0	0	62134.48	22481,29	-77392.5	110020.9	-174647	3794,872	-2427.9	-242211
창덕궁	-101642	1177832	0	0	17845.7	-18106,2	-23072,8	135153.5	-506514	52640.79	20466.43	-555071
종묘	-6473	102796	0	0	0	0	-5922	16683	-75815	0	0	0

2. 위의 회귀식에 여행날의 날씨를 대입하면 그 날씨에 최적화된 여행지를 선택할 수 있다.

03 추가 분석 - 여행지 키워드 분석

'한복? 땀복!'…폭염 속 대여 한복 직접 입어 보니

입력: 2018-07-25 23:06 | 수정: 2018-07-26 02:12

입고 있던 바지 위에 두겹 치마 속치마 바람 안 통해 땀 줄줄... 여름 소재 없어



최신기사

[르포] 사람 잡는 폭염에 관광객 '뚝'…전주한옥마을 직 격탄

송고시간 | 2018/08/03 16:03



폭염 탓에 한옥마을 관광객 65% ↓, 상인들 매출 하락에 울상



여행 키워드 분석 선택 이유

- ✓ 여행지에서 단지 유행하는 활동 을 하는 것 보다 날씨에 최적화된 활동을 추천한다면 만족도를 높 일 수 있다.
- ✓ 그러나 관광객 또는 관광업에서 이를 분석적으로 활용하지 않음.
- 관광객에게 새로운 패러다임 기반으로 관광지에서 키워드를 추천하고 날씨에 능동적으로 대처할 수 있는 분석이 필요

03 추가 분석 - 분석용 변수 선정

군집 개수 설정

R 에서 NbClust함수를 통해 2017. 8.25 ~ 2018.8.24 까지 1년간의 날씨를 분류할 군집 수를 설정

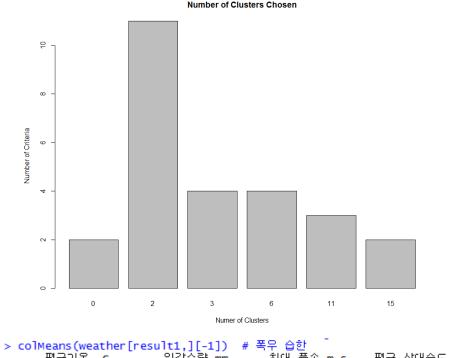


6개 군집 결정

KNN 을 통한 날씨 군집 구분

6개로 구분한 날씨를 각각 네이밍

- 1. 춥고, 건조한
- 2. 덥고, 폭우
- 3. 시원, 건조
- 4. 시원, 습한
- 5. 따뜻, 적당한 습도
- 6. 춥고, 적당한 습도



```
일강수량.mm.
                                    최대.풍속.m.s.
     평균기온..c.
         21.26667
                           83.66667
                                             4,70000
                                                             93.54444
                                                                             1006,64444
> colMeans(weather[result2,][-1]) # 서늘 약간 습한
                     일강수량.mm.
                                    최대. 풍속. m.s.
     평균기온..c.
                                                  평균.상대습도... 평균.해면기압.hPa.
         6.545000
                           3.558333
                                            4.065000
                                                             68.130000
                                                                            1020, 276667
> colMeans(weather[result3,][-1]) # 날씨 좋고 따뜻
                                    최대. 풍속. m.s.
     평균기온..c.
                                                  평균.상대습도... 평균.해면기압.hPa.
       19.4756757
                         0.5418919
                                           3.6827414
                                                            55.6094595
                                                                           1016.4189189
> colMeans(weather[result4,][-1]) # 날씨 좋고 시원 바람
                                    최대.풍속.m.s.
                     일강수량.mm.
                                                  평균.상대습도... 평균.해면기압.hPa.
     평균기온..c.
       12,2794872
                         0.4435897
                                           5.6846154
                                                            46.6564103
                                                                           1015.8923077
> colMeans(weather[result5,][-1]) # 춥고 건조
                     일강수량.mm.
                                    최대.풍속.m.s.
                                                  평균.상대습도... 평균.해면기압.hPa.
     평균기온..c.
      -2.17386364
                         0.06818182
                                          4.25821438
                                                           43.51250000
                                                                          1026.06022727
> colMeans(weather[result6,][-1]) # 업고 습한
                     일강수량.mm.
                                    최대.풍속.m.s.
     평균기온..c.
                                                  . 평균.상대습도... 평균.해면기압.hPa.
        24.435789
                           3.669474
                                            4.065263
                                                            71.224211
                                                                            1007.570526
```

03 추가 분석 – 분석용 변수 선정

모델 성능 평가

Trainset = 70%, Testset = 30% 로 나누어 성능 평가

오른쪽 그림과 같이 95.89%로 정분류

KNN 모델링이 적합하게 데이터를 나누는 것을 확인

■ 03 추가 분석 – 분석용 변수 선정

날씨와 해쉬태그 count를 mapping

날짜별 여행지역 해시태그와 함께 나온 해시태그 수집

EX) 2018년 7월 18일 #한강 #유람선 #치맥 #푸드트럭 834개 574개 382개 2018년 5월 8일 #한강 #공원 #운동 #피크닉 764개 645개 985개



<id, 날짜, 해시태그가 크롤링되어 저장된 모습>

오늘의 날씨가 6개의 군집 중 어디로 들어가는지 확인, 해당 군집의 최다 해시태그 추천

EX) 2018년 7월 23일 고온다습 1순위. 치맥 2순위. 유람선 3순위. 푸드트럭

04 결론 – 관광지 추천 제언

대한민국 구석구석 행복여행

로그인 회원가입 사이트맵 LANGUAGE 🖂 🤺 한국관광공사 🗵

검색어를 입력해주세요

Q 팝업열기 ※

구석구석 플러스

· 구석구석여행지찿기

구석구석 떠나볼까

- 축제·행사 · 한국관광 100선
- · 테마여행 • 여행플래너

· 최신기사

- · 콕콕! 지역정보
- 추천! 가볼만한 곳

레포츠

• 교과서 여행

어디로 갈까

• 관광지정보

- 음식점정보
 - 구석구석맛골목

· 우리고장 맛이야기

· 궁중음식 체험하기

- · 웰니스 관광 25선
- 방방곡곡 시티투어

무엇을 먹을까 어디서 잘까

- 숙박업소정보
- 베니키아 • 굿스테이

• 한옥스테이

• 캠핑장

- 이벤트 • 여행매거진

• 관광뉴스

• 관광정보지킴이

구석구석 소식

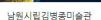
- 템플스테이
- 한국관광의 별 · 무장애여행
 - 생태관광
 - 산업관광

 - · K-Style Hub
 - · 내가 쓰는 여행기



영월에코빌리지 유스 호스텔

강원도 영월군 영월읍 동강로 716 문의: 예약문의 033-375-6800



전라북도 남원시 함파우길 65-14 문의: 안내 063-620-5660



문화시설

경상남도 합천군 가야면 가야산로 1742-17 문의:학예연구실 055-934-3150~2, .



유엔평화기념관 문화시설

부산광역시 남구 홍곡로320번길 106 문의: 유엔평화기념관 051-901-1400



노래박물관(남이섬)

강원도 춘천시 남산면 남이섬길 문의: 재단법인 노래의섬 031-580-8015



송도 구름산책로(스카 이워크)

부산광역시 서구 송도해변로 문의: 부산 서구 관광안내소 051-231-20...



대구종합유통단지관리

대구광역시 북구 유통단지로13길 9 문의:053-381-3690



정관박물관

부산광역시 기장군 정관읍 정관중앙로 122 문의: 정관박물관 051-720-6900~690...

문화시설

현재 관광지 추천 시스템 개선

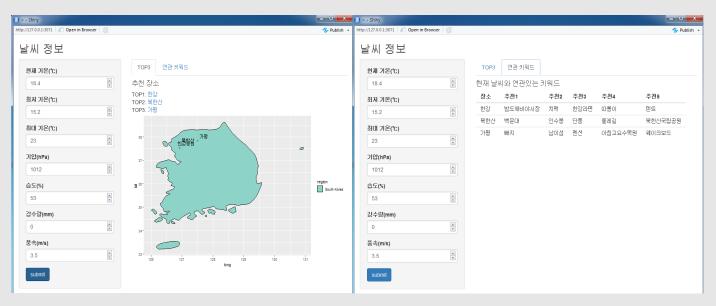
✓ 현재 한국관광공사의 관광지 추천 시스템에 새로운 지표로서 활용

관광 소비재, 관광 상품, 관광지 행사

04 결론 – 관광지 추천 제언

주제 1

날씨 변수를 통한 관광지 추천에 대한 새로운 패러다임 제시



R Shiny를 통한 날씨에 따른 추천 장소, 키워드 시각화

사용자들은 날씨에 기반하여 추천 관광지와 연관 키워드를 제공받게 되고 이는 수많은 사람들의 데이터에 의해 도출된 정보이므로 기존에 획일화된 추천보다 더 높은 정확성과 만족도를 보장한다.

04 결론 – 마케팅 활용 방안

주제 2

관광객을 증대시키는 지표로 활용







날씨에 민감한 지역에서 관광객 방문이 감소할 때 지역 행사에서 <mark>어떤</mark> 활동을 할지 판단할 때 활용

04 결론 – 마케팅 활용 방안

주제 2

관광객을 증대시키는 지표로 활용

지자체 행사축제 年1만1800건... 費用 톱10중 7건이 '수익 0'

조선일보 | 영암=김효인 기자 최희명 기자

 \square

입력 2015.09.10 03:00 | 수정 2015.09.11 18:47

[지방부채 100조원] [中] 우후죽순 '빚잔치 행사'

- 적자 뻔한테 열고 보자 223억 들인 오송 뷰티박람회, 제주 들불축제, 세종축제·...
- 같은 행사, 제 살 깎기 도자기축제, 경기도만 3곳... 국제보트쇼는 전국 3곳서... 행자부 "빚 행사 대책 세울것"

지난달 25일 방문한 전남 영암 코리아인터내셔널 서킷은 고요했다. 축구장 170개 넓이에 달하는 1 85만3000㎡ 규모 경기장과 12만명을 수용할 수 있는 스탠드는 모두 비어 있었다. 서킷을 둘러보니 3km 구간에서 갈라져 나온 5km 구간이 특히 깨끗했다. 앞선 주말에 3km 구간만 이용하는 아마추어 오토바이 대회가 열렸다고 한다. 굳게 닫힌 차량 정비 구역에는 2013년 포뮬러원(F1) 코리아 그랑 프리에 참가했던 '마크 웨버' '세바스티안 베텔' 등 유명 카레이서들 이름과 얼굴이 차고지 줄입구에 새겨진 채 그대로 남아 있었다. 이 경기장에서는 2013년 이후 F1 경기가 열리지 않았다.

전남도는 2010년부터 7년간 대회를 열기 위해 F1 조직위원회 및 F1 매니지먼트(FOM) 측과 계약했지만, 경영난을 이유로 작년부터 대회 개최를 포기했다. 경기장을 건설하는 데는 4285억원이 들었다. 국비가 1285억원이고, 전남도는 3000억원을 부담했다. 당시 전남도가 발행한 지방채 1230억원에 대한 이자만 향후 15년간 363억원에 달하지만, F1 경기는 F1 4차례 만에 누적적자 1902억원을



주식고수가 말하는 '실패하는 유형' 3가지!!

정치 주요뉴스

김무성 "장하성, 깜짝 놀랐다? 무책임 발언... 해임해야"



이해찬 "소득주도성장으로 4만달러 시디 열겠다" 현재 지역행사, 축제가 지역특색화를 위해서 수요를 고려하지 않고 공급하고 있는 Top-down방식

> 현 상황에서 날씨 및 연관 아이템 분석을 통해 수요를 감안한 행사 기획 및 상품 추천 개발 가능

감사합니다.

THANK YOU!